Dissertation

Concurrent rhythms of prosody and syntax in the neurophysiological tracking of sentence structure

Anastasia Glushko

Integrated Program in Neuroscience

McGill University, Montreal

April 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

LIST OF CONTENTS

ABSTRACT

The current doctoral thesis evaluates processing of sentence constituents using a novel electrophysiological approach to studying high-level language processing: namely, frequency tagging. The extant research results indicate it is debatable whether and how the hierarchical syntactic structure of language (Chomsky, 1959, 1995) is processed by language users. Principal difficulties in evaluating syntactic structure processing online are methodological in nature. They include the need to rely on studies of grammatical error processing as well as the frequent alignment of syntactic structure with other layers of sentential information (e.g., prosody and semantics). The first concern was recently addressed by Ding, Melloni, Zhang, Tian, and Poeppel (2016) who presented spoken sentences with isochronous structure (e.g., a two-syllable noun phrase followed by a two-syllable verb phrase) to their participants. This experimental design allowed them to 'tag' the frequencies at which phrases and sentences appeared in the speech stream: prominent EEG spectral power peaks were elicited at the exact ('tagged') frequencies of syntactic constituents. The results were taken as evidence for cortical tracking of hierarchical syntactic structures and the approach was said to be promising for studying syntactic processing in diverse populations. Crucially, however, the relative role of various layers of sentential information in the elicitation of observed effects in frequency tagging studies of sentence processing has yet to be understood. In Manuscript 1, I addressed this question by investigating concurrent processing of overt and covert (i.e., implicit, subvocalized) prosodic information on the one hand, and syntactic structure on the other using a frequency tagging paradigm (across three experiments). We applied several prosodic contours to two types of syntactic structures tagging contrasting frequencies in each case. The prosodic contours were applied either overtly, via manipulating the speech signal of the sentences, or covertly, by asking participants to imagine a specific prosodic contour. We demonstrated that overt and covert prosody processing does indeed contribute to the elicitation

of spectral power peaks at tagged sentence constituent frequencies. Further, we describe the interactive effects of syntax and prosody directly evident from the EEG responses at the tagged sentence constituent frequencies in perfectly grammatical sentences: When prosodic structure aligned with syntactic phrasing, the spectral power peak at phrase frequency was enhanced compared to a neutral prosodic condition without prosodic cues. When prosody and syntax were misaligned, responses at the sentence frequency were suppressed.

In Manuscript 2, we undertook a methodological investigation of the frequency tagging technique as applied to sentence processing, which included data simulations and (re-)analysis of real data. After providing an overview of the technique, we first address the role of non-syntactic and syntactic variables (other than prosody) in the elicitation of spectral power peaks at frequencies of sentence constituents in previous frequency tagging studies. Thereafter, we demonstrate potential caveats of various approaches to data analysis in these studies. Based on our investigation, we discuss the potential for use of the frequency tagging technique in studying sentence processing.

This dissertation contributes to our understanding of sentence processing by demonstrating concurrent and interactive low-frequency cortical tracking of prosodic and syntactic structures of sentences in the case of both overt and covert prosody. The work outlines the potential, as well as the constraints, of the frequency tagging technique as used in sentence processing research, and allows us to situate this approach within the spectrum of other neurolinguistic methods.

RÉSUMÉ

La présente thèse de doctorat évalue le traitement des constituants des phrases au moyen d'une nouvelle approche électrophysiologique permettant d'étudier les mécanismes sous-jacents de l'analyse linguistique de haut-niveau, que l'on nomme le marquage de fréquence ("frequency tagging" en anglais). Jusqu'ici, l'existence et le fonctionnement du traitement des structures syntaxiques de la phrase (Chomsky, 1959, 1995) ont été largement débattues. Les principales difficultés concernant leur évaluation sont méthodologiques: elles sont liées à la nécessité de se reposer sur le traitement d'erreurs grammaticales, et aussi à la superposition fréquente de la structure syntaxique avec d'autres niveaux d'information linguistique (ex. la prosodie et la sémantique). Le premier point a été récemment abordé par Ding, Melloni, Zhang, Tian et Poeppel (2016), qui ont présenté en modalité auditive des phrases avec une structure isochrone (ex. un syntagme nominal de deux syllabes, suivi d'un syntagme verbal également bisyllabique). Ce dessin expérimental a permis de "marquer" les fréquences, qui dans le signal de parole, correspondent aux syntagmes et aux phrases. En d'autres termes, des pics de densité spectrale du signal EEG ont été observés aux mêmes fréquences que les constituants syntaxiques. Ces données semblent attester du pistage cortical des structures syntaxiques, et l'approche expérimentale utilisée s'avérait prometteuse pour étudier le traitement de la syntaxe dans des populations diverses.

Cependant, les rôles respectifs des différents niveaux d'information linguistique sur les effets observés dans les études de marquage de fréquence sont assez mal compris. Dans le Manuscrit 1, nous avons étudié cette question en implémentant un paradigme de marquage de fréquence qui examine, à travers trois expériences, le traitement concomitant de la prosodie explicite et implicite d'une part, et de la structure syntaxique d'autre part. Nous avons imposé plusieurs contours prosodiques sur deux types de structures syntaxiques qui marquent des fréquences distinctes. Les contours prosodiques étaient appliqués soit explicitement, grâce à la

manipulation du signal acoustique, soit implicitement, en demandant aux participants d'imaginer un contour prosodique particulier. Nos résultats montrent que la prosodie explicite et implicite contribuent effectivement à l'induction de maximums de densité spectrale suivant le marquage de constituants de la phrase. De plus, nous observons une interaction entre la syntaxe et la prosodie: quand la structure prosodique est alignée sur la syntaxe, les pics observés à la fréquence des syntagmes étaient plus larges en comparaison à la condition sans indices prosodiques. Lorsque la prosodie et la syntaxe n'étaient pas alignées, les réponses à la fréquence des syntagmes étaient éliminées.

Le Manuscrit 2 est une évaluation méthodologique de la technique du marquage de fréquence appliquée au traitement de la phrase, incluant des simulations de données et des (ré-)analyses de données réelles. Nous avons d'abord questionné le rôle des variables syntaxiques et non-syntaxiques (autres que la prosodie) dans l'élicitation de pics de densité spectrale à la fréquence des constituants. Ensuite, nous avons mis en lumière des biais potentiels liés à plusieurs approches d'analyse de données dans les études existantes, et discuté le potentiel de la technique de marquage de fréquence pour étudier le traitement de la phrase.

Cette thèse démontre les effets coopératifs et intéractifs du pistage cortical des structures syntaxiques et prosodiques pour la prosodie implicite et explicite. Elle contribue à notre compréhension des mécanismes sous-jacents au traitement de la phrase. Le présent travail souligne le potentiel, mais aussi les contraintes associés à la technique du marquage de fréquence telle qu'utilisée pour la recherche sur la traitement du langage, et permet de situer cette approche sur le spectre des méthodes en neurolinguistique.

ACKNOWLEDGEMENTS

Many individuals and organizations have made my PhD work possible, and I can only hope that I will be able to give back as much as I have received in support and guidance in the years of my PhD.

My supervisor, Dr. Karsten Steinhauer, was the reason I moved across the ocean to pursue my PhD. Throughout my studies and to this day, I stand firmly behind this decision, having worked in his lab for more than five years. Karsten's vast expertise and his readiness to share it with his students are truly inspiring. Navigating a complex project can be challenging; yet, I was armed with the knowing that in the face of any difficulties I would encounter, Karsten would ultimately be on my side. The invaluable importance of such support in one's PhD cannot be overemphasized. His unwavering belief in my ability to pursue this and other ambitious projects was at times intimidating but made the past years of my life exciting and filled with professional and personal development. I am profoundly thankful for having worked with and learned from you, Karsten.

This project would not have been possible without the support of Dr. David Poeppel who encouraged me to pursue it and provided hours of consultation, generously sharing his exceptional expertise in the field. I am extremely grateful to him for the opportunity visit Max Plank Institute for Empirical Aesthetics (MPI) and learn from his team. David thank has become an important source of professional inspiration for me.

I am immensely grateful to Max Wolpert who joined this project at the very beginning of its development. He took part in the painstaking task of creating and piloting the experimental stimuli, collected data together with me, and spoke with me about the results -

which led to multiple insights. He also edited many parts of this dissertation. It was a rare pleasure to work with one of my closest friends.

This project was supported by the work of Ayesha Liaqat, former summer bursary student, and Katharina Gandras, former research assistant, who contributed to data collection and processing, and were exceptional at trouble-shooting technical issues in the most creative ways.

I wish to acknowledge the effort put in by the participants of my experiments. Many of them did not only take part in the study but also helped me with recruiting German native speakers in Montreal – I appreciate this a lot.

Each of my committee members, Dr. Aparna Nadig, Dr. Phaedra Royle, and Dr. Robert Zatorre, provided invaluable support since the start of my PhD, and their multifaceted expertise was very helpful in situating my research into a broader context.

Current and former members of my lab have become my most important colleagues and, above all, my good friends. These are the *LMs* – Dr. Lauren Fromont, Alexandre Herbay, and Max Wolpert, as well as Fayden Bokhari, Dr. Kristina Kasparian, and Dr. Stefanie Nickels. I have been inspired by all of you and learned from you. I am indebted to you the support you have given me, and I am beyond happy that we have become friends.

I am thankful to have worked at the MPI alongside and received advice from Dr. Alessandro Tavano, Dr. Xiangbin Teng, Dr. Johanna Rimmele, Dr. Oded Ghitza, and Dr. Yue Sun, as well as for the logistical support of Cordula Ullah. It is at the MPI where I met Dr. Nai Ding who was encouraging and shared part of his analysis code with me. The staff members and the graduate students from the MPI have been very welcoming, and David's team in Frankfurt became my second PhD home base.

I am grateful for the financial support for my PhD studies provided by CIHR, CRBLM, Mitacs, and McGill sponsors. I have received significant support from my university: both in terms of funding and while dealing with administrative matters. The Integrated Program in Neuroscience and McGill's Graduate and Postdoctoral Studies offices are highly skilled at making things run smoothly in one's PhD. The School of Communication Sciences and Disorders at McGill University also provided great logistical support for my project. A special thank you goes to Healthy Brains, Healthy Lives (and more directly to Dr. Falisha Karpati) - an impressive initiative realized by talented and dedicated people, it has contributed to my career in many ways through their various workshops and events.

Among the personal connections that have helped me keep in check with my values and intentions during my PhD studies are some of the closest people in my life: Aliona, J, Jay, Masha, Paul, and Toivo. I am deeply thankful for all the different ways you have supported me while I was working on my PhD. My Montreal gang – Alberto, Marcel, Océane, Mark, Polina, and Misha - thank you for making my immigrant experience much less difficult than it could have been.

Lastly and importantly, I am thankful to my mother who has been an inspiration for me throughout my life, has pushed for me to obtain the best education possible since I was a child, and has always supported me with all the resources she had. I am grateful for having been exposed to, and hopefully assimilated, her ethical standards, her passion for research, her healthy perfectionism, and her bravery. Спасибо, мамик!

FUNDING ACKNOWLEDGEMENT

The research conducted in the frame of the current thesis was funded by the SSHRC Insight grant Tracking Language Acquisition and Processing in Specific Language Impairment (435-2015-1280; PI: Dr. Phaedra Royle, Co-PI: Dr. Karsten Steinhauer) and the SSHRC Insight grant Re-evaluating the temporal dynamics of syntactic online processing: Towards an ecologically valid model (435-2013-0583; PI: Dr. Karsten Steinhauer, Co-PI: Dr. Phaedra Royle). The dissertation was also supported through the Vanier Canada Graduate Scholarship, CRBLM student stipend, as well as Mitacs Globalink funding received by Anastasia Glushko.

CONTRIBUTION TO ORIGINAL KNOWLEDGE

The thesis is the first work to use the frequency tagging paradigm while evaluating the concurrent roles of prosody and syntax in online sentence processing. Results provide evidence challenging the validity of several influential studies on sentence processing and their claims (e.g., Ding et al., 2016). Specifically, we have shown over three experiments and several data simulations, first, that prosodic structure processing is reflected in the spectrum of neural activity at low frequencies (Manuscript 1). Second, we demonstrated that a range of methodological concerns have to be addressed by frequency tagging research in sentence processing before this paradigm can be claimed to capture the neural signature of hierarchical phrase structure processing (Manuscript 2).

CONTRIBUTION OF AUTHORS

MANUSCRIPT 1

Anastasia Glushko: The author considered the potential of the frequency tagging technique for studying artificial grammar learning, thought of several possible experiments testing the validity of the claims in the original Ding and colleagues' (2016) study, created the experimental materials, arranged the experimental procedures, organized data collection, collected the data, analyzed the data, and prepared and revised the manuscript based on other authors' feedback.

David Poeppel: The author assisted in conceptual development of the project, provided support in data analysis, and gave feedback on data interpretation. In addition, he will edit the manuscript prior to its submission to a scientific journal.

Karsten Steinhauer: The author proposed prosodic processing as a starting point for the investigation of the alternative accounts of the original Ding and colleagues' (2016) study. He actively oversaw and gave feedback during all steps of the study development: i.e., creating the experimental materials, developing the experimental procedures, data collection, data analysis, and manuscript preparation.

MANUSCRIPT 2

Anastasia Glushko: The author conceived the methodological investigation of the Ding and colleagues' (2016) application of the frequency tagging technique to sentence processing. She generated an idea for all data simulations, performed them, and prepared the manuscript.

Max Wolpert: The author assisted in developing the stimuli for the pilot experiments that laid ground for the manuscript. He collected the pilot data, actively participated in numerous discussions of the methodological aspects of the frequency tagging technique, and edited the manuscript.

David Poeppel: The author provided expertise in research on neural oscillations in general and utilizing the frequency tagging technique in particular. He participated in several discussions of the methodological aspects of the data and will edit the manuscript prior to its submission to a scientific journal.

Karsten Steinhauer: The author initiated a large number of discussions on the methodological aspects of the frequency tagging technique, provided expertise in sentence processing research, and gave extensive feedback on the manuscript.

GENERAL INTRODUCTION

Understanding the structural relationships between single words is essential for language users. The extreme diversity of possible grammatical constructions within and across languages presents, on the one hand, a challenge for theoretical linguists attempting to parsimoniously describe the syntactic organization of language. On the other hand, an analogous challenge is faced by empirical research in psycho- and neurolinguistics investigating how native speakers develop the ability to process language syntax, and what the mechanisms of online syntactic phrase structure processing are.

Some of the most influential linguistic theories propose that the convoluted organization of language grammar can be described in terms of hierarchies produced by recursive merging of single linguistic units (Chomsky, 1981; revised in Chomsky, 1995). Phrases are formed by merging of two single words, larger phrases are composed of merging two smaller phrases, and eventually a sentence is formed in this hierarchical, recursive manner. That is, every sentence can be represented as a hierarchical tree structure, and an infinite amount of sentences can be produced from a limited number of single units via recursion. The recursive organization of language has been proposed as “the narrow language faculty” that is uniquely human and absent in the communication abilities of other species (Hauser, Chomsky, & Fitch, 2002; but see e.g., Pinker & Jackendoff, 2005). The matter of whether and how hierarchically organized syntactic constituents are processed online is, however, highly debatable (Pylkkänen, 2019; Ferreira, Bailey, & Ferraro, 2002; Frank & Bod, 2011; Christiansen & Chater, 2015), with empirical research on neurocognitive correlates of phrase structure processing facing multiple methodological challenges (for a review, see Uddén, de Jesus Dias Martins, Zuidema, & Fitch, 2019).

SYNTACTIC PHRASE STRUCTURE PROCESSING

Given the apparent essentiality of hierarchy to the organization of language, it is plausible that language users possess neurocognitive mechanisms for processing hierarchical syntax online. It is known from an extensive amount of behavioural and eye-tracking research that listeners and readers process sentence phrase structure (e.g., Fodor & Bever, 1965; Garrett, Bever, & Fodor, 1966; Holmes & O'Regan, 1981). This can be expressed, for instance, in their ability to be affected by the proximity of the syntactic boundaries in their ability to identify extraneous clicks (Fodor & Bever, 1965; Garrett, Bever, & Fodor, 1966). Eye fixations and regressive eye movements have been shown to follow the syntactic structure of sentences during reading (Holmes & O'Regan, 1981). These studies have laid the ground for the development of an influential syntax-based sentence processing theory - the garden-path model (Frazier, 1987). In the center of the latter is the way local syntactic ambiguities are processed: one of the interpretations is being preferred online, and sentence structure is reanalyzed at a later point if this initial interpretation is incompatible with subsequent words (so-called garden-path sentences).

Behavioural research on syntactic processing was followed up by neurolinguistic studies, showing that disambiguating words in garden-path sentences are associated with an elicitation of an event-related potential (ERP), the 'P600', that differed from the one typically driven by semantic analysis (the 'N400', Osterhout & Holcomb, 1992). Subsequently, the importance of syntactic processing for initial sentence analysis was further supported by reports of the 'early left anterior negativity' (ELAN; Hahne & Friederici, 1999; Friederici, Hahne, & von Cramon, 1998; Hasting & Kotz, 2008; but see Steinhauer & Drury, 2012). This effect appeared in response to sentences with violations of syntactic phrase structure and was elicited within the first 300 ms of the onset of the violation. Largely grounded on these findings, a neurophysiology-based model of online sentence processing was proposed (Friederici, 1995;

2002). In line with the garden-path model, it points to the primary role of syntactic analysis in sentence processing.

Building on behavioural and neurophysiological evidence for the major role of syntax in sentence processing, further studies have addressed the question of whether the syntactic relationships processed online are indeed hierarchical – and not merely linear – in nature (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Makuuchi, Balhmann, Anwander, & Friederici, 2009). An important line of evidence comes from neuroimaging research dissociating brain regions involved in complex sentence processing. First, it was shown that distinct brain areas are associated with syntactic structure complexity (i.e., as expressed in the presence of single or double embedded clauses in sentences, Makuuchi et al., 2009). Further, artificial grammar learning studies demonstrated that Broca’s area (composed of the pars opercularis and pars triangularis), often linked to syntactic processing in natural sentences (e.g., Bornkessel et al., 2005; for review, see Friederici, 2009; Grodzinsky & Friederici 2006), was activated during processing of learned, non-adjacent dependencies, while the frontal operculum (a phylogenetically older brain area) showed an activation pattern specific to processing of adjacent dependencies (Friederici et al., 2006). Involvement of Broca’s area in hierarchical processing was confirmed by studies of both natural language (Jeon & Friederici, 2013) and non-linguistic visually-based artificial grammars (Bahlmann, Schubotz, Mueller, Koester, & Friederici, 2009).

The claim that hierarchy is built online during sentence processing has faced various lines of criticism. The fact that some wrongful syntactic representations are maintained even after the temporal syntactic ambiguity is resolved led some researchers to propose that the full syntactic structure of the sentence is never constructed in the brain and, instead, a ‘good enough’ structure is used (Ferreira et al, 2002). The good-enough approach allocates an important weight to contextual information that supports the ambiguous syntactic structure. Other researchers have proposed that semantic rather than syntactic structure guides sentence analysis in real time (Pylkkänen, 2019). Finally, usage-based accounts of syntactic processing describe the nature of syntactic analysis as sequential rather than hierarchical (i.e., grounded in sequence processing abilities and not a dedicated recursion mechanism, Christiansen & Chater, 2015).

The heterogeneity of theoretical stances on the contribution of hierarchical syntactic phrase structure building to sentence processing is largely due to the methodologically-driven ambiguity of the empirical evidence available. Artificial grammar learning studies (such as Friederici et al., 2006; Bahlmann et al., 2009) have been criticized for using over-simplified syntactic structures that can be processed without using hierarchical processing (see e.g., Rohrmeier, Fu, & Dienes, 2012). In turn, when natural language stimuli are used, first, several layers of sentential information are simultaneously available for the processor (i.e., syntax, semantics, prosody), and it is challenging to study any of them in isolation. While this concern can be addressed by using error-based paradigms (for instance, studying ERP responses to syntactic violations), it is unclear to which extent processing of syntactic errors relies on the same mechanisms as processing grammatically correct sentences.

Recently, a new methodological approach with grammatically correct sentences has been used to study syntactic phrase structure (see detailed description of the method in Manuscript 2). In Ding and coauthors' (2016) study, participants listened to isochronously organized sentences, in which all words were of identical length (250 ms): phrases were formed of two words each, and sentences were two phrases long. This isochrony and the absence of pauses between words, phrases, or sentences meant that syntactic phrases and sentences appeared at stable frequencies: 2 Hz for phrases (two phrases were presented every second) and 1 Hz for sentences (every sentence was one second long). Researchers used MEG to record participants' neurophysiological activity, which was examined in the frequency domain. The main finding of the study was the elicitation of drastic spectral power peaks in the MEG data that were specific to the frequencies of sentences (1 Hz) and phrases (2 Hz). Given that overt prosodic cues (such as pitch and intensity changes as well as prosodic breaks) were neutralized in all stimuli materials, the authors concluded that slow neural oscillations 'concurrently' track hierarchical syntactic structure of sentences. While these findings were replicated several times (Ding et al., 2017; Makov et al., 2017; Sheng et al., 2019; Jin, Zou, Zhou, & Ding, 2018), their proposed syntactic nature has been questioned (Frank & Yang, 2018; Frank & Christiansen, 2018). In the next section, the role of a strong alternative account for the interpretation of these findings – namely, one based on covert prosody processing – will be discussed. The focus is placed on the neurophysiological signatures of prosodic phrase processing, both overt and covert, as evident from processing of fully correct sentences (similar to Ding et al., 2016) with controlled lexical and syntactic information.

THE ROLE OF PROSODY IN PHRASE PROCESSING

Prosody, or 'suprasegmental phonology', is a modulation of the speech stream that spreads beyond the level of a single phoneme and encompasses several linguistic phenomena including intonation (Féry, 2017). Prosodic structure's supporting role in sentence comprehension is evident from a large body of empirical research (e.g., Kjelgaard & Speer, 1999; for a review, see Cutler 1997). In syntactically ambiguous sentences, it is often prosodic phrasing that can help the language user process the sentence structure (Schafer, Speer, Warren, & White, 2000; Speer, Kjelgaard, & Dobroth, 1996). Neurophysiological responses to prosodic phrasing have been dissociated from syntactic phrasing using a unique, non-violation-based paradigm (Steinhauer, Alter, & Friederici, 1999). Contrasting two grammatically correct sentences with identical lexical items prior to, and directly following, the time window of interest, Steinhauer and colleagues (1999) showed that a slow positive shift is elicited at prosodic phrase boundaries, which they named the ‘Closure Positive Shift’ (CPS). The finding of the CPS was replicated in a variety of languages (reviewed by Bögels, Schriefers, Vonk, & Chwilla, 2011), and the paradigm was shown to be useful for studying language development in children (Männel, Schipke, & Friederici, 2013). The ERP component was presumed to be the signature of prosodic chunking.

The CPS is elicited irrespective of the specific prosodic cues used to mark the prosodic boundary. It can, for instance, be elicited in the absence of a boundary pause, as shown in the original study (i.e., purely due to pre-final lengthening and pitch rise; Steinhauer et al., 1999). Crucially, the CPS was later found to be elicited at positions of predicted prosodic boundaries in silent readers, that is, in the absence of overt prosodic phrase boundary markers (Steinhauer & Friederici, 2001; Steinhauer, 2003). This latter effect was linked to a phenomenon of implicit prosody previously studied using purely behavioural techniques (for a review, see Breen, 2014). Silent readers of sentences impose a covert prosodic contour on the materials read, and this imposition is driven by both syntactic phrasing and other factors (for a review, see Frazier, Carlson, & Clifton, 2006). During language processing, covert prosodic phrasing can be modulated by syntactic and semantic predictions, as reflected in the modulation of the CPS (Kerkhofs, Vonk, Schriefers, & Chwilla, 2007; Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010).

Prosodic phrase boundary processing has also been studied using the neural entrainment paradigm according to which neural oscillations can align their phase and frequency with properties of the stimulus. It has been shown that prosodic boundary processing follows a significant shift in the phase of the EEG delta (< 4 Hz) activity (Meyer et al., 2017). At the same time, coherence reduction between the prosodic contour and the delta phase was more pronounced in sentences that were interpreted in a way congruent with an alignment of the prosodic and syntactic boundaries (independent of the presence of an overt prosodic boundary). The reduction was smaller for the sentences interpreted as not having a syntactic phrase boundary at the location of the overt prosodic boundary. Meyer and colleagues (2017) considered two interpretations for these findings: either neural entrainment to sentence structure is driven by top-down syntactic processing (that may or may not be associated with prosodic phrasing, in line with Ding et al., 2016) or two types of delta oscillations are tracking different aspects of sentence structure (i.e., the 'acoustic' delta; Ghitza, 2017; and the syntactically-driven delta). The authors also speculated that CPS can 'mirror' delta entrainment. That is, tracking of sentence structure by slowly oscillating neural activity has been associated with both syntactic and overt prosodic processing. Their relative role in the elicitation of neurophysiological spectral power peaks at sentence constituent frequencies in recent frequency tagging studies (e.g., Ding et al., 2016, 2017) is yet to be determined, as is the possibility for delta activity to track covert prosodic phrases.

METHODOLOGICAL APPROACHES TO STUDYING NEURAL CORRELATES OF SENTENCE STRUCTURE PROCESSING

Ding and coauthors' studies (2016; 2017) used a novel neuroscientific technique for the field of sentence processing. Investigating its potential and the related considerations is one of the goals of the current thesis. Complementing behavioural research, neuroscientific methods are especially suitable for addressing certain questions about sentence processing: these include the interaction between various layers of sentential information and the hierarchical nature of online sentence processing. The known advantage of EEG and MEG is their good time resolution, meaning online sentence processing can be investigated with a high temporal accuracy. Both techniques also allow for qualification of distinct evoked transient responses associated with different cognitive mechanisms (such as the ELAN, the P600, and the CPS).

The most influential model in neurolinguistic sentence processing research is based on these temporal dynamics (Friederici, 2002). A drawback of this methodology, however, is that in most cases it is impossible not to rely on a violation-based paradigm while studying sentence processing (with CPS, garden-path studies, and semantic priming research being the exceptions; Steinhauer et al., 1999; Van Petten & Kutas, 1990; Osterhout & Holcomb, 1992).

Functional magnetic resonance imaging (fMRI) studies have poor temporal but good spatial resolution, and have often contrasted differently-organized correct sentences without relying on the violation paradigms (e.g., Makuuchi et al., 2009). The conceptual advantage of these techniques is their ability to address the question of whether two cognitive mechanisms are distinct, through hypothesizing that they would in these cases engage activity in distinct, non-overlapping brain networks (as in the case of processing recursive and non-recursive structures, Friederici et al., 2006; Bahlmann et al., 2009). The temporal dynamics of sentence processing are, however, not revealed by the fMRI technique.

The frequency tagging paradigm that was recently applied to sentence processing (Ding et al., 2016; for more details, see Manuscripts 1 and 2) both diverges from the violation paradigm and provides important information about the temporal dynamics of sentence processing. In this study, the changes in low-frequency neural activity evident from the analysis of spectral neural power (i.e., frequency domain analysis) and modulated by sentence structure can in principle be interpreted in a mechanistic neural entrainment paradigm. While fMRI research 'maps' behavioural phenomena to specific brain areas and EEG/MEG studies often similarly 'map' ERP components to cognitive and behavioural effects (Poeppel, 2012), neural entrainment presents a biologically-plausible explanatory hypothesis about the link between neural data and behaviour in language processing (Luo & Poeppel, 2007; Giraud & Poeppel, 2012; but see Rimmele, Morillon, Poeppel, & Arnal, 2018). Entrainment to acoustic edges in the speech stream (e.g., Gross et al., 2013), for example, has been previously shown to improve speech intelligibility (Riecke et al., 2018) and is considered to reflect attentional selection facilitating specific speech feature processing (for a recent review, see Obleser & Kayser, 2019).

In other words, the frequency tagging approach presents a promising experimental paradigm for the neuroscience of sentence processing. However, several methodological considerations should be addressed before the validity of the claims made by existing frequency tagging studies as those by Ding and co-authors (2016, 2017) is established. Importantly, as outlined above, several layers of sentential information are processed in real time by the language user: these involve syntax, prosody, and semantics. A semantic processing account has been previously proposed for the interpretation of Ding and coauthors' results (2016; 2017; Frank & Yang, 2018). In the current thesis, I will outline a prosodic account of their data. Further, an important claim of the study concerns the 'concurrent' processing of constituents at different levels of syntactic hierarchy. However, the hierarchical nature of processing cannot be easily established using simple constructions (see e.g., Rohrmeier et al., 2012; Udden et al., 2019). Moreover, as evident from frequency tagging research in other fields of neuroscience (e.g., Norcia, Appelbaum, Alex, Cottereau, & Rossion, 2015), spectral power peaks at multiples of the fundamental frequency (for example, the sentence rate; i.e., harmonics) are often seen in the spectrum of neural activity. This makes it difficult to interpret the phrase-level effects in sentences tested by Ding and coauthors (2016, 2017), undermining their claim that they uncovered the neural mechanisms responsible for hierarchical phrase structure processing. These and other methodological concerns are discussed in the present thesis.

THE STRUCTURE OF THE THESIS

The main parts of the current dissertation consist of two manuscripts. In Manuscript 1, the role of prosody (overt and covert) in the elicitation of cortical responses at frequencies of syntactic constituents is addressed across three experiments. It is demonstrated that both syntax and prosody modulate the spectral power peaks at sentence and phrase frequencies, but given empirical evidence available to date, it is more likely that the neural activity tracks (covert) prosodic rather than syntactic information per se. Manuscript 2 is dedicated to a methodological exploration of the frequency tagging technique as used in sentence processing research. Multiple data simulations demonstrate the caveats of the technique, and guidelines for its efficient use in language research are provided. Note that the target audience of Manuscript 1 is the community of researchers studying temporal dynamics of language processing using neurophysiological techniques, familiar with the frequency tagging technique. Manuscript 2 targets researchers in the area of neuro- and psycholinguistic research as well as, more broadly, cognitive science interested in either (i) using the frequency tagging technique or (ii) understanding the range of possible interpretations for outcomes of existing frequency tagging research. Due to this difference in the target audience, an in-depth characterization of the frequency tagging technique is not included in Manuscript 1 and is part of Manuscript 2 only. The order of the two manuscripts corresponds to the order in which the projects were conceived and realized. Finally, the General Discussion presents a summary of the theoretical and methodological implications of the current thesis and outlines possible directions for future research.

MANUSCRIPT 1

Overt and covert (imagined) prosody affects neurophysiological responses at frequencies of syntactic constituents

Anastasia Glushko[1], David Poeppel[34], and Karsten Steinhauer[12]

Affiliations:

[1]Centre for Research on Brain, Language and Music (Montreal, Canada)

[2]School of Communication Sciences and Disorders, McGill University (Montreal, Canada)

[3]Max Planck Institute for Empirical Aesthetics (Frankfurt, Germany)

[4]Department of Psychology, New York University (New York City, NY, USA)

ABSTRACT

Previous research suggested that slow cortical activity tracks sentence structure (Ding, Melloni, Zhang, Tian, & Poeppel, 2016). This was taken as evidence for hierarchical syntactic structure building during online sentence processing. As the authors argued that prosody could be ruled out as a contributing factor, here we investigated the role of both overt and covert, implicit prosody on these effects. Over three experiments, participants listened to series of sentences while their electroencephalography (EEG) was recorded. In the No Prosody experiment, prosodic cues in the sentence materials were neutralized and participants' task was to detect semantically implausible sentences. In the Prosody experiment, participants first listened to a series of sentences with overt prosodic phrase boundaries that either aligned or misaligned with the syntactic phrasing in the sentences (initial overt prosody trials). Immediately after each overt prosody trial, participants listened to a second series of sentences (covert prosody trial) with neutralized overt prosodic cues but were asked to imagine the prosodic contour present in the previous, overt prosody trial. In addition, participants' task was, again, to detect semantic outliers. Task effects were additionally investigated in the No Semantic Task experiment that did not include an outlier detection task. We found, first, that an EEG spectral power peak was elicited in one of the conditions of the No Prosody experiment while no syntactic constituents appeared at the corresponding frequency. We interpret this effect as elicited by covert prosodic grouping. Second, overt and covert prosody affected the EEG responses at the frequencies of syntactic constituents (sentences and phrases) in a manner dependent on whether prosodic and syntactic phrasing aligned or not. Effects of sentence structure on EEG were modulated by the task. We conclude that both overt and covert, implicit prosody can affect neural responses at sentence constituent frequencies whereas the role of syntax in their elicitation is yet to be confirmed by future research.

INTRODUCTION

Human language comprehension involves a variety of cognitive mechanisms for processing multiple types of information, from auditory perception to integration of words' semantic content with the grammatical structure of sentences. While many of these processing mechanisms have been shown to have parallels across the animal kingdom (ten Cate, 2017), syntactic phrase structure processing has been suggested as the essential element of human language that dissociates it from communication in other animals (Fitch & Hauser, 2004; Berwick, Friederici, Chomsky, & Bolhuis, 2013). Studying the neurocognitive mechanisms required for syntactic phrase structure processing while being essential for understanding human language abilities, is, however, often obstructed by the co-expression of phrasing in natural language through both syntax and prosody (e.g., Klatt, 1975; Nicol, 1996). Though they closely interact in natural language processing, prosody and syntax are distinct, in terms of their phenomenology and processing. Prosody refers to suprasegmental phonological features of language, including "tonal structure, pitch accents, phonological boundaries, duration and intensity" (Fery, 2017, p. 6). Specific prosodic features such as stressed or lengthened syllables and pitch contours are used to split linguistic input into prosodic constituents, such as phonological words and phrases at various levels of prosodic hierarchy (Selkirk, 1981). Importantly, however, as we will discuss in detail below, syntactic constituents at different hierarchical levels can be prosodic sister constituents at the same level (e.g., Frazier, Clifton, & Carlson, 2004).

Research on syntactic phrase structure processing utilized existing linguistic theories of language organization contrasting their predictions with the data from sentence processing studies. One very influential theoretical account that received some experimental support is the generative grammar model by Noam Chomsky (1959) who proposed that language consists of hierarchically organized minimal units (such as morphemes) merged into larger constituents (words) merged into even larger groups (phrases, and further, sentences). This perspective prompted the development of various influential psycholinguistic processing models (Fodor, 1975; Frazier & Fodor, 1978; Frazier, 1987) and experimental paradigms to explore the psychological reality of processing syntactic hierarchies while accounting for other cues (e.g., prosody, semantics, discourse). It has been shown, for instance, that listeners more accurately identify extraneous clicks if the click is presented closer to syntactic constituent boundaries, even when overt acoustic cues at syntactic boundaries are comparable to those at other positions (Fodor & Bever, 1965; Garrett, Bever, & Fodor, 1966). Eye-tracking studies demonstrated that readers, when faced with initially grammatically ambiguous sentences, make syntactically-motivated phrasing choices independent of how semantically implausible the parsing results would be (Rayner, Carlson, & Frazier, 1983), pointing to a primacy of syntactic over semantic information in real-time processing. Similarly, on-line studies using electroencephalography (EEG) and event-related brain potentials (ERPs) reported brain signatures for syntactic phrase structure processing within the first 300 ms after word onset, and thus apparently earlier than brain responses linked to semantic processing (Neville, Nichol, Bars, Foster, & Garrett, 1991; Friederici, Hahne, & Mecklinger, 1996; Hasting & Kotz, 2008; but see Steinhauer & Drury, 2012). Involvement of a traditionally language-related brain region, BA44 (part of the so-called “Broca’s area”), in phrase structure processing has been demonstrated via contrasting word lists and phrases (Zaccarella & Friederici, 2015) as well as non-hierarchical and recursively organized nonword sequences in the absence of overt prosodic cues (Friederici et al., 2006) using fMRI, again supporting the idea of hierarchical syntax processing.

These and similar studies providing apparent empirical support for Chomsky’s generative framework have investigated sentence processing indirectly by applying elaborate task demands (e.g., detection of click locations; Garrett et al., 1966) and relying on participants’ processing of linguistic anomalies (e.g., Friederici et al., 1996; Neville, Nichol, Bars, Foster, & Garrett, 1991). It can be argued, however, that the neurocognitive reality corresponding to online processing of syntactic phrases (and the boundaries between them) in grammatically *correct* sentences has rarely been studied directly. A recent magnetoencephalographic (MEG) research project was the first attempt to investigate syntactic phrase structure processing directly from listeners' comprehension of grammatical sentences using a "frequency tagging" approach (Ding, Melloni, Zhang, Tian, & Poeppel, 2016). The implication that it has provided ultimate neuroscientific evidence for Chomsky's hierarchical syntax theory through the finding of "concurrent neural tracking of hierarchical linguistic structures" (Ding et al., 2016, p.5) received immediate attention. In that study, participants listened to a series of 4-syllable sentences in Mandarin Chinese, all of which had their largest syntactic boundary ('|') after the second syllable (henceforth, 2+2 Syntax), separating a subject noun phrase (NP) from the verb phrase (VP) (illustrated by the English example *New plans | give hope*). The main analysis of the neural data was conducted in the frequency (i.e., not in the time) domain. As all syllables were exactly 250 ms long (4 syllables per second), their processing was reflected by a 4 Hz peak in the listeners' MEG signal. This effect can be, in principle, fully accounted by cortical tracking of speech envelope (e.g., Luo & Poeppel, 2007). Along with this 'bottom-up', stimulus-driven MEG power peak, however, native Mandarin Chinese listeners additionally elicited a 1-Hz peak reflecting sentence boundaries, as well as a 2 Hz peak that was taken to reflect the online processing of the syntactic boundary in mid-sentence position. These effects, in contrast to the 4 Hz peak, were interpreted as representing 'top-down' processing, or as being driven by mental representations. In line with this interpretation, no 2 Hz peak was found for the language stimuli in which the major syntactic boundary occurred after the first of the four words (henceforth, 1+3 Syntax; ibid). Similarly, English listeners who could identify only syllable boundaries but did not understand the Mandarin Chinese sentences (or their syntactic structure) only elicited 4-Hz peaks for both sentence types. Thus, the use of isochronous (i.e., equal in length and built of equally long phrases and words) sentence materials with uniform syntactic phrase structure (2-word NP + 2-word VP) allowed for 'tagging' the frequency of syntactic constituents in the neural recordings in real time, thereby uncovering the neural signatures of online syntactic phrasing. Importantly, the overt prosodic phrase boundary markers in the speech signal were 'neutralized': the pitch was flattened, all words had the same length independent of their position within the sentences, and there were no prosodic breaks at phrase or sentence boundaries. This served to rule out any (trivial) bottom-up phrase chunking based on acoustic cues.

Initially rather agnostic regarding the mechanistic nature of the frequency tagging effects in sentence processing (Ding et al., 2016), the authors went from qualifying them as neural activity at low frequency tracking hierarchical linguistic structure to equating them to neural entrainment (Ding et al., 2017; for reviews and definitions, see Giraud & Poeppel, 2012; Zoefel, ten Oever, & Sack, 2018; Obleser & Kayser, 2019). In addition to emphasizing the theoretical importance of their conclusions possibly pointing to endogenous neural oscillations aligning with, or entraining to, linguistic structure, Ding and coauthors (2017) noted that their experimental design was promising for studying sentence processing in children and clinical populations.

These results have since been replicated (Sheng et al., 2018; with EEG - Ding et al., 2017). Ding and colleagues' (2016) additionally ran an experiment with structured sequences of nonwords presented to participants before and after a training phrase during which participants learned the rules underlying the nonword sequences and were given a few examples of real sentences whose syntactic structure was said to be similar to that of the nonword sequences. While before the training phase the MEG power peak only arose at the frequency of a syllable, after learning, a peak at sentence frequency appeared, which was claimed to reflect the successful application of new grammar rules. Moreover, in an additional intracranial EEG experiment, Ding and colleagues (2016) extended their MEG results by showing that neural responses at the sentence and syntactic phrase frequencies (1 Hz vs 2 Hz) were spatially dissociable (i.e., seen at distinct electrodes placed at different locations on the cerebral cortex). Together, these results were taken as evidence for the “undeniable existence of hierarchical structure building operations in language comprehension” (Ding et al., 2016; p. 5), thus reflecting top-down neural mechanisms of syntactic phrase structure building.

From their careful control of stimuli and neutralized prosody, it is evident that the mechanisms driving the phrase and sentence-level effects in the studies by Ding and colleagues (2016; 2017) did not merely reflect stimulus-driven, ‘bottom-up’ processing of acoustic information present in the speech signal. However, ‘top-down’ sentence processing mechanisms (those driven by existing mental representations of the listeners) can still be very heterogeneous and include semantics, lexical transitional probabilities (see e.g., Frank & Christiansen, 2018; Frank & Yang, 2018), and prosody. Ding and colleagues emphasized that by avoiding *overt* prosodic cues in their spoken sentences, they had ruled out the possibility that their results were affected by prosodic processing. Yet prosodic processing is not limited to the detection of changes in pitch, sound intensity, and duration in the sound signal, but also heavily relies on previous experience and expectations (Itzhak, Pauker, Drury, Baum, & Steinhauer, 2010). These top-down aspects of prosodic processing are evident from a phenomenon known as implicit, or *covert*, prosody (Fodor, 2002).

Covert prosody denotes existing mental prosodic representations that are projected onto current linguistic input (Fodor, 1998; 2002). The phenomenon is commonly observed during silent reading where it is often discussed as a type of phonological recoding (Leinenger, 2014). Behavioural effects of these representations on language processing are seen from ample psycholinguistic data showing that covert prosodic events affect processing demands, and thus reading times. For instance, adults' silent reading times are increased when (1) the reanalysis of an ambiguous garden-path sentence requires implicit prosodic, and not only syntactic, revisions (Bader, 1998), and when (2) the number of stressed syllables in a word increases (Ashby & Clifton, 2005; for review of similar research, see Breen, 2014).

Pertinent to the potential effects of covert prosody on the MEG and EEG power peaks at frequencies of syntactic constituents reported in the studies by Ding and colleagues (2016, 2017) are the neurophysiological studies of covert prosody. A robust brain response of prosodic phrasing is the Closure Positive Shift (CPS; Steinhauer, Alter, & Friederici, 1999; Steinhauer & Friederici, 2001), a positive-going ERP waveform elicited by both overt and covert prosodic phrase boundaries. The prosodic nature of this neurophysiological response was confirmed through studies on delexicalized (e.g., hummed) sentence materials and music, in which the CPS was also elicited at prosodic phrase boundary positions (Steinhauer, 2003; Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005; Glushko, Steinhauer, DePriest, & Koelsch, 2016). Originally discovered in auditory sentence presentation with overt prosodic cues marking phrase boundaries (Steinhauer et al., 1999), the CPS has later been reported for comma positions in silent reading (Steinhauer & Friederici, 2001) and for predictable boundary positions in the absence of punctuation marks or overt prosodic cues in both written (Hwang & Steinhauer, 2011) and spoken language (Itzhak et al., 2010). The top-down nature of the covert prosody CPS was also illustrated by Steinhauer and Friederici's (2001) study, in which participants silently read correct sentences with ambiguous syntactic phrasing. Prior to reading the sentence, they were presented with a de-lexicalized example of a sentence with overt prosodic cues placed at one of the potential phrase boundary positions. Participants were instructed to apply this prosodic 'template' while silently reading the subsequent (unpunctuated) sentence. A CPS was shown to be evoked in these sentences at the specific boundary positions 'primed' by the prosodic template, that is, in a strictly top-down way. In another experiment, the same authors showed that when commas are placed at phrase boundary positions, the amplitude of the CPS elicited during reading is modulated by knowledge of language-specific comma rules (again, emphasizing its top-down nature; Steinhauer & Friederici, 2001; Steinhauer, 2003).

It has been shown experimentally that covert prosodic representations mediate sentence comprehension. In syntactically ambiguous sentences, covert prosodic processing can drive initial prosodic parsing (via prosodic principles such as, for instance, the optimal length of the resulting processing chunk; Frazier, 1998; Hirose, 2003; Hwang & Schafer, 2009) that in turn can affect syntactic parsing decisions. That is, and this is crucial, covert prosodic chunking may diverge from the chunking suggested by syntactic phrase hierarchy, and may result in erroneous initial syntactic parsing decisions (Watson & Gibson, 2004; Clifton, Carlson, & Frazier, 2006; for review, see Frazier, Clifton, & Carlson, 2006). While syntactic phrasing does affect the processing of prosodic boundaries (Buxó-Lugo & Watson, 2016), constituent length, semantic coherence, and information structure cues can lead to the placement of prosodic breaks at positions where major syntactic breaks are absent (Frazier, Clifton, & Carlson, 2004; see also discussion of such instances in Samek-Lodovici, 2005; Wagner & Watson, 2010; Shattuck-Hufnagel & Turk, 1996). For instance, Hwang and Steinhauer (2011) demonstrated that the presence of an early CPS in silent reading depended on the length (number of words) of the first noun phrase (i.e., it was governed by non-syntactic principles) and this, in turn, affected important syntactic parsing decisions later in the sentence. Overall, it is fair to assume that covert prosodic phrasing patterns will reflect the high variability of prosodic realizations seen in speech production with many prosodic boundaries being only optional (Allbritton, McKoon, & Ratcliff, 1996; Schafer, Speer, Warren, & White, 2000) and being inserted, for instance, driven by individual working memory capabilities (Swets, Desmet, Hambrick, & Ferreira, 2007). This absence of a one-to-one mapping of syntactic and covert prosodic phrasing (e.g., Frazier et al. 2006; Izhak et al., 2010, Hwang and Steinhauer, 2011) creates a potential problem for psycho- and neurolinguistic research aiming to eliminate prosodic processing demands.

In studying sentence processing, researchers "cannot escape prosody" (Fodor, 2002), and this extends to the presentation of auditory stimuli even when overt prosodic cues have been neutralized (Ding et al., 2016, 2017; Sheng et al., 2018). Moreover, the use of block designs in frequency tagging studies further facilitates the generation of top-down implicit prosodic contours, as the materials are highly predictable. The first sentence in a block carries structural (not restricted to syntactic) cues for the mentalization of an implicit prosodic contour. This contour can be applied to the following sentences in the block. In their '2+2 Syntax' sentences consisting of two phrases (a two-word NP and a two-word VP; e.g., *"New plans | give hope"*), the phrase-rate response could have been fully or partially accounted for by a covert prosodic phrase boundary. The generation of this boundary can be driven by either syntactic or prosodic constraints (i.e., length and balance constraints; Fodor, 1998), or by both.

Ding and colleagues' claim about the purely syntactic nature of their effects is partly based on the finding that increases of either exclusively phrase-rate or sentence-rate power were found at distinct intracranial electrodes (Ding et al., 2016). It is, however, possible that this dissociation of power increases at the phrase and the sentence frequencies across electrodes was of a non-syntactic nature as well. Prosodic structure has been previously described in hierarchical terms (Selkirk, 1981), with the highest level being the sentence, or more precisely, an utterance. Moreover, prosodic phrase processing depends on the relative strength of the prosodic boundary in the context of other boundaries within the same utterance (Frazier et al., 2006; Snedeker & Casserly, 2010; Steinhauer, 2011). It is conceivable that in the case of the 2+2 Syntax sentences, the differences in relative prosodic phrase boundary strength drove different levels and types of integration strategies (not only syntactic, but perhaps also semantic ones; e.g., Schafer, 1997; Osterhout, 1997), causing spatially non-overlapping phrase and sentence effects in frequency tagging studies (Ding et al., 2016).

The contribution of covert prosody to phrase-level brain responses and, by extension, the degree to which sentence- and phrase-level effects are driven by overlapping mechanisms of syntax and prosody processing, can be addressed via manipulation of the optional prosodic phrase boundary in 2+2 Syntax sentences. In the current study, we presented participants with different types of prosodic templates (similar to Steinhauer & Friederici, 2001) and asked them to mentally impose them onto the 2+2 Syntax sentences with neutralized prosody. These prosodic templates had boundary cues aligned with either the position of the syntactic phrase boundary (at the second word in a sentence; see W24 and W2 contours in Methods) or the first word of the sentence, driving the prosodic boundary away from the syntactic phrase boundary with contrastive focus (see W1 contour in Methods). We contrasted each of these settings with a baseline condition in which no prosodic template was applied to the stimuli and overt prosodic cues were neutralized; i.e., a default covert prosodic contour might have been applied by listeners, like in Ding and colleagues' study (see No Prosody in Methods). We predicted that applying a prosodic template aligned with syntactic phrasing would enhance the phrase-level EEG responses and facilitate syntactic phrase processing. In the case of the misaligned prosodic and syntactic phrasing, we predicted a reduction of phrase-level responses in the 2+2 Syntax sentences, with this effect amounting to the contribution of covert prosodic cues in the study by Ding and colleagues (2016; 2017). In other words, we predicted to not only see the isolated effects of prosodic phrasing (i.e., reflecting mere processing of a specific prosodic pattern), but to find an interaction between syntactic and prosodic phrasing.

Another crucial manipulation in the study by Ding and colleagues (2016) was the comparison of varying syntactic structures, including contrasting 2+2 Syntax sentences, on the one hand, with verb phrases in which the first word is a verb and the next three syllables represent its direct object, i.e., 1+3 Syntax, on the other hand (e.g., *fry to-ma-to*, *try oo-long tea*). This contrast, however, is characterized by both syntactic and covert prosodic differences. Most importantly in our current context, the placement of a prosodic boundary after two syllables is much less likely for 1+3 Syntax than for 2+2 Syntax constructions because of the highly collocated three-syllable noun phrase, representing a single semantic concept (the importance of semantic plausibility of prosodic chunks is evident from e.g., Frazier et al., 2006 and Itzhak et al., 2010). In other words, the Ding et al. (2016) study strongly confounded syntactic and prosodic structure, and the effects attributed by the authors to syntactic structure building could instead be argued to reflect prosodic processing. To address this issue, we used 1+3 Syntax constructions whose prosodic phrasing options were more comparable with the 2+2 Syntax ones: those were sentences such as *Hein is going to Bonn* (original: *Hein fährt nach Bonn*, lit.: *Hein goes to Bonn*) or *Lars likes the picture* (original: *Lars mag das Bild*), in which the major *syntactic* boundary between subject NP and VP appears after the first word, while an optional covert *prosodic* boundary can be placed after the second word in a sentence (similar to the 2+2 Syntax, due to prosodic balance constraints that favour 'equal sisters' of similar length; Fodor, 1998). These sentences were presented to participants in the same prosodic conditions as the 2+2 Syntax sentences.

Using these and other manipulations, the current study aimed to investigate the role of the often-neglected prosodic processing mechanisms in eliciting neural responses to sentence structure. We hypothesized that both overt and covert prosodic processing would affect EEG responses at the frequencies of syntactic constituents. Specifically, we predicted that when covert prosodic structure of the sentence was congruent with the syntactic phrasing, the responses at the frequencies of the 'prosody-supported' syntactic constituents would be increased compared to the condition with neutral, default prosody (similar to e.g., the behavioural effects in Hilton & Goldwater, 2019). In turn, we expected the incongruency between prosodic and syntactic structures of the sentences to cause decreased EEG responses at the frequencies of the corresponding sentence constituents. In addition, we hoped to clarify if covert prosodic processing may have contributed to Ding and coauthors' (2016, 2017) findings that were claimed to be syntactic in nature. To test these various hypotheses, we collected within-subject data from three experiments closely following the design of existing frequency tagging studies (Ding et al., 2016, 2017) to assess the effects of potential overt and covert prosodic boundaries on the EEG power at phrase boundary frequencies. The roles of syntactic and prosodic processing were first examined in sentences with neutralized prosody comparing sentences with comparable opportunities for insertion of optional prosodic breaks (expanding on the studies of Ding and colleagues, 2016, 2017; see No Prosody Experiment in Methods). In the second experiment (Prosody experiment), we presented listeners with sentences with overt prosodic contours that were aligned or misaligned with the syntactic phrasing. After listening to a trial of sentences with a specific overt prosodic contour, participants were asked to impose this contour onto the covert prosody trial, i.e., a trial of sentences with neutralized prosodic cues, similar to Steinhauer and Friederici's (2001) approach. This allowed us to investigate the impact of both overt and instructed, covert prosody on sentence processing as measured by frequency tagging. Both in the experiment with and the one without prosody, we asked participants to perform a task that required them to understand the sentences and to identify implausible outliers. Finally, in the last (No Semantic Task) experiment we tested potential task effects on the elicited EEG effects in a subset of participants (see Methods).

METHODS

PARTICIPANTS

We collected data from 37 adult German native speakers (age range: 19-45 years, mean age = 29; 19 women, 18 men), of whom the data from 25 participants (age range: 19-45 years, mean age = 28; 14 females, 11 males) were included in the analysis unless otherwise specified. The participants whose data were included in the main data analysis for the paper were the only ones exposed to all experimental conditions due to the significant length of the study (in all 37 participants, experiment was terminated after 5.5 hours in the lab, irrespective of whether a participant had been presented with all conditions). All participants were recruited and tested at McGill University in Montreal, most of them visiting Canada for work-and-travel purposes. They had acquired German language from birth and considered it their dominant language. The inclusion criteria for the study were the absence of neurologic or psychiatric disorders and hearing impairments, as well as normal or corrected vision. Participants provided written informed consent and received monetary compensation ($20/hour) for their time.

We assessed handedness using the Edinburgh Handedness Inventory ensuring all participants were right-handed (Oldfield, 1971). Participants filled out detailed in-house questionnaires about their language background and musical expertise to allow for the exploration of the inter-individual variability in the data. All parts of the study were approved by McGill's Faculty of Medicine *Institutional Review Board* (IRB) prior to data collection.

MATERIALS

Speech synthesis. In order to maximize replicability of our study, the four-word German sentences used in the experiment (in German language) were synthesized word-by-word with a built-in Apple synthesizer (the Anna voice). All words were monosyllabic, and their speech signals were exactly 320 ms long. The pitch of each word (and thus of the entire sentence) was flattened, and the intensity was normalized to 70 dB in Praat (Boersma & Weenink, 2019). The words were concatenated into 80 semantically plausible and 24 semantically implausible 4-

word sentences, which were further concatenated into *trials* each comprising 12 sentences (48 words). The semantically implausible 'outlier' sentences were arranged by combining the beginning and the end of two different semantically plausible sentences (e.g., *Das Zelt lacht lahm*; lit.: 'The tent laughs lamely') and were used as targets in the outlier detection task (see below). Each sentence was repeated 8 to 9 times within the same experimental block but never within the same trial. Each trial lasted for 15.36 seconds (12 sentences x 4 words x 320 ms; no pauses were introduced between words, phrases, and sentences), identical to the trials in Ding and colleagues (2017). For the two types of syntactic structure and for each type of prosodic contour used in the study (see below), we created 22 trials without any implausible outliers. In the experimental conditions that employed the outlier detection task, we added 8 additional trials with one outlier sentence each, but these were not subjected to subsequent data analysis.

Syntactic structure of the sentences. Sentences followed one of the two types of syntactic structures. In the case of the *2+2 Syntax* (40 sentences), sentences consisted of two syntactic phrases of equal length. The first phrase was a noun phrase (NP), consisting of a determiner and a noun, while the second one was a verb phrase (VP), most frequently comprised of a verb and an adverb (e.g., *Der Tisch steht da*; lit.: 'The table stands there', or 'The table is over there'). In rare cases, the verb phrase (VP) consisted of a particle verb with the corresponding particle replacing the adverb (e.g., *Mein Boot kippt um*; English: 'My boat tips over'). In the *1+3 Syntax* (40 sentences), the first phrase in each sentence included a one-word NP (i.e., a name), and the second phrase was represented by a 3-word VP (typically, a verb and its complement, e.g., a verb + a determiner/preposition + a noun; e.g., *Lars mag das Bild*; English: 'Lars likes the picture'; see Supplementary Materials *A* for the full list of sentences and additional details on their characteristics). The two types of syntactic structures were compared. Given that the current study used EEG (and not MEG, like the only frequency tagging study using 1+3 Syntax constructions with phrases of non-equal length), we ran a control experiment in a separate group of participants to establish that the EEG effects of the 1+3 grouping are analogous to the ones reported in Ding and coauthors' (2016) study. As described in detail in Supplementary Materials *B*, our results confirmed that 1+3 rhythm elicits an EEG spectrum similar to the MEG spectrum reported by Ding and colleagues (2016) and can, therefore, be contrasted with the 2+2 Syntax sentences in our main study. For both types of syntactic structure, the (acoustically unmarked) sentence boundary appeared once every 1.28 seconds (after four words) at a frequency of 1/1.28 (0.78) Hz (*sentence frequency*), and single words appeared every 320 ms (i.e., at a *word* frequency of 3.125 Hz). However, only in the 2+2 Syntax sentences (where the phrase boundary between the NP and the VP occurred after two words), syntactic phrases were isochronous and appeared at a constant frequency of 1.56 Hz (every 640 ms), that is, at *½ sentence frequency*.

Prosodic manipulations of the sentences. The sentences concatenated from words with neutralized prosody as described above constituted the *No Prosody* condition (henceforth, *NoP;* used in the No Prosody experiment) that was to be contrasted with the data from the *Overt* and *Covert Prosody* conditions (henceforth, *OvP* and *CovP* respectively; used in the Prosody experiment). As the general idea of our prosodic manipulation was to create prosodic patterns that would selectively support one syntactic structure (e.g., 2+2) while conflicting with the other one (e.g., 1+3), the most straightforward acoustic manipulation would have been to either insert pauses at a boundary position or increase the duration of pre-boundary syllables. This kind of prosodic manipulation changes the duration of pre-boundary words and has not only been found to be the most reliable boundary marker in natural speech, but has also been successfully used in previous studies to create cooperative and conflicting syntax-prosody pairings (e.g., Kjelgaard & Speer, 1999), including in EEG studies (Steinhauer et al., 1999; Bögels, Schriefers, Vonk, Chwilla & Kerkhofs, 2009; Pauker, Itzhak, Baum, & Steinhauer, 2011). However, in a frequency tagging study that crucially depends on the identical duration of all monosyllabic words (see above and Ding et al., 2016), durational manipulations are not an option. Instead, we manipulated pitch and intensity, two prosodic dimensions that also contribute to prosodic boundary marking (Streeter, 1978; Beckman, 1996; Männel, Schipke, & Friederici, 2013; Roll, Horne, & Lindgren, 2010). To this end, we synthesized artificial pitch and sound intensity contours in Matlab R2019a (Mathworks, 2011) by adding a number of sine waves of specific frequencies into a complex wave while controlling for the presence or absence of modulations at critical frequencies. Prototypes of these contours were initially derived from original speech samples produced by a native speaker who was instructed to emphasize different parts of the sentences. The resulting artificial prosodic contours were then imposed onto the sentences with neutralized prosody (No Prosody) in Praat (Boersma & Weenink, 2019), thereby creating the Overt Prosody condition (see Figure 1).

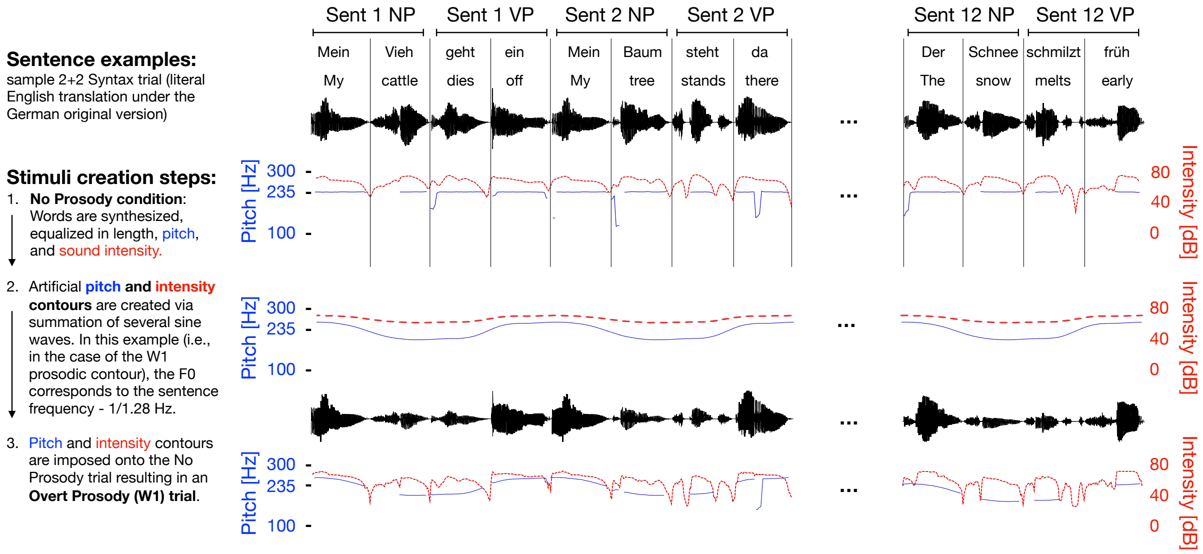


Figure 1. Stimuli development scheme. Single words were synthesized and concatenated into trials (12 sentences each). In the No Prosody experiment (1), prosodic cues were neutralized: i.e., there were no pauses between words within trials, all words were 320 ms long, pitch was flattened, and sound intensity was constant across words. Artificial prosodic contours (2) were imposed on the trials with neutralized prosody to create stimuli for the overt prosody trials for the Prosody and the No Semantic Task experiments. Pitch contour is depicted in blue (note that infrequent sudden drops of pitch values typically reflect unavailability of pitch information due to unvoiced phonemes), and sound intensity is represented by red lines. (3). Audio files for all stimuli are available upon request.

Apart from the No Prosody conditions with neutralized prosody (Figure 2a), three different prosodic contours varying in the position and frequency of prosodic modulations were created to differentially support – or conflict with – the two syntactic structures in the study (Figure 2b-d). The prosodic features manipulated were sound intensity and pitch because duration increases or pauses, while common at phrase boundaries in natural speech, are incompatible with the frequency tagging experimental design. On the other hand, lengthening the pre-boundary syllables and inserting prosodic breaks at boundary positions if implemented, would have emphasized that the prosody marks prosodic phrase chunking and not the informational prominence of a specific word or constituent. As we discuss below, information structure indeed most likely contributed to the results in some of our experimental conditions.

In the first prosodic contour (Figure 2d), the maxima of sound intensity and pitch were placed on Words 2 and 4 (W24 contour). That is, the fluctuations of pitch and intensity appeared *not* at the sentence frequency, but at the ½ sentence frequency (1.56 Hz).

In the two remaining prosodic contours, which we will refer to as 'Word 1' (*W1*) and 'Word 2' (*W2*), the maximum pitch and sound intensity values were reached at either the first word in the sentence (W1) or the second one (W2). Thus, the prosodic cues were modulated at the sentence frequency (0.78 Hz). Additional harmonics descending in amplitude of the 0.78 Hz frequency were added to increase naturalness of the speech signal, but no pitch or sound intensity changes were introduced at the ½ sentence frequency (1.56 Hz; see Figure 2b-c). The 0.78 Hz sine wave with added harmonics (avoiding the 1.56 Hz frequency) were then shifted in time resulting in the sound intensity and pitch maxima being placed on either the first (W1) or the second (W2) word of each sentence.

We imposed each of the three prosodic contours onto all experimental sentences of both 2+2 and 1+3 Syntax structures, thereby creating different degrees of congruency between the syntactic and the prosodic phrasing of the sentences. We considered the W24 prosodic contour to be congruent with the syntactic phrasing of the 2+2 Syntax sentences: the prosodic boundary cues were placed at words wrapping up the syntactic phrases. Our relatively simple (but universally applicable) pitch and intensity manipulation can be viewed as somewhat ambiguous between (a) boundary tones marking phrases and (b) pitch accents emphasizing individual words, similar to a narrow focus (Beckman, 1996). However, as a focus-pattern emphasizing Words 2 and 4 is linguistically implausible in our sentences (both in 2+2 and in 1+3 structures), its interpretation in favor of boundaries separating 2-word phrases is more likely. We predicted an enhancement of ½ sentence rate EEG responses in the case of syntax-prosody congruency (2+2 Syntax). We expected this effect to be stronger than any analogous effect in the 1+3 Syntax sentences. This is because the 1+3 Syntax sentences with the W24 prosodic contour present the case of the syntax-prosody incongruency: prosodic changes are not placed at the phrase-final position (i.e., the second word in 1+3 Syntax sentences does not wrap up a syntactic phrase) and syntactic phrases are not repeated at the frequency of ½ sentence, at which the prosodic changes occur. In this case, we expected to see reduced brain activity at the frequency of the sentence due to participants hindered ability to syntactically chunk sentences. That is, we predicted that prosodic and syntactic effects would be non-additive and that prosodic and syntactic analyses would interact during online sentence processing.

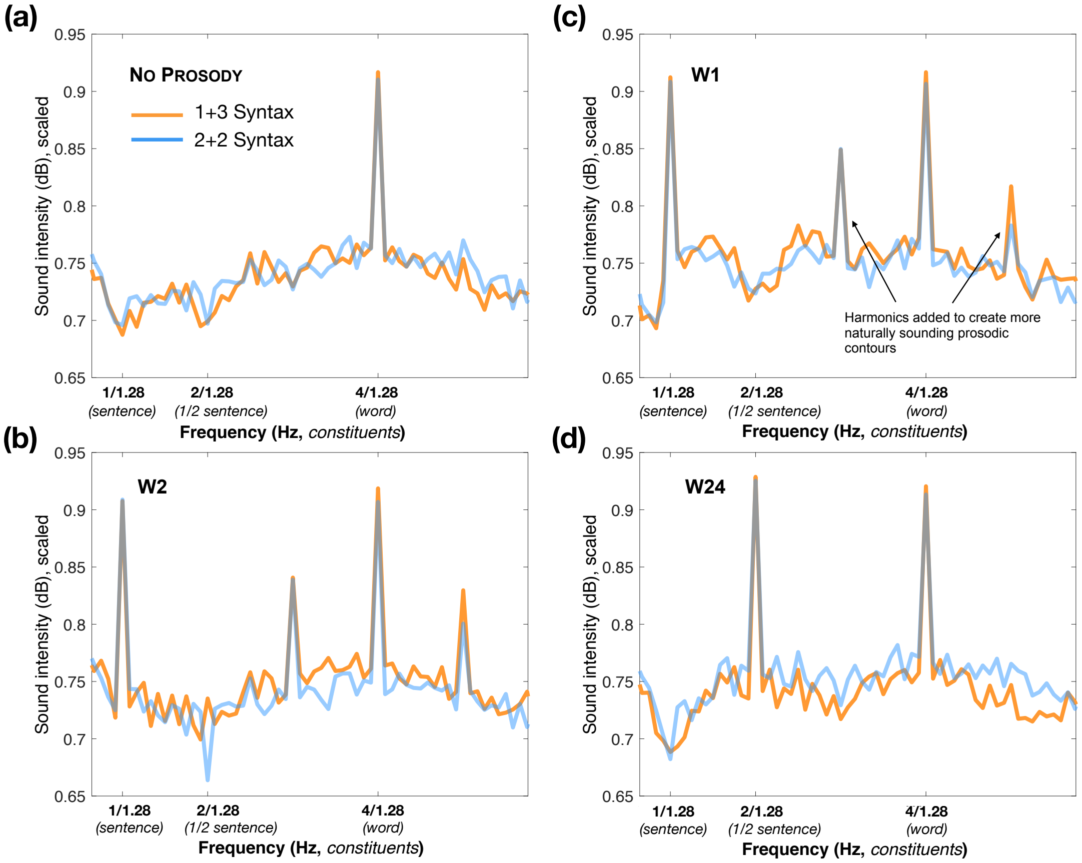


Figure 2. Spectrum of stimulus sound intensity for the various prosodic contours. (a) In the No Prosody condition, systematic sound intensity fluctuations reflect its rise within each word in a sentence and its fall at word boundaries, resulting in a single peak at the word frequency (4/1.28, or 3.125 Hz) in the sound intensity spectrum. (b+c) In both the W1 and W2 prosodic contours, sound intensity is modulated at the sentence frequency, resulting in an additional intensity peak at 1/1.28 (0.78) Hz. The other two (smaller) peaks reflect harmonics introduced to increase the naturalness of the stimuli. Note that W1 and W2 have virtually the same spectrum, even though the positions of sound intensity and pitch maxima are on different words. (d) In contrast to W1 and W2, the W24 contour displays a ½ sentence rate intensity modulation in addition to the word/syllable frequency peak, but no peaks at other frequencies (i.e., no sound intensity increase was present at the sentence frequency). Note that the pitch contours were created using identical fundamental frequency and harmonics as the sound intensity contours, meaning their spectra are qualitatively identical. An example of the pitch contours used can be seen in Figure 1, and the pitch spectra per condition are represented in Supplementary Materials *D*.

Note that the W1 and W2 prosodic manipulations, while being aligned or misaligned with the two types of syntactic structures used in our study, modulated the information structure of our sentences by placing a narrow focus on either the first or the second word in each of them. Information structure refers to the way information is distributed within the sentence: e.g., which information is new, not given to the listener or the reader, which alternatives to the information are implied, and what the main topic of the utterance is (Féry, 2017; Wagner & Watson, 2010). Information structure can be expressed via prosodic prominence (the acoustic correlate of which often is a pitch accent) of the parts of the sentence that are most relevant on a discourse level. Given that among other factors information structure potentially contributed to the results from the W1 and W2 prosodic conditions, we report these data in Supplementary Materials *C*.

We tested all experimental sentences for intelligibility in all Overt Prosody conditions in a total of 20 pilot participants who did not subsequently participate in the EEG recordings, while the No Prosody intelligibility data were collected from every participant at the beginning of the main EEG experiment (see Supplementary Materials D).

In the CovP condition, we used the No Prosody sentences presented in the context of the OvP sentences (see Procedure below). Contrasting the OvP and the CovP conditions allowed us to identify the role of 'overt' prosody (acoustically realized in OvP sentences) and self-generated/superimposed 'covert' prosody (in CovP sentences).

PROCEDURE

Every participant visited the lab for 5-6 hours, including a 3.5-4.0 hour period of EEG recording involving three experiments with multiple breaks throughout the EEG session. During the EEG cap setup, participants filled out behavioural questionnaires. After that, they performed a stimuli familiarization task. The experimenter explained to the participants that the stimuli were synthesized and the speech rate was relatively high, which is why some of the sentences might possibly be difficult to understand right away. To avoid any comprehension problems during the EEG study, participants had an opportunity to read through the full list of sentences (including the semantic outliers) prior to the experiment and then performed a computerized sentence intelligibility task (note that exposing participants to the stimuli prior to the main experiment was done in previous research as well; Jin et al., 2018). In this task, participants listened to every sentence (with a maximum of two replays) and typed in what they heard. Using this task, we were able to verify that all participants understood the vast majority of the sentences: on average, they correctly typed in 100 out of 104 sentences (for individual results, see Supplementary Materials A). Following the behavioural task, the main series of EEG experiments started.

We conducted three experiments (see Figure 3 for experimental flow). Every participant started with the No Prosody experiment that served to establish a baseline for (syntax and, potentially, default covert prosody) processing 1+3 and 2+2 Syntax sentences. Next participants engaged in the No Semantic Task (see below) and the Prosody experiments (whose order was counter-balanced across participants). At the end of the study, we repeated the No Prosody experiment (with a randomized trial order different from the one at the beginning of the experiment) to control for the familiarity with the sentences between the No Prosody and the Prosody experiments and participants' fatigue. Trials in the Prosody experiment were presented in blocks containing trials with the same prosodic and syntactic structure. The order of the blocks was counter-balanced across participants. The data from the two runs of the No Prosody experiment were averaged after ensuring the main patterns were unaffected by whether the data were recorded at the beginning or at the end of the experiment (see Results). This order allowed for minimal influence of the prosodic contours from the Prosody experiment on the processing of sentences in the No Prosody experiment. The only exception to this was the fact that 13 out of 25 participants listened to the sentences with the

W2 prosodic contour prior to listening to the second run of the No Prosody experiment (see Figure 3).

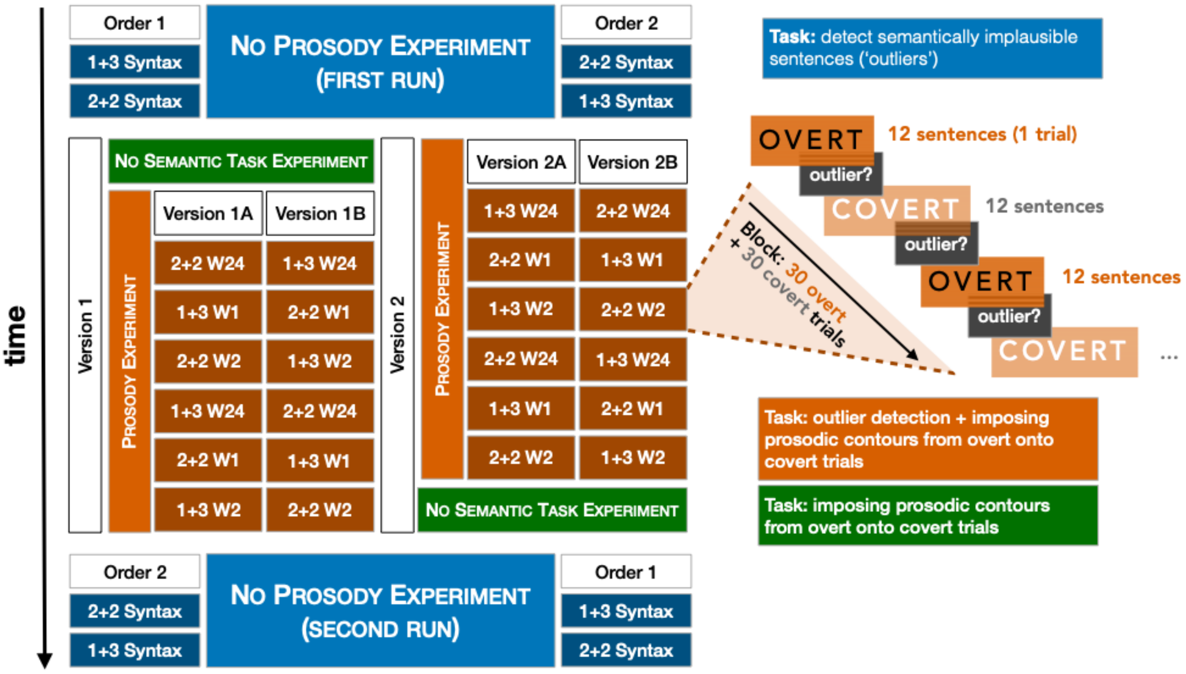


Figure 3. Experimental flow. On the left – the order of the three experiments conducted. In blue – the No Prosody experiments with the corresponding 1+3 and 2+2 Syntax blocks. In dark orange – the Prosody experiments with each block containing trials with sentences with the same syntactic structure and intonation. In green – the No Semantic Task, always represented by one block comprised of sentences with the same syntactic and intonational structure (see main text for details). On the right – the tasks used in the experiments and the scheme of experimental flow within the sample block of the Prosody experiment (2+2 W2 = 2+2 Syntax sentences with W2 prosodic contour). Starting with an overt trial (12 sentences), it continues with an outlier detection prompt, and then with the covert trial, during which participants listen to sentences with neutralized prosody while imposing onto them the intonational contour they attended to in the preceding overt trial.

In the **No Prosody experiment**, participants listened to trials containing the sentences with neutralized prosody. At the end of each trial (i.e., after listening to 12 consecutive sentences), they had to indicate via button press if that trial contained a sentence that did not make sense (an 'outlier') or if all sentences were plausible. Trials consisting of the sentences with the same syntactic structure (1+3 or 2+2 Syntax) formed a block, with the order of blocks being counter-balanced across participants.

In the **Prosody experiment,** participants listened to every syntactic structure and every prosodic contour developed for the study (they were fully crossed). We planned to compare these data to the 'default' syntactic and prosodic processing assessed in the No Prosody experiment. Trials were presented to the participants in pairs: all 12 sentences in the respective first trial would have a specific intonation contour, or 'sentence melody' (e.g., 2+2 Syntax sentences with an *Overt* W24 contour), whereas the sentences in the respective second trial (*Covert Prosody*) would always have neutralized prosody. However, since participants were instructed to silently impose the prosodic contour from the first trial (here: W24) during the second trial (see Figure 5b), the *Covert Prosody* condition would inherit its prosodic characterization form the first trial (here: 2+2 Syntax sentences with *Covert* W24 prosody). Participants were told that because the sentence melody was the same for all sentences within a given first trial (with bottom-up *Overt* prosody), they just had to try staying with the same intonational contour while listening to the second trial (top-down *Covert* Prosody). After each trial (with overt or covert prosody) participants had to indicate by button press if that trial contained a semantic outlier sentence or not.

The structure of the **No Semantic Task experiment** was similar to the Prosody experiment. However, in this experiment, participants were only presented with trials that did not have semantic outliers in them and only went through one of the experimental conditions (pseudorandomly selected for each participant: for example, Syntax 1+3, Prosodic Contour W24; see Figure 3). After the participant listened to the first trial, they would click 'Next', following which the second trial would start playing. This condition was introduced to investigate the effect of the semantic outlier detection on the EEG data in the Prosody experiment. We predicted that the data will be characterized less by the effects of syntax-prosody congruency and more by the independent processing of prosodic changes.

EEG RECORDING AND PROCESSING

EEG data were recorded at a 500 Hz sampling rate using 64 cap-mounted electrodes (extended International 10-20 electrode organization System, Jasper, 1958; Waveguard^{TM} original ANT Neuro EEG system), referenced online to the right mastoid. Matlab (Mathworks, 2019) and EEGLAB (version 14_1_0b; Delorme & Makeig, 2004) were used for EEG data preprocessing (the code is available upon request). Offline, we re-referenced the data to linked mastoids, removed bridged electrodes (the values were interpolated from the neighbouring electrodes after extracting epochs from the data), and performed resampling of the continuous datasets to 250 Hz. We filtered the data separately with a low- (20 Hz cut-off, filter order = 152) and a high-pass (0.2 Hz cut-off, filter-order = 2266) FIR filters using the Kaiser window (β = 5.65326). We removed eye movement artifacts using Independent Component Analysis (ICA; Lee, Girolami, & Sejnowski, 1999) run on the strongly high-pass filtered copies of the original datasets (1 Hz cut-off, filter order = 454). Note that we used these datasets for ICA decomposition only, for which we cut them into dummy epochs that underwent automatic artifact removal. Epochs at any time point exceeding the threshold of |400| μV were removed along with (1) the time points associated with activity on any of the electrodes that deviated from the mean amplitude at that electrode by 2 SDs and (2) the time points during which activity across electrodes deviated from the mean of activity at all electrodes by 6 SDs. We copied the results of the ICA decomposition back onto the continuous data from which the activity of the components accounting for eye movements was then subtracted.

For frequency tagging analysis, the data were cut into 14.08-sec long epochs time-locked to the beginning of the second (rather than the first) sentence in each trial to avoid transient noise associated with the processing of the beginning of a given trial. Epochs containing signal crossing the |50| dB threshold in the 0-5 Hz frequency range were removed. The mean of each epoch was subtracted from each data point in it, after which EEG was averaged across trials resulting in one average for each participant, electrode, and experimental condition. Following Ding and colleagues (2017), three frequency domain measures were calculated: 1) evoked power assessed using the fast Fourier transform (FFT) of time-domain EEG responses averaged across trials (i.e., the FFT of the ERP, representative of the power of brain activity synchronized with the speech input), 2) inter-trial coherence (i.e., the coherence of phase angles across trials, which can change differently from the evoked power), and 3) induced power (i.e., representing the power spectrum of the EEG activity not synchronized to the speech signal and calculated by subtracting the evoked power from the FFT of the time domain of each trial). The data were zero-padded resulting in single bin values calculated at the frequency of 0.065 Hz. Statistical analysis was done only for evoked power and the ITC, whereas induced power was only computed for visual comparison with the other two measures (see Supplementary Materials *E*).

Due to the uneven distribution of noise across the frequency domain, the evoked power was further transformed by subtracting from every frequency bin value in the spectrum the mean of the response magnitude at the neighbouring bins comprising 0.5 Hz prior as well as 0.5 Hz following that frequency bin (8 bins on each side of the target one). The resulting data can be seen as normalized EEG magnitude across the frequency spectrum.

STATISTICAL ANALYSIS

Behavioural and EEG data were analyzed using linear mixed-effects models (lme4 package in R, Bates, Maechler, Bolker, & Walker, 2015). The R code for these analyses is available upon request. All models were chosen following forward-directed model comparison based on the Akaike Information Criterion (AIC). The final models are presented in the Results section. For the analysis of *behavioural responses* in the No Prosody and Prosody experiments, we fitted two generalized binomial models. The first one tackled the effect of Prosody Type (Overt, Covert, or No Prosody). The fixed effects tested for inclusion into this model were Prosody Type, Syntax (1+3 vs. 2+2), and Item Type (Correct vs. Outlier). Here, and in the second model (see below), random intercepts for each participant and item were included by default, prior to model comparison. All fixed effects included in the best model were then tried as random slopes when appropriate while the model was converging successfully and was not reaching singularity. The second model was fitted on the data excluding the No Prosody condition to tap into the effects of Prosodic Contour (accent on Word 1 vs. Word 2 vs. Words 2 & 4). The fixed effects tried were Contour (W1 vs. W2 vs. W24), Syntax (1+3 vs. 2+2), Prosody (Overt vs. Covert), and Item Type (Correct vs. Outlier). Additionally, d-prime values were calculated to form one of the predictors of the EEG data (following the findings by Ding and coauthors, 2017).

Prior to the linear mixed-effect models analysis, EEG magnitude at the sentence (0.78 Hz) and the ½ sentence (1.56 Hz) frequencies was tested using a one-tailed paired t-tests against the evoked power at the neighbouring frequencies (as described above, 8 frequency bins on each side). This was done separately for each experimental block (i.e., each type of prosodic contour, syntactic structure, and task). All p-values were Bonferroni-corrected for multiple comparisons. We extended this analysis by directly and systematically comparing the ITC and the normalized EEG magnitude at the target frequencies in different experimental conditions using linear mixed-effects models. For normalized EEG magnitude, i.e., the difference between the EEG magnitude at the target and at the neighbouring frequencies (scaled between 0 and 1) were modelled to investigate effects of experimental manipulation as well as of scalp topography. In the case of ITC modelling, raw ITC values for target frequencies were modelled. Four main models were fit to predict the EEG data in the No Prosody and the Prosody experiments: two of them included No Prosody and Overt Prosody conditions only (one model for evoked power and one for the ITC), and the other two included No Prosody and Covert Prosody conditions (again, an evoked power and an ITC model). The fixed effects tested for these models were Prosodic Contour, Syntax, Frequency (Sentence vs. ½ of a sentence), d-prime values, as well as the following scalp distribution factors: Anteriority (Frontal vs. Central vs. Posterior), and Laterality (Left vs. Medial vs. Right; see Supplementary Materials *F* for ROI definition). Potential side effects of familiarity and alertness of the participants were investigated by building additional models for normalized EEG magnitude (hereafter, EEG magnitude) and ITC analysis. In these models, only the data from the No Prosody experiment were used, and in addition to the fixed effects of Syntax, Frequency, Anteriority, and Laterality, we studied the factor ExperimentPart (Beginning vs. End).

From the No Semantic Task experiment, the data were analyzed for the 1+3 Syntax W1 (N=10) and the 1+3 Syntax W24 (N=10) conditions (i.e., the conditions for which the data of at least 10 participants were available). The data from the No Semantic Task were contrasted with the corresponding conditions in the Prosody experiment. Four models were fitted: one per condition (W1, W24) and per EEG measure (EEG magnitude, ITC). Similarly to the Prosody experiment analysis, we tried including Frequency, Anteriority, and Laterality into the model. Additionally, we included factors Task (Dual task vs. No Semantic Task) and Prosody (Overt vs. Covert).

The random effect structure by default included random intercepts for each participant, and all fixed effects included in the best model were tried as random slopes while the models were non-singular and converging successfully. The p-values were computed using the Satterthwaite's method (lmerTest package; Kuznetsova Brockhoff, & Christensen, 2017). All models were visually checked for normality and homoscedasticity of the residuals. Residuals crossing the mean+|2.5*SD| threshold were removed when their distribution was skewed. For the ITC measurements, the data were transformed using the Box-Cox method (Box & Cox, 1964). The absence of multicollinearity was ensured based on the condition number (Belsley et al., 1980, cited in Baayen, 2008) and the variance inflation factor (Craney & Surles, 2002). While the Contour fixed effect was contrast-coded with treatment contrasts (with the No Prosody condition as a baseline), other categorical predictors with more than three levels were coded with sum contrasts. The post-hoc analysis of interactions was performed by comparing the least-squares means and their standard errors (lsmeans and lstrends functions from the lsmeans package; Lenth, 2016).

We additionally investigated the relationships between our EEG effects (i.e., the ½ sentence and the sentence rate EEG responses within and between different conditions) using Pearson’s correlation to investigate our interpretations regarding the (dis)similarity of some findings (see Results).

RESULTS

PERFORMANCE ON THE BEHAVIOURAL TASK

On average, participants were 73% accurate in assessing the semantic acceptability of the sentences in trials (comparable to results of Ding et al., 2017). The mean d-prime of 1.061 (SD = 0.918) reflects a positive ratio of hits compared to false alarms in the dataset. Performance was higher on trials without outliers than on those with semantically implausible sentences ($\beta = 1.491$, SE = 0.381, $p < .001$). Among trials without outliers, participants performed better at assessing acceptability of those with overt or covert prosody compared to trials in the No Prosody conditions (OvP vs. NoP: $\beta = 0.339$, SE = 0.125, $p = .007$; CovP vs. NoP: $\beta = 0.429$, SE = 0.126, $p < .001$). The type of intonation contour or the syntactic structure of sentences did not significantly affect performance on the behavioural task.

EEG RESULTS

NO PROSODY EXPERIMENT

EEG magnitude values at sentence and ½ sentence frequencies in 1+3 and 2+2 Syntax were significantly larger than noise (see Supplementary Materials G1; Figure 4). The sentence-level effects were in line with our predictions for both types of sentences. At the same time, according to the syntax-only view, the peak in the EEG spectrum at the ½ sentence frequency would be expected only in the case of the 2+2 Syntax (corresponding to the phrase frequency). To further investigate the (unexpected) ½ sentence rate peak in the 1+3 Syntax (right column of Figure 4), we analyzed the data from a larger subset of participants (N=36), each of whom was presented with at least one No Prosody block at the beginning of the experiment. We again found the ½ sentence peaks in both 1+3 and 2+2 Syntax conditions (p values < .001; see Supplementary Materials *H*). Including factor Syntax (1+3 vs 2+2) did not improve the fit of the linear mixed effect model predicting the EEG magnitude or the ITC across the ½ sentence and the sentence frequencies. We therefore concluded that at least the ½ sentence EEG peak in the 1+3 Syntax condition was likely to be induced by mechanisms other than syntactic processing, specifically, by participants having placed a covert prosodic boundary in the middle of the 1+3 Syntax sentences. We further tested this hypothesis by investigating the correlations between the different EEG peaks in the No Prosody experiment and by comparing their scalp distributions.

We found no correlation between the ½ sentence peaks in the two types of syntactic constructions (EEG magnitude: $r^2 = 0.067$, $p = .451$; ITC: $r^2 = -0.093$, $p = .659$), while the sentence peaks were positively correlated (EEG magnitude: $r^2 = 0.495$, $p = .012$; ITC: $r^2 = 0.461$, $p = .02$). The two ½ sentence peaks also had different scalp distributions: the effect in the 2+2 Syntax was posterior (frontal - posterior: $\beta = – 0.012$, SE = 0.002, $p < .001$; central - posterior: $\beta = - 0.01$, SE = 0.002, $p < .001$), while in the 1+3 Syntax it was broadly distributed. Hence, it is likely that the ½ sentence peaks in the two syntactic conditions were of different nature. Additionally, to test a post-hoc hypothesis that the ½ sentence peak in the 1+3 Syntax condition might be reflecting a harmonic of sentence frequency, we predicted that the two peaks should be correlated, which, however, was not the case (EEG magnitude: $r^2 = 0.162$, $p = .440$; ITC: $r^2 = 0.187$, $p = .371$).

As the No Prosody experiment was run twice (once at the beginning and once at the end of the experiment), we additionally studied the effects of familiarity and alertness on results. While the EEG magnitude at the sentence and the ½ sentence frequencies was reduced at the end compared to the beginning of the experiment ($\beta = -0.01$, SE = 0.0004, $p < .001$) and this reduction was more pronounced in the case of the 1+3 Syntax sentences (ExperimentPart × Syntax: $\beta = -0.004$, SE = 0.0004, $p < .001$), crucially, the Syntax × Frequency × ExperimentPart interaction did not improve the fit of the model. The ITC model including ExperimentPart was not improved by adding the factor Syntax to it.

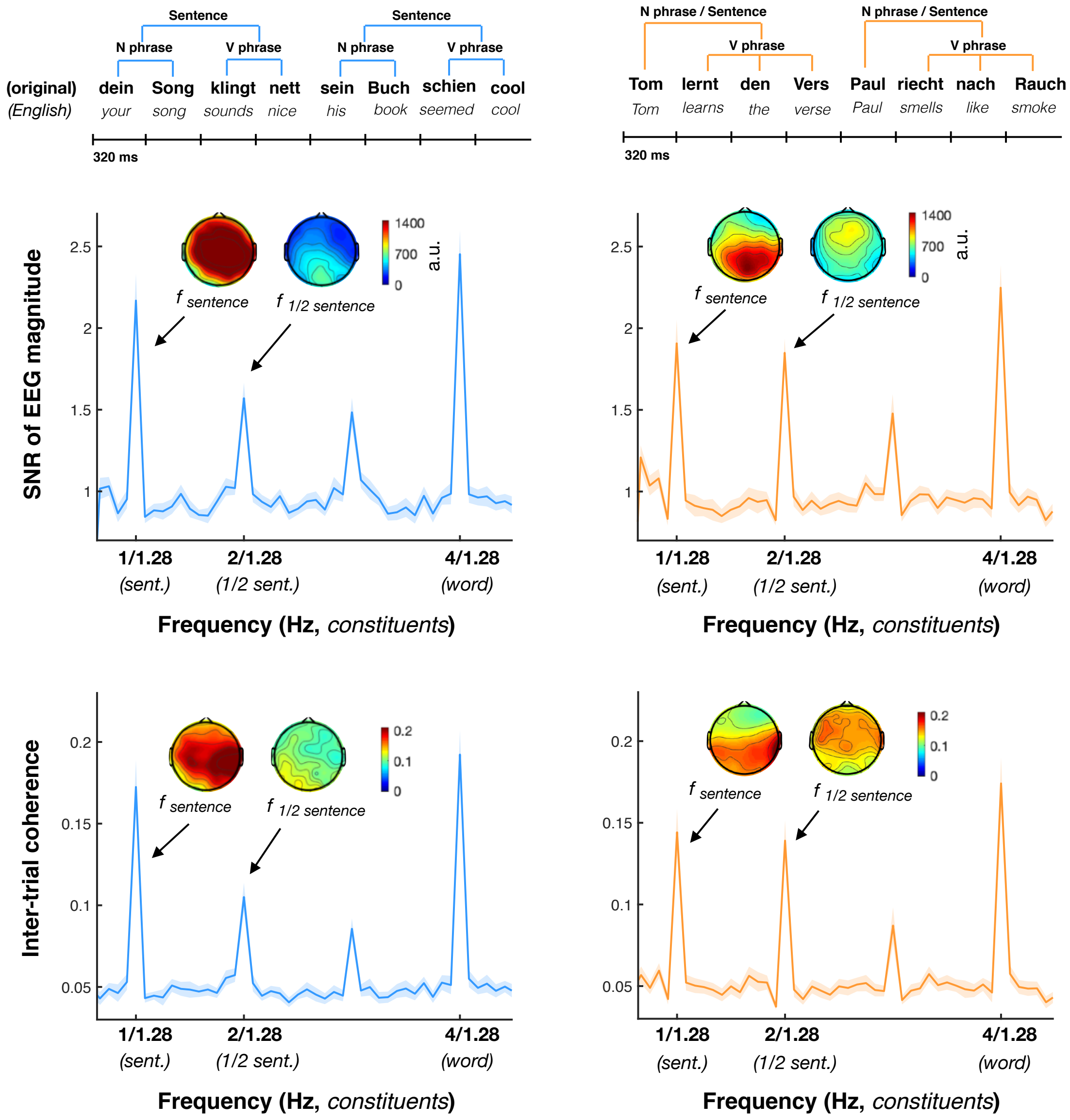


Figure 4. EEG spectrum in the No Prosody experiment. Left panel (from top to bottom): sample sentences with the 2+2 Syntax structure, the EEG magnitude spectrum recorded while participants were listening to the 2+2 Syntax sentences, and the corresponding ITC spectrum. Right panel: same for 1+3 Syntax. Note that the main syntactic boundary in the 1+3 Syntax condition (i.e., the one between first and second words) is not reflected in the spectrum, because the phrases forming the 1+3 Syntax condition are non-isochronous. The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. For plotting purposes, the EEG magnitude spectrum in this and other figures was converted to signal-to-noise (SNR) values reflecting the proportion of signal (EEG magnitude at the target frequency) and noise (mean EEG magnitude within 0.5 Hz on each side of the target frequency). For instance, the ~1.5 SNR at the ½ sentence frequency in 2+2 Syntax means that the EEG responses at this frequency were 1.5 times greater than the EEG magnitude at frequencies within 0.5 Hz to the left and the right from it. Scalp maps depict the scalp distribution of the EEG signal (EEG magnitude in arbitrary units, and ITC) at the sentence and the ½ sentence frequencies.

PROSODY EXPERIMENT

The EEG data from the Prosody experiment are depicted in Figures 5 and 6 (as well as in Supplementary Materials C). A direct comparison can be drawn between the sound intensity spectrum of the stimuli (i.e., upper panel in each row of each figure) and their respective EEG measurements: the EEG magnitude (middle panel) and the ITC (bottom panel). The most prominent and relevant peaks in the EEG spectrum are marked on the horizontal axes in each figure, and correspond to sentence frequency, ½ sentence frequency, and word frequency. The voltage maps reflect scalp distributions of the difference in EEG spectrum peak values between conditions. In the following, we will focus on sentence and ½ sentence frequency to investigate the effect of prosody on phrase structure processing.

As in the case of the No Prosody sentences, EEG magnitude at sentence and ½ sentence frequencies in every experimental condition in the Prosody experiment was significantly larger than noise (for statistics, see Supplementary Materials G2). We further analyzed the results by comparing peaks across experimental conditions to test our hypothesis about the effects of the syntax-prosody congruency level on the EEG responses at frequencies of syntactic constituents (see Supplementary Materials *I* for statistical model summaries). Including the Contour × Syntax interaction improved the predictive ability of Overt-vs-No and Covert-vs-No prosody models for EEG power and ITC, and this effect was statistically significant in all of them. Moreover, the Contour × Syntax × Frequency interaction contributed to improving the fit of all models aside from the Covert-vs-No prosody model of the ITC, and reached significance in the case of the ITC Overt-vs-No prosody and EEG magnitude Covert-vs-No prosody models. Together, these two interactions confirm that the prosodic contours had different effects on EEG peaks elicited by the 2+2 as compared to the 1+3 Syntax sentences.

Figure 5 illustrates the comparison between the No Prosody and the W24 prosodic contour in 2+2 sentences, depicting first the sound intensity spectrum of the speech signals (first row) and then the spectra for EEG signal (rows 2 and 3). The left panel (a) represents the overt prosody condition, the right panel (b) covert prosody. When participants listened to 2+2 Syntax sentences with an overt W24 prosodic contour, the EEG magnitude across the sentence and ½ sentence frequencies was the *highest* compared to all other conditions (W24 - W1: β = 0.005, SE = 0.001, p = .001; W24 - W2: β = 0.005, SE = 0.001, p = .001; W24 - NoP: β = 0.01, SE = 0.001, p < .001). As can be seen from Figure 5a, the enhancing effect of the W24 prosodic contour on ITC in 2+2 Syntax was specific to the ½ sentence frequency (W24 - no: β = 0.069, SE = 0.014, p < .001). The interaction of the Contour × Syntax × Frequency effect with the fixed effect of Anteriority did not improve the fit of the Overt-vs-No prosody models.

Given the substantial impact the W24 pattern had on the ½ sentence peak in the sound spectrum (Figure 5a, first row), it could be argued that the corresponding EEG changes at this frequency (in rows 2 and 3) might, in principle, be driven by bottom-up acoustic changes (though see absence of these changes in the 1+3 Syntax with the W24 prosodic contour below). Crucially, however, the covert prosody condition resulted in similar EEG changes as the overt prosody condition. That is, when participants were presented with prosodically neutralized versions of the exact same 2+2 Syntax sentences - but were asked to simply imagine the congruent W24 prosodic contour in absence of any prosodic cues in the speech signal (see Figure 5b, first row) –, we once again observed very similar EEG effects (Figure 5b, rows 2 and 3). Both the sentence and the ½ sentence EEG magnitude peaks were *larger* for the sentences with the covert W24 prosodic contour compared to all other contours (W24 - W1: β = 0.01, SE = 0.001, p < .001; W24 - W2: β = 0.015, SE = 0.001, p < .001; W24 - no: β = 0.009, SE = 0.001, p < .001). ITC was largest for the W24 condition across frequencies as well (W24 - NoP: β = 0.011; SE = 0.003; p = .001; W24 - W1: β = 0.018, SE = 0.004; p < .001; W24 - W2: β = 0.03, SE = 0.004, p < .001). The interaction of the Contour × Syntax × Frequency effect with the fixed effect of Anteriority improved the fit of the Covert-vs-No prosody EEG magnitude model and reached statistical significance as well. Interestingly, the scalp distributions of the effects and ½ sentence and sentence rates were different: the ½ EEG magnitude enhancement was only seen at central (β = -0.11, SE = 0.22, p <.001) and posterior (β = -0.11, SE = 0.22, p = .004) electrodes, while the sentence rate enhancement reached significance only at frontal channels (β = -0.16, SE = 0.003, p <.001).

In contrast to the 2+2 Syntax, when 1+3 Syntax sentences were presented with the overt W24 contour (i.e., the prosodic contour most *incongruent* with the 1+3 Syntax structure), this combination elicited EEG magnitudes across sentence and ½ sentence frequencies that were significantly *smaller* than in the No Prosody and other prosodic contours (W24 - W1: β = -0.007, SE = 0.001, p < .001; W24 - W2: β = -0.011, SE = 0.001, p < .001; W24 - NoP: β = -0.012, SE = 0.001, p < .001). Importantly, an analogous effect was observed when the W24 prosodic contour was applied to the sentences covertly, or imagined by the participants (e.g., W24 - NoP: β = -0.012, SE = 0.001, p < .001; see Figure 6). Incongruent prosody also resulted in the suppression of ITC across sentence and ½ sentence frequencies compared to the no prosody condition in 1+3 Syntax sentences (OvP W24 - NoP: β = -0.049, SE = 0.014, p = .011; CovP W24 - NoP: β = -0.022, SE = 0.003, p = <.001).

As depicted in Figure 6, syntax-prosody incongruency effects were largely driven by cortical responses at the sentence frequency in both Overt-vs-No prosody (ITC: W24 - NoP: β = -0.087, SE = 0.014, p <0.001) and Covert-vs-No prosody models (EEG magnitude: W24 - NoP: β = -0.018, SE = 0.001, p < .001). The CovP effect had a centro-posterior distribution. A much smaller but still significant suppression of ½ sentence rate responses by incongruent prosody was seen for covert prosody as well (EEG Magnitude: CovP W24 - NoP: β = -0.007, SE = 0.001, p < .001; note, however, that when tested separately at different levels of the Anteriority factor, this effect did not reach significance neither at frontal, nor at central or posterior electrodes).

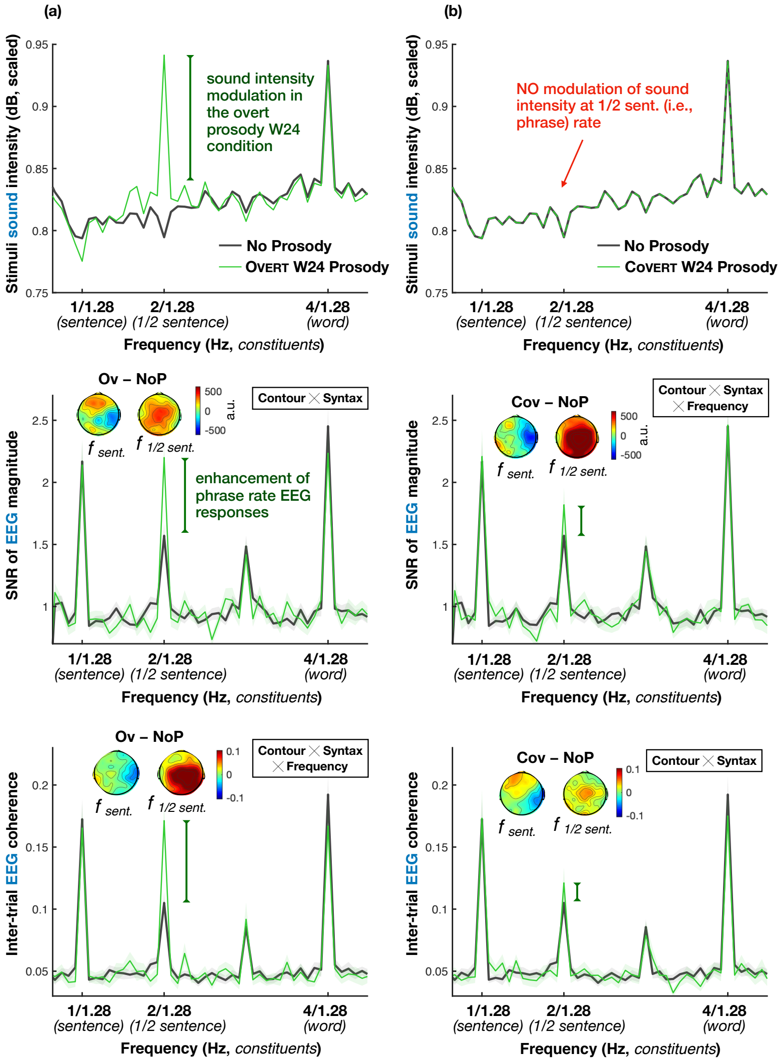


Figure 5. EEG results for the 2+2 Syntax sentences with the prosodic contour fully congruent with their syntactic structure (W24, i.e., pitch and intensity maxima placed on the second and the forth words of each sentence; thin green lines) plotted against the data from the same 2+2 Syntax sentences in the No Prosody experiment (bold grey lines): overt prosody (left column) and covert prosody (right column) conditions. The top row represents spectrum of sound intensity envelopes of the sentences, the middle and the bottom row depict the SNR of EEG magnitude and the ITC spectra respectively. The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the difference between EEG signal (EEG magnitude in arbitrary units, and ITC) in the Prosody and the No Prosody experiments (separately for overt and covert prosody conditions and for sentence and ½ sentence frequencies). In the right upper corner of each EEG spectrum graph we indicated the highest-order statistically significant effect in the corresponding linear mixed-effects model. Key effects are marked with vertical lines (green for Prosody > No Prosody): when a congruent prosodic contour is applied to the sentences (whether overtly or covertly), the EEG responses at the frequency of syntactic constituents supported by prosody (i.e., the ½ sentence frequency) are enhanced compared to the condition with no overt or instructed prosody (No Prosody).

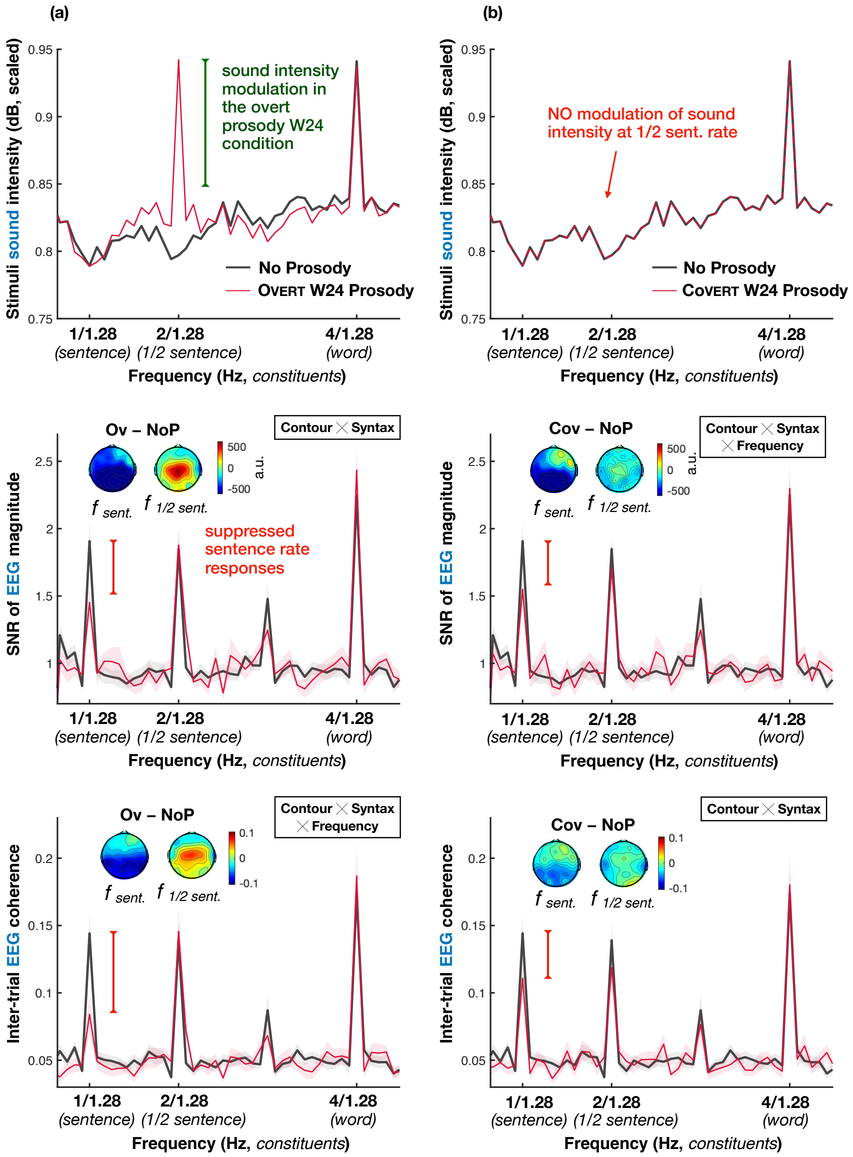


Figure 6. EEG results for the 1+3 Syntax sentences with the prosodic contour least congruent with their syntactic structure (W24, i.e., pitch and intensity maxima placed on the second and the forth words of each sentence; thin red lines) plotted against the data from the same 2+2 Syntax sentences in the No Prosody experiment (bold grey lines): both overt prosody (left column) and covert prosody (right column) conditions. The top row represents the spectrum of the sound intensity envelopes of the sentences, the middle and the bottom rows depict the SNR of EEG magnitude and ITC spectra respectively. The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of difference between EEG signal (EEG magnitude in arbitrary units, and ITC) in the Prosody and No Prosody experiments, separately for overt and covert prosody conditions and for the sentence and the ½ sentence frequencies. In the right upper corner of each EEG spectrum graph we indicated the highest-order statistically significant effect in the corresponding linear mixed-effects model. Prominent significant effects are marked with vertical lines (red for Prosody < No Prosody). When an incongruent prosodic contour was overtly or covertly applied to the sentences, EEG responses at the frequencies of syntactic constituents not supported by prosody (i.e., the ½ sentence frequency) were diminished compared to the condition with no overt or instructed prosody.

The effect of the W24 prosodic contour was as predicted. When prosody was congruent with the syntactic structure of sentences it was applied to, cortical responses at the syntactic constituent rates were enhanced compared to sentences with neutralized prosody. When, on the other hand, this prosodic contour was incongruent with syntactic phrasing, we saw reduced cortical responses in this condition, again, compared to sentences with neutralized prosodic cues. These effects were specific to the frequency of syntactic constituents that were prosodically emphasized (in the case of the congruent contours) or de-emphasized (in the case of incongruent contour). Moreover, the effects of overt and covert prosody were found to be strikingly similar. This extends to the fact that participants with a larger enhancement of ½ sentence rate responses by the W24 contour overtly applied to 2+2 Syntax sentences were also the ones with the larger effect in the CovP condition. Similarly, larger suppression of the sentence-level responses by overt W24 prosody in the 1+3 Syntax sentences was associated with larger suppression in the CovP trials (see Figure 7).

The W24 prosodic contour creates the strongest cases of both syntax-prosody congruency (with 2+2 Syntax) and incongruency (with 1+3 Syntax): temporally as well as rhythmically, the prosodic phrasing aligned or misaligned with syntactic phrasing in the stimuli. We will now turn to the case of partial syntax-prosody alignment effects. As illustrated above in Table 1, the W1 and W2 prosodic contours were only partially congruent or incongruent with the syntactic phrasing of sentences used in the present study, such that their differential effects on the two syntactic structures were expected to be less strong than those of the W24 contour. While not affecting our interpretation of the data from the W24 condition, as noted in the Methods section, the results from the W1 and W2 conditions were largely driven by factors beyond the duality of phrasing-based syntax-prosody interface and were therefore not included in the main text of the manuscript (see Supplementary Materials *C*).

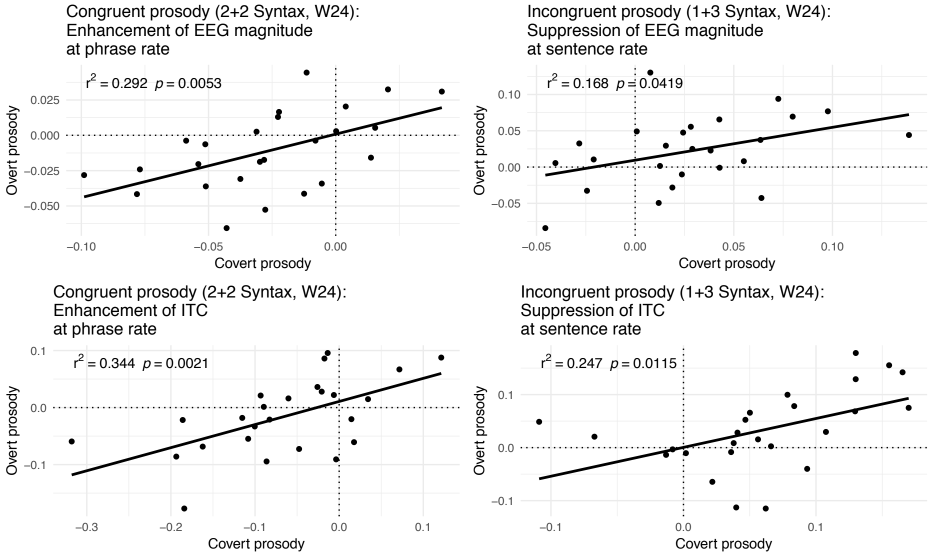


Figure 7. Correlations between effects of overt and covert W24 prosodic contour congruent with the 2+2 Syntax and incongruent with the 1+3 Syntax sentences. On the left panel: the ½ (phrase) rate enhancement of EEG responses in the case of syntax-prosody congruency (2+2 Syntax W24 prosodic contour) contrasted with the No Prosody EEG spectrum peak. On the right panel: the sentence rate suppression of EEG responses in the case of syntax-prosody incongruency (1+3 Syntax W24 prosodic contour). The values represent differences in EEG magnitude (in arbitrary units, scaled, top row) and ITC (bottom row) between the Prosody and the No Prosody experiments at either ½ sentence frequency (in the case of 2+2 Syntax sentences on the left) or sentence frequency (in the case of 1+3 syntax sentences on the right).

NO SEMANTIC TASK EXPERIMENT

The No Semantic Task experiment served to investigate the effect of top-down processing on the results of the Prosody experiment. The EEG magnitude at ½ sentence frequency was significantly larger than noise in the overt W24 (t = -8.25, p < .001; for more details, see Supplementary Materials G3) and covert W24 (t = -3.597, p = .006). This was not the case for the EEG magnitude at the sentence frequency. Note, however, that in this sample of participants, the EEG Magnitude in the OvP condition of the Prosody experiment was not significantly larger than noise either.

The full results of the linear mixed-effect model comparing EEG responses between experimental conditions are outlined in Supplementary Materials *I*. As seen from Figure 8, when the Syntax 1+3 sentences were presented with the overt W24 prosodic contour, the EEG magnitude for both sentence and ½ sentence peaks was larger when participants did not have to detect semantic outliers (prosodic task - prosodic+semantic task: β = 0.025, SE = 0.004, p < .001). Note, however, that since the OvP sentence-level effects were not significantly larger than noise in either NoSemTask or Prosody experiments in this subset of participants, the only interpretable differences in the OvP conditions are at the ½ sentence rate. In the CovP condition we found that the EEG magnitude at sentence frequency was smaller in NoSemTask compared to the Prosody experiment (prosodic task - prosodic+semantic task: β = - 0.053, SE = 0.006, p < .001). The difference did not reach significance at ½ sentence frequency. Factor Task was not included into the ITC model based on model comparison. For results from the W1 condition, see Supplementary Materials *C*.

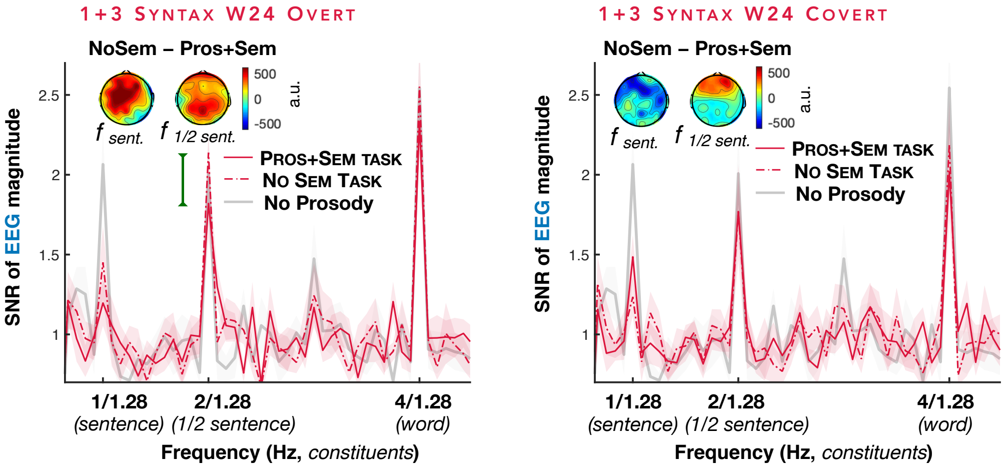


Figure 8. EEG Magnitude results for the No Semantic Task experiment: 1+3 Syntax sentences with overt (left panel) and covert (right panel) W24 prosodic contour. For ITC results, see main text and Supplementary Materials *I*. Data from 10 participants per condition are plotted against the data from the same participants in the No Semantic Task (dashed-dotted red lines), the Prosody (solid red lines), and the No Prosody experiments (bold grey lines). The No Prosody condition is added to the graph to demonstrate the absence of the qualitative differences between the data of the subset of participants and the data in the main No Prosody experiment (however, the No Prosody condition was not included in the linear mixed effects models, for which only the comparison between the Prosody and the No Semantic Task experiment was quantified). The lines in the spectrum plots reflect group averages of the EEG magnitude SNR in the frequency domain, with the shaded area depicting standard errors of the mean. Scalp maps represent the scalp distribution of difference between the effects in the No Semantic Task and Prosody experiments (EEG magnitude in arbitrary units). Prominent significant effects are marked with vertical lines (green for No Sem Task > Pros+Sem Task).

DISCUSSION

Over three experiments, we studied how neural responses to linguistic phrases and sentences are modulated by prosody – both overt, explicitly present in the speech signal, and covert, implicitly imposed on speech with neutralized prosodic cues. We used the frequency tagging approach, a newly applied method to the study of sentence processing. Although results from recent frequency tagging studies (Ding et al., 2016, 2017) have been interpreted as evidence for cortical tracking of hierarchical syntactic structure, our data show that other top-down mechanisms can partially account for these effects. Novel contributions of our study include, first, the unconfounding of syntactic and covert prosodic chunking in the No Prosody Experiment. Humans impose implicit, or covert, prosodic contours on the language materials presented to them (Fodor, 1998; 2002), and this is reflected by the brain activity at the frequency at which prosodic changes occur. The second novel aspect of our study is that we experimentally combined syntactic structures with congruent and incongruent prosodic contours (overtly and covertly). We showed that covert prosodic structure interacts with syntactic phrasing of the sentence modulating cortical responses at the frequencies of syntactic constituents in a non-trivial, not additive but interactive way. Our findings challenge the claims of previous research on cortical tracking of syntactic structure in grammatical sentences (Ding et al., 2016; 2017).

IMPLICIT PROSODY IS PROCESSED BY DEFAULT WHEN OVERT PROSODY IS NOT AVAILABLE

In the No Prosody experiment, participants listened to sentences with contrasting major syntactic phrase boundary locations (1+3 and 2+2 Syntax, respectively) and reliably detected semantic outliers among both types of sentences. Even though all overt prosodic cues were neutralized, the pattern of cortical responses to the 1+3 and 2+2 Syntax sentences pointed to implicit prosodic processing and could not be explained exclusively by syntactic processing. Specifically, auditory presentation of both types of syntactic structure elicited ½ sentence rate

EEG magnitude and ITC peaks, although only the 2+2 Syntax condition contained 2-word syntactic phrases appearing isochronously at the ½ sentence rate. Based on this feature of our experimental design that successfully unconfounded prosodic and syntactic boundary positions, as well as the scalp distributions and the inter-individual variability in the magnitudes of the two ½ sentence rate responses, we propose that implicit, covert prosody processing is the most plausible mechanism to account for the ½ sentence rate effect in the 1+3 Syntax condition.

The scalp topography of the ½ sentence EEG response peaks differed between the two conditions: the 1+3 Syntax sentences elicited a broadly distributed peak, while the peak for 2+2 Syntax sentences was more posterior. Further, while the sizes of the presumably non-prosodic sentence EEG peaks did correlate with each other, the 1+3 Syntax ½ sentence peak size did not correlate with its counterpart, the ½ sentence peak in 2+2 Syntax sentences. These results are congruent with our interpretation that an optional covert prosodic boundary was automatically processed by listeners in 1+3 structures at the end of the second word, splitting the sentence into prosodic phrases of equal lengths (see same-size-sister constraint discussed by Fodor, 1998; de la Crus-Pavía & Elordieta, 2015). Note that placement of a prosodic boundary in the absence of major syntactic boundaries is not uncommon (e.g., to create tension, one may place a boundary within a simple prepositional phrase as in *The Oscar for best picture goes to… 'Parasite'*), even when splitting a preposition from its NP complement, or a verb from its direct object (e.g., Gee & Grosjean, 1983; see also participants applying analogous phrasing when asked to split sentences into "natural groups" in Martin, 1970).

Assuming the ½ sentence EEG peak in the 1+3 Syntax condition reflects covert prosodic processing, and given that it is larger than the corresponding response in the 2+2 Syntax condition at frontal and central (but not posterior) electrodes, we would expect to find similar, fronto-central effects elicited by covert prosodic phrasing. Indeed, the sentence frequency EEG responses are especially enhanced at frontal electrodes in the 1+3 Syntax sentences with W1 contour condition in the No Semantic Task compared to the Prosody experiment (see Supplementary Materials *C*). In other words, we see coherent scalp distribution differences in the EEG responses to sentence constituents that are either more prosodically (more frontal effects) or more syntactically (more posterior effects) driven.

In previous research, the 1+3-like syntactic structures did not match the 2+2 sentences on the possibility of inserting an optional covert prosodic boundary in the middle of the sentence (Ding et al., 2016). Instead, the three last syllables in a 1+3 structure represented a single 3-syllable noun (for example, *fry to-ma-to,* where *tomato* is a single word in Mandarin, similar to English; see Table S1 in Ding et al., 2016). This was the case for all 1+3 stimuli in their Chinese experiment, whereas 1+3 structures were not tested in their English experiment. In contrast, their 2+2 Syntax sentences allowed for a prosodic break in the middle of the sentence, aligning with the syntactic boundary. In the English materials, those were sentences with a two-word noun phrase and a two-word verb phrase (for example, *dry fur rubs skin*) while in most cases for the Chinese materials, the phrase boundary coincided with the single word boundary (i.e., most sentences were formed by two-syllable nouns followed by two-syllable verbs: e.g., *bro-thers ar-gue*). This means that the ½ sentence rate MEG responses in the 2+2 (but not in the 1+3) Syntax condition of that study could possibly reflect either covert prosodic or syntactic differences (or both) from the 1+3 condition, i.e., the results are confounded by covert prosody processing. In our current study, the stimulus materials provided listeners with an opportunity to insert an optional (covert) mid-sentence prosodic boundary into 1+3 Syntax sentences as well as 2+2 sentences, so the confound between syntax and prosody effects was largely eliminated. The finding that participants' EEG responses were characterized by a peak at the ½ sentence rate in both sentence types supports our proposal that apparent phrase structure effects in previous studies may reflect a mix of syntactic and prosodic chunking at best.

Since the last two words in all 1+3 Syntax sentences of the present study formed a single syntactic unit (typically a verb complement), it is conceivable that both syntactic and prosodic processing contributed to results for both sentence types in the No Prosody experiment. However, the ratio of these contributions arguably differed between the syntactic structures. The ½ sentence EEG peak for the 2+2 Syntax sentences may be primarily syntactically driven, similar to the more accurate detection of prosodic phrase boundaries at syntactically justified positions in the behavioural study by Buxó-Lugo and Watson, 2016 (see Introduction). In contrast, the ½ sentence rate peak in the 1+3 Syntax condition was likely driven mainly by non-syntactic principles such as the Uniformity/Balance principle (e.g., Gee & Grosjean, 1983; McCarthy & Prince, 1996; Fodor, Nickels, & Schott, 2018), which is in line with previous reports of prosodic boundaries placed at positions that did not correspond to major syntactic boundaries (e.g., see variation of pitch and edge tones produced for identical syntactic structures in Schafer and colleagues, 2000, and constituent length effects in de la Crus-Pavía and Elordieta, 2015). Our claim that the ½ sentence rate peak in the 1+3 Syntax condition was not elicited because of syntactic chunking mechanisms is based on the fact that the syntax-based grouping of the 1+3 Syntax sentences into a 2+2-like structure was virtually impossible, both in terms of syntactic theory and in terms of psycholinguistic parsing models, as we will explain next.

First, the contrast between the 2+2 and the 1+3 Syntax sentences in both the previous research (Ding et al., 2016) and our own was created based on the organization of the hierarchically highest syntactic phrases within sentences (Chomsky, 1965). In our own materials, the largest syntactic boundary ('|') in both 2+2 and 1+3 structures generally separated a subject NP from its VP (2+2 [*Der Tisch*]$_{\mathbf{NP}}$ | [*steht da*]$_{\mathbf{VP}}$; 1+3: [*Max*]$_{\mathbf{NP}}$ | [*geht* [*zum*

Turm]$_{\mathbf{PP}}$]$_{\mathbf{VP}}$), whereas the syntactic boundary after the verb in 1+3 sentences is a minor boundary within the VP and characterized only by a ‘left edge’ (the opening of a syntactic node). The appropriateness of this comparison is highlighted by the recent report of greater brain activity at phrase-final positions with larger amount of closing syntactic nodes (Nelson et al., 2017). The frequencies of the hierarchically highest syntactic phases within 1+3 Syntax sentences did not correspond to the ½ sentence frequency and were not tagged in our study.

Second, a large body of psycholinguistic research shows that principles of late closure and minimal attachment govern sentence parsing during reading (i.e., when overt prosodic cues are not available to help resolve closure ambiguities; Frazier & Fodor, 1978; Ferreira & Clifton, 1986; Frazier, 1987; for review of other initial choice theories, see Mitchell, 1994). In self-paced reading and eye-tracking studies (e.g., Frazier & Rayner, 1982), participants take more time processing sentences with early closure sentences (for example, EC: *While Anastasia was eating lunch was served to other students*) than late closure constructions (LC: *While Anastasia was eating lunch other students were still working*), and this ‘garden-path’ effect is specific to the location in the sentence at which the syntactic ambiguity is resolved (in the example above, at *was served*). In the domain of psycholinguistics, this processing difference is often explained as an initial parsing preference that produces a structure with fewer syntactic nodes (minimal attachment) and the maximal integration of new words into the preceding phrase (late closure). That is, the noun phrase *lunch* will initially be attached to the verb phrase *Anastasia was eating* as the object of the verb *was eating*. Further, when the forthcoming information implies a different parsing (as in our EC example), the sentence needs to be reanalyzed: the cognitive cost of this reanalysis is reflected in slower reading times. Interestingly, prosodic boundaries (or commas in written language) can prevent or even reverse these garden path effects (e.g., Pauker et al., 2011; Drury et al., 2016). Similarly, in our study, after processing the first two words of a 1+3 Syntax sentence (e.g., *Hein goes* in in the sentence *Hein goes to Bonn*), listeners were much more likely to preferentially attach the complement (*to Bonn*) to the current VP (that was opened with *goes*) rather than process a syntactic phrase boundary ahead of time. On the other hand, a *prosodic* boundary after the first two words, breaking up the syntactic structure, is clearly acceptable (see also Frazier et al. 2002, for empirical evidence and similar conclusions).

Thirdly, while the minimal attachment and late closure principles are helpful in locally *ambiguous* garden-path sentences with an optional (in-)transitive verb (*was eating [lunch]*), for verbs that are strictly transitive (e.g., *to like*), grammaticality constraints require that a subsequent NP needs to be integrated into the VP. The immediate influence of transitivity biases on parsing decisions and ERPs has previously been demonstrated (Itzhak et al., 2010). Importantly, 17 out of the 40 sentences used in the experiment contained verbs of this type, which required a complement in order to constitute a full VP. Thus, syntactic chunking at the ½ sentence rate could not take place as it would result in ungrammatical constituents (for example, *[*Lars likes*] [*the picture*], see Supplementary Materials *A*). In sum, while a prosodic boundary in mid-sentence position is possible in all of our sentences, they cannot be syntactically driven in 1+3 sentences.

One last possibility for a non-prosodic interpretation of the ½ sentence rate peak in the 1+3 Syntax sentences is that it represents the harmonic of the sentence frequency signal, or a different feature of transforming non-isochronous signal into the frequency domain. We know that multiples of the fundamental frequency in the spectrum can be associated with increases of EEG power in the frequency domain, if the signal is not strictly sinusoidal (Zhou, Melloni, Poeppel, and Ding, 2016). If the slow EEG wave elicited by sentence-level processing had 'sharper' edges in it, higher-frequency harmonics (i.e., instances of increased power at the multiples of sentence frequency) would be seen in the frequency transform of the data. However, in our control experiment on non-syntactic rhythmic grouping, the spectrum of auditory materials with 1+3 grouping was not characterized by a peak in brain activity at the ½ constituent rate (see Supplementary Materials *B*); neither was this the case in the original study by Ding and colleagues (2016). In absence of a systematic relationship between sentence and ½ sentence frequency peaks, there is little support for the hypothesis that harmonics would account for the ½ sentence peak in 1+3 structures.

As neither syntactic structure nor low-level harmonic effects can explain the results in our No Prosody experiment, the elicitation of the EEG response at the ½ sentence frequency in 1+3 Syntax sentences strongly points to a role for covert prosodic phrasing in absence of syntactic phrasing. This notion is perfectly consistent with the dominant view in both theoretical phonology (e.g., Selkirk, 2000) and psycholinguistics (Frazier, et al., 2004, 2006; Fodor et al., 2018), according to which syntactic and prosodic phrasing often dissociate. This shift from a syntactic towards a prosodic interpretation of ½ sentence frequency peaks does not only challenge Ding and coauthors' (2016) original account for their EEG findings. More importantly, it also undermines their larger claims that syntactic constituents are concurrently tracked by the brain and that frequency tagging presents a reliable technique for studying syntactic parsing. In our data, hardly distinguishable EEG spectra were elicited during presentation of sentences with drastically different syntactic structures.

TOP-DOWN PROSODIC AND SYNTACTIC PROCESSING MECHANISMS INTERACT

In the No Prosody experiment discussed in the previous section, we explored the possibility that covert prosodic cues affect EEG responses to sentence structure "by default", without any overt acoustic cues that could have biased participants towards a specific prosodic contour. In other words, any impact of prosodic processing could not have resulted from acoustically driven 'bottom-up' cues, but must instead reflect 'top-down' prosodic phrasing strategies. The next level of our exploration involved active manipulation of covert prosodic contours. In the 'covert prosody' part of the Prosody experiment (CovP conditions), we explicitly asked participants to imagine (or ‘project’) a specific prosodic contour while listening to speech with neutralized overt prosodic cues. That is, instead of letting participants activate their subjective default for a prosodic pattern, we now actively promoted the use of a specific covert prosodic contour, which was provided by the preceding 12 sentences of the ‘overt prosody’ trial. The ‘overt-prosody’ part of the Prosody experiment (OvP conditions) allowed us to contrast the ‘induced top-down’ activation of prosody (in CovP conditions) with an overt acoustic (‘bottom-up’) version of the same prosodic pattern and to compare the resulting brain activation between these two conditions in relation to the No Prosody experiment. Importantly, the additional task of imagining a prosodic contour alongside the original semantic outlier detection task did not hinder participants’ performance, which demonstrates that participants were still able to pay attention to sentence-level information. As we will discuss in detail below, the EEG differences between the Prosody and No Prosody experiments do not merely reflect prosodic processing differences alone, but stem from interactions between prosodic and syntactic structures of the sentences.

In the Prosody experiment, we presented participants with overt and covert prosodic contours that were fully congruent or incongruent with the syntactic phrase structure of the sentences. We predicted, based on previous research on syntactically ambiguous sentences (Speer, Kjelgaard & Dobroth, 1996; Kjelgaard & Speer, 1999; Schafer et al., 2000; Dinctopal & Fodor, 2015), that congruent, cooperating prosodic structure would facilitate syntactic parsing. In turn, incongruent, ‘conflicting’ prosody was expected to render sentence processing more difficult and require additional cognitive resources (e.g., Speer et al., 1996; Steinhauer et al., 1999). It has also been previously shown that processing speed of sentences with neutral prosody (the closest to our No Prosody condition) lies in-between the speeds of processing conflicting and cooperating prosody sentences (Speer et al., 1996; Dinctopal & Fodor, 2015). In our experiment, we used the EEG responses to the sentences in the ‘neutral’, No Prosody condition (NoP) as a baseline for studying the effects of congruent and incongruent prosodic contours.

The strongest syntax-prosody congruency (2+2 Syntax sentences with W24 prosodic contour compared to the NoP sentences) induced an increase of the EEG responses at the ½ sentence rate, which corresponds to both the prosodic and the syntactic phrase rate. This was unsurprising and could be explained merely by the presence of acoustic stimulation at the ½ sentence frequency in the OvP condition, and thus by bottom-up mechanisms. Crucially, if this EEG effect was indeed driven by bottom-up mechanisms due to acoustic changes of the speech signal, we would expect similar EEG effects for the same acoustic manipulation (W24) when applied to 1+3 Syntax. However, there was no such enhancement of the ½ sentence rate EEG responses in the 1+3 Syntax sentences (although it was still significantly greater than noise) when an identical W24 prosodic contour was overtly applied, which demonstrates that these effects are not purely bottom-up in nature. Although we observed a ½ sentence peak for both syntactic conditions in the No Prosody experiment, the W24 prosodic contour only enhanced the peak for the 2+2 Syntax condition, that is, in the case of prosody ‘cooperating’ with the syntax. We interpret this as follows: covert prosodic principles for both 1+3 and 2+2 Syntax sentences elicited a ½ sentence peak in the No Prosody experiment, but when prosodic boundary placement was facilitated by an overt, explicit prosodic contour in the Prosody experiment, syntactic phrase chunking was only strengthened in the 2+2 sentences (i.e., in the case of syntax-prosody congruency). In other words, the same acoustic change applied to syntactically different sentences induced different cortical effects, strongly suggesting that syntactic processing was modulated through prosodic processing in the OvP condition.

When OvP W24 contour was applied to 1+3 Syntax sentences (creating the case of an incongruency between syntactic and prosodic phrase structures), not only it did *not* have a boosting effect at the EEG responses at the ½ sentence frequency, but instead it resulted in suppressed EEG responses at the sentence frequency compared to the No Prosody experiment. Note that only the processing of isochronous constituents is likely to be reflected in the EEG spectrum in our experimental design (based on Ding and colleagues' finding of an absent ½ sentence peak in their 1+3 structure, as well as our own corresponding nonword data; see Supplementary Materials *B*). Taking this into account, the sentence rate presents the only frequency at which syntactic constituents can be tagged in 1+3 Syntax sentences. In 1+3 sentences with closure ambiguities (*Hein goes to Bonn*), guided by externally imposed prosody, participants may have initially processed a closure after the second word in a sentence (*goes*). In contrast to the default covert prosodic grouping in the NoP experiment, the externally imposed prosody presumably had a different set of prosodic cues, including both pronounced pitch and intensity changes, which was interpreted as the end of the syntactic clause. Following the mid-sentence externally imposed prosodic boundary, when participants encountered the prepositional phrase (*to Bonn*), they were forced to assign to it the function of a verb complement (making it part of the VP). The need for re-opening the VP prevented participants from efficiently chunking each sentence, most likely by taking up cognitive resources at the same time when sentence chunking was supposed to occur. Unfortunately, the behavioural task used in the current study was not sensitive enough to capture these processing difficulties, and future research could address this issue and investigate the relationship between the behavioural and the EEG data. Our interpretation of the EEG effect is, however, consistent with – and reminiscent of – the relatively low acceptability judgements and a P600 effect found for sentences with intransitively biased verbs at boundary positions in Itzhak and colleagues (2010). Our findings are also in line with the prosodic account of initial parsing choices in locally ambiguous sentences suggesting that listeners settle on a specific syntactic structure guided by prosodic cues (Kjelgaard & Speer, 1999; or for more complex accounts integrating prosody and other non-syntactic variables affecting parsing choices, see e.g., DeDe, 2010;

Itzhak et al., 2010). In our present data, processing difficulties resulting from temporary prosody-syntax conflicts were reflected in the suppressed EEG responses at the sentence rate. For syntactically unambiguous sentences (such as *Lars likes the picture,* where the verb is obligatorily transitive), the strong, externally imposed prosodic boundary cue placed on the second word in the sentence acted as conflicting prosody (based on syntactic analysis). It hindered integration of the 3-word phrase, thereby delaying processing to the point that chunking sentences online within a trial of twelve sentences was complicated. Again, Itzhak and coauthors (2010) demonstrated in their study that even more so than for intransitively biased verbs, transitively biased verbs at prosodic boundary positions trigger a CPS response followed by a P600 components at the point of disambiguation (and these sentences were associated with low acceptability judgements).

Recall that in the No Prosody experiment when syntactic and covert prosodic sentence structures were controlled for, the EEG data did not clearly distinguish between 1+3 and 2+2 syntactic structures. In our OvP conditions, however, the differential effects of the W24 prosody on the respective EEG responses point to an interaction between syntax and prosody and, therefore, strongly suggest the contributions of both syntax (whether direct or through prosodic processing; see section *Prosody, Syntax, Or Both, And Why This Matters* below) and prosody to the size of the EEG peaks at frequencies of sentence constituents.

COVERT PROSODY MODULATES CORTICAL TRACKING MIMICKING EFFECTS OF OVERT PROSODY

We further instructed participants in the Prosody experiment to covertly apply a prosodic contour while listening to speech with neutralized overt prosodic cues. With this paradigm, we wanted to know whether we could replicate the effects observed in the OvP condition even for sentences with neutralized prosodic cues. Indeed, EEG responses to sentences with covertly projected contours were strikingly similar to the ones induced by the sentences with overt prosodic cues in the acoustic signal. Imagining a prosodic contour fully congruent with the syntactic phrase structure of the 2+2 Syntax sentences resulted in enhanced EEG responses at the phrasal rate (with a centro-posterior scalp distribution) and this time also at sentence frequency (frontally). This latter effect at sentence frequency was not seen in the OvP condition and might imply that facilitated parsing might have led participants to mentally project a stronger prosodic boundary at the end of the sentence.

The W24 prosodic contour applied to the 1+3 Syntax sentences did not elicit enhanced EEG responses but instead, the EEG responses were reduced. This suggests that, once again, while a covert prosody-syntax match (in 2+2 sentences) facilitated parsing, a covert prosody-syntax *mismatch* rendered parsing more difficult. Similar to the OvP condition, incongruent covert prosody *suppressed* EEG responses at the sentence rate. As in the OvP condition, we attribute this effect to reduced processing speed. The fact that covertly applying an incongruent prosodic contour hindered sentence chunking might have resulted in some of the participants actively avoiding insertion of a covert W24 contour, which could have resulted in a slight reduction of the EEG magnitude at the ½ sentence frequency (note, however, that the effect size was relatively small for this finding). Overall, however, the effects of overt and covert prosodic contours are noticeably similar, which was further supported by the finding of a correlational relationship between principal EEG effects in the overt and the covert prosody conditions.

In the covert prosody part of our study, we demonstrated that top-down effects of covert prosody mimic effects of overt prosodic cues likely relying on the same top-down chunking mechanism guided by experience and expectations. We showed that syntactic and prosodic processing interact even in the absence of overt prosodic cues. Similar effects of externally imposed covert prosody have been previously reported by Steinhauer and Friederici (2001) who ‘primed’ silent readers of sentences with closure ambiguities using a specific prosodic contour. In their data, the syntax-prosody mismatch ERP responses (i.e., the P600 component) were elicited in conditions where the mismatch was driven by the prosodic contour played to the participants prior to written sentence presentation. The behavioural slow-down of processing due to covert prosody has previously been shown for silent reading of English sentences presented without an optional comma between the subordinate and the main clauses (Staub, 2007; for analogous results in Mandarin Chinese, see Ren & Yang, 2010). Extending these findings, our data show that covert prosody effects can be elicited not only in silent reading ERP studies but also in *auditory* experiments with neutralized prosody, affecting EEG responses at frequencies of syntactic constituents (incl. the studies by Ding and coauthors, 2016, 2017). Top-down effects beyond syntactic processing were also apparent in the conditions reflecting only partial syntax-prosody congruency (i.e., W1 and W2 prosodic contours). As expected, however, for these the overall picture was more complex due to the ambiguous status of those prosodic manipulations, which affected information structure (see Supplementary Materials *C*).

MODULATION OF CORTICAL RESPONSES TO SENTENCE STRUCTURE BY TASK REQUIREMENTS

The semantic outlier detection task in the Prosody and No Prosody experiments encouraged participants to integrate melodic (prosodic) and linguistic (syntactic and semantic) cues. The higher-level processing required for this task should also help integrate information across sentence constituents. It is, therefore, not surprising that EEG responses peaked at the sentence rate even when syntax and prosody were completely incongruent. This finding is in line with the idea that (at least part of) the brain responses at the sentence rate are partially independent of prosody (Ding et al., 2016, 2017) and may reflect syntactic or semantic integration. It is also compatible with previous research on the role of attention in frequency tagging responses to linguistic chunks (e.g., word tracking is suppressed but still evident in a purely sensory task in Ding et al., 2018). However, it is not clear to what extent the elicitation of EEG peaks at the sentence frequency in the case of syntax-prosody incongruency, as well as the interactions between syntactic phrasing and prosody in our study overall, were dependent on semantic processing (for a semantic interpretation of previous frequency tagging research, see Frank & Yang, 2018).

For the No Semantic Task experiment, we predicted that the EEG spectrum would more closely reflect the acoustic properties of the stimulus materials, given the lack of a demand for higher level task processing. In line with this prediction, we found an enhanced, likely bottom-up, ½-sentence frequency peak in the 1+3 Syntax W24 OvP condition, - an effect we interpret as resulting from the prevalent focus on acoustic change (and not higher-level linguistic) processing. When participants imposed a covert prosodic contour in the No Semantic Task experiment, the sentence peak was reduced in comparison to the condition with the semantic outlier detection task. This, again, suggests that imposing covert prosody without an additional need to focus on sentence meaning drove participants' processing further away from higher-level linguistic structure towards non-linguistic, melodic processing.

Task modulation of the EEG responses at frequencies of sentence constituents is unsurprising given the top-down nature of the neurophysiological power peaks elicited in our results and previous studies (Ding et al, 2016; 2017; Ding et al., 2018). In terms of their cognitive correlates, these peaks are dependent on both non-prosodic (syntactic and/or semantic) and prosodic (overt and covert) processing, as supported by the comparison of the Prosody and No Semantic Task experiments. In terms of the neural mechanisms, frequency tagging effects at sentence and ½ sentence frequency may represent both entrainment of slow oscillations to sentence structure (as originally proposed by Ding et al., 2016) and transient neural responses evoked by changes in semantic, syntactic, or prosodic stimuli characteristics (see Glushko et al., in preparation), both of which can be task-modulated.

Neural oscillations are often discussed as a mechanism for attentional prediction (Arnal & Giraud, 2012), indicating that orientation towards prosodic processing (driven by task) can be associated with slow neural activity that selectively tracks the prosodic contour. Among the transient ERP responses, those associated with non-prosodic (syntactic and semantic) processing have been previously reported to be suppressed or otherwise modulated when attention is shifted to non-linguistic information (for review, see Kuperberg, 2007). With respect to the CPS, an ERP seen at prosodic phrase boundary positions, less is known about whether and how it is modulated by task requirements. Steinhauer and colleagues (1999) found no significant task effects (comprehension vs. prosody judgement + comprehension tasks) on CPS or the associated N400/P600 pattern reflecting garden-path processing. Scalp distribution differences were reported for ERPs associated with prosodic phrase boundaries that were elicited with stimuli ranging from grammatical sentences to jabberwocky, pseudoword, and hummed materials (Pannekamp et al., 2005).

Our results acquired from a relatively small group of participants in the No Semantic Task experiment should be interpreted with caution and replicated as well as expanded on by future research. From the present data (see also Supplementary Materials *C*), we can conclude that task-induced modulations of neural data in frequency tagging studies show promise for shedding light on the prosodic, semantic, and syntactic contributions to EEG activity peaks at sentence constituent frequency, and on the relationship between these effects and transient ERP responses.

PROSODY, SYNTAX, OR BOTH, AND WHY THIS MATTERS

Given that prosodic and syntactic sentence structures are closely aligned with one another (although not in a strict way; Selkirk, 2000; Selkirk, 2002; Frazier et al., 2004; Fodor et al., 2018), a question arises of whether we can disentangle the tracking of the two by slow neural activity. When the current study was conceived, we considered three scenarios. First, only syntactic phrases might be tracked by slow EEG waves (leading to neurophysiological power peaks at low frequencies, such as 1 or 2 Hz), and prosodic chunking could rely on different mechanisms (as reflected in transient EEG responses, such as the CPS; Steinhauer, 2003). The second option is that prosodic chunks are tracked by slow EEG waves while syntactic processing is realized at higher (e.g., gamma; Nelson et al., 2016) frequencies and through other mechanisms (such as a top-down phase reset underlying temporal predictions, as proposed by Rimmele, Morillon, Poeppel, & Arnal, 2018). Finally, the third option is that both syntax and prosody are cortically tracked at low frequencies contributing to the frequency peaks at sentence and ½ sentence frequencies (in line with the proposal that there are qualitatively different delta oscillations; Meyer et al., 2017). All three scenarios could in principle encompass the interaction between syntactic and prosodic processing online reported in our Prosody experiment as well as in previous research (for review, see Cutler 1997). We now turn to discussing how the data from the current study fits with each of these three theoretical possibilities.

The first, syntax-only, scenario cannot account for our data. Specifically, recall that the ½ sentence EEG peak is present in the 1+3 Syntax grouping and is in fact more pronounced than the same effect in the 2+2 Syntax sentences in fronto-central scalp areas. A syntax-only account would predict stronger syntactic chunking at ½ sentence rate in 2+2 than in 1+3 Syntax sentences. The parsimonious prosody-only account, in contrast, could plausibly account for these effects: the scalp distribution differences between the 1+3 and 2+2 Syntax ½ sentence EEG peaks could reflect involvement of partially non-overlapping neuronal groups in prosodic phrasing motivated by syntactic boundaries (2+2 Syntax) and prosodic phrasing motivated exclusively by *length* or *balance* prosodic constraints (1+3 Syntax). The third option, still, is not out of the question. It is possible that in addition to the prosodic effects already accounted for by the prosody-only view above, 2+2 Syntax phrases are tracked by slow neural activity as reflected in the posterior scalp areas. At these posterior scalp channels, concurrent tracking of prosody and syntax could possibly have reached a ceiling value, so that no greater responses at the ½ sentence rate were seen for 2+2 Syntax compared to the 1+3 Syntax sentences. We can, in other words, conclude that either only prosodic or both prosodic and syntactic phrases are tracked by cortical activity as reflected from the slow EEG activity at frequencies of sentence constituents.

For both proposed applications of the frequency tagging technique (namely, improving the mechanistic understanding of neurocognitive language processing and language testing in clinical populations) the distinction between syntax and prosody as tracked by slow neural activity is crucial. Even if prosodic phrasing was virtually always syntactically motivated, we know that syntax and prosody processing are psycholinguistically dissociable (e.g., Steinhauer, Alter, & Friederici, 1999). Development of a comprehensive neurolingusitic theory of sentence processing requires pinning down the neural mechanisms (low vs. high frequency, oscillatory vs. transient) corresponding to these psycholinguistic phenomena. In assessing language processing in clinical populations, the distinction between syntax and prosody becomes relevant, for instance, in the cases when prosodic processing might be impaired, and the severity or the presence of the syntactic impairment is variable (e.g., in autism, DePriest, Glushko, Steinhauer, & Koelsch, 2016).

OPEN QUESTIONS

The current study raises several questions for further research that concern both the role of overt and covert prosody in sentence processing, as well as methodological details of the frequency tagging technique. First, we showed that cortical effects of overt and covert prosodic processing as well as syntactic parsing within correct sentences are not simply additive: their interactions exist on the level of phrasing and involve information structure encoding as well. These effects should be appropriately described in a wider variety of sentence types by future research. Such studies should take into account constrains on prosodic phrasing that are beyond the syntax-prosody interface: such as information structure and semantic plausibility (Selkirk, 2000; Selkirk, 2002; Frazier et al., 2004). The latter one is especially relevant in the context of recent proposals highlighting the prevalent role of semantic rather than syntactic merging in online sentence processing (Pylkkänen, 2019).

Second, our No Semantic Task experiment as well as previous research (Ding et al., 2018; Jin, Zou, Zhou, & Ding, 2018) show that task requirements affect results of frequency tagging studies of language, qualitatively changing the pattern of cortical responses at the frequencies of interest. One of the limitations of our study is the use of a simple outlier detection task with outliers being repeated several times throughout the experiment. The behavioral measures of this task were not sensitive enough to capture the differences between processing sentences with supporting and conflicting prosody that were reported in previous research (see e.g., Hilton & Goldwater, 2019). Task effects should be considered in future studies, with the type of chunking motivated by the task and the level of processing triggered by it being the focus of researchers' attention. An appropriate task would engage higher-level sentence processing while not encouraging a specific type of within-sentence constituent grouping. For instance, comprehension question after each trial targeting variable sentence constituents in different sentences within trials would potentially encourage a deeper semantic analysis while avoiding to modulate their default chunking.

Third, in both the current study and in one of the experiments by Ding and colleagues (2016), participants were provided with template sentences and explicitly instructed to impose their structure (syntactic frames in Ding and coauthors' study, prosodic contours in our experiment) onto the auditory stimuli (nonword sequences in Ding and coauthors' study, sentences in our project). In both cases, one could argue that participants imposed *either* syntactic *or* covert prosodic structure on the new materials. For Ding and co-authors (2016), this would mean that their claim that the results were driven by rule-based, syntactic mechanisms might be wrong. Covert prosodic grouping could have been used to process nonword sequences following real-sentence examples in their study (note as well that the nonword sequences were presented with pauses between them during the training phase). For our study, this could mean that effects of syntax-prosody congruency in the case of covert prosody are due to participants having built a strong syntactic template of the 2+2 Syntax structure, due to the listening to the OvP condition in which syntactic and prosodic phrasing were fully congruent (i.e., 2+2 Syntax W24 prosody). Thus, one might argue that it was this syntactic template that primed the processing of the CovP sentences: the template had an effect in the congruent prosody (2+2 Syntax W24) condition because syntactic phrases were facilitated through prosody in OvP sentences. In other words, the CovP effects in the syntax-prosody congruency conditions (the strongest case of which being 2+2 Syntax W24 prosody) would ultimately reflect syntactic, or structural, priming (Bock, 1986; Pickering & Ferreira, 2008), and not the application of a covert prosodic contour on the stimuli. Future research should consider this possibility and use de-lexicalized prosodic contours as templates for experiments with instructed prosody. In the current study, however, it is clear that syntactic priming, i.e., the syntactic facilitation effect (to our knowledge, negative syntactic priming has never been reported) could not have caused the effects of syntax-prosody incongruency (e.g., effects in the 1+3 Syntax W24 condition). That is, it is unlikely that a grammatical sentence was perceived as having a violated structure because a 'broken' syntactic template was formed during the presentation of previous sentence. We know that while grammaticality illusions (i.e., ungrammatical sentences perceived as grammatical; e.g., Lago, Shalom, Sigman, Lau, & Phillips, 2015), are common, the reverse — ungrammaticality illusions, in which a perfectly grammatical sentence is perceived as an ungrammatical one — are extremely rare, much less robust, and might be attributed to statistical artifacts (Parker & Phillips, 2016; Phillips, Wagers, & Lau, 2006). When overt prosody in our study was non-cooperating (like in the 1+3 Syntax condition with W24 contour) and partially disrupted participants' ability to chunk sentences, it is, therefore, implausible that this corrupted syntactic structure was then imposed onto sentences in the CovP condition.

CONCLUSIONS

From our three experiments, we conclude that the evidence for the syntactic nature of effects in existing frequency tagging and similarly designed studies (Ding et al., 2016, 2017; Sheng et al., 2018) is insufficient. Covert prosody can strongly affect brain responses at syntactic constituent frequencies and this has to be accounted for before this experimental paradigm can be used “to objectively assess language processing in children and difficult-to-test populations” (Ding et al., 2016, p. 5). Moreover, covert prosodic patterns, which humans use in a top-down manner while processing language with no overt prosodic cues (i.e., during silent reading or in experiments with artificially neutralized overt prosody) interact, in complex ways, with other (syntactic, semantic, and possibly information structural) layers of sentence processing. The current study showed how these interactions can instantiate themselves during processing of correct sentences with prosody and syntax that are congruent or incongruent, as well as with and without attention placed on sentence meaning. Given the increasing interest in decoding naturalistic (written) language processing (Alday, 2019), fully categorizing interactions that cannot be derived from the language stimuli by relying purely on bottom-up mechanisms, and the morphology of the associated cortical responses, is a crucial task for future research; a concern already expressed by prosody researchers, such as Myers, Lense, and Gordon, 2019.

REFERENCES

Alday, P. M. (2019). M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457-473. https://doi.org/10.1080/23273798.2018.1546882

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714-735. https://doi.org/10.1037/0278-7393.22.3.714

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390-398. https://doi.org/10.1016/j.tics.2012.05.003

Ashby, J., & Clifton Jr, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *96*(3), B89-B100. https://doi.org/10.1016/j.cognition.2004.12.006

Baayen, R. H. (2014). Experimental and psycholinguistic approaches to studying derivation. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology* (pp. 95-117). Oxford: Oxford UP. https://doi.org/10.1093/oxfordhb/9780199641642.013.0007

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In *Reanalysis in sentence processing* (pp. 1-46). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9070-9_1

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7.

Beckman, M. E. (1996). The parsing of prosody. *Language and cognitive processes*, *11*(1-2), 17-68. https://doi.org/10.1080/016909696387213

Belsley, D. A., Kuh, E., & Welsch R.E. (1980). Regression diagnostics. Identifying influential data and sources of collinearity. *Wiley series in probability and mathematical statistics*, Wiley, New York. https://doi.org/10.2307/3150985

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, *17*(2), 89-98. https://doi.org/10.1016/j.tics.2012.12.002

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, *18*(3), 355-387. https://doi.org/10.1016/0010-0285(86)90004-6

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. [Computer program]. Version 6.0.52, retrieved from http://www.praat.org/

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject-and object-control verbs. *Journal of Cognitive Neuroscience*, *22*(5), 1036-1053. https://doi.org/10.1162/jocn.2009.21269

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, *8*(2), 37-50. https://doi.org/10.1111/lnc3.12061

Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of memory and language*, *90*, 1-13. https://doi.org/10.1016/j.jml.2016.03.001

Chomsky, N. (1959). On certain formal properties of grammars. *Information and control*, *2*(2), 137-167. https://doi.org/10.1016/S0019-9958(59)90362-6

Chomsky, N. (1965). Aspects of the theory of syntax. *Cambridge, MA: MITPress*.

Clifton, C., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*, *13*(5), 854-861. https://doi.org/10.3758/BF03194009

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), 391-403. https://doi.org/10.1081/QEN-120001878

de la Cruz-Pavía, I. & Elordieta, G. (2015). Prosodic phrasing of relative clauses with two possible antecedents in Spanish: a comparison of Spanish native speakers and L1Basque bilingual speakers. *Folia Linguistica 49*(1), 185-204. https://doi.org/10.1515/flin-2015-0006

DeDe, G. (2010). Utilization of prosodic information in syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, *39*(4), 345-374. https://doi.org/10.1007/s10936-009-9139-x

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dinctopal, N., & Fodor, J. (2015). Interplay of syntactic parsing strategies and prosodic phrase lengths in processing Turkish sentences. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), 1-10.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*, 1-9. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 1-7. https://doi.org/10.1038/nn.4186

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. *Journal of Neuroscience*, *38*(5), 1178-1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Drury, J. E., Baum, S. R., Valeriote, H., & Steinhauer, K. (2016). Punctuation and implicit prosody in silent reading: an ERP study investigating English garden-path sentences. *Frontiers in psychology*, *7*, 1-12. https://doi.org/10.3389/fpsyg.2016.01375

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348-368. https://doi.org/10.1016/0749-596X(86)90006-9

Féry, C. (2017). *Intonation and prosodic structure*. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781139022064

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377-380. https://doi.org/10.1126/science.1089401

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press, Cambridge, MA.

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414-420. https://doi.org/10.1016/S0022-5371(65)80081-0

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*(2), 285-319.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Speech Prosody 2002 International Conference,* Aix-en- Provence, France.

Fodor, J.D., Nickels, S. & Schott, E. (2018). Center-Embedded Sentences: What's Pronounceable is Comprehensible. In R.G. de Almeida, and L.R. Gleitman (Eds.), *On Concepts, Modules, and Language: Cognitive Science at Its Core* (pp. 139– 168).

New York: Oxford University Press. https://doi.org/10.1093/oso/9780190464783.003.0007

Frank, S. L., & Christiansen, M. H. (2018). Hierarchical and sequential processing of language: A response to: Ding, Melloni, Tian, and Poeppel (2017). Rule-based and word-level statistics-based processing of language: insights from neuroscience. Language, Cognition and Neuroscience. *Language, Cognition and Neuroscience*, *33*(9), 1213-1218. https://doi.org/10.1080/23273798.2018.1424347

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PloS one*, *13*(5), 1-11. https://doi.org/10.1371/journal.pone.0197304

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (p. 559–586). Lawrence Erlbaum Associates, Inc.

Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, *27*(2), 123-146. https://doi.org/10.1023/A:1023241830722

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291-325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210. https://doi.org/10.1016/0010-0285(82)90008-1

Frazier, L., Carlson, K., & Clifton Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244-249. https://doi.org/10.1016/j.tics.2006.04.002

Frazier, L., Clifton Jr, C., & Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua*, *114*(1), 3-27. https://doi.org/10.1016/S0024-3841(03)00044-5

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, *103*(7), 2458-2463.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219. https://doi.org/10.1037/0278-7393.22.5.1219

Garrett, M., Bever, T., & Fodor, J. (1966). The active use of grammar in speech perception. *Perception & Psychophysics, 1*(1), 30–32. https://doi.org/10.3758/BF03207817

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, *15*(4), 411-458. https://doi.org/10.1016/0010-0285(83)90014-2

Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: shared processing mechanisms and effects of musical expertise. *PloS one*, *11*(5), 1-27. https://doi.org/10.1371/journal.pone.0155300

Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(7), 1207-1219. https://doi.org/10.1162/jocn.2008.20083

Hilton, C., & Goldwater, M. (2019). Linguistic Syncopation: Alignment of Musical Meter to Syntactic Structure and Its Effect On Sentence Processing. In A.K. Goel, C.M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 435-441). Montreal, QC: Cognitive Science Society.

Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, *32*(2), 167-195. https://doi.org/10.1023/A:1022448308035

Hwang, H., & Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *Journal of Psycholinguistic Research*, *38*(2), 151-175. https://doi.org/10.1007/s10936-008-9091-1

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: the interplay between implicit prosody and syntax in Korean "garden path" sentences. *Journal of Cognitive Neuroscience*, *23*(11), 3555-3575. https://doi.org/10.1162/jocn_a_00001

Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, *21*(1), 8-13. https://doi.org/10.1097/WNR.0b013e328330251d

Jasper, H. H. (1958). The 10/20 international electrode system. *EEG and Clinical Neurophysiology*, *10*(2), 370-375.

Jin, P., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature communications*, *9*(1), 1-15. https://doi.org/10.1038/s41467-018-07773-y

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153-194. https://doi.org/10.1006/jmla.1998.2620

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*(3), 129-140. https://doi.org/10.1016/s0095-4470(19)31360-9

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133-149. https://doi.org/10.1016/j.jml.2015.02.002

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, *11*(2), 417-441. https://doi.org/10.1162/089976699300016719

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin*, *140*(6), 1534–1555. https://doi.org/10.1037/a0037830

Lenth, R. (2016). Least-Squares Means. R Package 'lsmeans'. Journal of Statistical Software, 69(3), 1-33. https://doi.org/10.18637/jss.v069.i01

McCarthy, J. J., & Prince, A. (1996). Prosodic Morphology 1986. *Linguistics Department Faculty Publication Series, 13*. Retrieved from https://scholarworks.umass.edu/linguist_faculty_pubs/13

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3-and 6-year-olds. *Developmental Cognitive Neuroscience*, *5*, 86-94. https://doi.org/10.1016/j.dcn.2013.01.003

Martin, E. (1970). Toward an analysis of subjective phrase structure. *Psychological Bulletin*, *74*(3), 153. https://doi.org/10.1037/h0029856

The MathWorks. (2011). *MATLAB User's Guide*. The MathWorks, Inc., Natick, MA.

Mitchell, D. C. (1994). Sentence parsing. In *Handbook of Psycholinguistics*, ed. by Morton Ann Gernsbacher (pp. 375–409). San Diego, CA: Academic Press.

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the Envelope: Developments in Neural Entrainment to Speech and the Biological Underpinnings of Prosody Perception. *Brain Sciences*, *9*(3), 70. https://doi.org/10.3390/brainsci9030070

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., ... & Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, *114*(18), 3669-3678. https://doi.org/10.1073/pnas.1701590114

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151-165. https://doi.org/10.1162/jocn.1991.3.2.151

Nicol, J. L. (1996). What can prosody tell a parser? *Journal of Psycholinguistic Research*, *25*(2), 179-192. https://doi.org/10.1007/BF01708571

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494-522. https://doi.org/10.1006/brln.1997.1793

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*(3), 407-421. https://doi.org/10.1162/0898929053279450

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321-339. https://doi.org/10.1016/j.cognition.2016.08.016

Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience*, *23*(10), 2731-2751. https://doi.org/10.1162/jocn.2011.21610

Phillips, C., Wagers, M., & Lau, E. F. (2010). Grammatical illusions and selective fallibility in real-time language comprehension. In: J. Runner (ed.), *Experiments at the Interference, Syntax & Semantics*, vol. 37 (pp. 147–180), Bingley, UK: Emerald Publications. https://doi.org/10.1163/9781780523750_006

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358-374. https://doi.org/10.1016/S0022-5371(83)90236-0

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, *22*(10), 870-882. https://doi.org/10.1016/j.tics.2018.08.003

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain Research*, *1330*, 114-123. https://doi.org/10.1016/j.brainres.2010.03.020

Samek-Lodovici, V. (2005). Prosody–syntax interaction in the expression of focus. *Natural Language & Linguistic Theory*, *23*(3), 687-755. https://doi.org/10.1007/s11049-004-2874-7

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of psycholinguistic research*, *29*(2), 169-182. https://doi.org/10.1023/A:1005192911512

Schafer, A.J. (1997). *Prosodic Parsing: The Role of Prosody in Sentence Comprehension. (Unpublished doctoral dissertation)*. University of Massachusetts.

Selkirk, E. O. (1981). On the nature of phonological representation. *Advances in psychology,* 7, 379-388. https://doi.org/10.1016/S0166-4115(08)60213-7

Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In: Horne, M. (Ed.), *Prosody: Theory and Experiment*. Kluwer Academic Publishers, Dordrecht, pp. 231–262. https://doi.org/10.1007/978-94-015-9413-4_9

Selkirk, E. (2002). The syntax-phonology interface. *International Encyclopedia of the Social and Behavioral Sciences*, ed. by Neil J. Smelser and Paul B. Baltes, 15407-15412.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193-247. https://doi.org/10.1007/BF01708572

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., ... & Gao, J. H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, *29*(8), 3232-3240. https://doi.org/10.1093/cercor/bhy191

Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, *25*(7-9), 1234-1264. https://doi.org/10.1080/01690960903525499

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249-271. https://doi.org/10.1007/BF01708573

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and language*, *86*(1), 142-164. https://doi.org/10.1016/S0093-934X(02)00542-4

Steinhauer, K. (2011). Combining Behavioral Measures and Brain Potentials to Study Categorical Prosodic Boundary Perception and Relative Boundary Strength. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS XVII), Hongkong (China), p. 1898-1901.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135-162. https://doi.org/10.1016/j.bandl.2011.07.001

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267-295. https://doi.org/10.1023/A:1010443001646

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191. https://doi.org/10.1038/5757

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, *64*(6), 1582-1592. https://doi.org/10.1121/1.382142

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64. https://doi.org/10.1037/0096-3445.136.1.64

ten Cate, C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review*, *24*(1), 91-96. https://doi.org/10.3758/s13423-016-1091-9

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905-945. https://doi.org/10.1080/01690961003589492

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and cognitive processes*, *19*(6), 713-755. https://doi.org/10.1080/01690960444000070

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology*, *6*, 1-9. https://doi.org/10.3389/fpsyg.2015.01818

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, *10*, 1-8. https://doi.org/10.3389/fnhum.2016.00274

Zoefel, B., Ten Oever, S., & Sack, A. T. (2018). The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Frontiers in Neuroscience, 12*, 1-13. https://doi.org/10.3389/fnins.2018.00095

MANUSCRIPT 1: SUPPLEMENTARY MATERIALS

SUPPLEMENTARY MATERIALS *A*

Full list of 1+3 Syntax sentences (English translations in parenthesis)
Max baut das Haus. (Max is building the house.)
Hein fährt nach Bonn. (Hein is going to Bonn.)
Karl fand den Grund. (Karl is looking for the reason.)
Fritz fliegt nach Neuss. (Fritz is flying to Neuss.)
Franz fragt nach Tee. (Franz is asking for tea.)
Ben fuhr mit mir. (Ben went with me.)
Till geht nach Rom. (Till is going to Rome.)
Kurt geht zum Turm. (Kurt is going to the Tower.)
Jan giert nach Lob. (Yann is craving for praise.)
Lex greift nach Karl. (Lex is attacking Karl.)
Chris kennt das Buch. (Chris knows the book.)
Dirk kommt nach Laos. (Dirk comes to Laos).
Bert läuft zum Tor. (Bert is going to the gate.)
Tom lernt den Vers. (Tom is learning the verse.)
Lars mag das Bild. (Lars likes the picture.)
Phil maß das Bad. (Phil is measuring the bathroom.)
Hans mied den Rum. (Hans is avoiding the rhum.)
Tim muss nach Kiel. (Timm has to go to Kiel.)
Klaus plant das Fest. (Klaus is planning the celebration.)
Knut putzt das Klo. (Knut is cleaning the toilet.)
Nick reist nach Lund. (Nick is travelling to Lund.)
Alf rennt zum Fluss. (Alf is running to the river.)
Paul riecht nach Rauch. (Paul is smelling like smoke.)
Mark ringt nach Luft. (Mark is gasping for air.)
Gert roch den Gin. (Gert smelled the gin.)
Ron ruft nach Lars. (Ron is calling Lars.)
Rolf rührt den Teig. (Rolf is mixing the dough.)
Jörg ruht seit neun. (Jorg has been resting since nine.)
Joern sah zum Baum. (Jorn looked at the tree.)
Ken schleicht zum Sitz. (Ken is sneaking to the seat.)
Kai spart das Geld. (Kai is saving the money.)
Rick spricht zum Volk. (Rick is speaking to the people.)
Sven stinkt nach Schweiss. (Sven stinks of sweat.)
Horst strebt nach Glück. (Horst is striving for happiness.)
Jens sucht das Mehl. (Jens is looking for flour.)
Nils wäscht das Hemd. (Nils is washing the shirt.)
Ralf wies zum Rad. (Rail is pointing to the bike.)
Kim wischt den Flur. (Kim is wiping the hall.)
Wim zahlt das Bier. (Wim is paying for the beer.)
Phil zieht nach Prag. (Phil is moving to Prague.).

Full list of 2+2 Syntax sentences
Der Bär trinkt erst. (The bear is drinking first.)
Mein Baum steht da. (My tree is [standing] there.)
Das Bett riecht frisch. (The bed smells fresh.)

Das Blatt fliegt her. (The leaf is flying here.)
Mein Boot kippt um. (My boat tips over.)
Mein Boss spricht klar. (My boss speaks clearly.)
Dein Brot schmeckt schlecht. (Your bread tasks bad.)
Sein Buch schien cool. (His book seemed cool.)
Der Busch wächst schnell. (The bush is growing quickly.)
Ihr Chor sang schlimm. (Her choir sang poorly.)
Der Clown lacht lahm. (The clown is laughing lamely.)
Der Dieb lügt krass. (The thief is lying blatantly.)
Sein Fall lief schief. (His fall went wrong.)
Sein Feld blüht schön. (His field is blossoming beautifully.)
Der Fisch schwimmt flink. (The fish is swimming nimbly.)
Mein Freund reist gern. (My friend likes to travel.)
Dein Holz brennt schnell. (Your wood burns fast.)
Sein Hund hört nicht. (His dog does not listen.)
Ihr Kerl spielt falsch. (Her boy is playing incorrectly.)
Ihr Kind schläft dort. (Her child is sleeping there.)
Ihr Kleid ist naß. (Her dress is wet.)
Ihr Koch backt nie. (Her cook never bakes.)
Mein Kopf juckt rechts. (My head is itching on the right.)
Sein Licht scheint schwach. (His light seems weak.)
Ihr Mann sieht kaum. (Her husband barely sees.)
Das Paar wohnt links. (The couple lives on the left.)
Sein Pferd trabt schwer. (His horse is trotting heavily.)
Mein Pool roch fies. (My pool smelled nasty.)
Das Reh fällt hin. (The deer fell in.)
Das Schaf war naß. (The sheep was wet.)
Dein Schal weht rum. (Your scarf is blowing around.)
Der Schnee schmilzt früh. (The snow metls early.)
Dein Schrank wog viel. (Your dresser weighed a lot.)
Dein Sohn ritt fort. (Your son rode away.)
Dein Song klingt nett. (Your song sounds nice.)
Das Team malt selbst. (The team paints itself.)
Mein Vieh ging ein. (My kettle is dying off.)
Der Wolf biss oft. (The wolf bit often.)
Das Zelt blieb hier. (The tent stayed here.)
Sein Ziel hing hoch. (His goal was [hung] high.)

Outlier sentences: 1+3 Syntax

Max baut den Rauch. (Max is building the smoke.)
Rolf rührt das Haus. (Rolf is mixing the house.)
Jan giert nach Kiel. (Yann is craving Kiel.)
Lex greift nach Prag. (Lex is attacking Prague.)
Dirk kommt nach Tee. (Dirk is coming to tea.)
Kim wischt das Glück. (Kim is washing happiness.)
Nils wäscht das Mehl. (Nils is washing the flour.)
Hein fährt nach Grund. (Hein is going to the reason.)
Knut putzt den Vers. (Knut is cleaning the verse.)
Kurt geht nach Lob. (Kurt is going to the praise.)
Gert roch das Fest. (Gert smelled the celebration.)
Tom lernt das Hemd. (Tom is learning the shirt.)

Outlier sentences: 2+2 Syntax

Dein Brot sang schlimm. (Your bread sang poorly.)
Der Bär blüht schön. (The bear is blossoming beautifully.)
Mein Boot trinkt erst. (My boat is drinking first.)
Dein Schal trabt schwer. (Your scarf is trotting heavily.)
Das Zelt lacht lahm. (The tent is laughing lamely.)
Der Busch schläft dort. (The bush is sleeping there.)
Dein Holz lügt krass. (Your wood is lying blatantly.)
Sein Fall biss oft. (His fall bit often).
Der Wolf schmilzt früh. (The wolf melts early.)
Sein Feld wohnt links. (His field lives on the left.)
Sein Licht juckt rechts. (His light is itching on the right.)
Das Bett spielt falsch. (The bed is playing incorrectly.)

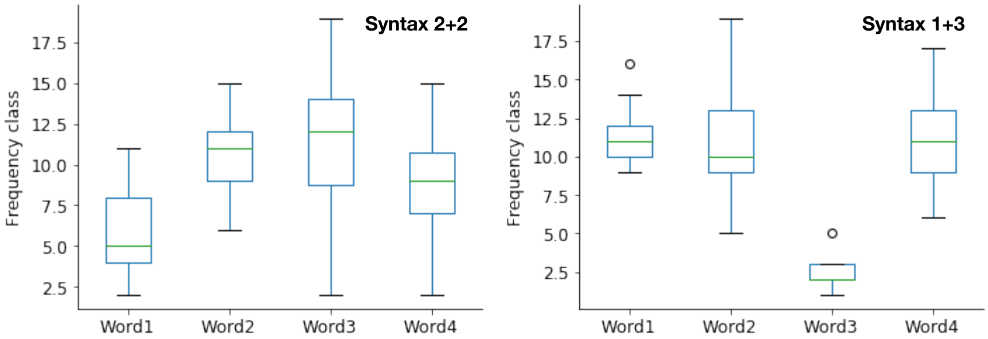


Figure A1. Word frequencies for correct sentences in the experiment: on the left: 2+2 Syntax; on the right: 1+3 Syntax

Word form frequencies were extracted from a German news subcorpus based on material from 2012 (1,000,000 sentences; Eckart & Quasthoff, 2013). Instead of using raw word form frequencies, we extracted frequency classes that are meant to be a more robust measure of word frequency and are logarithmically scaled. Class 0 represents the most frequent word in the corpus, and the classes increase with extremely rare words having a class of 20 or higher.

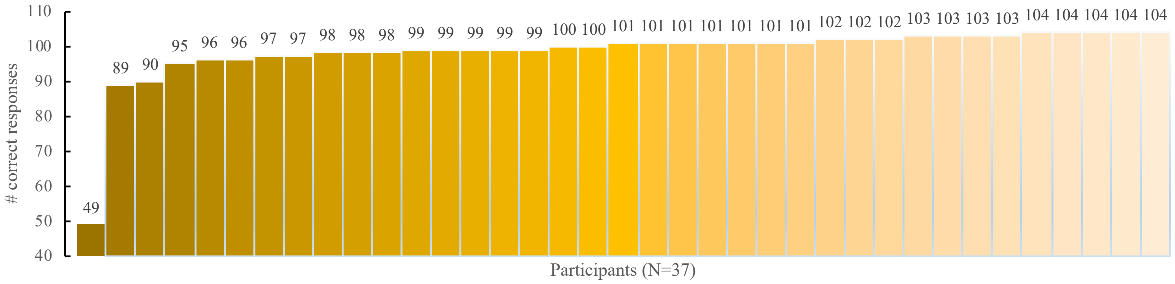
References

Eckart, T., & Quasthoff, U. (2013). Statistical corpus and language comparison on comparable corpora. In *Building and using comparable corpora* (pp. 151-165). Springer, Berlin, Heidelberg.

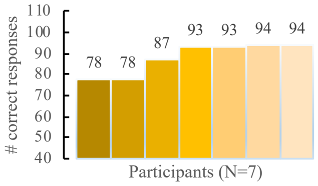
Stimuli intelligibility across OvP conditions and for the NoP sentences.

Participants were presented with all 104 sentences one-by-one. They could listen to every sentence maximum twice and had to type in what they understood. Response was considered correct if the sentence typed in by the participant was identical to the one presented, or if there were minor orthographical error and/or if one proper name was mistaken by another one while not changing the meaning of the sentence. The EEG (and behavioural) data from the participant who scored 49 out of 104 on the NoP sentences were not analyzed.

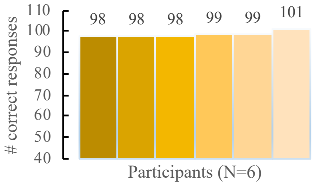
(1) Sentence intelligibility: NoP



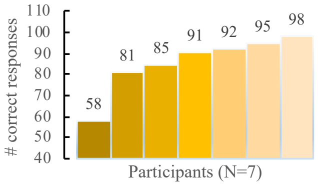
(2) Sentence intelligibility: W24



(3) Sentence intelligibility: W2



(4) Sentence intelligibility: W1



SUPPLEMENTARY MATERIALS *B*

Control experiment: Non-syntactic grouping

Ding and colleagues (2016) investigated the processing of 1+3 Syntax structures (for example, fry$_1$ | to-ma-to$_3$; see main text for details) using frequency tagging with MEG. To ensure that the results from this condition will be comparable to their data when EEG is used, we conducted a control experiment using non-syntactic rhythmic grouping. The reason for using non-syntactic grouping is that with natural language materials, many layers of information (e.g., semantics, syntax, prosody) can contribute to chunking of words into larger constituents. In our case, only one type of manipulation (rhythmic grouping differences) was ensured by using nonword materials. We compared processing of 2+2 and 1+3 grouping created by manipulating the average fundamental frequency at which the nonword was presented, resulting in artificially "sung" stimuli. Sequences made of four monosyllabic nonwords each contained two chunks: one presented with the average pitch of 123.47 Hz, corresponding to the B of the second octave, and the other one with the pitch of 164.81 Hz, corresponding to the E of the third octave (see below).

Methods

Participants. Eleven German native speakers participated in the experiment (age range: 20-35 years, mean age = 26; 6 women, 5 men). Participant inclusion criteria were identical to the ones in the main study.

Materials. We recorded a German native speaker pronouncing fifteen monosyllabic non-words that were phonotactically correct in German. The pitch in them was flattened, and the intensity was normalized in Praat (Boersma & Weenink, 2019) to 70 dB. The length of each nonword was manually adjusted to 250 ms. The nonwords were concatenated into sequences of four in a pseudorandom manner: within each sequence, no vowel or consonant was repeated, and no real words were created through concatenation. The 4-nonword sequences were concatenated into trials comprising 12 sequences each (48 nonwords). Each sequence lasted for one second, making trials 12 seconds long. No pauses were introduced at any time point within trials. Given the length of each nonword in this experiment was 250 ms, the nonword frequency corresponded to 4 Hz, the ½ sequence frequency to 2 Hz, and the sequence frequency to 1 Hz.

For each 4-nonword sequence, pitch was manipulated in a way specific to either the 1+3 or the 2+2 grouping. In the case of the 1+3 grouping, the first nonword was assigned a pitch of 123.47 Hz, and nonwords 2, 3, and 4 were assigned a pitch of 164.81 Hz. In the 2+2 grouping, the pitch of 123.47 Hz was imposed in Praat on the first two nonwords, and the pitch of 164.82 on nonwords 3 and 4. Note that sound intensity was controlled for, and sound intensity fluctuations were only observed at the nonword (i.e., the syllabic) rate (see left panel of Figure B1 in Supplementary Materials B).

To keep participants' attention, similar to the main study, we used an outlier detection task. Thirty trials per grouping were presented to each participant. In twenty two of these trials, all nonword sequences had the same (either 1+3 or 2+2) grouping. The 8 trials contained rhythmic outliers: either a 1+3 grouping sequence was inserted in a trial of 2+2 grouping sequences, and vice versa.

Procedure. This experiment was part of a series of pilot tests administered with the same group of participants. In total, participants came to the lab for approximately 4 hours. After filling out behavioural questionnaires (identical to the ones in the main study), participants were brought into the EEG booth, with the EEG session lasting for approximately 2 hours with several major breaks between tests. Participants were told that the sequences of syllables would form patterns. After each trial they were asked to respond if the main pattern within a trial has ever been broken or not. Trials were presented in blocks of 30, with blocks containing correct and outlier trials for one type of grouping (either 1+3 or 2+2).

EEG recording and processing. See main study for details. In contrast to the main study, single bin values in this experiment were calculated at the frequency of 0.0833 Hz (different from the main study due to the sequences of nonwords being shorter in this experiment).

Statistical analysis. We computed the average accuracy and the d-prime for the behavioural data in this pilot experiment in R (R Core Team, 2018). The EEG Magnitude was analyzed separately for the 1+3 and the 2+2 grouping using one-tailed paired t-tests contrasting the response at either the ½ sequence or the sequence frequency with the noise (average EEG magnitude at neighbouring bins comprising 0.5 Hz prior as well as 0.5 Hz following the target frequency). The p-values were corrected for multiple comparisons using the Bonferroni method.

Results and discussion

The accuracy on the outlier detection task was 66.1%, with the mean d-prime of 1.175 meaning participants did distinguish between the two types of rhythmic grouping. The EEG findings were in line with the behavioural data. The EEG responses at the sequence rate (1 Hz) were significantly different from noise in both 1+3 and 2+2 grouping types ($t = -6.169$, $p < .001$ and $t = -4.562$, $p = .001$, respectively. At the ½ sequence rate, as predicted based on the MEG results in Ding and colleagues' study (2016), the EEG magnitude was only significant from noise in the 2+2 grouping ($t = -3.106$, $p = .011$). These results suggest that 1+3 grouping can be used in EEG frequency tagging studies as a control condition for constructions with 2+2 grouping.

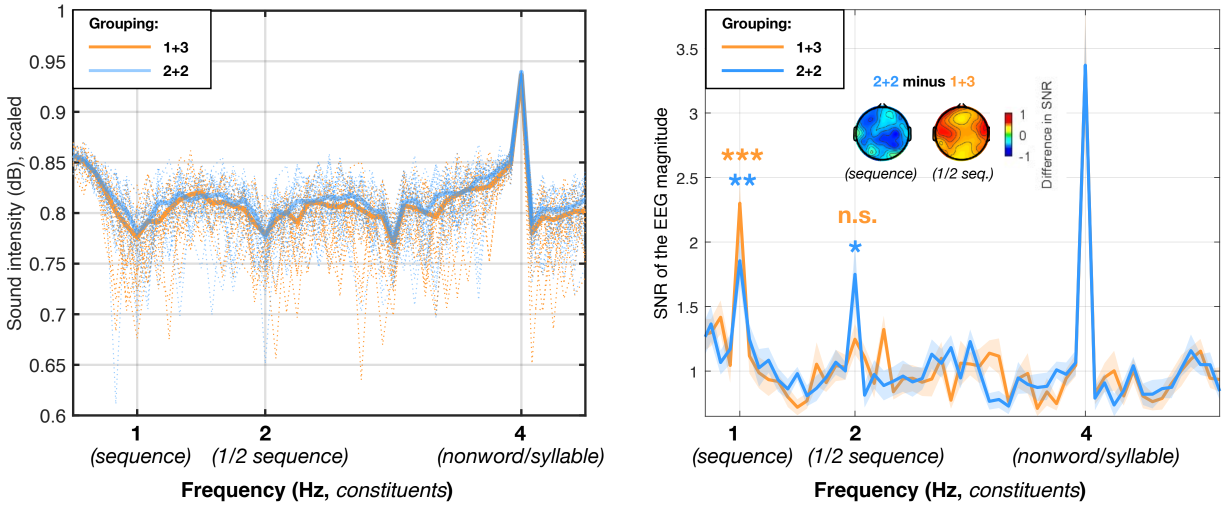


Figure B1. Sound intensity envelope and the EEG magnitude spectra for 1+3 and 2+2 nonword grouping based on pitch changes ("sung" groupings). On the left is the sound intensity envelope calculated per trial, converted to a logarithmic decibel scale, and scaled from 0 to 1. The envelope was computed for each trial separately (thin dotted lines) and then averaged across trials (thick solid lines). On the right is the EEG magnitude spectrum from 11 participants in this experiment. Note that while the EEG responses at the frequency of a full sequence are significantly greater than noise in both conditions, while at the ½ sequence frequency this is only the case for the 2+2 grouping. The scalp maps depict distribution of the differences between EEG magnitude SNR in the two conditions (2+2 minus 1+3 grouping) across electrodes, separately for the sequence and the ½ sequence frequency. $^{*}p<0.05$ $^{**}p<0.01$ $^{***}p<0.001$.

References

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, *19*(1), 158.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SUPPLEMENTARY MATERIALS *C*

W1 and W2 prosodic manipulations

Methods

As summarized in Table C1, the various prosodic contours were either congruent with a given syntactic phrasing because of its (a) temporal alignment with the syntactic phrase boundary (i.e., the temporal shift of the prosodic contour within the sentence), (b) because of their rhythmic correspondence (prosodic fluctuations appear at the frequency of the syntactic phrase), or (c) both. We considered a prosodic contour temporally aligned with the syntactic boundary in the cases where the accentuated word was the last one within the syntactic constituent. The rhythmic / frequency-based congruency manipulation allowed us to use frequency tagging and analyze the data in the frequency domain. For instance, we made sure that the W1 prosodic contour was not only temporally incongruent with the syntactic phrasing of the 2+2 Syntax sentences, but that the effects of this incongruency would also be evident from the frequency domain (with two syntactic phrases, there is still only one pitch and intensity peak per sentence in 2+2 Syntax materials with the W1 prosodic contour). We included the W2 condition to be contrasted with the W24 prosodic contour. Both highlight the boundary between words two and three in each sentence (i.e., the boundary between the NP and the VP in 2+2 Syntax sentences), meaning that both represent the case of temporal congruency between syntactic phrasing and prosody for 2+2 Syntax sentences (see Table C1).

We used the W2 prosodic contour to further test the syntax-prosody interactions in a condition where initial syntactic boundary processing would be *initially required* for the elicitation of prosody-syntax congruency effects. The W2 contour was not characterized by the fluctuations of sound intensity or pitch at the ½ sentence rate (see Figure 2b in the main text). If this prosodic contour was applied to randomly arranged nonwords, one would not predict an EEG power peak at the ½ sentence frequency. However, if it is applied to a 2+2 Syntax sentence, the initial syntactic structure building aligned with the prosodic boundary would trigger enhanced processing of the two phrases in a sentence (through the emphasis of the syntactic boundary that is part of the ½ sentence, or a phrase, rhythm). We would again not expect to see such effects in the 1+3 Syntax condition. This finding in the W2 condition would present exhaustive evidence of the rather non-robust phrase-level effects (see inter-individual differences in Ding et al., 2017) cannot be fully explained by covert prosodic phrasing.

Table C1. Summary of the degree of syntax-prosody congruency across experimental conditions. '+' – congruency. '-' – incongruency. Both the frequency- and time-based congruency at the level of a sentence (to the left from the slash) and the phrase (to the right from the slash) were considered. For example, the W1 contour is only congruent with the syntactic structure of the 1+3 Syntax sentences rhythmically at the level of the whole sentence and temporally at the level of syntactic phrases. The colour coding indicates the degree of congruency at the critical level (i.e., the level of the phrase): full (green), partial (yellow), and no congruency (red).

	1+3 Syntax			2+2 Syntax		
Prosodic contour	W1	W2	W24	W1	W2	W24
Stimulation frequency	$+^{s}/-^{p}$	$+^{s}/-^{p}$	$-^{s}/-^{p}$	$+^{s}/-^{p}$	$+^{s}/-^{p}$	$-^{s}/+^{p}$
Boundary timing	$-^{s}/+$	$-^{s}/-^{p}$	$+^{s}/-^{p}$	$-^{s}/-^{p}$	$-^{s}/+^{p}$	$+^{s}/+^{p}$
Total congruency score	partial	no	no	no	partial	full

Results

In the case of the partial (temporal but not frequency-based) prosody-syntax congruency (2+2 Syntax with W2 prosodic contour), the EEG magnitude was slightly increased across the ½ sentence and the sentence frequencies when participants listened to sentences with overt prosodic cues compared to the NoP sentences (OvP W2 - NoP: $\beta = 0.004$, SE = 0.001; $p = .004$; see Figure C1). In contrast, the W2 contour overtly applied to 1+3 Syntax sentences (creating a partial incongruency between prosody and syntax) did not produce this effect (see Figure C2).

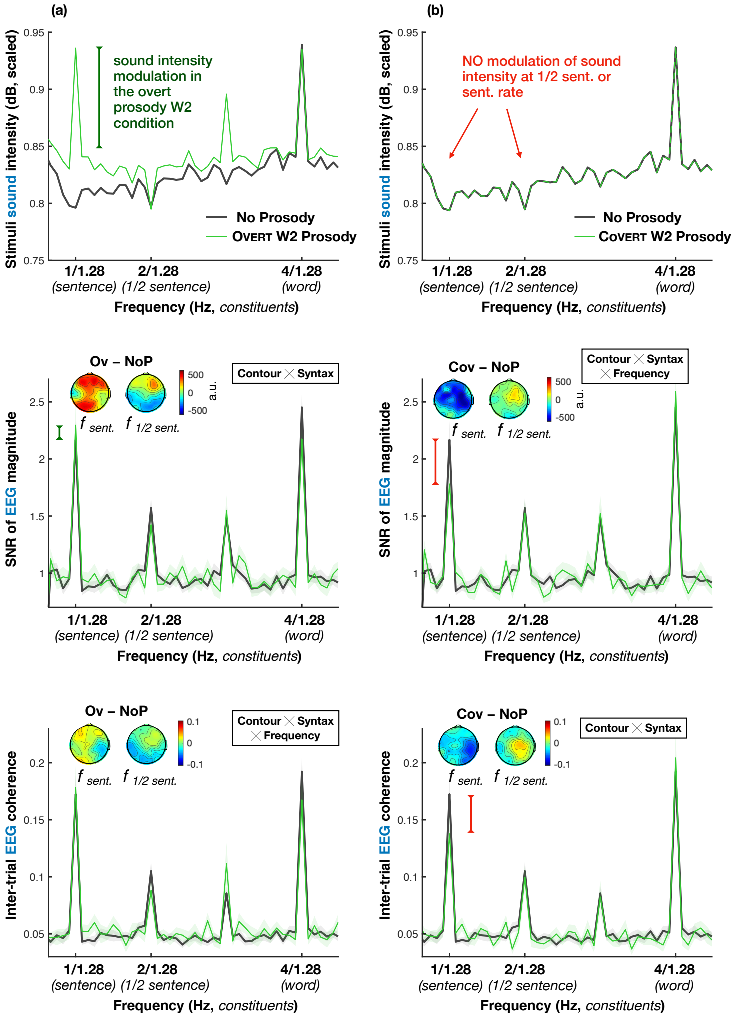


Figure C1. EEG results for the 2+2 Syntax sentences with the prosodic contour partially congruent with their syntactic structure (W2, i.e., pitch and intensity maxima placed on the second word of each sentence; thin green lines) plotted against the data from the same 2+2 Syntax sentences in the No Prosody experiment (bold grey lines): the overt prosody (left column) and the covert prosody (right column) conditions. The top row represents the spectrum of the sound intensity envelope of the sentences, the middle and the bottom row depict the EEG magnitude SNR and the ITC spectra respectively. The lines in the spectrum plots reflect grouop averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the difference between EEG signal (the EEG magnitude is in arbitrary units) in the Prosody and the No Prosody experiments (separately for overt and covert prosody conditions and for the sentence and the ½ sentence frequencies). In the right upper corner of each EEG spectrum graph we named the highest-order statistically significant effect in the corresponding linear mixed-effects model. Prominent significant effects are marked with vertical lines (green for Prosody > No Prosody, red for Prosody < No Prosody): when the temporally congruent prosodic contour was applied to the sentences (whether overtly or covertly), the EEG magnitude was enhanced at the sentence frequency in the overt prosody condition compared to the condition with no overt or instructed prosody (No Prosody). In covert prosody, there is a decrease of the EEG responses at the sentence rate compared.

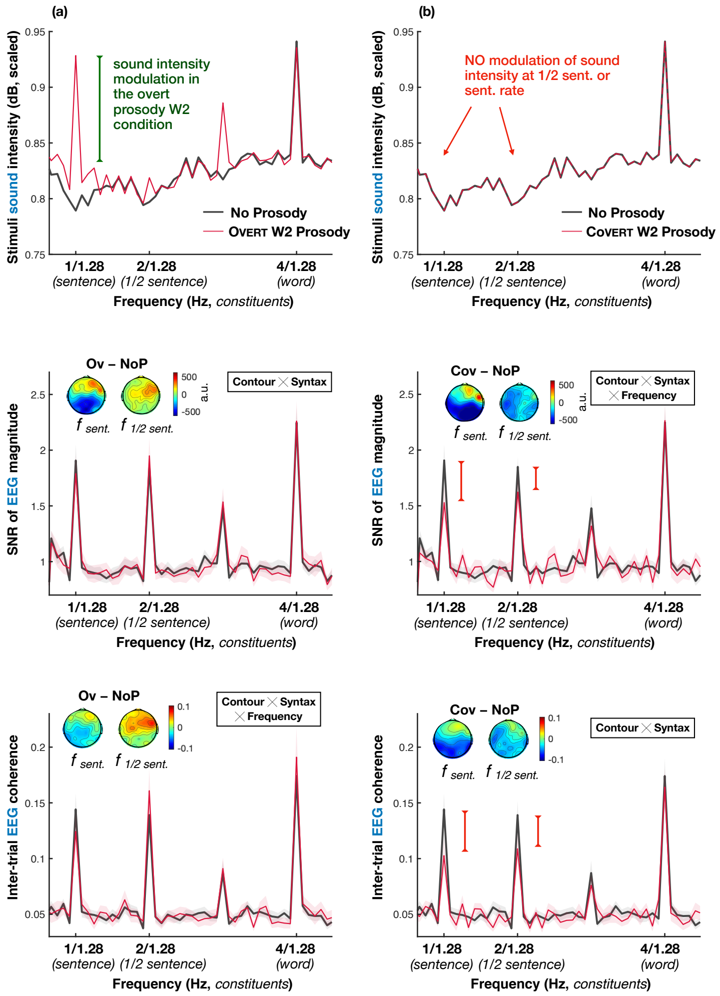


Figure C2. EEG results for the 1+3 Syntax sentences with the prosodic contour partially incongruent with their syntactic structure (W2, i.e., pitch and intensity maxima placed on the second word of each sentence; thin red lines) plotted against the data from the same 1+3 Syntax sentences in the No Prosody experiment (bold grey lines): the overt prosody (left column) and the covert prosody (right column) conditions. The top row represents the spectrum of the sound intensity envelope of the sentences, the middle and the bottom row depict the EEG magnitude SNR and the ITC spectra respectively. The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the difference between EEG signal (EEG magnitude is in arbitrary units) in the Prosody and the No Prosody experiments (separately for overt and covert prosody conditions and for the sentence and the ½ sentence frequencies). In the right upper corner of each EEG spectrum graph we named the highest-order statistically significant effect in the corresponding linear mixed-effects model. Key significant effects are marked with vertical lines (red for Prosody < No Prosody): when the temporally congruent prosodic contour was covertly applied to the sentences, the EEG magnitude was suppressed across the sentence and the ½ sentence frequencies compared to the condition with no overt or instructed prosody (No Prosody).

When the W2 prosodic contour was applied covertly, the EEG measures were smaller across the sentence and the ½ sentence frequencies for the 1+3 Syntax and, unexpectedly, the sentence frequency in the 2+2 Syntax condition (however, only at central electrodes in the case of the EEG magnitude: $\beta = -0.015$, SE = 0.002, $p < .001$), but importantly, as seen from Figure C1, the ½ sentence rate in 2+2 Syntax sentences was least affected (W2 - NoP, 2+2 Syntax: EEG magnitude at ½ sentence rate: $\beta = -0.0003$, SE = 0.001; $p = 1$; ITC across ½ and sentence rates: $\beta = -0.023$, SE = 0.003, $p < .001$). We assumed that the rather unexpected suppression of the sentence rate EEG magnitude in the 2+2 Syntax sentences (and the analogous effect in the 1+3 Syntax sentences) with covert W2 prosodic contour reflected a special case of the interaction between syntax and prosody. That is, the absence of the temporal alignment of the prosodic boundary with the sentence boundary hindered sentence-level tracking when no bottom-up processing of the fundamental (sentence) frequency was required (i.e., in covert prosody in contrast to overt prosody when bottom-up sound intensity and pitch changes appeared at the sentence frequency). Unsurprisingly, unlike the effects in the W24 condition, the effects of the overt and covert W2 prosodic contour were only correlated at the ½ sentence peak in the 2+2 Syntax sentences with the W2 contour (ITC: $r^2 = 0.406$, $p < .001$). This is in line with the idea that the sentence rate effects of the W2 contour in overt and covert prosody reflect partially non-overlapping mechanisms (see Discussion).

The effects of the W1 prosodic contour did not fully align with our original predictions that accounted purely for the effects of syntax-prosody (in-)congruencies. In the 1+3 Syntax sentences, EEG magnitude was smaller in the W1 compared to the No Prosody condition, and the effect seemed to be driven largely by the sentence frequency in overt prosody and exclusively so in covert prosody where the effect reached significance at central and posterior electrodes (OvP W1 - NoP: $\beta = -0.005$, SE = 0.011, $p < .001$; CovP W1 - NoP, sentence frequency: central: $\beta = -0.014$, SE = 0.002, $p < .001$, posterior: $\beta = -0.019$, SE = 0.002, $p <$

.001; see Figure C4). The ITC was lower for 1+3 Syntax W1 sentences compared to No Prosody but only when covert prosody was applied (CovP W1 - NoP: $\beta = -0.025$, SE = 0.003, $p < .001$). That is, the prosody temporally aligned with the syntactic phrasing had an inhibitory effect on the EEG responses that was mostly confined to the frequency of a full sentence.

When the W1 prosodic contour was temporally misaligned with the syntactic phrase boundary (i.e., in 2+2 Syntax sentences), it had an enhancing effect at the EEG amplitudes, which was more pronounced at the ½ sentence rate (OvP W1 - NoP: $\beta = 0.004$, SE = 0.001, $p = .003$; CovP W1 - NoP, ½ sentence: $\beta = 0.006$, SE = 0.001, $p < .001$; see Figure C3). In the covert condition, a small suppression of the sentence rate EEG magnitude was seen in the 2+2 sentences at central ($\beta = -0.012$, SE = 0.002, $p < .001$) and posterior electrode sites ($\beta = -0.013$, SE = 0.002, $p < .001$).

We outline detailed proposed interpretations of the W1 effects in the Discussion. One of our lines of reasoning with this respect was the fact that W1 is the prosodic contour that modulates the information structure of the sentences much more so than W2 and W24 contours. Creating a focus on the first word in every sentence, it drives participants' attention away from the other words that are, in contrast to the first word, relevant to the task performed by participants. This interpretation is in line with the result pattern behind the significant d-prime x Frequency x Syntax x Contour interaction. Recall that given the strong relationship between sentence-level tracking effects in a recent frequency tagging study with EEG (Ding et al., 2017), we included d-prime as one of the predictors into our model as a control variable. We, therefore, with caution interpret this four-way interaction, in the frame of which one effect was consistent across EEG magnitude and ITC models and confirmed by correlation analysis on raw (and not predicted by the model) data. Namely, the sentence-level EEG effects of processing sentences the overt W1 contour had a negative relationship with the performance on the behavioural outlier detection task (see Figure C6).

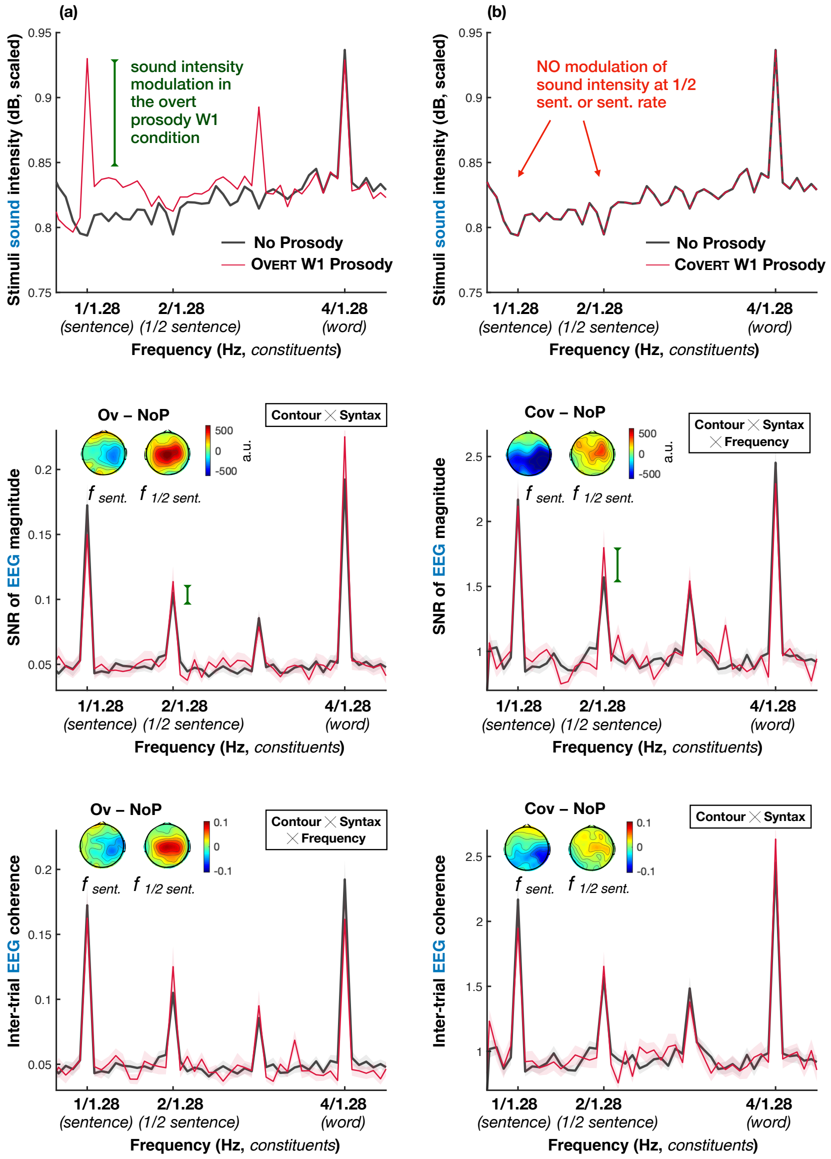


Figure C3. EEG results for the 2+2 Syntax sentences with the prosodic contour incongruent with their syntactic structure (W1, i.e., pitch and intensity maxima placed on the first word of each sentence; thin red lines) plotted against the data from the same 2+2 Syntax sentences in the No Prosody experiment (bold grey lines): the overt prosody (left column) and the covert prosody (right column) conditions. The top row represents the spectrum of the sound intensity envelope of the sentences, the middle and the bottom row depict the EEG magnitude SNR and the ITC spectra respectively. The lines in the spectrum plots reflect averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the difference between EEG signal in the Prosody and the No Prosody experiments (separately for overt and covert prosody conditions and for the sentence and the ½ sentence frequencies). In the right upper corner of each EEG spectrum graph we named the highest-order statistically significant effect in the corresponding linear mixed-effects model. Key significant effects are marked with vertical lines (green for Prosody > No Prosody, the EEG magnitude was enhanced at the ½ sentence frequency in the prosody condition compared to the condition with no overt or instructed prosody (No Prosody). In covert prosody, there was a decrease of the EEG responses at the sentence rate compared to the No Prosody condition.

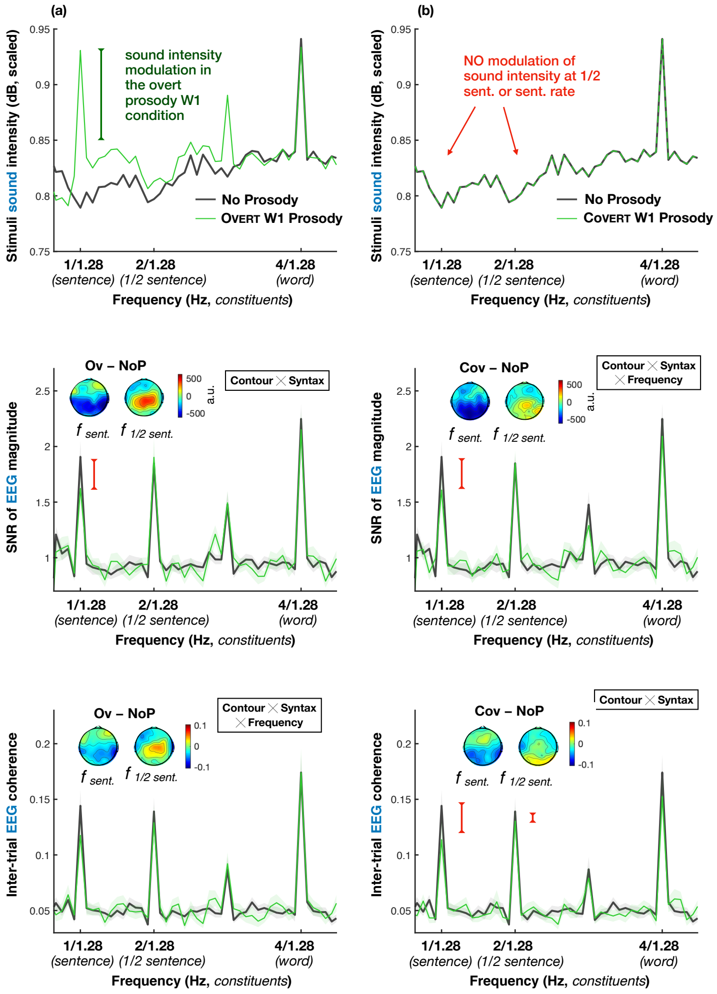
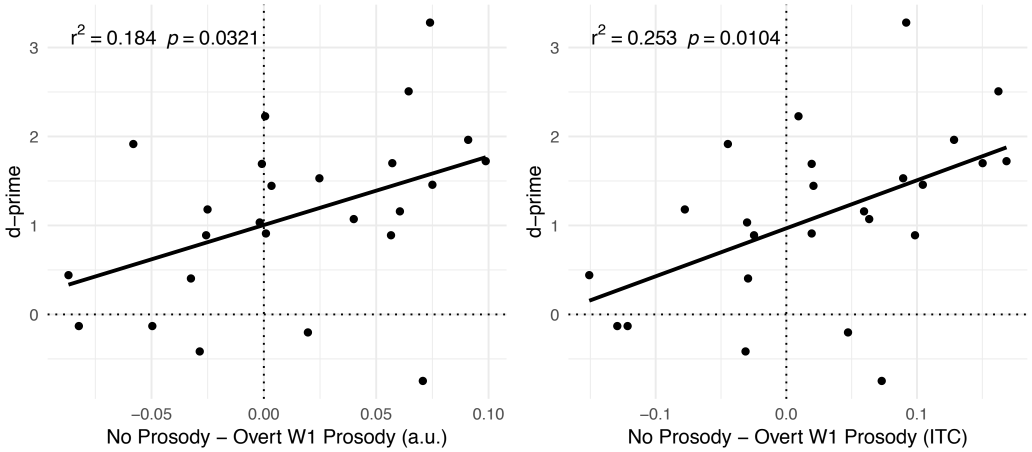


Figure C4. EEG results for the 1+3 Syntax sentences with the prosodic contour incongruent with their syntactic structure (W1, i.e., pitch and intensity maxima placed on the first word of each sentence; thin green lines) plotted against the data from the same 1+3 Syntax sentences in the No Prosody experiment (bold grey lines): the overt prosody (left column) and the covert prosody (right column) conditions. The top row represents the spectrum of the sound intensity envelope of the sentences, the middle and the bottom row depict the EEG magnitude SNR and the ITC spectra respectively. The lines in the spectrum plots reflect averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the difference between EEG signal (EEG magnitude is in arbitrary units) in the Prosody and the No Prosody experiments (separately for overt and covert prosody conditions and for the sentence and the ½ sentence frequencies). In the right upper corner of each EEG spectrum graph we named the highest-order statistically significant effect in the corresponding linear mixed-effects model. Key significant effects are marked with vertical lines (red for Prosody < No Prosody): when the congruent prosodic contour was applied to the sentences (whether overtly or covertly), the EEG response were suppressed at the sentence frequency (in covert prosody, across the sentence and the ½ sentence frequency) in the prosody conditions compared to the condition with no overt or instructed prosody (No Prosody).


(a) 1+3 Syntax:
EEG Magnitude at sentence frequency

(b) 1+3 Syntax:
ITC at sentence frequency

(c) 2+2 Syntax:
EEG Magnitude at sentence frequency

(d) 2+2 Syntax:
ITC at sentence frequency

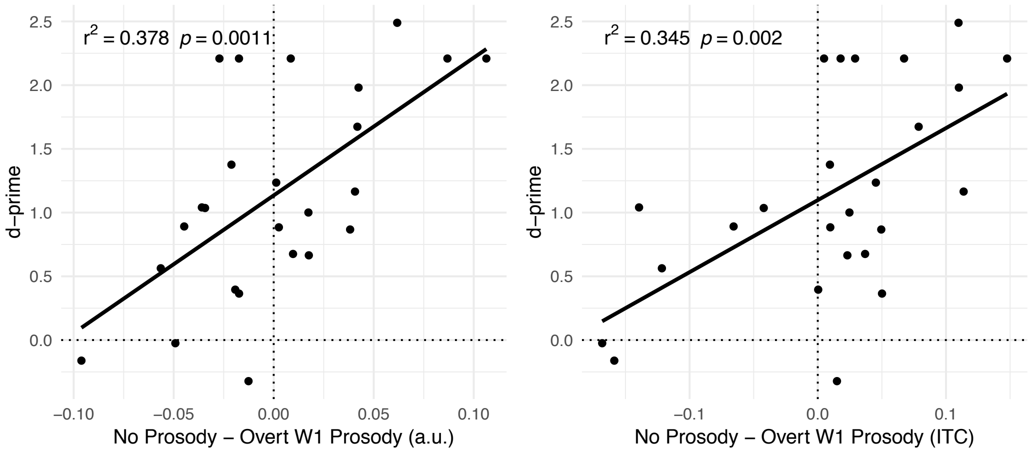


Figure C5. Correlation between the suppression of EEG responses in the overt W1 condition compared to the No Prosody sentences at the sentence rate, on the one hand, and the performance on the behavioural outlier detection task on the other hand (d-prime values). The EEG effects were quantified by subtracting the EEG responses at the sentence frequency in the W1 condition from the EEG responses at the sentence frequency in the No Prosody condition. 1+3 Syntax sentences are represented in (a) and (b). Responses to 2+2 Syntax sentences can be seen in (c) and (d). Both scaled EEG Magnitude (left panel) and ITC (right panel) values are shown. Larger suppression of the sentence rate effects in the Overt W1 condition was associated with better accuracy on the task.

The enhancement of the ½ sentence rate responses in the 2+2 Syntax sentences was also an unexpected result, that as will be addressed in the Discussion, we interpreted as projection of focus to a larger constituent, i.e., the two-word noun phrase, whose syntactic and/or semantic chunking is therefore facilitated. This would mean that the nature of the mechanisms underlying the elicitation of the ½ sentence EEG peak in the W1 condition, on the one hand, and in the W2 and the W24 conditions (that reflect the congruency between prosodic and syntactic phrase chunking) on the other hand, are partially non-overlapping. We followed-up on this idea by investigating the similarity between the effect in the fully congruent (W24) prosodic contour and the temporally congruent (W2) contour on the ½ sentence rate EEG effects. We predicted for these effects to be more change similarly across participants, while the relationship between the W24 and the W1 effects should be less strong. This different is most appropriately tested on the effects of covert prosody, for which bottom-up acoustic processing (that would be different across the three prosodic contour) could not confound the results. Our prediction was confirmed: we found that on the one hand, the EEG peak at the ½ sentence rate in 2+2 Syntax with covert W2 and W24 prosodic contours positively correlated (see Figures C7a-b). On the other hand, there was no statistically significant correlation between the phrase rate enhancement caused in 2+2 Syntax sentences by the W1 contour and the effect a fully congruent W24 prosodic contour had on the EEG responses to these sentences (Figures C7c-d).

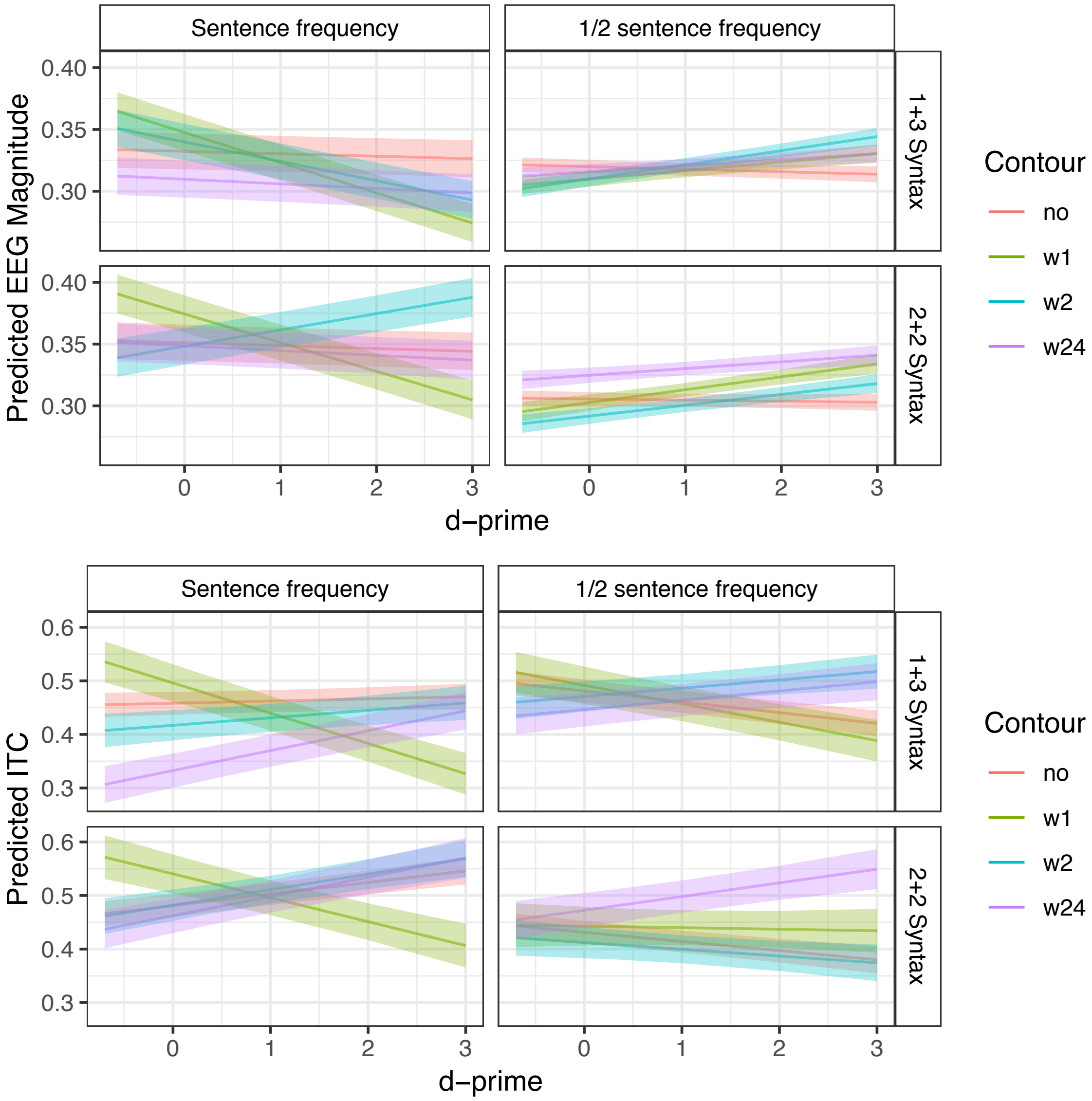
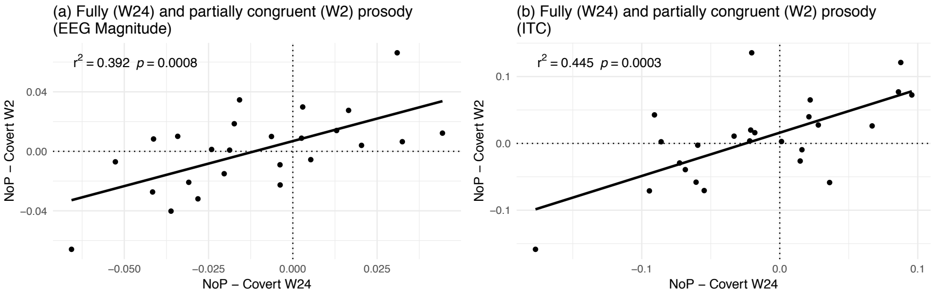
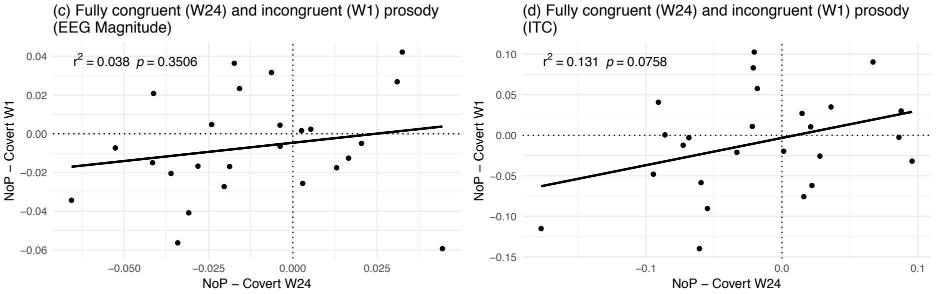


Figure C6. Visualization of the predicted EEG Magnitude (top panel) and ITC (bottom panel) for the four-way interaction included into Overt Prosody vs. No Prosody linear mixed-effects models. Two effects were consistent across EEG Magnitude and ITC: 1) the negative relationship between the d-prime values and the sentence rate EEG responses to the W1 prosodic contour across both 1+3 and the 2+2 Syntax sentences; and 2) the positive relationship between the d-prime values and the ½ sentence rate EEG responses to the W24 prosodic contour across the 1+3 and the 2+2 Syntax sentences. While (1) was also seen in the raw data (see Figure C7), the effect of the W24 contour did not reach significance in the raw data.



(a) Fully (W24) and partially congruent (W2) prosody (EEG Magnitude)

(b) Fully (W24) and partially congruent (W2) prosody (ITC)



(c) Fully congruent (W24) and incongruent (W1) prosody (EEG Magnitude)

(d) Fully congruent (W24) and incongruent (W1) prosody (ITC)

Figure C7. Correlations between the effects of different prosodic contours on the EEG responses 2+2 Syntax sentences at the ½ sentence rate. On the top panel – the two prosodic contours fully or partially congruent with the syntactic phrasing of the sentences show a significant correlation of the EEG peak at the phrase (½ sentence) rate. On the bottom panel – the ½ sentence rate enhancement in the prosodic contour incongruent with the syntactic phrasing (W1) does not correlate with the same effect in the strongest prosody-syntax congruency condition (W24). EEG Magnitude values are plotted on the left panel, and ITC on the right panel.

No Semantic Task

In the case of the W1 prosodic contour overtly or covertly imposed on the Syntax 1+3 sentences, the EEG measurements were larger at the sentence rate (i.e., the frequency of prosodic contour modulation) for the no semantic task condition (prosodic task - prosodic+semantic task: EEG magnitude: $\beta = 0.015$, SE = 0.004, p = .001; ITC: $\beta = 0.016$, SE = 0.006, p = .044). This difference was largely driven by the medial electrodes, and the effect was more pronounced at frontal sites. At the ½ sentence frequency, the EEG measurements were smaller in the absence of the semantic task (EEG magnitude: $\beta = -0.04$, SE = 0.004, p < .001; ITC: -0.037, SE = 0.006, p < .001) and in fact was the only EEG response in the No Semantic Task experiment that did not reach statistical significance when EEG magnitude was tested against noise.

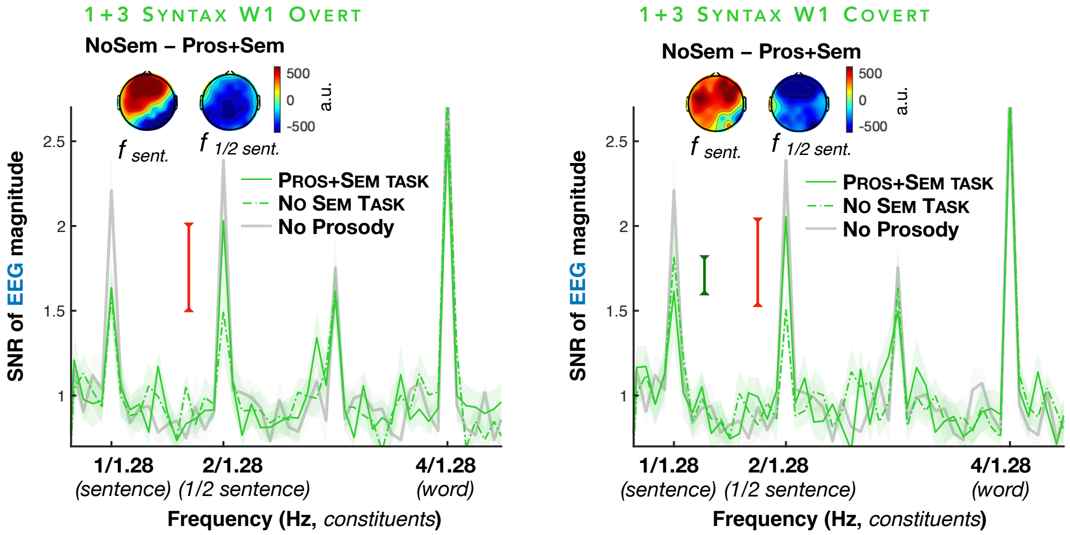


Figure C8. EEG Magnitude results for the No Semantic Task experiment: 1+3 Syntax sentences with overt (left panel) and covert (right panel) W1 prosodic contour. For ITC results, see main text and Supplementary Table 2. Data from 10 participants per condition are plotted against the data of the same participants in the No Semantic Task (dashed-dotted green lines), Prosody (solid green lines), and the No Prosody experiments (bold grey lines). The No Prosody condition is added to the graph to demonstrate the absence of the qualitative differences between the data of the subset of the participants and the data in the main No Prosody experiment (however, the No Prosody condition was not included in the linear mixed effects models, for which only the comparison between the Prosody and the No Semantic Task experiment was quantified). The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. Scalp maps represent the scalp distribution of the difference between the effects in the No Semantic Task and the Prosody experiments (in arbitrary units). Prominent significant effects are marked with vertical lines (red for No Sem Task < Pros+Sem Task, green for No Sem Task > Pros+Sem Task).

Discussion

The role of non-prosodic structure building in phrase tracking. We further explored the interaction between syntax and prosody using the W2 prosodic contour that did not systematically change in sound intensity or pitch at the ½ sentence rate but created a phrase boundary aligned with the syntactic boundary of the 2+2 Syntax sentences in time (i.e., spitting the sentence into two parts of equal length). Similar to the W24 prosodic contours, for the W2 condition, we predicted that cortical responses to structural elements of the sentence would be enhanced in the case of congruent prosody (in the case of the 2+2 Syntax sentences, but not for the 1+3 Syntax sentences). If this was the case, we could confirm that the peak of EEG responses at the phrase frequency reported by us and Ding and coauthors in the 2+2 Syntax sentences are definitively to at least some extent driven by non-prosodic mechanisms. We, therefore, aside from the enhancement of EEG responses at the frequency of acoustic changes (i.e., the sentence frequency), predicted to see the effect of syntax-prosody congruency.

In fact, the effect of the syntax-prosody congruency was seen at the *sentence* frequency in the OvP data. It is difficult to attribute the enhanced sentence rate EEG responses for the 2+2 Syntax sentences with the W2 contour to purely bottom-up processing of the acoustic stimuli envelope given the responses to the incongruent, 1+3 Syntax sentences, were not enhanced. However, at the ½ sentence frequency in the OvP W2 condition the EEG responses did not differ from the effects in the No Prosody experiment (potentially caused by minor differences in the sound intensity spectra between the 1+3 and 2+2 Syntax sentences in the W2 condition). Yet in the covert prosody, the principal predicted syntax-prosody congruency vs. incongruency effect at the ½ sentence can still be detected. In the case of the covertly applied W2 contour, we found a general reduction of the EEG responses across the sentence and the ½ sentence rates in both syntactic conditions, with the magnitude of the EEG spectrum being most preserved at the phrasal rate for 2+2 Syntax. The overall reduction of EEG spectrum peaks across sentence and ½ sentence frequencies in the CovP condition might reflect an unfortunate effect of the order of experiments: the W2 condition was on average closer towards the end of the experiment than other conditions, which could have caused participants losing their focus of attention more so than for other experimental blocks (especially when they had to perform both the semantic and the prosodic tasks in the CovP condition). However, the differences between conditions at the ½ sentence rate in the CovP condition are still in line with our prediction that cooperating prosody will have a positive effect on cortical responses to sentence structure. Like in the W24 prosodic contour, this effect in the case of the W2 covert contour is more pronounced at specific frequency of the constituent accentuated by the prosody. This is true even in the absence of acoustic changes at this (½ sentence) frequency highlighting the undeniable role of non-prosodic chunking mechanisms in phrase tracking in our and previous frequency tagging studies (Ding et al., 2016, 2017).

The suppression of the sentence frequency analogous to the effect seen in the 1+3 Syntax sentences with the W24 prosodic contour in the OvP condition was not seen in the sentences with the W2 contour. This is not surprising given that the processing of the acoustic changes at the sentence frequency could partially mask effects of syntax-prosody incongruency.

One unexpected result arising from the covert W2 prosodic contour was the reduction of the sentence rate responses in the 2+2 Syntax condition. A possible explanation is that because no prosodic cue clearly marked the sentence boundary, the sentence-level chunking in both 1+3 and 2+2 Syntax conditions was weaker in the covert prosody condition. In other words, although this prosodic contour was congruent with the sentence frequency in terms of its rhythmic properties, it was incongruent with the sentence boundary in terms of its timing, and the latter showed potentially more influence on the EEG responses at the sentence rate. While we did not predict this, this result, first, shows that phrase-level prosody-syntax congruency effects cannot be accounted for solely by the difference in their subharmonics (i.e., the sentence rate that changes differently from its harmonic, ½ sentence rate, in the 2+2 Syntax sentences with the W2 covert prosodic contour). Second, these data suggest that future frequency tagging studies should account for and report acoustic characteristics of the stimuli in the *time* (and not only the frequency) domain. Third and importantly, we found that temporal alignment of prosodic and syntactic boundaries is a significant predictor of cortical activity at the frequencies of both phrases (in the CovP W2 condition) and sentences (in the CovP W24 condition) when no prosodic (overt or covert) change is initially present at this frequencies.

This shows the undeniable contribution of non-prosodic mechanisms to the EEG responses to sentences and phrases within them in our and previous research (Ding et al., 2016, 2017).

Beyond the duality of the syntax-prosody interface. The last prosodic manipulation used in the Prosody experiment was the W1 prosodic contour that created a prosody-syntax incongruency with the syntactic phrasing within the 2+2 Syntax sentences, while not being incongruent with the 1+3 Syntax. We predicted for incongruent prosody and syntax that the ½ sentence EEG peak in the 2+2 Syntax sentences would decrease from the NoP to the W1 manipulation (whether overt or covert conditions), which would not be the case for the 1+3 Syntax sentences. However, this was not the case, and the phrase-rate responses in the 2+2 Syntax sentences were in fact enhanced when this incongruent prosodic contour was used. Meanwhile in the 1+3 Syntax sentences, the EEG responses were decreased by W1 prosody (with the effect being largely confined to the sentence rate). While we remain cautious about over-interpreting these effects, we offer two post-hoc interpretations for them that should be considered in the present and future studies and that go beyond the simplest notion of syntax-prosody alignment.

One possibility is related to cognitive mechanisms mediated by information structure of the sentence, i.e., mechanisms that go beyond the congruency between prosodic and syntactic phrasing. Prosodic prominence reflecting information structure of the utterance, while acoustically being evident from a single word, can, for instance, functionally highlight a larger constituent, a phenomenon called "focus projection". It can be that the first word in the W1 2+2 Syntax ("the" in "The table stands there" – *The table is over there*) sentences carrying a pitch accent ("the" in "The table is over (lit.: "stands") there") projects prosodical prominence to the minimal informational unit it belongs to – the noun phrase ("the table"). The ½ sentence rate enhancement of EEG responses by the W1 contour in 2+2 Syntax sentences also has a much weaker (and statistically non-significant) correlational relationship with the effects of the

W24 contour at the same frequency than the effects in the W2 condition. This supports the idea that these effects (elicited by W2 and W24 conditions vs. elicited in the W1 condition) at least partially rely on non-overlapping mechanisms. We can conclude from these data that syntactic chunking coexists with compositional, informational chunking motivated by information structure (expressed through prosodic prominence).

In the case of the 1+3 Syntax sentences, prosodic prominence placed on the first word created the most discourse-wise inappropriate information structure of the utterance: the first word in the 1+3 Syntax sentences is the least relevant one and carries the least information. This mismatch between prosody (intensity and pitch accent on word one) and the information in the sentence might have contributed to a slightly weakened ability of participants to integrate words in full sentences (hence the decrease of the sentence-rate responses). Moreover, better ability to attend to higher-level information in the sentence and perform on the task was seen in participants with a larger suppression of sentence-rate responses in 1+3 Syntax W1 condition. The statistical models fitted here in fact predicted enhanced or unchanged sentence rate responses in poor performers in the W1 condition (see Figures C6-C7), meaning that bottom-up processing (while not paying enough attention to the semantic task) at the sentence rate trades-off with higher-level cognitive demands (i.e., that cause slowing down of sentence processing and decreased sentence rate EEG responses. The post-hoc interpretation highlighting the potential role of information structure in our effects is rather speculative, it is one of the possibilities allowing us to account for the results from the W1 conditions that cannot be accounted for by either syntactic parsing, or prosodic prosody, or the combination of the two.

The second possibility for interpreting the W1 effects, or, specifically, the increased ½ sentence EEG peak in the 2+2 Syntax sentences is that the prosodic contour we created was imprecisely encoded by the participants with a ½ sentence peak harmonic being processed in addition to the fundamental sentence rate prosodic fluctuations. If the prosodic contour encoded by the participants in fact had a ½ sentence rate harmonic in it (that was not present in the acoustics of the OvP stimuli), the increase in the EEG responses at this frequency would not be surprising. The idea that the prosodic contour could have been imprecisely encoded goes back to the pilot experiments to inform the design of the current study. We recorded a native speaker producing 2+2 and 1+3 Syntax sentences with the accent placed on either the second (similar to the W2 contour) or the first (similar to the W1 contour used in the current study) words in the sentence. When the speaker produced the accent on the first word in 2+2 syntax sentences, we found intensity fluctuation at the ½ sentence frequency (note that this is not the case for the W2 contour; see Figure C9). It is possible that participants covertly processed this, potentially more natural version of the W1 contour in the current experiment, in which the sound intensity was high at the first harmonic of the sentence frequency.

Task effects (see also "Modulation of cortical responses to sentence structure by task requirements" in the main text). The sentence-rate EEG magnitude and ITC peaks in the No Semantic Task experiment for the W1 contour showed analogous activity to the ½-sentence rate cortical responses in the W24 condition when semantic task was absent, reflecting enhanced processing of the prosodic contour (i.e., melodic processing). With no requirements placed on the processing of higher-level sentence information, we see the effect of the ½-sentence rate responses in the Prosody experiment being practically eliminated in the No Semantic Task experiment. This can reflect a more precise processing of the W1 prosodic contour given that integration of prosody with semantics was not prompted by the task in the No Semantic Task experiment.

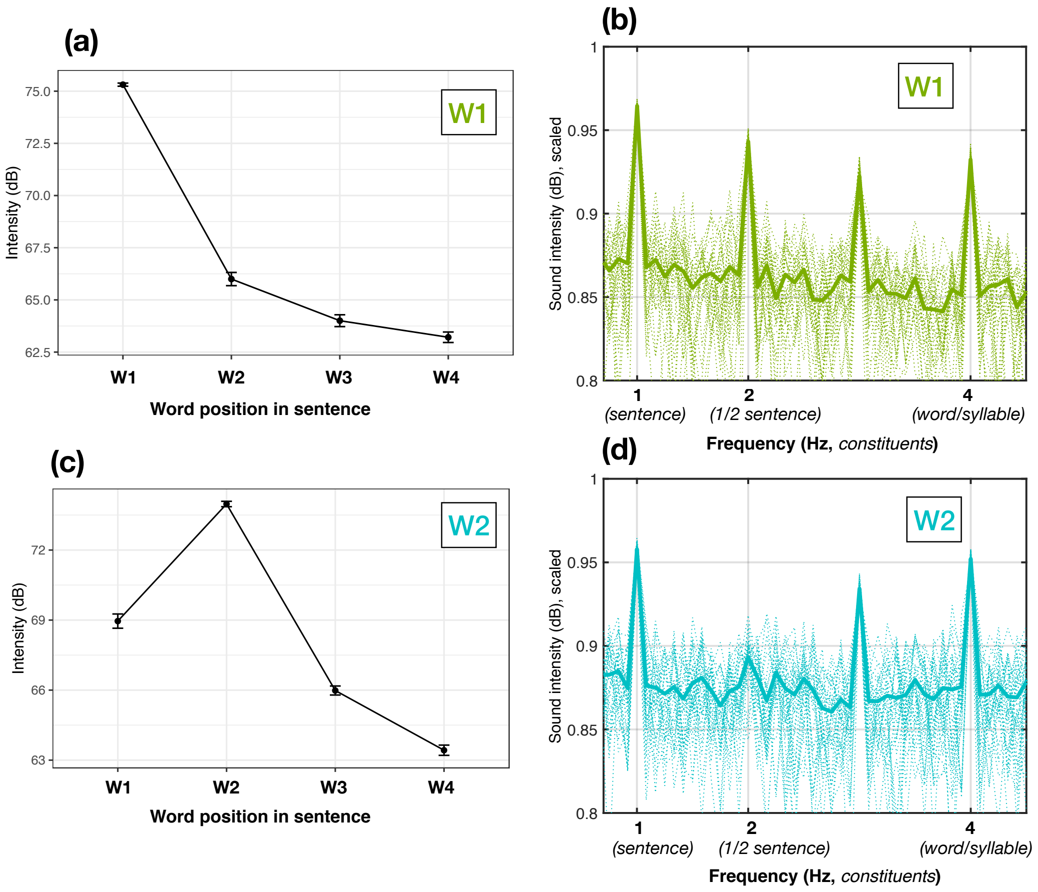
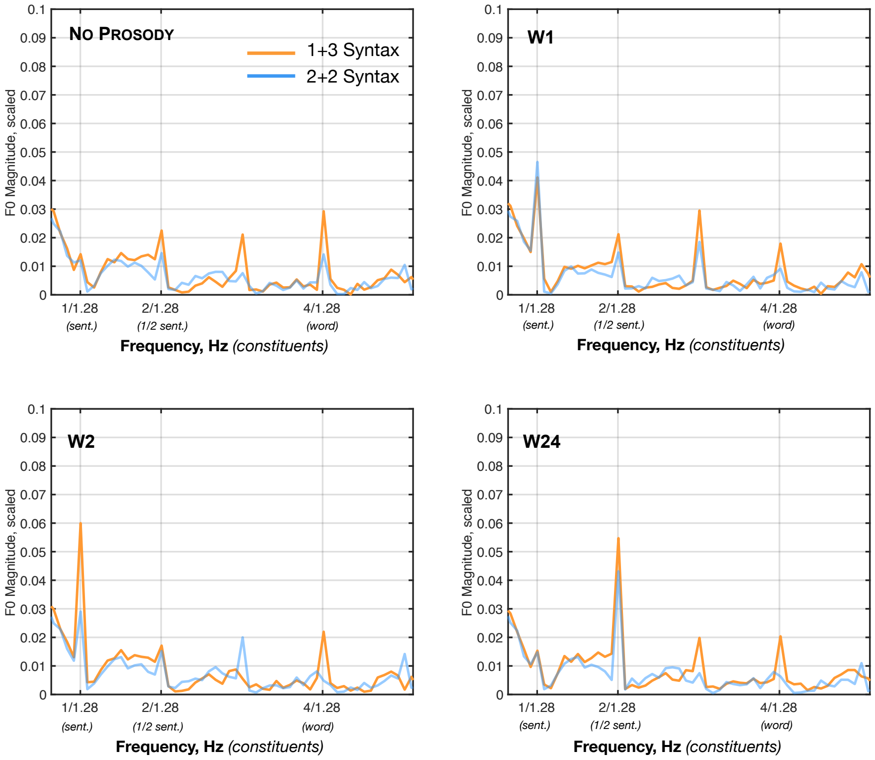


Figure C9. Sound intensity from sentences with naturally produced prosody (pilot data for the main experiment). For our pilot experiment, we created 72 1+3 Syntax sentences largely overlapping with the set of sentences used for the study reported in this manuscript. These sentences were concatenated from words recorded separately and adjusted in pitch (which was flattened), intensity, and length (for each word to be 250 ms long). Several sentences with natural prosody were recorded by asking the native speaker to accentuate either the first (W1) or the second (W2) word of the sentence. Further, we manually adjusted these sentences to be one second long (with each word lasting for 250 ms) and extracted the pitch and intensity contours from these ones with the most representative prosodic contours. These pitch and intensity contours were then imposed on all the sentences concatenated from words with neutralized prosody. On the left are the mean sound intensity values (with standard error bars) for each word position in the 72 sentences used. On the right is the scaled sound envelope spectrum for individual trials separately (thin dotted lines) and in averaged form (thick solid lines). The upper panel represents the 1+3 Syntax sentences with the W1 prosodic contour, for which in the sound envelope spectrum (b), we found a prominent ½ sentence sound intensity peak. The bottom panel represents the analysis of the same 1+3 Syntax sentences with the W2 contour (note no intensity peak at the rate of 2 Hz).

References

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience*, *11*, 1-9. https://doi.org/10.3389/fnhum.2017.00481

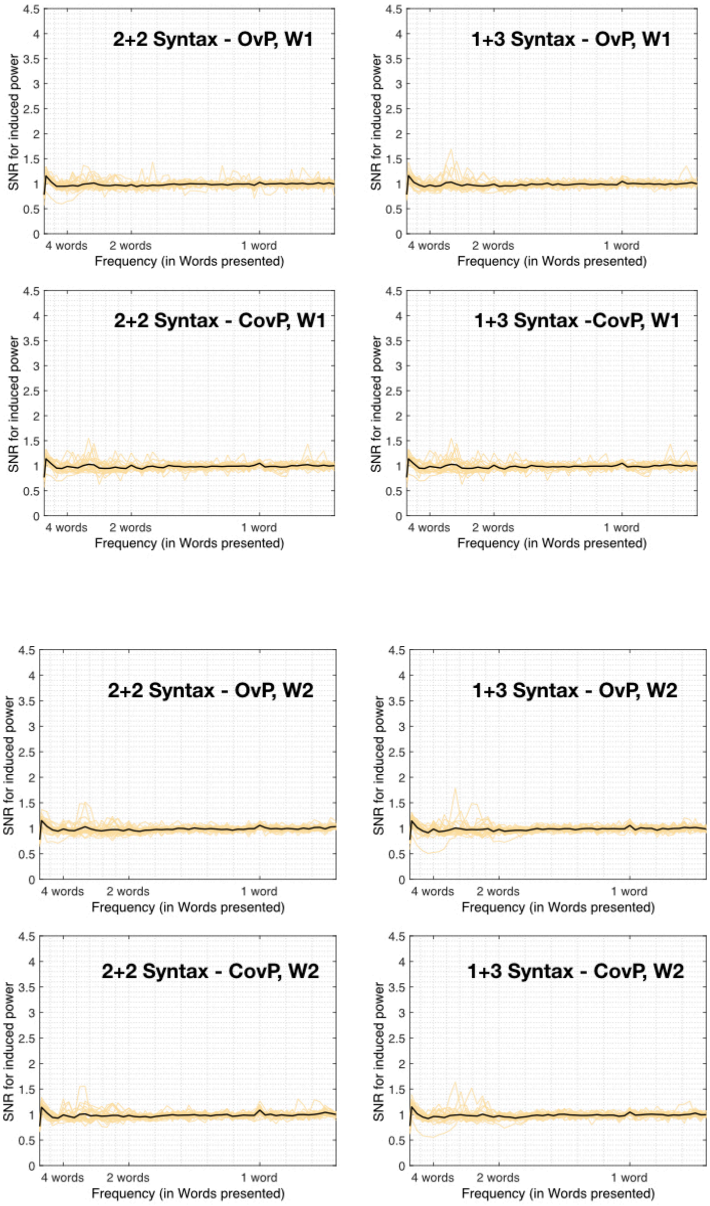
Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, *19*(1), 1-7. https://doi.org/10.1038/nn.4186

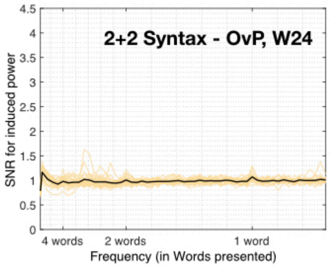


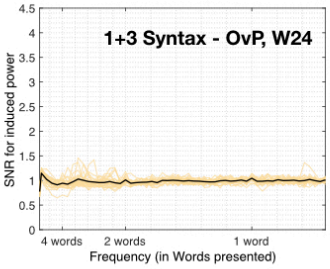
Supplementary Materials *D*. Spectrum of stimulus fundamental frequency (F0) fluctuations for the various prosodic contours. (a) No Prosody sentences: flattened pitch contour. (b+c) In both the W1 and W2 prosodic contours, F0 is modulated at the sentence frequency, resulting in an additional peak at 1/28 (0.78) Hz. The other two (smaller) peaks, similar to sound intensity spectrum, reflect harmonics introduced to increase the naturalness of the stimuli. (d) In contrast to W1 and W2, the W24 contour displays a ½ sentence rate F0 but no major peaks at other frequencies (i.e., no F0 increase was present at the sentence frequency). Note that F0 approximation can be somewhat imprecise due to the fact that F0 values can only be extracted from voiced phonemes and are, therefore, non-continuous. When continuous pitch information is extracted from audio files, missing pitch values are replaced by zeros, and the frequency transform of these continuous pitch values has minor peaks additional to those seen in the intensity spectrum (e.g., the peaks at ½ sentence frequency in W1 and W2).

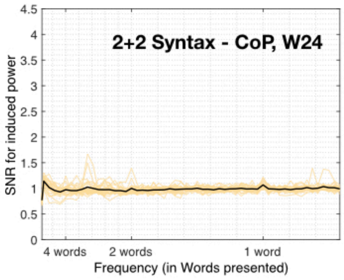
SUPPLEMENTARY MATERIALS *E*

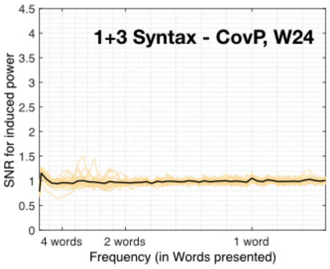
Induced power across experimental conditions. The absence of EEG power peaks at the frequencies of 4 words (sentence frequency) and 2 words (½ sentence frequency) reflects that most EEG responses at these frequencies were aligned in time across trials.

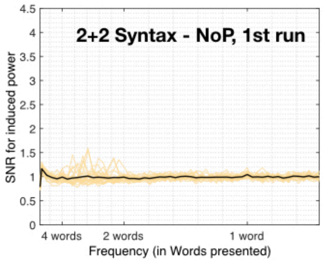


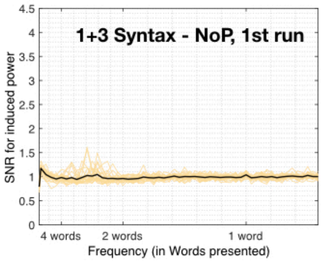

2+2 Syntax - OvP, W24
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

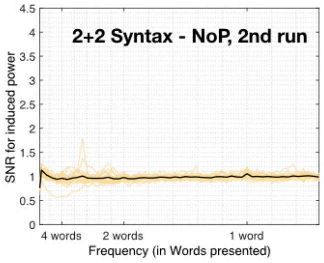

1+3 Syntax - OvP, W24
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

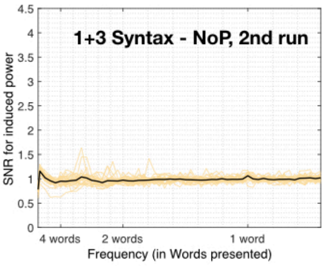

2+2 Syntax - CoP, W24
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


1+3 Syntax - CovP, W24
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


2+2 Syntax - NoP, 1st run
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


1+3 Syntax - NoP, 1st run
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


2+2 Syntax - NoP, 2nd run
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5


1+3 Syntax - NoP, 2nd run
SNR for induced power
Frequency (in Words presented)
4 words
2 words
1 word
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

SUPPLEMENTARY MATERIALS *F*

Distribution of electrodes among the scalp distributions factors used in statistical analysis.

Electrode	Anteriority	Laterality
AF3	Frontal	Medial
AF4	Frontal	Medial
AF7	Frontal	Left
AF8	Frontal	Right
C1	Central	Medial
C2	Central	Medial
C3	Central	Medial
C4	Central	Medial
C5	Central	Left
C6	Central	Right
CP1	Central	Medial
CP2	Central	Medial
CP3	Central	Medial
CP4	Central	Medial
CP5	Central	Left
CP6	Central	Right
CPz	Central	Medial
Cz	Central	Medial
F1	Frontal	Medial
F2	Frontal	Medial
F3	Frontal	Medial
F4	Frontal	Medial
F5	Frontal	Left
F6	Frontal	Right
F7	Frontal	Left
F8	Frontal	Right
FC1	Central	Medial
FC2	Central	Medial
FC3	Central	Medial
FC4	Central	Medial
FC5	Central	Left
FC6	Central	Right
FCz	Central	Medial
Fp1	Frontal	Medial
Fp2	Frontal	Medial
Fpz	Frontal	Medial

FT7	Central	Left
FT8	Central	Right
Fz	Frontal	Medial
O1	Posterior	Medial
O2	Posterior	Medial
Oz	Posterior	Medial
P1	Posterior	Medial
P2	Posterior	Medial
P3	Posterior	Medial
P4	Posterior	Medial
P5	Posterior	Left
P6	Posterior	Right
P7	Posterior	Left
P8	Posterior	Right
PO3	Posterior	Medial
PO4	Posterior	Medial
PO5	Posterior	Left
PO6	Posterior	Right
PO7	Posterior	Left
PO8	Posterior	Right
POz	Posterior	Medial
Pz	Posterior	Medial
T7	Central	Left
T8	Central	Right
TP7	Posterior	Left
TP8	Posterior	Right

SUPPLEMENTARY MATERIALS *G*

T-test results: comparing signal and noise EEG magnitude values in No Prosody (G1), Prosody (G2), and No Semantic Task (G3) experiments. Bonferroni-corrected p-values are reported.

Table G1. T-test results for the No Prosody experiment.

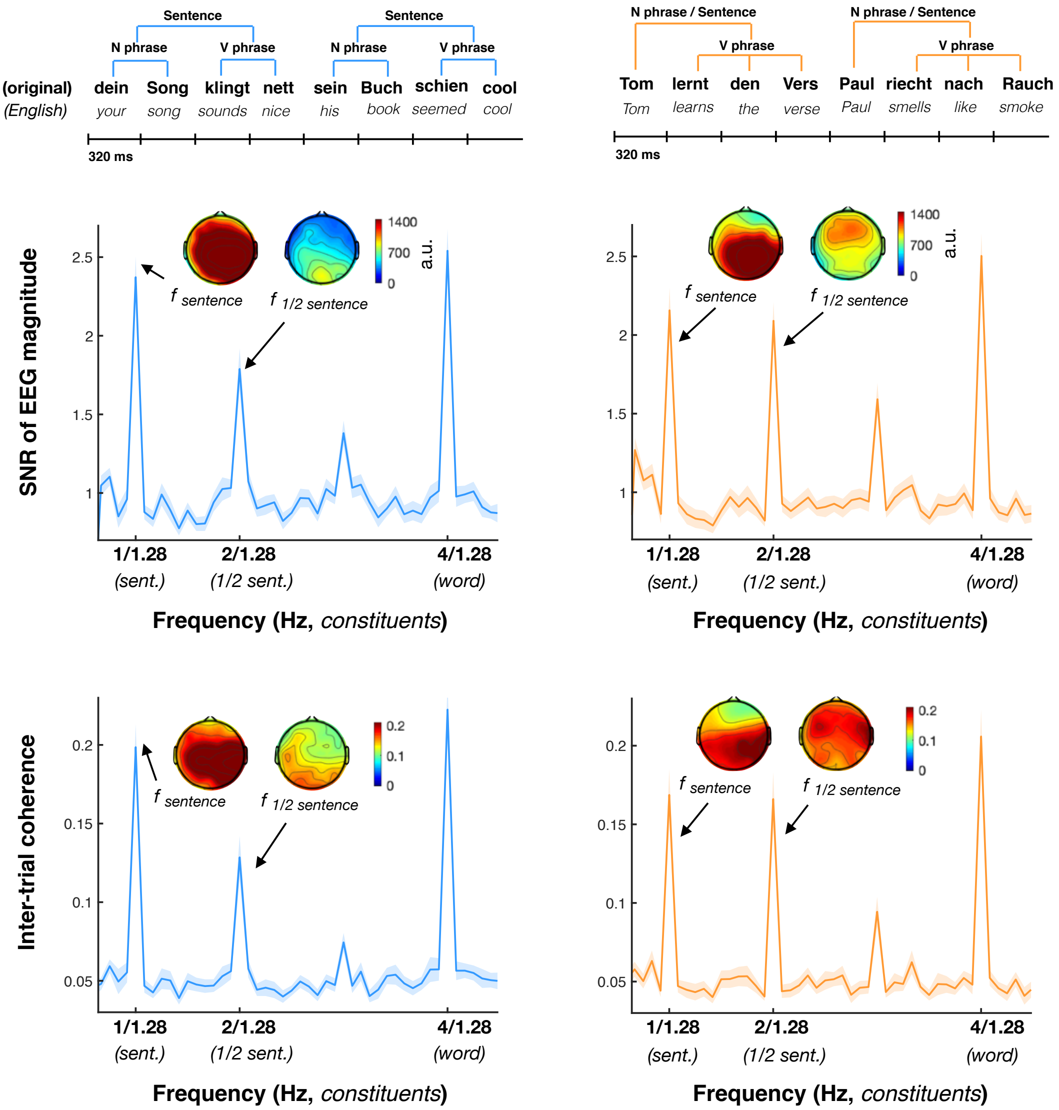
Condition	Frequency	*t*-value	*p*-value
1+3 Syntax, Beginning	Sentence	-7.135	<.001
1+3 Syntax, Beginning	1/2 Sentence	-8.104	<.001
1+3 Syntax, End	Sentence	-2.855	.017
1+3 Syntax, End	1/2 Sentence	-4.177	.001
2+2 Syntax, Beginning	Sentence	-7.630	<.001
2+2 Syntax, Beginning	1/2 Sentence	-4.408	<.001
2+2 Syntax, End	Sentence	-4.770	<.001
2+2 Syntax, End	1/2 Sentence	-3.727	.002

Table G2. T-test results for the Prosody experiment.

Condition	Frequency	*t*-value	*p*-value
1+3, CovP, W24	Sentence	-2.633	.015
1+3, CovP, W24	1/2 Sentence	-5.638	<.000
1+3, CovP, W1	Sentence	-3.314	.003
1+3, CovP, W1	1/2 Sentence	-7.826	<.000
1+3, CovP, W2	Sentence	-2.956	.007
1+3, CovP, W2	1/2 Sentence	-4.044	<.000
1+3, OvP, W24	Sentence	-2.510	.019
1+3, OvP, W24	1/2 Sentence	-5.911	<.000
1+3, OvP, W1	Sentence	-4.007	.001
1+3, OvP, W1	1/2 Sentence	-5.921	<.000
1+3, OvP, W2	Sentence	-3.690	.001
1+3, OvP, W2	1/2 Sentence	-5.734	<.000
2+2, CovP, W24	Sentence	-5.885	<.000
2+2, CovP, W24	1/2 Sentence	-5.812	<.000
2+2, CovP, W1	Sentence	-3.207	.004
2+2, CovP, W1	1/2 Sentence	-5.208	<.000
2+2, CovP, W2	Sentence	-5.605	<.000
2+2, CovP, W2	1/2 Sentence	-4.105	<.000
2+2, OvP, W24	Sentence	-5.446	<.000
2+2, OvP, W24	1/2 Sentence	-8.063	<.000
2+2, OvP, W1	Sentence	-6.169	<.000
2+2, OvP, W1	1/2 Sentence	-5.106	<.000
2+2, OvP, W2	Sentence	-6.564	<.000
2+2, OvP, W2	1/2 Sentence	-3.624	.001

Table G3. T-test results for the No Semantic Task experiment.

Condition	Frequency	*t*-value	*p*-value
1+3, CovP, W24, NoSem	Sentence	-1.671	.129
1+3, CovP, W24, Sem+Pro	Sentence	-3.244	.010
1+3, OvP, W24, Sem+Pro	Sentence	-1.189	.265
1+3, OvP, W24, NoSem	Sentence	-1.587	.147
1+3, CovP, W24, NoSem	1/2 Sentence	-2.686	.025
1+3, CovP, W24, Sem+Pro	1/2 Sentence	-3.488	.007
1+3, OvP, W24, Sem+Pro	1/2 Sentence	-3.597	.006
1+3, OvP, W24, NoSem	1/2 Sentence	-8.250	.000
1+3, CovP, W1, NoSem	Sentence	-3.064	.013
1+3, CovP, W1, Sem+Pro	Sentence	-2.021	.074
1+3, OvP, W1, Sem+Pro	Sentence	-2.819	.020
1+3, OvP, W1, NoSem	Sentence	-4.129	.003
1+3, CovP, W1, NoSem	1/2 Sentence	-4.035	.003
1+3, CovP, W1, Sem+Pro	1/2 Sentence	-4.791	.001
1+3, OvP, W1, Sem+Pro	1/2 Sentence	-4.846	.001
1+3, OvP, W1, NoSem	1/2 Sentence	-1.908	.089



EEG spectrum in the first run of the No Prosody experiment (N=36). On the left panel (from top to bottom): sample sentences with the 2+2 Syntax structure, the SNR of the EEG magnitude spectrum recorded while participants were listening to the 2+2 Syntax sentences, and the corresponding ITC spectrum. On the right panel – the 1+3 Syntax condition. Note that the main syntactic boundary in the 1+3 Syntax condition (i.e., the one between the first and the second words) is not reflected in the spectrum, because the phrases forming the 1+3 Syntax condition are non-isochronous. The lines in the spectrum plots reflect group averages, with the shaded area depicting standard errors of the mean. Scalp maps depict the scalp distribution of the EEG signal (EEG magnitude is in arbitrary units) at the sentence and the ½ sentence frequencies.

SUPPLEMENTARY MATERIALS *I*

Results of the linear mixed-effects models fitted for the analysis

(1) Prosody experiment

	Overt vs. No Prosody Models		*Covert vs. No Prosody Models*	
	EEG Magnitude	*ITC*	*EEG Magnitude*	*ITC*
Constant	0.324*** (0.004)	0.457*** (0.010)	0.325*** (0.005)	0.501*** (0.010)
Anteriority1	0.003*** (0.001)	0.013*** (0.002)	0.004*** (0.001)	0.012*** (0.002)
Anteriority2	-0.004*** (0.001)	-0.020*** (0.002)	-0.006*** (0.001)	-0.021*** (0.002)
d	-0.002*** (0.001)	-0.002 (0.002)	-0.004*** (0.0004)	-0.006*** (0.002)
Syntax1:Anteriority1		0.002 (0.002)	-0.001 (0.001)	
Syntax1:Anteriority2		-0.002 (0.002)	-0.001 (0.001)	
Freq:Anteriority1	0.005*** (0.001)	0.015*** (0.002)	0.005*** (0.001)	0.008*** (0.001)
Freq:Anteriority2	-0.004*** (0.001)	-0.015*** (0.002)	-0.004*** (0.001)	-0.005*** (0.001)
Cont_w1:Anteriority1	-0.001 (0.001)	-0.004 (0.003)	-0.001 (0.001)	-0.003 (0.003)
Cont_w2:Anteriority1	0.001 (0.001)	-0.00001 (0.004)	-0.002** (0.001)	-0.003 (0.003)
Cont_w24:Anteriority1	-0.002 (0.001)	-0.004 (0.003)	-0.001 (0.001)	-0.003 (0.003)
Cont_w1:Anteriority2	0.007*** (0.001)	0.019*** (0.004)	0.004*** (0.001)	0.006 (0.003)
Cont_w2:Anteriority2	0.0001 (0.001)	0.002 (0.003)	0.004*** (0.001)	0.015*** (0.003)
Cont_w24:Anteriority2	-0.001 (0.001)	-0.007 (0.004)	0.005*** (0.001)	0.009** (0.003)
Cont_w1:Syntax1	**-0.004** (0.001)**	**-0.005 (0.004)**	**-0.003*** (0.001)**	**-0.013*** (0.003)**
Cont_w2:Syntax1	**0.004*** (0.001)**	**-0.008* (0.004)**	**-0.003*** (0.001)**	**-0.024*** (0.004)**
Cont_w24:Syntax1	**-0.011*** (0.001)**	**-0.045*** (0.004)**	**-0.011*** (0.001)**	**-0.005 (0.003)**
Freq	0.013*** (0.003)	0.003 (0.002)	0.015*** (0.004)	0.022** (0.008)
Hemi1	-0.006*** (0.001)	-0.014*** (0.003)	-0.006*** (0.0004)	-0.012*** (0.001)
Hemi2	0.007*** (0.001)	0.009*** (0.002)	0.008*** (0.0003)	0.009*** (0.001)
Syntax1	-0.0002 (0.001)	0.007*** (0.002)	0.0002 (0.0005)	0.009*** (0.002)
Cont_w1	0.005*** (0.001)	0.026 (0.015)	-0.004*** (0.001)	-0.012*** (0.004)
Cont_w2	-0.005*** (0.001)	-0.017 (0.012)	-0.009*** (0.001)	-0.037*** (0.004)
Cont_w24	-0.004** (0.001)	-0.039** (0.014)	-0.001* (0.001)	-0.001 (0.003)
Anteriority1:d	0.0002 (0.0004)			
Anteriority2:d	-0.002*** (0.0004)			
Freq:Hemi1	-0.004*** (0.001)	-0.012*** (0.003)	-0.003*** (0.0004)	-0.010*** (0.001)
Freq:Hemi2	0.003*** (0.001)	0.004 (0.002)	0.002*** (0.0003)	-0.001 (0.001)
Freq:Syntax1	-0.007*** (0.001)	-0.015*** (0.002)	-0.008*** (0.0005)	-0.019*** (0.001)
Hemi1:Syntax1	0.002*** (0.0004)	0.006** (0.002)	0.001** (0.0004)	0.002 (0.001)
Hemi2:Syntax1	-0.002*** (0.0004)	-0.002 (0.002)	-0.002*** (0.0003)	-0.005*** (0.001)
Cont_w1:Hemi1	-0.0002 (0.001)	-0.002 (0.004)		
Cont_w2:Hemi1	0.003*** (0.001)	-0.007 (0.004)		
Cont_w24:Hemi1	-0.002 (0.001)	-0.005 (0.004)		
Cont_w1:Hemi2	0.002 (0.001)	0.008* (0.003)		
Cont_w2:Hemi2	0.001 (0.001)	0.003 (0.003)		
Cont_w24:Hemi2	0.003** (0.001)	0.010** (0.003)		
Freq:Cont_w1	0.015*** (0.001)	0.028*** (0.004)	-0.006*** (0.001)	-0.018*** (0.002)
Freq:Cont_w2	0.008*** (0.001)	-0.0001 (0.004)	-0.003*** (0.001)	-0.007** (0.002)
Freq:Cont_w24	-0.007*** (0.001)	-0.029*** (0.004)	-0.003*** (0.001)	-0.006** (0.002)
Anteriority1:Hemi1	-0.005*** (0.001)	-0.009** (0.003)	-0.004*** (0.001)	-0.010*** (0.002)
Anteriority2:Hemi1	0.004*** (0.001)	0.008* (0.003)	0.004*** (0.001)	0.009*** (0.002)
Anteriority1:Hemi2	0.005*** (0.0005)	0.001 (0.002)	0.004*** (0.0004)	0.003* (0.001)
Anteriority2:Hemi2	-0.003*** (0.001)	0.002 (0.003)	-0.002*** (0.001)	0.001 (0.002)
Freq:Syntax1:d	0.0004 (0.0005)	-0.004* (0.001)	0.002*** (0.0003)	0.003*** (0.001)
d:Freq	-0.0003 (0.001)			
Freq:Hemi1:Syntax1	0.003*** (0.0004)	0.006*** (0.001)	0.003*** (0.0004)	0.008*** (0.001)
Freq:Hemi2:Syntax1	-0.003*** (0.0004)	-0.006*** (0.001)	-0.003*** (0.0003)	-0.005*** (0.001)
Syntax1:d	-0.0002 (0.0005)	-0.004* (0.002)	-0.0003 (0.0003)	-0.006*** (0.001)
Hemi1:d		0.005*** (0.002)		
Hemi2:d		-0.003** (0.001)		
Freq:d		0.017*** (0.001)	-0.002*** (0.0004)	-0.004** (0.001)
Cont_w1:Syntax1:d	-0.001 (0.001)	-0.006* (0.003)		0.004 (0.003)
Cont_w2:Syntax1:d	-0.006*** (0.001)	0.008** (0.003)		0.013*** (0.003)
Cont_w24:Syntax1:d	0.0001 (0.001)	0.003 (0.003)		-0.011*** (0.002)
Cont_w1:d	-0.006*** (0.001)	-0.032*** (0.005)		-0.004 (0.003)
Cont_w2:d	0.006*** (0.001)	0.014** (0.005)		0.004 (0.003)
Cont_w24:d	0.002** (0.001)	0.032*** (0.004)		-0.004 (0.002)
Syntax1:Anteriority1:Freq		-0.007*** (0.001)	-0.002*** (0.001)	
Syntax1:Anteriority2:Freq		-0.004*** (0.001)	-0.004*** (0.001)	
Anteriority1:Freq:Cont_w1	-0.006*** (0.001)	-0.020*** (0.003)	-0.001 (0.001)	
Anteriority2:Freq:Cont_w1	0.006*** (0.001)	0.020*** (0.003)	0.005*** (0.001)	
Anteriority1:Freq:Cont_w2	-0.002* (0.001)	-0.011*** (0.003)	-0.002** (0.001)	

Anteriority2:Freq:Cont_w2	0.001 (0.001)	0.003 (0.003)	0.002* (0.001)	
Anteriority1:Freq:Cont_w24	-0.004*** (0.001)	-0.014*** (0.003)	-0.002* (0.001)	
Anteriority2:Freq:Cont_w24	0.007*** (0.001)	0.031*** (0.003)	0.005*** (0.001)	
Freq:Hemi1:Cont_w1	0.003* (0.001)	0.014*** (0.004)		
Freq:Hemi2:Cont_w1	-0.004*** (0.001)	-0.016*** (0.003)		
Freq:Hemi1:Cont_w2	-0.00003 (0.001)	0.006 (0.004)		
Freq:Hemi2:Cont_w2	-0.0003 (0.001)	-0.003 (0.003)		
Freq:Hemi1:Cont_w24	0.006*** (0.001)	0.019*** (0.004)		
Freq:Hemi2:Cont_w24	-0.007*** (0.001)	-0.016*** (0.003)		
Anteriority1:Freq:Hemi1	-0.001 (0.001)	-0.001 (0.002)	-0.002** (0.001)	
Anteriority2:Freq:Hemi1	0.002** (0.001)	0.006** (0.002)	0.002*** (0.001)	
Anteriority1:Freq:Hemi2	-0.001* (0.0005)	-0.007*** (0.001)	0.0002 (0.0004)	
Anteriority2:Freq:Hemi2	0.001 (0.001)	0.004* (0.002)	-0.0002 (0.001)	
Freq:Syntax1:Cont_w1	**-0.0002 (0.001)**	**-0.005 (0.004)**	**0.0003 (0.001)**	
Freq:Syntax1:Cont_w2	**0.002 (0.001)**	**-0.013*** (0.003)**	**0.001* (0.001)**	
Freq:Syntax1:Cont_w24	**0.001 (0.001)**	**-0.008* (0.003)**	**-0.003*** (0.001)**	
d:Freq:Cont_w1	-0.016*** (0.001)	-0.032*** (0.003)		
d:Freq:Cont_w2	-0.005*** (0.001)	-0.006* (0.002)		
d:Freq:Cont_w24	-0.004*** (0.001)	-0.008*** (0.002)		
d:Freq:Syntax1:Cont_w1	**0.0002 (0.001)**	**0.009*** (0.003)**		
d:Freq:Syntax1:Cont_w2	**-0.008*** (0.001)**	**-0.007** (0.002)**		
d:Freq:Syntax1:Cont_w24	**-0.0002 (0.001)**	**0.006* (0.002)**		
Syntax1:Hemi1:d		-0.002 (0.001)		
Syntax1:Hemi2:d		-0.003* (0.001)		
Anteriority1:Cont_w1:Hemi1		-0.003 (0.005)		
Anteriority2:Cont_w1:Hemi1		-0.001 (0.006)		
Anteriority1:Cont_w2:Hemi1		-0.006 (0.005)		
Anteriority2:Cont_w2:Hemi1		-0.001 (0.006)		
Anteriority1:Cont_w24:Hemi1		-0.004 (0.005)		
Anteriority2:Cont_w24:Hemi1		0.002 (0.006)		
Anteriority1:Cont_w1:Hemi2		0.016*** (0.004)		
Anteriority2:Cont_w1:Hemi2		-0.015** (0.005)		
Anteriority1:Cont_w2:Hemi2		0.005 (0.004)		
Anteriority2:Cont_w2:Hemi2		-0.001 (0.005)		
Anteriority1:Cont_w24:Hemi2		0.012** (0.004)		
Anteriority2:Cont_w24:Hemi2		-0.008 (0.005)		
Freq:Hemi1:d		-0.0002 (0.001)		
Freq:Hemi2:d		-0.003* (0.001)		
Syntax1:Cont_w1:Anteriority1		-0.003 (0.003)	0.0004 (0.001)	
Syntax1:Cont_w2:Anteriority1		0.002 (0.003)	-0.001 (0.001)	
Syntax1:Cont_w24:Anteriority1		-0.002 (0.003)	0.001 (0.001)	
Syntax1:Cont_w1:Anteriority2		0.004 (0.003)	-0.002* (0.001)	
Syntax1:Cont_w2:Anteriority2		0.003 (0.003)	0.003** (0.001)	
Syntax1:Cont_w24:Anteriority2		0.015*** (0.003)	0.002* (0.001)	
Syntax1:Anteriority1:Hemi1			-0.0003 (0.001)	
Syntax1:Anteriority2:Hemi1			-0.0004 (0.001)	
Syntax1:Anteriority1:Hemi2			-0.001* (0.0004)	
Syntax1:Anteriority2:Hemi2			0.001 (0.001)	
Freq:Syntax1:Cont_w1:Anteriority1			0.002 (0.001)	
Freq:Syntax1:Cont_w2:Anteriority1			0.001 (0.001)	
Freq:Syntax1:Cont_w24:Anteriority1			0.001 (0.001)	
Freq:Syntax1:Cont_w1:Anteriority2			0.0003 (0.001)	
Freq:Syntax1:Cont_w2:Anteriority2			0.003** (0.001)	
Freq:Syntax1:Cont_w24:Anteriority2			0.001 (0.001)	
Freq:Syntax1:Anteriority1:Hemi1			0.001 (0.001)	
Freq:Syntax1:Anteriority2:Hemi1			-0.0001 (0.001)	
Freq:Syntax1:Anteriority1:Hemi2			-0.001** (0.0004)	
Freq:Syntax1:Anteriority2:Hemi2			-0.0003 (0.001)	
Observations	30,309	30,748	30,313	30,732
Log Likelihood	49,909.500	15,502.900	51,994.510	16,544.180
Akaike Inf. Crit.	-99,654.990	-30,775.790	-103,829.000	-32,990.360
Bayesian Inf. Crit.	-98,972.820	-29,817.430	-103,163.500	-32,582.040

*p**p***p<0.001

(2) No Semantic Task experiment

	1+3 Syntax W24 Models		*1+3 Syntax W1 Models*	
	EEG Magnitude	*ITC*	*EEG Magnitude*	*ITC*
Constant	0.358*** (0.014)	0.501*** (0.015)	0.310*** (0.011)	0.470*** (0.014)
Hemi1	-0.012*** (0.002)	-0.010** (0.003)	-0.006** (0.002)	-0.001 (0.003)
Hemi2	0.016*** (0.002)	0.006* (0.003)	0.010*** (0.002)	0.007* (0.003)
Task1	-0.0001 (0.002)		-0.006*** (0.001)	-0.005* (0.002)
Anteriority1:Hemi1		-0.008 (0.004)	-0.004 (0.003)	
Anteriority2:Hemi1		-0.002 (0.005)	0.003 (0.003)	
Anteriority1:Hemi2		0.009* (0.003)	0.009*** (0.002)	
Anteriority2:Hemi2		-0.008 (0.004)	-0.009*** (0.003)	
Freq1:Task1	-0.007*** (0.001)		0.014*** (0.001)	0.013*** (0.002)
Freq1	-0.021* (0.011)	-0.037*** (0.011)	0.011 (0.010)	-0.008 (0.011)
Prosody1	-0.006*** (0.001)		0.003* (0.001)	
Hemi1:Task1	-0.002 (0.002)		0.005* (0.002)	0.010** (0.003)
Hemi2:Task1	0.007*** (0.002)		0.001 (0.002)	0.001 (0.003)
Anteriority1:Freq1			0.005** (0.002)	
Anteriority2:Freq1			-0.007*** (0.002)	
Anteriority1:Task1			0.003 (0.002)	0.005 (0.003)
Anteriority2:Task1			0.007*** (0.002)	0.021*** (0.003)
Hemi1:Freq1:Task1			-0.001 (0.002)	0.001 (0.003)
Hemi2:Freq1:Task1			0.008*** (0.002)	0.007** (0.003)
Anteriority1:Freq1:Task1			0.001 (0.002)	
Anteriority2:Freq1:Task1			0.008*** (0.002)	
Hemi1:Freq1	0.0004 (0.002)		0.005* (0.002)	0.009** (0.003)
Hemi2:Freq1	-0.007*** (0.002)		-0.007*** (0.002)	-0.014*** (0.003)
Task1:Prosody1	-0.013*** (0.001)			
Freq1:Prosody1	0.013*** (0.001)			
Task1:Freq1:Prosody1	-0.007*** (0.001)			
Anteriority1		-0.00002 (0.003)	0.009*** (0.002)	0.022*** (0.003)
Anteriority2		-0.013*** (0.003)	-0.019*** (0.002)	-0.035*** (0.003)
Observations	4,880	4,960	4,960	4,960
Log Likelihood	4,416.874	2,961.485	4,951.937	2,783.170
Akaike Inf. Crit.	-8,797.748	-5,894.971	-9,845.875	-5,526.339
Bayesian Inf. Crit.	-8,680.876	-5,803.842	-9,657.109	-5,396.156

*p**p***p<0.001

NEURAL TRACKING OF SENTENCE CONSTITUENTS: WHAT DO WE REALLY KNOW?

In Manuscript 1, we addressed the questions of (1) whether prosodic processing is reflected in the neurophysiological spectral power peaks at frequencies of sentence constituents, and (2) whether the interaction between syntactic and prosodic processing is reflected in the EEG responses at these frequencies. We demonstrated both the effects of prosody *per se* and their interaction with syntactic structure processing.

The next question logically arising from our data relates to the biologically and psycholinguistically plausible mechanistic interpretation of these and other results of frequency tagging research in sentence processing. Do they reflect neural entrainment of endogenous neural oscillations to the structure of the sentences? Or, can they reflect a series of transient EEG responses? Can additional non-syntactic variables account for the results in frequency tagging studies of sentence processing? And related to these considerations, what are the advisable practices for the future use of this novel technique in sentence processing research? These questions are addressed in Manuscript 2. To avoid superfluous overlap between the two manuscripts, in the frame of the current thesis, Manuscript 1 data are not summarized in detail in Manuscript 2. This will change in the final version of Manuscript 2 that will be submitted for publication in a peer-reviewed journal.

MANUSCRIPT 2

Frequency tagging of sentence processing in neurophysiological data: remarks on methodology

Anastasia Glushko[1], Max Wolpert[1], David Poeppel[2,3], Karsten Steinhauer[1,4]

Affiliations:

[1]Centre for Research on Brain, Language and Music (Montreal, Canada)

[2]Max Planck Institute for Empirical Aesthetics (Frankfurt, Germany)

[3]Department of Psychology, New York University (New York City, NY, USA)

[4]School of Communication Sciences and Disorders, McGill University (Montreal, Canada)

ABSTRACT

Recent research used the frequency tagging technique and reported that peaks of neurophysiological power arise at frequencies of sentence constituents when listeners process isochronously organized sentences. This was claimed to reflect that humans process online syntactic hierarchies online, with the assumed underlying neural mechanism being the neural entrainment of slow neural oscillations to syntactic structure of the stimuli. We conducted a methodological investigation of the frequency tagging approach in its use for studying sentence processing. We report two data simulations and outline methodological considerations, concluding that results reported in sentence-level frequency tagging studies to date can be interpreted in the absence of the notion of syntactic hierarchy or slow oscillatory activity. Among other factors, word-level information such as word frequency as well as linear word chunking driven by transitional probabilities between words can modulate spectral power peaks at low frequencies. Taking into account these and other concerns, we discuss advisable methodological practices for experiment creation and data analysis that should be considered by future frequency tagging studies of sentence processing.

INTRODUCTION

Researchers have used numerous experimental techniques in an attempt to understand how humans produce and perceive language. Psycholinguistic research addresses this question through studying the cognitive mechanisms of language processing, while the field of neurolinguistics focuses on how cognition maps onto the biological substrate of the brain. Many of the insights in both psycho- and neurolinguistics, especially those tapping into the temporal dynamics of language processing, have been produced by studies deploying so-called "violation-based" experimental paradigms (for instance, contrasting the processing of *The spoilt child throws the toys on the floor* and **The spoilt child throw the toys on the floor*, Hagoort, Brown, & Groothusen, 1993). While violation-based paradigms have been enormously popular, their ecological validity has been criticized (e.g., Baayen, 2014): it is unclear to what extent the processing of linguistic errors reflects mechanisms involved in natural language processing consisting largely of grammatically correct sentences.

More recently, several new approaches have been introduced to complement violation paradigms. One such methodology, dubbed "frequency tagging", has been implemented for studying sentence processing in a more ecological format. The first frequency tagging study of syntactic processing (Ding, Melloni, Zhang, Tian, & Poeppel, 2016) used entirely grammatical spoken sentences to measure temporal processing dynamics[1]. This work received substantial interest in the community – the original paper has been cited almost 400 times in the past 4 years. However, although frequency tagging has promising potential and has a strong history in the field of sensory neuroscience (Bohórquez & Özdamar, 2008; Brown & Norcia, 1997;

---

[1] Neuroimaging studies using functional magnetic resonance imaging often do contrast correct sentences of different types (for example, sentences with object- vs. subject-relative clauses; Peelle, McMillan, Moore, Grossman, & Wingfield, 2004; for more examples, see Kaan & Swaab, 2002) but have a very limited time resolution restricting the range of research questions that can be addressed by this research.

Snyder, 1992), there are still a number of methodological issues that constrain extending this technique to the study of language processing.

The present paper aims to assess the research potential of the frequency tagging methodology for language research, and attempts to address related challenges faced by the field. We start with a brief review of the frequency tagging technique and its use in language research to date (focused on studying multiple-word processing – i.e., syntax, semantics, and prosody). We then turn to discussing restrictions that frequency tagging places on the choice of stimulus materials and experimental design in language studies. In a separate section, we highlight in more detail several data analysis opportunities that are useful in uncovering the dynamics of sentence processing using this methodology. We then conclude by considering possible applications and tentative best practices for future frequency tagging studies.

WHAT IS FREQUENCY TAGGING?

The frequency tagging technique has been extensively used in research on sensory processing to capture steady-state evoked potentials (SSEP) during repetitive, rhythmic presentation of visual (e.g., Regan, 1965; Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015), auditory (e.g., Rees, Green, & Kay, 1986; Picton, John, Dimitrijevic, & Purcell, 2003), and somatosensory stimuli (e.g., Snyder, 1992; Tobimatsu, Zhang, & Kato, 1999). The basic experimental design dates back to the 1930s (Adrian & Matthews, 1934; cited in Norcia et al., 2015), and follows a similar procedure across sensory modalities: sensory stimuli (for example, flickering lights) are presented or modulated at a specific frequency (often above 10 Hz), and the processing of this stimulus stream is reflected in a non-transient brain response at the same frequency, i.e., SSEPs. Originally, SSEPs were described in the time domain as a repetitive pattern of EEG responses echoing in their shape and timing the changes in the stimulus. The advantages of their investigation in the frequency domain (as power increases at specific frequencies) were, however, immediately evident (Regan, 1982). These advantages include, first, the high signal-to-noise ratio (SNR) of the technique, and second, its capacity to clearly separate responses at distinct frequencies (and hence, associated with processing distinct features in the stimuli) that might be difficult to dissociate from one another in the time domain (e.g., due to their partial or complete overlap in time). Both the high SNR and the ease of detecting the processing of a specific stimulus from the brain signal, due to the effects being confined to a known frequency bin, lead the SSEPs to become a promising tool to assess sensory abilities in clinical contexts (e.g., hearing ability testing, Cone-Wesson, Dowell, Tomlin, Rance, & Ming, 2002; Canale, Lacilla, Cavalot, & Albera, 2006, or developing brain-computer interfaces, e.g., Gao, Xu, Cheng, & Gao, 2003). Moreover, the passive nature of the paradigm meant that there was essentially no need to rely on participants' behavioural responses, and hence the paradigm was deemed suitable for studying special populations (e.g., newborns and patients with limited communication abilities, e.g., Nakabayashi, Kurokawa, & Yamamoto, 2001; Luts, Desloovere, & Wouters, 2006).

Frequency tagging studies have shifted from first being a means to investigate low-level sensory feature processing, to more complex processing, such as integration of face features (Boremanse, Norcia, & Rossion, 2014), musical beat perception (Nozaradan, Peretz, Missal, & Mouraux, 2011), and statistical learning (Buiatti, Peña, & Dehaene-Lambertz, 2009; Kabdebon, Pena, Buiatti, & Dehaene-Lambertz, 2015). A major common finding across these studies is that frequency tagging can not only capture processing of low-level stimulus characteristics (i.e., bottom-up processing) but can reflect top-down mechanisms driven by experience or higher-level integrative processing mechanisms. Buiatti and colleagues (2009), for instance, presented participants with nonword sequences with different transitional probabilities between distinct syllables (i.e., high between some syllables, forming a chunk, and low between others, forming between-chunk boundaries). They found increases of EEG power corresponding to the frequency at which syllable chunks were presented. Nozaradan, Peretz, and Mouraux (2012) further showed that perceived musical beats elicit neurophysiological spectral power peaks at frequencies that are not tied to a corresponding change in acoustic cues (see also Nozaradan, Schönwiesner, Keller, Lenc & Lehmann, 2018).

This potential of the frequency tagging technique to capture top-down processing and dissociate it from stimulus-driven, bottom-up cognitive mechanisms is attractive for language research. Physical characteristics of speech do not carry the complex meaning and structure that make language our primary means of communication. Instead, they serve as signs, which we process relying on top-down cognitive mechanisms heavily grounded in our experience with language use (Tomasello, 2000). Frequency tagging seems to provide a tool for tracking these top-down cognitive mechanisms as they are unveiled through cortical dynamics. Moreover, this paradigm would allow for a shift from violation-based paradigms (e.g., Neville, Nichol, Barss, Foster, & Garrett, 1991; Osterhout & Mobley, 1995; Petersson, Forkstam, & Ingvar, 2004) towards directly studying processing of semantically and grammatically legal sentences directly. Both these advantages became the main rationale for the first frequency tagging study of sentence processing (Ding et al., 2016) that targeted one of the most controversial topics in language research: the existence of cognitive representations of the hierarchical organization of language (Chomsky, 1959; Berwick, Friederici, Chomsky, & Bolhuis, 2013; Christiansen & Chater, 1999; Frank, Bod, & Christiansen, 2012).

Ding and colleagues (2016) presented Mandarin Chinese native speakers with spoken stimuli consisting of isochronous four-syllable linguistic constructions (i.e., sentences or phrases). Syllables in every construction were of identical length (250 ms). The stimuli were presented in sequences of twelve constructions each, with no pauses between syllables, words, or any larger constituents. All constructions in a trial of twelve had identical syntactic organization, which allowed the authors to tag syntactic constituent frequencies, similar to frequency tagging in SSEP studies on sensory processing. In one of the conditions, four syllables formed a sentence in which the first two syllables made up a noun phrase, and the last two a verb phrase (for example, "bro-thers ar-gue", henceforth 2+2 Syntax). Because the syllables, phrases, and sentences within a trial were isochronous, syllables appeared at a frequency of 4 Hz (4 syllables per second), phrases at a frequency of 2 Hz (2 phrases per second), and sentences at a frequency of 1 Hz (1 sentence per second). While participants listened to the sentence sequences, their MEG was recorded, and clear peaks of MEG spectral power were found at constituent frequencies. Because no acoustic cues were present at phrase or sentence boundaries, the authors concluded that the MEG responses reflected the tracking of the hierarchical sentence syntactic structure, the neurocognitive reality of which has been debated for decades.

Although the frequency tagging approach steps away from violation paradigms, it still relies on contrasting highly constrained experimental conditions and does not represent fully naturalistic language assessment (for review of the latter, see Alday, 2019). In the study by Ding and colleagues (2016), several syntactic constructions were contrasted with the 2+2 Syntax sentences. One of them was represented by four-syllable verb phrases, in which the first word was a verb followed by a phrase boundary, followed by a three-syllable verb complement ('fry to-ma-to', henceforth 1+3 Syntax). The other two control conditions were two-syllable phrases (not embedded into a larger 4-syllable construction) and random strings of four syllables not chunked in any larger units. The MEG power peaks were again consistent with the specific frequencies tagged in these materials, whereas the "untagged" frequencies (such as the 2 Hz frequency in the 1+3 Syntax constructions and the 1 Hz in the two-syllable phrases) were not characterized by MEG power increases.

Ding and colleagues' (2016) original experiments (that also included testing Chinese and English materials in English native speakers without knowledge of Chinese) were followed up with an electrocorticography (ECoG) study that identified neural sources that only partially overlapped for cortical responses to phrase and sentence frequencies. This result demonstrated that some of the neurocognitive mechanisms of phrase and sentence tracking might be distinct (this was also shown using a combined MEG and structural magnetic resonance paradigm by Sheng and colleagues, 2018). The later EEG version of the study (Ding et al., 2017) replicated the emergence of spectral power peaks at syntactic constituent frequencies in 2+2 Syntax sentences. By referring to analysis of time-domain data and data simulations, researchers have claimed that the spectral power peaks at sentence constituent frequencies represent slow oscillations at the respective frequencies (and *not* transient responses at constituent boundaries or amplitude-modulated faster, gamma oscillations; Zhou, Melloni, Poeppel, and Ding, 2016; Zhang & Ding, 2017). The authors proposed that low-frequency neural activity is ‘tracking’ sentence constituents, and the tracking is evident at the latest during the presentation of the second sentence (Zhang & Ding, 2017). Moreover, it demonstrated a positive relationship between sentence level cortical responses and performance on a behavioral task aimed at identifying trials with semantically implausible sentences (Ding et al., 2017). The EEG frequency tagging responses to linguistic structure have further been shown to be dependent on the task and strongly modulated by participants’ attention (Ding et al., 2018; Jin, Zou, Zhou, & Ding, 2018; Makov et al., 2017). A similar link between performance and the frequency tagged responses to phrases was shown to be stronger in correctly than incorrectly comprehended sentences in a slightly different paradigm by Keitel, Gross, and Kayser (2018). In the latter study, the authors moved away from full isochronicity in their design and used phrases of differing length while investigating sound envelope tracking within the frequency range of 0.6-1.3 Hz.

Summarizing the results of the first frequency tagging studies of sentence processing, their neurophysiological data seemed to reflect top-down processes relevant for speech comprehension and specific to the frequencies of sentence constituents. The claim of the original studies (Ding et al., 2016; 2017) was that they had presented evidence for cortical tracking of hierarchical sentence structure. However, as we will outline in the next section of the paper, the contribution of syntactic and other processing mechanisms (e.g., Frank & Yang, 2018; Glushko, Poeppel, & Steinhauer, in preparation) to the cortical responses at frequencies of sentence constituents has been questioned and needs further investigation. Further, continuing development of accessible tools for EEG analysis of frequency tagging data in language research would be essential for efficient use of this paradigm on populations with low communication abilities (one of the potential advantages of the frequency technique). Given the high interest in the frequency paradigm from the language research community (incl. its use in bilingualism and sleep research, Blanco-Elorrieta, Pylkkänen, & Poeppel, 2019; Makov et al., 2017), we see the need to outline potential methodological constraints for studying sentence-level processing using frequency tagging and suggest some solutions for the limitations of existing research. Note that throughout the paper, we will often refer to EEG data but unless otherwise specified, the methodological details discussed are applicable to MEG data as well.

EXPERIMENTAL DESIGN CONSIDERATIONS

CONTROL OF NON-SYNTACTIC VARIABLES. As highlighted above, neurolingusitic sentence processing research has often relied on violation paradigms. The need for such a seemingly unnatural experimental manipulation comes from the necessity to create so-called “minimal contrasts” between experimental conditions. Presenting participants with language materials that differ from one another only on one dimension (for example, grammatical complexity of the sentence) while keeping other dimensions (such as sentence length and semantic complexity) constant or balanced allows for the characterization of the architecture of distinct neurocognitive constructs and their subcomponents (e.g., Steinhauer & Drury, 2012). For instance, in the statistical learning study by Buiatti and colleagues (2009), the specific nonwords across the two (chunk vs. random) conditions in the experiment were the same but presented in a different order. This created a minimal contrast between the conditions, and the authors were able to attribute the EEG differences at chunk frequency to chunk learning, or segmentation. In contrast, when Ding and colleagues (2016, 2017) contrasted 2+2 and 1+3 Syntax constructions, the lexical items used in the two experimental conditions were different, as were the possibilities for insertion of covert prosodic boundaries (see Glushko et al., in preparation). In other words, syntactic differences between the conditions were confounded with differences in other language characteristics (e.g., semantics and prosody), resulting in ambiguous data. Such ambiguities can and should be avoided. Below, we will discuss a number of important linguistic dimensions that have to be considered and controlled for while developing the stimuli for a frequency tagging study on sentence processing.

Word frequency. One of the variables known to affect word-level EEG responses is word frequency: words appearing less frequently in the input are associated with smaller amplitude N400s – a negative event-related potential elicited by any word, and being modulated by a variety of semantic factors assumed to reflect lexical semantic processing (Halgren et al., 2002).

Word frequency can and does correlate with syntactic structure: for example, syntactic structure dictates possible syntactic category choices, and syntactic categories can, on average, have contrasting word frequencies (e.g., function words are more frequent than content words). Given this alignment of two conceptually different linguistic features, the question arises of whether neural responses modulated by word frequency alone can produce spectral power peaks at sentence constituent frequencies. We tested this hypothesis, first, by simulating word-level EEG responses using a 4 Hz sinusoid and amplitude-modulating this simple sine wave using the word frequency corresponding to each word in the stimuli used by Ding and colleagues (2016; see Figure 1). If we use Ding and colleagues' (2016) stimuli materials to modulate word-level power of cortical responses purely using the frequency of every word in their experiment, we can see that both the sentence and the phrase spectral peaks reach significance as well (Figure 1). Next, we used the same logic to simulate data for more realistic word-level responses consisting of small early onset components and a following N400, the resulting spectra are strikingly similar to the ones in previous studies (e.g., Ding et al., 2017; see Figure 2).

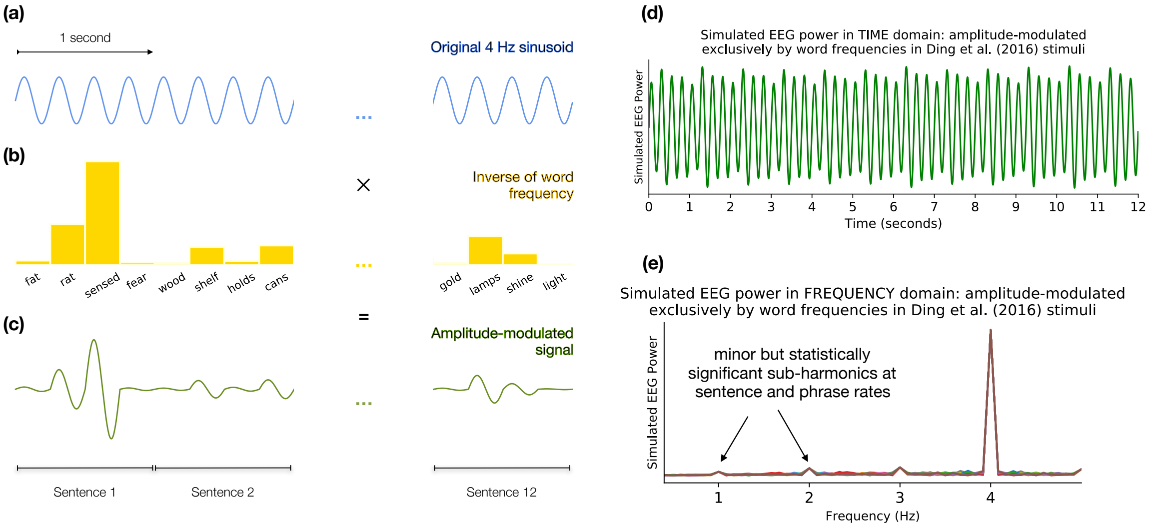
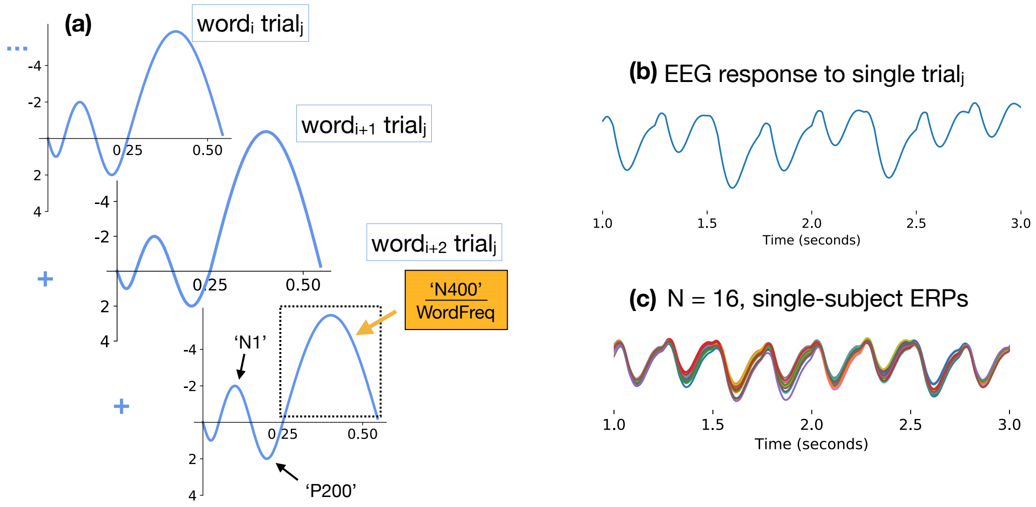


Figure 1. Schematic simulation of the word-level MEG/EEG responses using English materials from the original Ding and coauthors' (2016) study based purely on modulating amplitudes of simulated word-level (4 Hz) responses depending on the word frequency of the corresponding words. The amplitude modulation of each EEG trial was produced in the following way. Each trial's data were concatenated from the data for single words, where each word was a 250-ms-long, 4-Hz pure sinusoid sampled at 256 Hz. All simulated time-domain cortical responses for separate words initially had identical amplitudes (a). These amplitudes were later modulated by the inverse of the log-transformed word frequency (b). Word frequencies were retrieved from the Sketch Engine application (Kilgarriff et al., 2014). Multiplication of the 4 Hz responses with the word frequency values resulted in (c), an amplitude-modulated 4 Hz sinusoid. The simulated data in (d) are created following the design of the original Ding et al. (2016; 2017): simulated data from 22 trials, with 12 sentences per trial, were created for 16 participants, all having a pseudorandom arrangement of the 22 trials used (sentences were randomly assigned to trials as long as they repeated an approximately equal amount of times). Single-subject time-domain averages were transformed into frequency domain using fast Fourier transform (e). Using paired, one-tailed t-tests, we found that the sentence-, phrase-, and word-level peaks were all significantly greater than noise (all *p*-values < .001). Python code used to generate the figures in this article is available upon request.



(d) N = 16, single-subject spectra

(e) N = 16, single-subject spectra

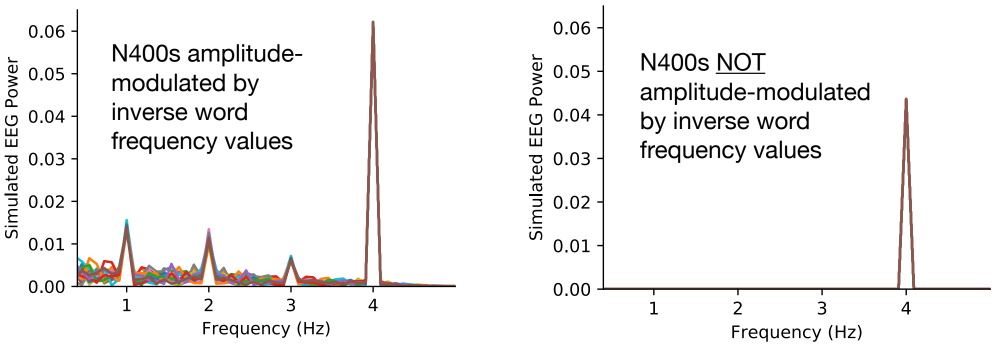


Figure 2. Schematic simulation of the (m)ERP responses in Ding and coauthors' studies (2016; 2017). This simulation was produced by creating ERP profiles for each word that were composed of small-amplitude auditory onset components (N1, P200) and a following N400 response (a) lasting altogether for 550 ms. The N400 elicited by a specific word spilled over to the next word (starting 250 ms after the previous one), resulting in the summation of the ERPs for adjacent words with a 300 ms overlap. The N400 was the only component amplitude-modulated by the inverse of the associated word frequency. Time-domain data were low-pass filtered at 25 Hz (similar to Ding et al., 2016). A sample of the resulting trial's signal can be seen in (b). We produced 22 trials for 16 participants, creating 16 single-subject ERPs. These were transformed into frequency domain using fast Fourier transform. Prominent peaks at the frequencies of sentence constituents (1 and 2 Hz) can be seen in (d) even though neither hierarchical features nor slow oscillatory activity played any role in our simulation. Importantly, in a control simulation (e), we reproduced the simulated data following steps (a)-(d) with one exception: the N400 amplitudes were not modulated by inverse word frequency values. In this case, the peaks at 1 and 2 Hz disappear (e).

Word frequency is not the only measure of the occurrence and co-occurrence of words. Frank and Yang (2018) used a computational model without hierarchical features that was based purely on distributional semantics (e.g., the sharedness of the context in which words appear): that is, they simulated the data in Ding and colleagues' (2016) study without using any notion of syntax or hierarchy. They were able to closely replicate Ding et al.'s results showing the emergence of prominent spectral power peaks at both sentence and phrase frequencies, accounting for the main results of that study with non-syntactic and non-hierarchical features. It is still unclear to what extent semantic processing contributed to Ding and coauthors' (2016, 2017) data, but Frank and Yang's (2018) simulation demonstrates that semantics cannot be discounted in explaining their low-frequency spectral peaks.

Frank and Yang's (2018) and our own simulation results are not surprising. Zhou and colleagues (2016) extensively discussed the various spectral patterns that can be produced by EEG data in the absence of slow waves themselves. A 20 Hz sinusoid ("carrier frequency") amplitude-modulated at 1 Hz ("modulation frequency") resulted in spectral peaks at the fundamental, 20 Hz frequency, as well as at the frequencies at the distance of the modulation frequency (1 Hz) of the fundamental frequency (i.e., 19 and 21 Hz). Similarly, if the carrier 4 Hz word-level EEG response is amplitude-modulated at the frequency of 2 Hz, spectral power peaks will emerge at 2, 4, and 6 Hz (see Figure 3). Given existing evidence of multiple word-level, non-syntactic factors (such as imageability or concreteness, Swaab, Baynes, & Knight, 2002; Barber et al., 2013) can affect word-level EEG responses, these must be taken into account by future studies.

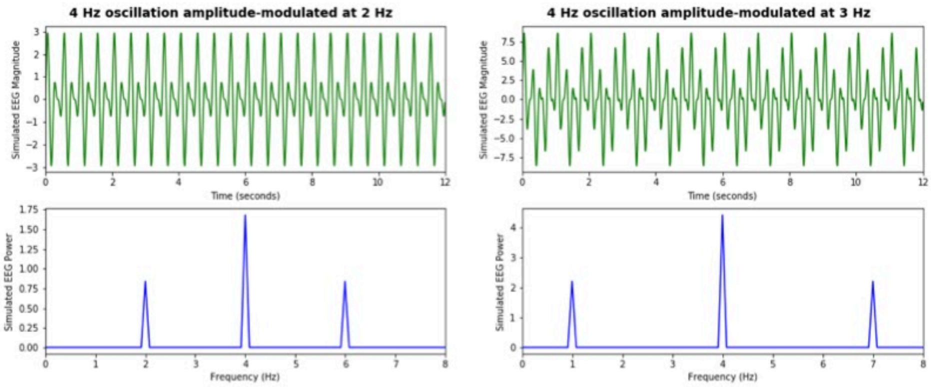
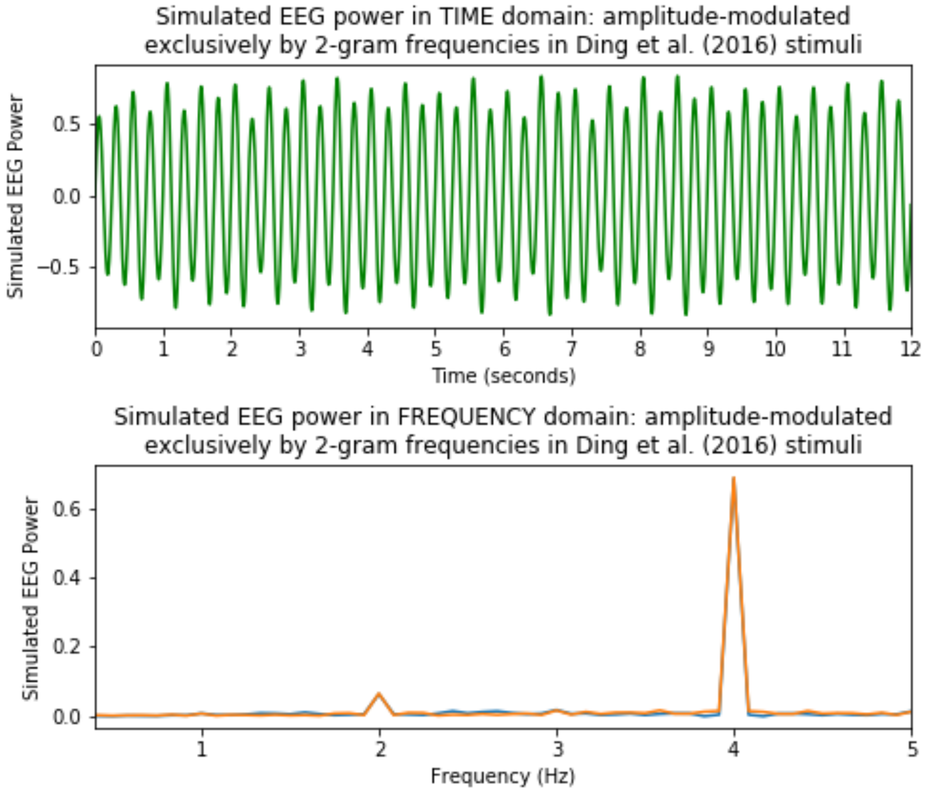


Figure 3. Amplitude-modulation of a simulated "word-level" 4 Hz oscillation at the frequency of 2 Hz (left panel) and 3 Hz (right panel). The upper row represents the simulated time domain response, the bottom row depicts the corresponding frequency spectra. In both cases, a spectral peak is seen at the carrier frequency (4 Hz). In the case of the 2 Hz modulation (left), an additional peak is produced at the frequency of 2 Hz (i.e., the phrase frequency in 1-second long 2+2 Syntax sentence in Ding et al., 2016; 2017). In the case of the 3 Hz modulation (right), an additional peak is produced at the frequency of 1 Hz. Note that a 1 Hz peak (but not the 2 Hz peak) is also present in the spectrum of the 4 Hz carrier wave amplitude-modulated at both 1 Hz and 2 Hz.

Transitional probabilities. Transitional probabilities are also relevant for research on sentence-level processing mechanisms. The probability of a word's occurrence given the words preceding it is decoded by the brain (DeLong, Urbach, & Kutas, 2005) and can, just like word frequency, be correlated with syntactic structure (Reali & Christiansen, 2005). The relevance of transitional probabilities to the study of language processing emerged from an old debate between frequentist and generativist approaches to language acquisition and processing (e.g., Chater & Christiansen, 2015). Researchers have emphasized the importance of counterbalancing transitional probabilities across and within syntactic constituents in frequency tagging research (Tavano, 2017), especially to distinguish between hierarchical, rule-based processing and linear, transitional probability-based chunking. Ding and colleagues (2016) presented evidence that when transitional probabilities between words are manipulated within a short experiment (i.e., when calculations are specific to the probability of the transition appearing within the experiment and not in the speech corpus), MEG responses at sentence constituent frequencies are unaffected. While within-experiment transitional probabilities can have an effect on processing, transitional probabilities acquired through natural language input over the lifespan should be taken into account as well. This point is demonstrated in Figure 4, where we simulated word-level responses to Ding and colleagues' (2016) experimental materials that were amplitude-modulated by the inverse frequency of each word's co-occurrence metric given the word immediately preceding it (i.e., 2-gram frequency). In contrast to the word frequency simulation, the sentence-level (1 Hz) spectral peak is not evident from the data, but a robust, phrase-level (2 Hz) power peak emerges when 2-gram frequencies alone are taken into account.


Simulated EEG power in TIME domain: amplitude-modulated exclusively by 2-gram frequencies in Ding et al. (2016) stimuli

Simulated EEG power in FREQUENCY domain: amplitude-modulated exclusively by 2-gram frequencies in Ding et al. (2016) stimuli

Figure 4. Schematic simulation of word-level MEG/EEG responses to the English materials used in Ding and coauthors' (2016) study based purely on modulating amplitudes of these responses depending on 2-gram frequency extracted using a Bing web search. The simulation procedure was kept the same as in the case of word frequency (see Figure 1), but only the data from 2 participants (22 12-sentence long trials each) are presented. The amplitude of the simulated cortical response to each word was multiplied with the inverse scaled frequency of this word occurring immediately after the preceding word (the frequency was capped at the 80th percentile to avoid outliers).

Note that while in this demonstration we amplitude-modulated single-word responses, it is conceivable that pairs of words (or even longer word sequences) with high transitional probabilities between single words will be processed as a chunk, with the chunk as a unit eliciting a low-frequency brain response. This low frequency response would not reflect hierarchical syntactic processing but rather linearly organized, experience-based processing of input statistics. Attributing frequency tagging effects to processing of hierarchical sentence structure is feasible only when hierarchical and linear grouping are unconfounded.

Phonotactics and acoustic information. Phonological and acoustic structure can also correlate with syntactic organization of sentences. In our own pilot data, we recorded a German native speaker producing 2+2 Syntax structures such as “[Dein Lied] [klingt nett]” (English: “[Your song] [sounds nice]”) and carefully controlled pitch, intensity, and length within each word, following Ding and colleagues’ (2016) procedures. However, as seen from Figure 5, a peak of sound intensity was seen at the frequency of the ½ sentence, which would have confounded the interpretation of the EEG data with the effects at the ½ sentence rate reflecting (at least partially) bottom-up processing of acoustic stimuli characteristics. We presumed that this issue was related to German being relatively regular with respect to its inflectional morphology and further increased the number of irregular verbs, controlled for the potential rhythmicity produced by repeating sounds of the same quality (for example, a fricative repeating at the end of every two words), reduced the variability of some edge phonemes, and, in contrast, *increased* the variability of others. A major factor in creating a successful list of trials (Figure 5, right panel) was having a large number of sentence arrangements within trials, as well as multiple arrangements of trials within stimuli lists, and later selecting the lists with appropriate acoustic characteristics. Importantly, the order of the sentences within trials contributes to rhythmic acoustic changes becoming evident in the frequency transformation of the sound intensity envelope. Taking into account these factors, we were able to create 22 trials that did not show any such confounds with acoustic information and could, therefore, be used in our frequency tagging study of prosodic processing (Glushko et al., in preparation).

While sound envelope values can align with syntactic structure, phonological features, as with transitional probabilities between words, can correlate with sentence syntax as well (Gahl & Garnsey, 2004). Syntactic structure may result in systematic phonological regularities in the speech signal that may be picked up by neural generators in a bottom-up fashion (Di Liberto, Crosse, & Lalor, 2018; Bonte et al., 2005). These effects may not always be measurable through the sound intensity envelope and need to be ruled out in studies that make strong claims about top-down mechanisms such as syntactic parsing. Moreover, even when bottom-up phonological features of the stimuli are controlled for, covert prosodic processing that does not always have a one-to-one alignment with syntactic structure can evidently modulate neural responses at sentence constituent frequencies (for an in-depth investigation, see Glushko et al., in preparation). Covert prosody could also plausibly produce amplitude-modulations of auditory onset responses, such as the P200, driven by the fact that content words are more likely to receive stress within a phonological word than function words (Gordon & Roettger, 2017).

To summarize, non-syntactic variables can conceivably produce brain responses that in the frequency domain are very similar to responses elicited in previous frequency tagging studies (Ding et al., 2016; 2017; Sheng et al., 2019; Jin et al., 2018; Makov et al., 2017). Similarly to Frank and Yang (2018), these effects were produced in the absence of slow oscillations but rather by faster amplitude-modulated EEG responses.

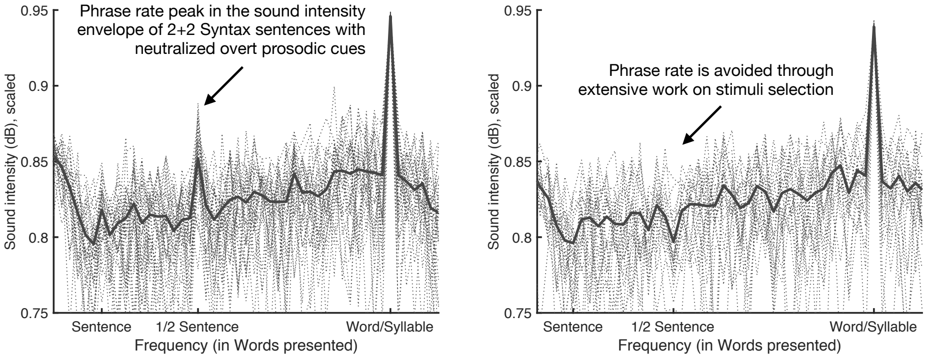


Figure 5. Sound envelope spectrum for 2+2 Syntax German stimuli (for example, [Your song | sounds nice]). On the left: sound intensity envelope spectrum of stimuli created closely following procedures for creating English and Chinese stimuli in Ding and colleagues' (2016) studies. Twenty-two trials (12 sentences each) were arranged from 60 base sentences repeated 4-5 times. On the right: sound intensity envelope spectrum of stimuli created through elaborated control of their phonological features and selected from a large variety of trials, many of which still maintained a phrase rate (i.e., ½ sentence rate) sound envelope spectral peak. Twenty-two trials (12 sentences each) were eventually created from 40 base sentences repeated 6-8 times.

CHOICE OF SYNTACTIC STRUCTURES. Aside from semantic relatedness, and statistical and acoustic characteristics of the stimuli, the selection of the syntactic constructions under investigation should be given careful consideration as well. One of the claimed advantages of the frequency tagging technique for language processing is its capability to capture "concurrent" processing of different timescales during the presentation of sentences with multiple phrases: this is meant to be reflected in spectral power peaks elicited at the frequencies of sentences, phrases, words, and syllables (Ding et al., 2016, 2017). However, the isochronicity of the materials (when, for instance, two words form a phrase, and two of these phrases form a sentence) results in cortical responses to lower level constituents (e.g., phrases at 2 Hz in Ding et al., 2016) to be quantified at the multiples of the frequency at which higher level constituents appear (e.g., sentences at 1 Hz in Ding et al., 2016).

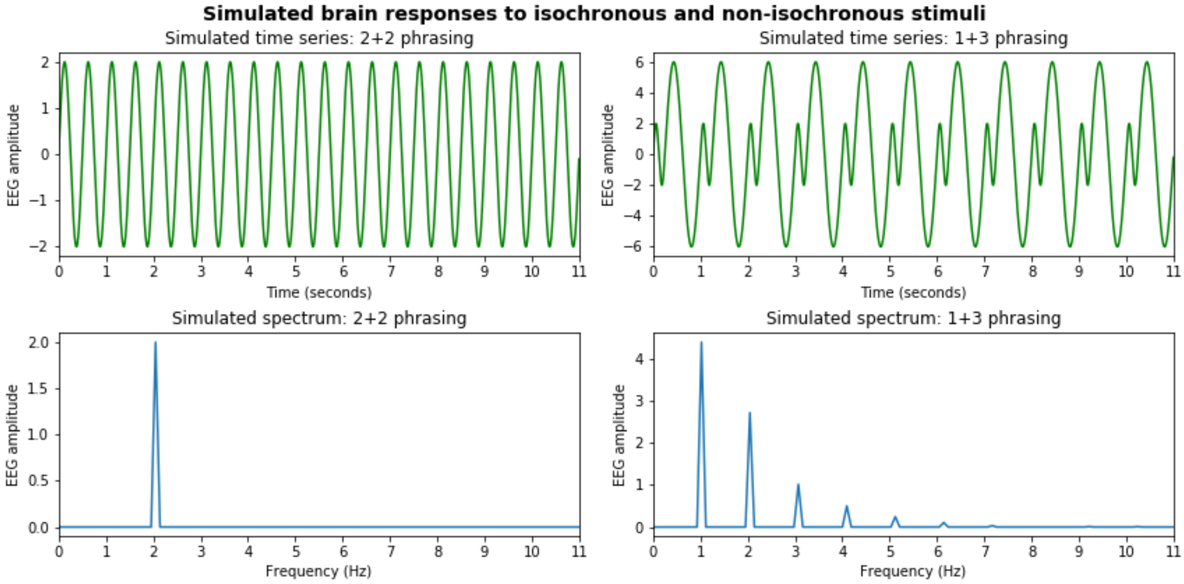


Figure 6. Schematic representation of the EEG tracking of isochronous and non-isochronous stimuli tracking. The top row represents simulated time series data for either two phrases presented isochronously at the rate of 2 Hz (on the left; such as in the case of the 2+2 Syntax, but note that here only the response to phrases is modelled) or two phrases presented non-isochronously (on the right; such in the case of 1+3 Syntax – "fry to-ma-to"). Time series for isochronous phrasing (on the left) are a continuous 2 Hz sinusoid. The time series for non-isochronous phrasing (on the right) are produced by concatenating a 250 ms long 4 Hz sinusoid with a 750 ms long 100/75 Hz sinusoid. The corresponding frequency transforms (bottom row) both have a spectral peak at the 2 Hz frequency.

Spectral power frequency peaks at the frequencies harmonic to the fundamental frequency (that is multiples of the fundamental frequency, i.e., harmonics at 2, 3 and 4 Hz for a given fundamental frequency of 1 Hz) are commonly elicited if the data are not strictly sinusoidal but include sharp edges (Zhou, Melloni, Poeppel, & Ding, 2016) which is indeed the case for brain activity (Cole & Voytek, 2017). Two potential solutions are conceivable for unconfounding responses corresponding to the fundamental frequency of phrase tracking and the harmonics of the sentence tracking response. First, one could study phrase and sentence processing in separate experimental blocks (such as contrasting sequences of four-word sentences, two-word phrases, and single words; as implemented by Ding et al., 2016). In this case, however, the question of whether phrases and sentences are tracked "concurrently" by the brain would remain unresolved. An alternative, second solution would be to use a control condition with sentences that match the target structure at the sentence level (for example, 2+2

Syntax) structure at the sentence level but do not include phrases repeating at the harmonic to the (i.e., multiple of the) sentence frequency. Ding and colleagues (2016), for instance, presented participants with 1+3 Syntax constructions in Mandarin Chinese, such as "[fry] [to-ma-to]" (where each syllable is 250 ms long). These were verb phrases with no syntactic constituents appearing at the rate of 2 Hz (i.e., repeating every 500 ms). Yet while in this one study so far, no significant spectral power peak was elicited at the frequency harmonic (2 Hz) to the whole constituent fundamental (1 Hz), we recommend that such non-isochronous materials be used after clear predictions are made about the associated cortical responses in the time domain. For instance, according to a recent theoretical proposal, processing of higher-level linguistic characteristics can be realized through a top-down, predictive phase-reset of neural oscillations that themselves only dictate temporal constraints on processing (Rimmele, Morillon, Poeppel, & Arnal, 2018). We may assume that in the 1+3 Syntax constructions the first phrase (250 ms long) is tracked by a response at the 4 Hz frequency, and the processing of the second phrase (750 ms long) is realized through a predictive phase-reset of delta oscillations that continue tracking the next phrase resulting in neural activity at 100/75 (= 1.333) Hz frequency (i.e., analogous to the mechanism proposed for the 2+2 Syntax). In that case, the frequency transform of these data would not be consistent with Ding and colleagues' (2016) findings (see Figure 6). On the one hand, the observation that non-isochronous time domain data have non-trivial representations in the frequency domain should be taken into account when considering using non-isochronous stimuli. One the other hand, when non-isochronous materials, such as Ding and colleagues' (2016) 1+3 Syntax verb phrases do not result in the frequency similar to the ones in Figure 6 (right panel), this should be considered when interpreting the mechanisms underlying phrase tracking in 2+2 Syntax sentences. I.e., is cortical tracking *only* seen for isochronous stimuli, or is it instantiated *differently* in the case of non-isochronous stimuli? What does this mean for our understanding of natural language processing, where the signal is only pseudorhythmic? These questions should be addressed by future research before strong conclusions regarding hierarchical syntactic processing can be drawn. With respect to selecting syntactic constructions, making clear predictions for, and investigating the time domain responses associated with non-isochronous materials is recommended.

EXPERIMENTAL PROCEDURE. With respect to experimental procedures in frequency tagging studies at the sentence level, two concerns arise: the number of trials in the experiment and the choice of task. With respect to the first one, 22 trials, each 12-sentences long were used by a few studies so far (Ding et al., 2016; 2017, Sheng et al., 2019). One issue present at least in English, as shown above, is that multiple stimuli characteristics have to be experimentally controlled, making stimuli creation difficult. In our own experiment, we used 40 original sentences (instead of the 60 used by Ding and colleagues', 2016), largely due to controlling for acoustic confounds (see Figure 5). Additionally, while each sentence is repeated multiple times across trials in all sentence processing frequency tagging studies so far, repeated stimuli are known to elicit brain activity different from first presentation data (Grill-Spector, Henson, & Martin, 2006). From our own data (Glushko et al., in preparation), we could see that the averages from just 11 trials resulted in significant EEG responses at sentence and phrase frequencies (for both ITC and EEG magnitude, *p* values < .001, calculated using one-tailed t-tests comparing the EEG responses at sentence and phrase frequencies with EEG responses at neighbouring frequencies; see Figure 7). For the sentence-rate EEG responses, only one participant out of 35 in the first half and two participants for the second half of the block had a signal-to-noise ratio (SNR) lower than 1 for EEG power. However, for the phrase-rate responses, 12 out of 35 participants (34%) had an SNR below 1 for EEG magnitude in the first half, and 6 out of 35 (17%) in the second half of the block. That is, responses at the first harmonic of the sentence frequency (phrase rate) are less robust than the responses at the sentence rate (as also shown by Ding et al., 2017, where only about half of the participants had a phrase level response significantly greater than noise). To sum up, when stimuli development is complicated by the need to control for a variety of linguistic and acoustic characteristics of the stimuli, using a smaller number of trials in a frequency tagging sentence processing experiment seems acceptable if sentence-level and not phrase-level responses are to be interpreted. For phrase-level responses, future research should tap into the factors accounting for their high inter-individual variability and apparent enhancement with increased exposure to the stimuli.

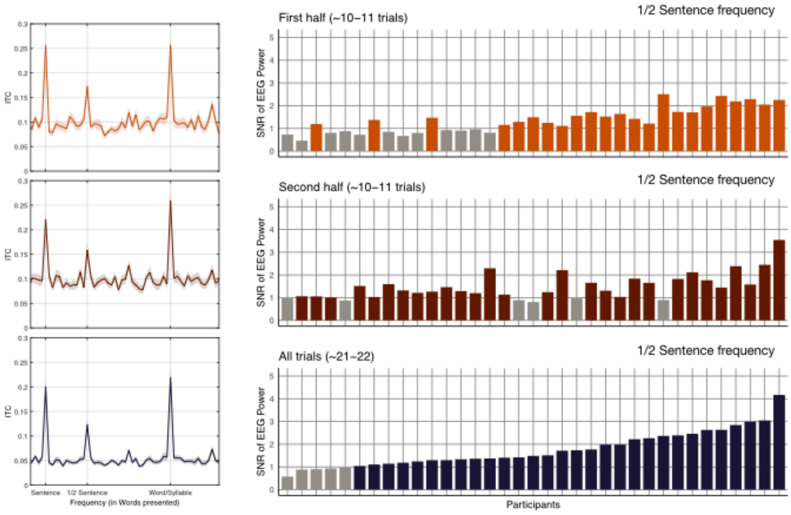


Figure 7. Effects of reducing the number of trials in a frequency tagging study. The data are from Glushko and coauthors (in preparation), specifically the 2+2 Syntax condition presented with neutralized prosody (No Prosody experiment, similar to Ding et al., 2016, 2017). We now split the EEG trials for each participant into the first and the second halves. ITC averages and single-subject signal-to-noise ratios (SNR) of EEG power are presented for (a) the first half of the experimental block (10-11 trials per subject), (b) the second half of the experimental block (10-11 trials per subject), and taken together - (c) for the whole duration of the experimental block (21-22 trials per subject). Grey bars represent subject data with SNR < 1. While the total EEG power during the beginning of a single trial has been previously demonstrated not to be different from the power at the end of the trial (Zhang & Ding, 2017), the ½ sentence EEG peak seems to arise at least in some participants throughout the experiment. This difference is not observed in data from the sentence frequency peak.

Finally, task effects present a separate concern. As shown by Jin and coauthors (2018) in their frequency tagging paradigm, EEG activity phase-locks to the most relevant word based on the task. Ding and colleagues (2018) also demonstrated that EEG responses to words decrease if the task is purely sensory (we saw similar effects in our own data for higher-level sentence constituents; Glushko et al., in preparation). This indicates that the basic task requirements are, first, that it must trigger higher-level linguistic processing and, second, that it should not create confounding phase-locking effects (for example, if two adjacent words need to be integrated to perform a task, this might contribute to processing them as a chunk).

DATA ANALYSIS: TRIAL- AND AVERAGE-BASED ANALYSIS

Frequency tagging studies focusing on relatively low frequencies (< 10 Hz) vary in which data the frequency transform is obtained from. While some authors (e.g., Nozaradan et al., 2018; Boremanse, Norcia, & Rossion, 2014; Ding et al., 2018; in sentence processing - Sheng et al., 2018; Jin et al., 2018; Makov et al., 2017) average the data across trials in the time domain (similar to ERPs) and apply the Fourier transform on this average, others (e.g., Buiatti et al., 2009; Ding et al., 2016, 2017 in sentence processing) apply the Fourier transform to the EEG of each trial separately and then average across trials in the frequency domain.

The advantages of averaging the EEG data in the time domain are an undoubtedly higher signal-to-noise ratio and the fact that only activity phase-locked to the stimuli is analyzed. However, the conclusions that can be derived from the frequency transform performed on the event-related potential (the EEG average across trials) are qualitatively different from the ones that can be drawn from the frequency spectrum of raw EEG.

The difference between the data averaged in the time domain and the data from single trials is the effect inter-trial latency jitter has on the time-domain averages. Latency jitter does not qualitatively change the spectrum of time-invariant, continuous oscillations (such as, for instance, the idealized concurrent entrainment responses at 1 Hz, 2 Hz, and 4 Hz in the case of Ding and colleagues' 2016 study). However, it is debatable whether effects reported in these studies are truly oscillatory, especially given the likely contribution of non-syntactic variables to EEG responses at the frequencies of syntactic constituents in previous studies (e.g., Ding et al., 2016, 2017; Frank & Yang, 2018). The alternative is that effects at the frequencies of syntactic constituents are produced or modulated by a series of evoked responses. Zhou and colleagues (2016) reject this possibility, emphasizing that the spectrum of transient responses repeated at a slow rate is generally different from the spectrum of slow oscillations (see Figures 2 and 6b in their paper). Specifically, slow oscillations produce prominent spectral peaks at the fundamental frequency (for example, at the sentence frequency of 1 Hz; see Figure 8c), and repeating transient responses (see Figure 9a-b below) produce low-power response increases, most of which were found at frequencies other than the repetition rate (Figure 8c; in our example, a cycle of a 4 Hz wave is repeated at the frequency of 1 Hz). At least in principle, this qualitative difference between the spectra allows us to make a distinction between slow waves and fast transient responses repeated at a slow frequency.

This distinction has two problems: first, slow waves tracking syntactic structure might not be strictly sinusoidal. From previous frequency tagging studies on sentence processing, we know that spectral peaks can reach significance at frequencies beyond those at which syntactic constituents are presented (i.e., harmonics; e.g., Figures 2c and Figure 3 Ding et al., 2016). That is, often we cannot count on the absence of harmonics (e.g., at 2, 3, and 4 Hz) beyond the frequencies of syntactic constituents (at 1 Hz) if we attempt to test the contribution of slow oscillations to neural responses at those frequencies. What is left is the qualitative difference between the spectra of (i) transient, infrequently repeated responses and (ii) slow oscillations. This is where the processing of the data becomes important: fast transient responses that are temporally smeared due to latency jitter across trials can produce spectra with the highest power placed at a slow frequency (Figure 9d-f). The spectrum of a 4 Hz transient response repeated every second can have its highest power at 2 Hz if the responses are averaged across trials with latency jitter (Figure 9f). This example does not represent a full spectrum of transient responses elicited by sentence processing, yet it demonstrates that latency jitter should be considered by studies performing the frequency transform on data averaged in the time domain (for other relevant demonstrations, see Figure IC in Box 1 in Obleser & Kayser, 2019; Figures 2.6 and 8.7 in Luck, 2014).

It is worth noting that averaging the data in the time domain, however, does seem to drastically increase the signal-to-noise ratio of the frequency tagging in sentence processing studies. In our own pilot data, at least some of the effects evaluated using frequency transforms of single trials were hardly distinguishable from noise. In already published, openly available EEG data (Jin et al., 2018), the sentence-level effect that was statistically significant when an ERP (trial average) was transformed into the frequency domain did not reach significance when an average of frequency transforms of single trials was used (see Figure 10).

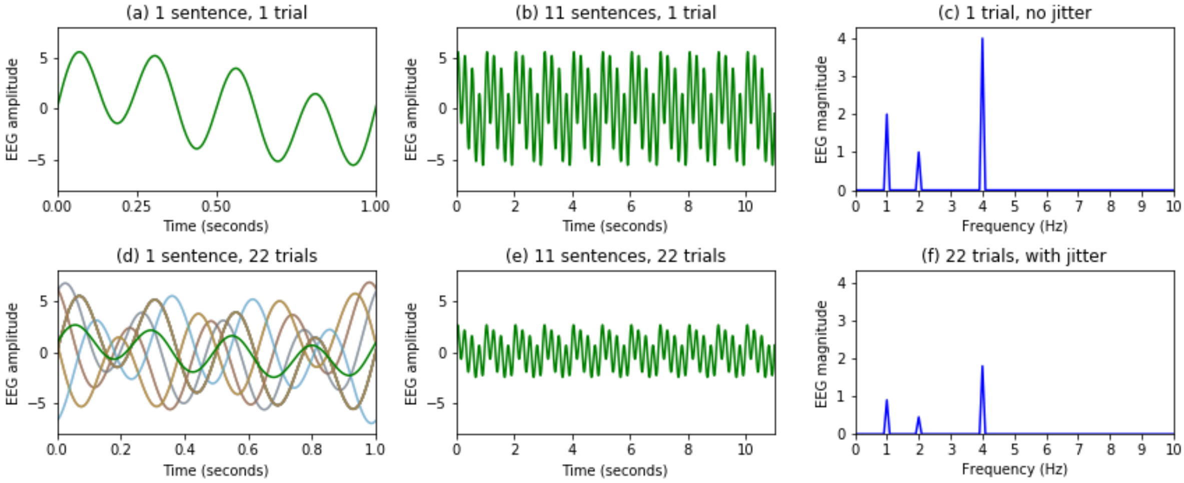


Figure 8. Effects of averaging across trials (prior to FFT) on complex time-invariant responses with random phase shifts (i.e., latency jitter). Trials were created by summation of the 1 Hz, the 2 Hz, and the 4 Hz sinusoids. Random phase shifts for all trials were introduced (d-f). The top row represents the time domain response to one trial, the bottom row demonstrates the effect of averaging 22 trials with latency jitter (the green line in (d) and (e) represents the signal average). As can be seen from the comparison of (c) and (f), responses are suppressed due to latency jitter but importantly, they are qualitatively identical to the frequency transform of a single trial.

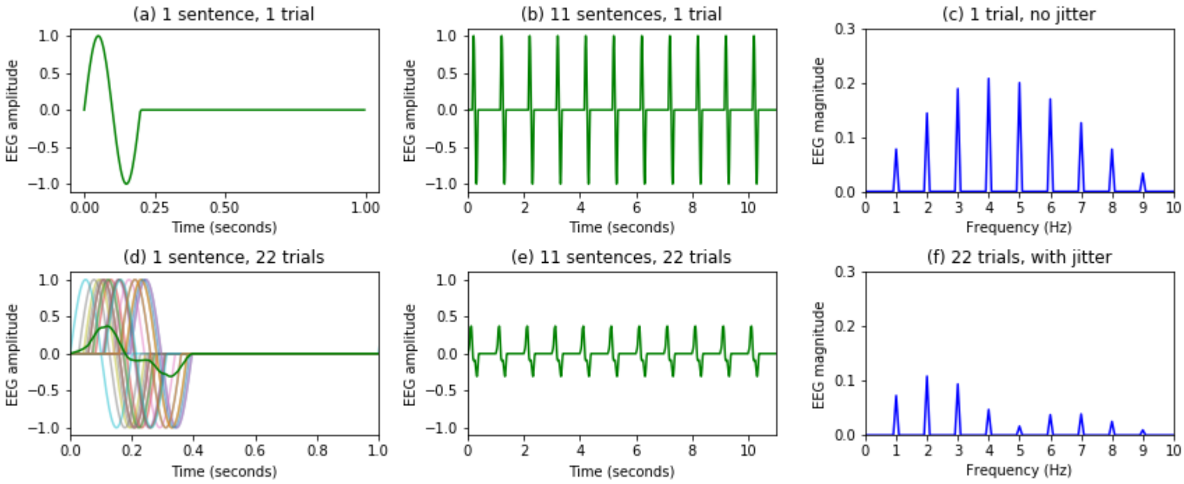


Figure 9. Effects of averaging across trials (prior to FFT) on transient responses (4 Hz) repeated at a slow rate (1 Hz) with random phase shifts (i.e., latency jitter). Random phase shifts in 50% of the trials were introduced (d-f). The top row represents the time domain response to one trial, the bottom row demonstrates the effect of averaging 22 trials with latency jitter (the green line in (d) and (e) represents the signal average). The comparison of (c) and (f) demonstrates that the spectrum of rhythmically repeated transient responses changes qualitatively due to latency jitter (which is only the case when the response averaged in the time domain undergoes frequency transform).

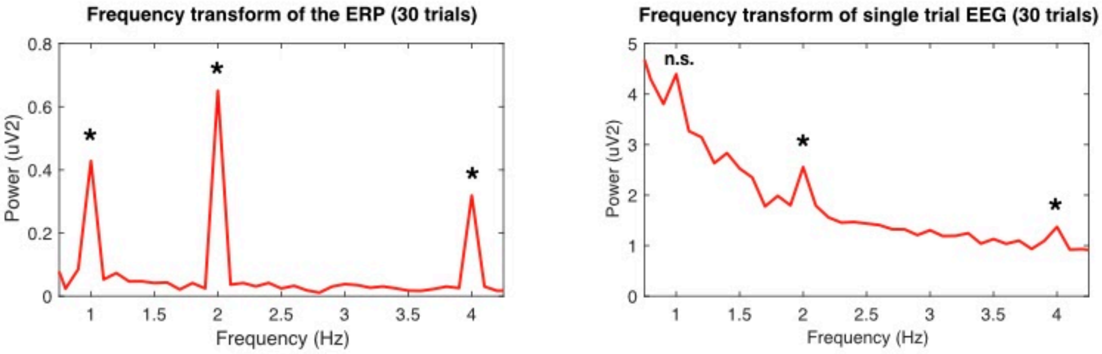


Figure 10. Jin and colleagues' (2018) data reanalyzed using frequency transform applied to single trial EEG data (on the right) in comparison to the original result based on the frequency transform of the ERP (on the left). We accessed the openly available dataset used for producing Figure 1 in Jin and coauthors' (2018) study (corresponding to the left panel in the current figure). Here, we are presenting the geometric means of EEG responses for 15 participants at the Cz electrode and the corresponding statistical results. Using one-tailed t-tests and correcting for multiple comparisons, we show that the sentence-level peak does not reach significance when single trial EEG data is transformed into the frequency domain. That is, the results of studies using single trial frequency transforms (Ding et al., 2016, 2017) were not fully replicated.

IMPLICATIONS FOR FUTURE RESEARCH

While the frequency tagging method is promising in other domains of cognitive neuroscience, the advantage of this technique for sentence processing research remains questionable and needs further investigation. First, stimuli materials should be carefully selected given that the neurophysiological data spectrum at low frequencies can be modulated by amplitude changes in EEG/MEG responses at higher frequencies that are different from sentence structure processing per se. These include word frequency, transitional probabilities between words, semantic relatedness (Frank & Yang, 2018), and acoustically or phonologically driven processing effects. The spectrum of neurophysiological data is often ambiguous due to the complexity and aperiodicity of the EEG/MEG signal. When we move away from sinusoids modulated in an extremely restricted way (such as in demonstrations by Zhou et al., 2016), it becomes clear that amplitude-modulated word-level responses can produce spectra reminiscent of the ones in frequency tagging studies claiming to have presented 'undeniable' evidence for psycholinguistic processing of hierarchical sentence structure (Ding et al., 2016; p. 5). Moreover, overt and, importantly, covert prosody effects can be reflected in the spectral power peaks at low frequencies, plausibly representing true slow neural activity (Glushko et al., in preparation), just as observed for word grouping based on transitional probabilities (see e.g., Buiatti et al., 2009; Getz, Ding, Newport, & Poeppel, 2018). Last but not least, syntactic structure itself is an important factor. Studying sentence hierarchy using isochronous hierarchical structures is only possible through the use of appropriate control conditions (i.e., comparing 2+2 and 1+3 Syntax sentences in Ding et al., 2016). However, spectra of neurophysiological responses to processing non-isochronous phrases are yet to be described. In this respect, our suggestion for future research is to focus on studying constituents that are not embedded in larger structures in an isochronous way (for example, by studying two-word constructions; or four-word chunking in sentences ignoring spectral responses at higher frequencies), while also avoiding using non-isochronous materials unless the time domain responses to them are understood.

A distinct methodological issue in the use of the frequency tagging technique is the trade-off between SNR and our ability to analyze unambiguous spectra generated from single trials. With a trial length of 12 seconds and 30 trials used in a span of a short experiment, participants listen to 360 sentences in every single experimental condition, which is much more than what is needed to create reliable ERP responses. A smaller number of trials (i.e., 11) significantly reduces the SNR of responses to at least some sentences constituents even when responses are averaged in the time domain. When 30 trials with 12 sentences per trial are used, the signal-to-noise ratio is increased but still admittedly low when EEG is not averaged in the time domain to avoid related artifacts. Given the constraints that stimulus creation presents for sentence-level frequency tagging experiments, achieving an appropriate signal-to-noise ratio presents a significant challenge. The preferable approach to averaging in frequency tagging experiments seems, therefore, to use data averaged in the time domain while performing additional control analyses. These analyses can include data simulation with latency jitter and taking into account predicted responses to various characteristics of the stimuli. These simulations can be compared to real data in terms of the relative sizes of the neurophysiological responses at low frequencies and their harmonics. Notably, latency jitter can be minimized by controlling for a variety of the psycholinguistic and low-level variables discussed in previous sections of this article. With this respect, a creation of an openly accessible pool of balanced stimuli materials presents one forward-looking solution.

CONCLUSIONS

Recent research utilizing the frequency tagging technique has shown increased spectral MEG and EEG power at sentence constituent frequencies. We demonstrated that these effects are quite ambiguous and can reflect both slow neural responses and faster, amplitude-modulated responses. The amplitude of faster responses can be dependent on non-syntactic variables such as word frequency and the statistical probability of a word appearing given the word preceding it (along with prosody and semantics whose contribution has been discussed by previous studies; Glushko et al., in preparation; Frank & Yang, 2018). These and other variables discussed in the current article have to be controlled for in future frequency tagging studies. Furthermore, data analysis in frequency tagging research on sentence processing should include a consideration of the spectral power peaks at frequencies harmonic to the ones tagged in the study. Our results suggest that previous claims regarding slow oscillations entraining to the hierarchical structure of sentences are premature.

REFERENCES

Adrian, E. D., & Matthews, B. H. (1934). The Berger rhythm: potential changes from the occipital lobes in man. *Brain*, *57*(4), 355-385. https://doi.org/10.1093/brain/57.4.355

Alday, P. M. (2019). M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457-473. https://doi.org/10.1080/23273798.2018.1546882

Baayen, R. H. (2014). Experimental and psycholinguistic approaches to studying derivation. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology*. Oxford: Oxford UP, 95-117. https://doi.org/10.1093/oxfordhb/9780199641642.013.0007

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*(1), 47-53. https://doi.org/10.1016/j.bandl.2013.01.005

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, *17*(2), 89-98. https://doi.org/10.1016/j.tics.2012.12.002

Blanco-Elorrieta, E., Ding, N., Pylkkänen, L., & Poeppel, D. (2019). Understanding requires tracking: noise and knowledge interact in bilingual comprehension. *bioRxiv*, 609628, 1-13. https://doi.org/10.1101/609628

Bohórquez, J., & Özdamar, Ö. (2008). Generation of the 40-Hz auditory steady-state response (ASSR) explained using convolution. *Clinical Neurophysiology*, *119*(11), 2598-2607. https://doi.org/10.1016/j.clinph.2008.08.002

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, *116*(12), 2765-2774. https://doi.org/10.1016/j.clinph.2005.08.012

Boremanse, A., Norcia, A. M., & Rossion, B. (2014). Dissociation of part-based and integrated neural responses to faces by means of electroencephalographic frequency tagging. *European Journal of Neuroscience*, *40*(6), 2987-2997. https://doi.org/10.1111/ejn.12663

Brown, R. J., & Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Research*, *37*(17), 2401-2408. https://doi.org/10.1016/S0042-6989(97)00045-X

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, *44*(2), 509-519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Canale, A., Lacilla, M., Cavalot, A. L., & Albera, R. (2006). Auditory steady-state responses and clinical applications. *European Archives of Oto-Rhino-Laryngology and Head & Neck*, *263*(6), 499-503. https://doi.org/10.1007/s00405-006-0017-y

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137-167. https://doi.org/10.1016/S0019-9958(59)90362-6

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157-205. https://doi.org/10.1207/s15516709cog2302_2

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, *6*, 1-18.

Cole, S. R., & Voytek, B. (2017). Brain oscillations and the importance of waveform shape. *Trends in Cognitive Sciences*, *21*(2), 137-149. https://doi.org/10.1016/j.tics.2016.12.008

Cone-Wesson, B., Dowell, R. C., Tomlin, D., Rance, G., & Ming, W. J. (2002). The auditory steady-state response: comparisons with the auditory brainstem response. *Journal of the American Academy of Audiology*, *13*(4), 173-187.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121. https://doi.org/10.1038/nn1504

Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. *Eneuro*, *5*(2). 1-13. https://doi.org/10.1523/ENEURO.0084-18.2018

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in human neuroscience*, *11*, 1-9. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 1-7. https://doi.org/10.1038/nn.4186

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *Journal of Neuroscience*, *38*(5), 1178-1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Frank, S. L., & Christiansen, M. H. (2018). Hierarchical and sequential processing of language: A response to: Ding, Melloni, Tian, and Poeppel (2017). Rule-based and word-level statistics-based processing of language: insights from neuroscience. Language, Cognition and Neuroscience. *Language, Cognition and Neuroscience*, *33*(9), 1213-1218. https://doi.org/10.1080/23273798.2018.1424347

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PloS one*, *13*(5), 1-11. https://doi.org/10.1371/journal.pone.0197304

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, *279*(1747), 4522-4531. https://doi.org/10.1098/rspb.2012.1741

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 748-775. https://doi.org/10.1353/lan.2004.0185

Gao, X., Xu, D., Cheng, M., & Gao, S. (2003). A BCI-based environmental controller for the motion-disabled. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 137-140. https://doi.org/10.1109/TNSRE.2003.814449

Glushko, A., Poeppel, D., & Steinhauer, K. (in preparation). Imagined prosody affects neurophysiological responses at frequencies of syntactic constituents.

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1). https://doi.org/10.1515/lingvan-2017-0007

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14-23. https://doi.org/10.1016/j.tics.2005.11.006

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483. https://doi.org/10.1080/01690969308407585

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*(3), 1101-1116. https://doi.org/10.1006/nimg.2002.1268

Jin, P., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature Communications*, *9*(1), 1-15. https://doi.org/10.1038/s41467-018-07773-y

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, *6*(8), 350-356. https://doi.org/10.1016/S1364-6613(02)01947-2

Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25-36. https://doi.org/10.1016/j.bandl.2015.03.005

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*(3), 1-19. https://doi.org/10.1371/journal.pbio.2004473

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press, Cambridge, MA. https://doi.org/10.1073/pnas.0703993104

Luts, H., Desloovere, C., & Wouters, J. (2006). Clinical application of dichotic multiple-stimulus auditory steady-state responses in high-risk newborns and young children. *Audiology and Neurotology*, *11*(1), 24-37. https://doi.org/10.1159/000088852

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*(32), 7772-7781. https://doi.org/10.1523/JNEUROSCI.0168-17.2017

Nakabayashi, M., Kurokawa, A., & Yamamoto, Y. (2001). Immediate prediction of recovery of consciousness after cardiac arrest. *Intensive Care Medicine*, *27*(7), 1210-1214. https://doi.org/10.1007/s001340100984

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151-165. https://doi.org/10.1162/jocn.1991.3.2.151

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of Vision*, *15*(6), 1-46. https://doi.org/10.1167/15.6.4

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*(49), 17572–17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*(28), 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011

Nozaradan, S., Schönwiesner, M., Keller, P. E., Lenc, T., & Lehmann, A. (2018). Neural bases of rhythmic entrainment in humans: critical transformation between cortical and lower-level representations of auditory rhythm. *European Journal of Neuroscience*, *47*(4), 321–332. https://doi.org/10.1111/ejn.13826

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language*, *34*(6), 739-773. https://doi.org/10.1006/jmla.1995.1033

Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. *Brain and Language*, *91*(3), 315-325. https://doi.org/10.1016/j.bandl.2004.05.007

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, *28*(3), 383-407. https://doi.org/10.1016/j.cogsci.2003.12.003

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, *42*(4), 177-219. https://doi.org/10.3109/14992020309101316

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*(6), 1007-1028. https://doi.org/10.1207/s15516709cog0000_28

Rees, A., Green, G. G. R., & Kay, R. H. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hearing Research*, *23*(2), 123-133. https://doi.org/10.1016/0378-5955(86)90009-2

Regan, D. (1965). A study of the visual system by the correlation of light stimuli and evoked electrical responses (Doctoral dissertation). Retrieved from: https://spiral.imperial.ac.uk/bitstream/10044/1/16967/2/Regan-D-1965-PhD-Thesis.pdf

Regan, D. (1982). Comparison of transient and steady-state methods. *Annals of the New York Academy of Sciences*, *388*(1), 45-71. https://doi.org/10.1111/j.1749-6632.1980.tb19349.x

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., ... & Gao, J. H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, *29*(8), 3232-3240. https://doi.org/10.1093/cercor/bhy191

Snyder, A. Z. (1992). Steady-state vibration evoked potentials: description of technique and characterization of responses. *Electroencephalography and Clinical Neurophysiology/Evoked potentials Section*, *84*(3), 257-268. https://doi.org/10.1016/0168-5597(92)90007-X

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135-162. https://doi.org/10.1016/j.bandl.2011.07.001

Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: an ERP study. *Cognitive Brain Research*, *15*(1), 99-103. https://doi.org/10.1016/S0926-6410(02)00219-7

Tavano, A. (2017). Internal rhythms to sentences [Video]. Retrieved from https://www.aesthetics.mpg.de/institut/veranstaltungen/vergangene-veranstaltungen/neural-oscillations-in-speech-and-language-processing-berlin-2017.html

Tobimatsu, S., Zhang, Y. M., & Kato, M. (1999). Steady-state vibration somatosensory evoked potentials: physiological characteristics and tuning function. *Clinical neurophysiology*, *110*(11), 1953-1958. https://doi.org/10.1016/S1388-2457(99)00146-7

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive linguistics*, *11*(1/2), 61-82. https://doi.org/10.1515/cogl.2001.012

Zhang, W., & Ding, N. (2017). Time-domain analysis of neural tracking of hierarchical linguistic structures. *Neuroimage*, *146*, 333-340. https://doi.org/10.1016/j.neuroimage.2016.11.016

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, *10*, 1-8. https://doi.org/10.3389/fnhum.2016.00274

GENERAL DISCUSSION

The current thesis consists, first, of three empirical experiments systematically testing the nature of the interactions between prosody (overt or covert) and syntax as evident from the EEG responses at frequencies of sentence constituents (Manuscript 1). Second, we addressed various methodological issues associated with the use of the frequency tagging technique in sentence processing research (Manuscript 2). On top of the discussion points included in the manuscripts, we conclude by turning to 'big picture' issues, namely (1) to the biologically-plausible, mechanistic interpretations of the effects seen in our and previous studies of sentence processing using frequency tagging; (2) the potential of the frequency tagging paradigm for different types of sentence processing research; and (3) future directions pointed to by the results of the present research project.

DO BRAIN OSCILLATIONS ENTRAIN TO SYNTACTIC AND PROSODIC SENTENCE STRUCTURES?

In our study, we showed that EEG responses at low frequencies are modulated by both changes in syntactic and prosodic structure of the sentence. Previous frequency tagging studies of sentence processing have claimed that it is slow cortical oscillations that entrain to syntactic phrase structure of the sentences (Ding et al., 2016; 2017). We address several issues associated with this claim one by one.

Given the empirical data available to us so far and the performed data simulations, more research is needed to definitively claim that frequency tagging results reflect entrainment to syntactic structure in the narrow sense (i.e., entrainment of endogenous neural oscillations to an external stimulus; Obleser & Kayser, 2019). As noted by Obleser and Kayser (2019) in their comprehensive review of neural entrainment in the auditory domain, "temporal (e.g., phase) alignment [with the stimulus changes] may or may not be an instantiation of true entrainment in a strict sense" (p. 918). In fact, the responses might represent a series of evoked responses. In the case of the typical frequency tagging design in sentence processing, we have seen that transient EEG responses can in fact produce spectral power peaks in the frequency domain at the frequencies of syntactic constituents: when a word-level EEG response is amplitude-modulated or when it is repeated at the beginning of every sentence (especially when average ERP analysis is performed). In other words, whether the effects represent true neural entrainment is still to be determined by future research with stronger applications of the frequency tagging paradigm and by complementing frequency tagging experiments with other techniques.

In addition to the cross-examination of the overall mechanistic claim regarding neural entrainment, the question arises whether entrainment to *hierarchical* sentence structure is evident from the data, i.e., if different levels of syntactic hierarchy (i.e., sentences and phrases) are *concurrently* tracked by slow brainwaves. In frequency tagging studies, *concurrent* tracking can be investigated when contrasting different types of syntactic constructions with embeddings (i.e., the 2+2 Syntax and the 1+3 Syntax condition in our and the first, Ding and colleagues' (2016) study). Two distinct issues arise here. First, as outlined in Manuscript 2, this will most likely require the use of non-isochronous materials, the neural tracking characteristics of which is a matter for future research. Based on Ding and colleagues (2016) findings, it is conceivable that either the phrases are not tracked by slow cortical activity in the case of 1+3 Syntax sentences, or the mechanism of phrase tracking in the 2+2 Syntax sentences and its link to natural language processing should be re-evaluated. Second, in the only study available to date that unconfounded syntax and prosody in their potential effect on the ½ sentence rate EEG power peak (Manuscript 1), we saw that covert prosody alone can drive the elicitation of EEG responses at the ½ sentence rate (i.e., the phrase rate in the case of the 2+2 Syntax). This means that even if sentence-level neural tracking is driven by syntactic mechanisms (which is yet to be determined given a variety of alternative accounts voiced in the current thesis and other research; see Frank & Yang, 2018; Frank & Christiansen, 2018), the "concurrent" phrase-level tracking could in fact represent concurrent tracking of (covert) prosodic phrasing. That is, current claims of hierarchical syntactic structure being reflected in the spectral power peaks at frequencies of sentence constituents are premature.

One last issue with respect to the distinction between neural tracking of prosodic and syntactic structure is of a theoretical character. The need to distinguish between prosodic and syntactic effects on neural tracking of phrase structure is often challenged: if covert prosodic processing is largely determined by syntactic analysis (during silent reading or while listening to sentences with neutralized overt prosody), what is the theoretical relevance of understanding whether it is syntactic phrase structure or prosodic phrasing (driven by syntactic analysis) that is being tracked by neural activity? In my perspective, this distinction is a crucial one.

As emphasized throughout the current thesis, prosodic structure can diverge from syntactic structure in natural language, and that means that for an accurate description of natural language, the distinct mechanisms of prosodic tracking should be described. Avoiding to highlight the possibly essential role of prosody and other mechanisms (such as distributional semantics and statistical features of linguistic input as acquired by the language user throughout life) in the cortical tracking of sentence constituents by low frequency cortical responses has implications for the development of neurolinguistic research.

Consider two possible uses of the frequency tagging technique for language research. First, as in the original studies by Ding and colleagues' (2016, 2017), frequency tagging can be utilized to make theoretical claims about how syntactic hierarchies are processed by the brain. A mechanistic interpretation (i.e., neural entrainment) is provided as interpretation of the effects implying that neural oscillations align with hierarchical syntactic phrasing, and this allows language users to parse sentences. This mechanism is sufficient to account for syntactic phrase processing realized by the brain. However, if it is prosodic phrasing that the neural oscillations align with, we would expect additional neural mechanisms to be involved in the processing of syntactic structure. For instance, it has been previously shown that the envelope of high-gamma oscillations (fast, 70-150 Hz waves) is aligned with syntactic phrases (Nelson et al., 2017)[2], resembling the idea of phrase tracking by slow oscillatory activity but potentially distinct from that mechanism (Zhou et al., 2016). Similarly, it is possible that syntactic structure processing is reflected in repeated transient neurophysiological responses that can but do not necessarily have a prominent effect on the spectral power peaks at low frequencies. Understanding the underlying logic of neurocognitive sentence processing requires the knowledge of how syntactic and prosodic processing is realized by the brain, which includes the question of whether the associated neural mechanisms are distinct and if and how they interact. Avoiding to mention the crucial role of prosody in studies utilizing Ding and colleagues' (2016) paradigm can cause a chain of misinterpretations in the literature, similar to the effect that strong claims about the syntactic nature of ELAN had on several branches of neurolinguistic research, including serious implications for language acquisition and bilingualism (Steinhauer & Drury, 2012).

With respect to the second potential application of the frequency tagging technique (and related to the previous point), i.e., its use in special populations, the distinction between syntax and prosody becomes important specifically for the cases where prosodic and rhythmic chunking abilities can be impaired. Examples include developmental language disorder (DLD; Cumming, Wilson, & Goswami, 2015) and autism spectrum disorder (ASD; McCann, & Peppé, 2003; DePriest, Glushko, Steinhauer, & Koelsch, 2017). In other words, both syntactic

---

[2] Note, however, that in this silent reading study covert prosody effects were not accounted for.

and prosodic processing are likely to interact during online sentence processing and the results of this interaction are integrated to ensure efficient sentence use. In many special populations, however, it becomes essential to dissociate between the two. Importantly, these are the cases in which the methodological advantages of the frequency tagging technique could be especially relevant.

FREQUENCY TAGGING: A PROMISING METHOD FOR LANGUAGE RESEARCH?

The current thesis was originally motivated by the potential of the frequency tagging technique in studying learning of recursive sequences in developmental language disorder. This project grew over time and is further continuing beyond the current dissertation. Its original goal was to develop a new approach to studying artificial grammar learning across cognitive domains in DLD. In the following, we will outline the potential of the technique for (1) addressing theoretical questions in psycho- and neurolinguistics of sentence processing; (2) studying artificial grammar and complex sequence learning; (3) language research in special populations.

Some of the central questions of psycho- and neurolinguistic research are how various linguistic cues interact and are integrated during online sentence processing and whether hierarchical sentence structure is built online based on top-down representations of language grammar. As discussed above, the latter question can hardly be directly addressed by the frequency tagging paradigm. It is, however, possible to contrast the frequencies tagged by various layers of sentential information (for instance, prosody and syntax, like it is done in Manuscript 1) to tap into the interaction of these different types of information and its effect on sentence structure tracking. One consideration here is the limited diversity of syntactic structure that can possibly be used in a frequency study paradigm. The limitation is driven, on the one hand, by the longest possible length of the syntactic constituent that can be analyzed.

As of now, analysis of neurophysiological responses at frequencies below 0.5 Hz is challenging due to a large amount of noise in the data at low frequencies. While the responses to low frequency constituents in frequency studies of visual processing have been often analyzed by investigating the harmonics of the fundamental frequency (Norcia et al., 2015), this is not straightforward in sentence processing research. Based on our current understanding of syntax, each two single constituents are merged into a larger one, and responses to these smaller constituents will likely overlap with the harmonic of the response to the larger constituent. Moreover, the role of harmonics in sentence processing frequency tagging studies is still to be described, and while first efforts have been made into resolving this question (e.g., Tavano, 2017), harmonics do not seem to be elicited at least in some syntactic structures (see 1+3 Syntax constructions in Ding et al., 2016). On the other hand, and related to the former concern, cortical responses to phrases embedded into larger syntactic constituents are yet to be understood: non-isochronicity of these constituents and the overlap with the harmonics elicited by higher-level constituents remains an unresolved issue. To sum up, the advantages of the frequency tagging technique for sentence processing research are (1) its potential for a mechanistic interpretation of neurophysiological responses to sentence processing, (2) its divergence from the violation paradigm, and (3) the possibility of studying the effects of the interaction between various types of sentential information on neurophysiological responses to sentence structure. It is, however, limited by the nature of the stimuli that can be developed.

Beyond its use in studying sentence processing per se, the frequency tagging paradigm has been recognized as a useful tool for investigating the more general mechanisms. An important line of research on processing hierarchical structures presents experiments in artificial grammar learning. Frequency tagging has already been used for studying simple sequence learning either resembling word segmentation in first language acquisition (e.g., Buiatti et al., 2009) or probabilistic finite-state grammar learning (Getz, Ding, Newport, &

Poeppel, 2018). Using nonword stimuli has the advantage of eliminating some of the methodological drawbacks of the frequency paradigm associated with effects of semantic processing (see Manuscript 2). One great potential of frequency tagging for artificial grammar research is that virtually no behavioural responses are necessary to track learning progress. That is, tracking of learning progress at different stages can be done without using a task that necessarily makes subsequent learning less implicit, opening the window to the use of explicit learning strategies. With respect to the complexity of structures, while the issue with embedded structure tracking remains, fundamental frequency of longer constituents can in principle be tracked for constituents up to 10 monosyllabic nonwords long (with each syllable lasting for 200 ms), which means complex structure tracking could be studied using this paradigm. More research is needed to clearly understand the spectral representation of various variants of grouping, but when this is achieved, frequency tagging could additionally allow us to shed light on what specific grouping underlies learning of a grammar, which is often unclear in current research (see e.g., Rohrmeier et al., 2012).

Finally, there is potential for the frequency tagging technique as used in sentence processing research to be used with special populations (as also stated by Ding et al., 2016). The advantages of the technique are its high signal-to-noise ratio (SNR) and, again, the fact that it can be utilized without relying on behavioural responses. The high SNR (when the data are averaged in the time domain before Fourier transform; see Manuscript 2) allows for the minimal length of the experiment (10-15 minutes per condition, meaning about 30 min for a small experiment). Not relying on behavioural responses makes this approach relevant for research in children. Overall, once the basic methodological concerns are addressed (Manuscript 2), there are myriad applications for clinicians and educators.

FUTURE DIRECTIONS

Further examination of the interactions between prosody and syntax in the elicitation of neurophysiological response at low frequency is needed to ensure appropriate understanding of neural mechanisms involved in sentence processing. These effects should be replicated utilizing syntactic constructions different from the ones presented in our experiments. Dissociating prosodic and syntactic phrasing can be done by acquiring behavioural responses from participants who would state which of the two possible interpretations they committed to, like it was elegantly done by Meyer and coauthors (2017). While Meyer and coauthors used the classical contrast between low and high attachment structures (for example, *The client sued the murderer with the corrupt lawyer*), this paradigm can be applied to shorter sentences as well (e.g., parentheticals such as *Peter, says Mary, smells*, or assertions such as *Children, study hard!*).

Future directions include the use of more naturalistic prosodic contours, as well as computationally determining the portions of the spectral power peaks accounted for by the various layers of sentence processing. Note as well that covert prosody research is typically done on silent reading (e.g., Steinhauer & Friederici, 2001), and the contours used by silent readers can differ from the default prosodic contours acquired from auditory language input (Jun, 2010). Given that frequency tagging in sentence processing research is used on materials with neutralized overt prosody that are likely to elicit covert prosodic processing, more psycholinguistic investigation of covert prosody as evident from studies on auditory materials should be conducted. Further, minimizing the effect of covert prosody can be done by using unblocked stimuli (i.e., each trial containing sentences with different structures, and responses to either single sentence or pairs of sentences can be analyzed) or stimuli with non-prosodic auditory interference (for instance, by applying a 3 Hz pitch amplitude modulation on the 1-second-long sentence materials). Next, a careful examination of the neurophysiological data in the time domain can clarify the degree of independence of cortical responses at frequencies of different sentential constituents: this work is under development and will be continued beyond this thesis.

Part of this dissertation is dedicated to outlining the crucial methodological considerations for frequency tagging research in language processing. Stimuli development constitutes a significant challenge given the multiple constraints on syntactic, semantic, and prosodic structures of the materials. Creating a pool of well-controlled materials that is openly available for researchers is one of my future goals.

CONCLUSIONS

The current thesis provides an examination of how different types of sentential information (namely, prosody and syntax) are tracked by neurophysiological activity during online sentence processing. We showed that the effects of prosody and syntax on the EEG responses at the frequencies of sentence constituents interact, whether the prosody is overt (present in the speech signal) or covert (imposed top-down based on prior knowledge and expectations). Our experiments raise questions regarding the mechanistic model of phrase structure processing as proposed by previous frequency tagging studies: we demonstrated that the claims that it is syntactic structure *per se* that is tracked by neural activity need support from future studies. In fact, as evident from our No Prosody experiment, in the absence of a syntactic boundary, covert prosodic phrase structure can be tracked by EEG responses at low frequencies. To date, such evidence hardly exists for syntactic tracking. This finding and the extensive methodological analysis of the frequency tagging technique as applied to sentence processing research make a substantial contribution to our understanding of the role of neural entrainment in sentence processing and advance the newly developed field of frequency tagging research of language processing.

REFERENCES

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., & Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research*, *1298*(7), 161–170. https://doi.org/10.1016/j.brainres.2009.08.017

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Linguistics and Language Compass*, *8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). *Investigating the Neural Correlates of Continuous Speech Computation with Frequency-Tagged Neuroelectric Responses*, *44*(2), 509–519. Retrieved from https://doi.org/10.1016/j.neuroimage.2008.09.015

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137–167. https://doi.org/10.1016/S0019-9958(59)90362-6

Chomsky, N. (2010). *Lectures on Government and Binding*. *Lectures on Government and Binding*. Dordrecht, Netherlands: Foris. https://doi.org/10.1515/9783110884166

Chomsky, N. (1995). *The minimalist program.* MIT Press, Cambridge, MA.

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, *6*, 1-18. https://doi.org/10.3389/fpsyg.2015.01182

Cumming, R., Wilson, A., & Goswami, U. (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: a new look at a perceptual hypothesis. *Frontiers in Psychology*, *6*, 1–16. https://doi.org/10.3389/fpsyg.2015.00972

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Depriest, J., Glushko, A., Steinhauer, K., & Koelsch, S. (2017). Language and music phrase boundary processing in Autism Spectrum Disorder: An ERP study. *Scientific Reports*, *7*(1), 1-12. https://doi.org/10.1038/s41598-017-14538-y

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*, 1–9. https://doi.org/10.3389/fnhum.2017.00481

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Féry, C. (2017). *Intonation and prosodic structure*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/9781139022064

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414–420. https://doi.org/10.1016/S0022-5371(65)80081-0

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, *22*(6), 829–834. https://doi.org/10.1177/0956797611409589

Frank, S. L., & Christiansen, M. H. (2018). Hierarchical and sequential processing of language: A response to: Ding, Melloni, Tian, and Poeppel (2017). Rule-based and word-level statistics-based processing of language: insights from neuroscience. *Language, Cognition and Neuroscience*, *33*(9), 1213-1218. https://doi.org/10.1080/23273798.2018.1424347

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PLoS ONE*, *13*(5), 1–11. https://doi.org/10.1371/journal.pone.0197304

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The Psychology of Reading* (p. 559–586). Lawrence Erlbaum Associates, Inc.

Frazier, L., Clifton Jr, C., & Carlson, K. (2004). Don't break, or do: prosodic boundary preferences. *Lingua*, *114*(1), 3-27.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*(3), 259–281. https://doi.org/10.1006/brln.1995.1048

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences*, *13*(4), 175–181. https://doi.org/10.1016/j.tics.2009.01.001

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Friederici, A. D., Hahne, A., & Von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language*, *62*(3), 311–341. https://doi.org/10.1006/brln.1997.1906

Garrett, M., Bever, T., & Fodor, J. (1966). The active use of grammar in speech perception. *Perception & Psychophysics*, *1*(1), 30–32. https://doi.org/10.3758/BF03207817

Getz, H., Ding, N., Newport, E. L., & Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition*, *181*, 135–140. https://doi.org/10.1016/j.cognition.2018.08.019

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, *32*(5), 545–561. https://doi.org/10.1080/23273798.2016.1232419

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, *16*(2), 240–246. https://doi.org/10.1016/j.conb.2006.03.007

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), 1-14. https://doi.org/10.1371/journal.pbio.1001752

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. https://doi.org/10.1162/089892999563328

Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(7), 1207–1219. https://doi.org/10.1162/jocn.2008.20083

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). Neuroscience: The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 417–430. https://doi.org/10.1016/S0022-5371(81)90533-8

Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., & Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *NeuroReport*, *21*(1), 8–13. https://doi.org/10.1097/WNR.0b013e328330251d

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211–225. https://doi.org/10.1016/j.cognition.2005.04.006

Jeon, H. A., & Friederici, A. D. (2013). Two principles of organization in the prefrontal cortex are cognitive hierarchy and degree of automaticity. *Nature Communications*, *4*(1), 1–8. https://doi.org/10.1038/ncomms3041

Jin, P., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature Communications*, *9*(1), 1–15. https://doi.org/10.1038/s41467-018-07773-y

Jun, S. A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes*, *25*(7), 1201–1233. https://doi.org/10.1080/01690965.2010.503658

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2007). Discourse, syntax, and prosody: The brain reveals an immediate interaction. *Journal of Cognitive Neuroscience*, *19*(9), 1421–1434. https://doi.org/10.1162/jocn.2007.19.9.1421

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*(32), 7772–7781. https://doi.org/10.1523/JNEUROSCI.0168-17.2017

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8362–8367. https://doi.org/10.1073/pnas.0810928106

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds.

Developmental Cognitive Neuroscience, *5*, 86–94. https://doi.org/10.1016/j.dcn.2013.01.003

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, *27*(9), 4293–4302. https://doi.org/10.1093/cercor/bhw228

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., ... & Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, *114*(18), 3669-3678. https://doi.org/10.1073/pnas.1701590114

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of vision*, *15*(6), 1-46. https://doi.org/10.1167/15.6.4

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806. https://doi.org/10.1016/0749-596X(92)90039-Z

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, *29*(1–2), 34–55. https://doi.org/10.1080/02643294.2012.710600

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*(6461), 62-66.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*(2), 161-169. https://doi.org/10.1016/j.cub.2017.11.033

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, *22*(10), 870-882. https://doi.org/10.1016/j.tics.2018.08.003

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive context-free grammars. *PloS One*, *7*(10). https://doi.org/10.1371/journal.pone.0045885

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169-182. https://doi.org/10.1023/A:1005192911512

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., ... & Gao, J. H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, *29*(8), 3232-3240. https://doi.org/10.1093/cercor/bhy191

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249-271. https://doi.org/10.1007/BF01708573

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142–164. https://doi.org/10.1016/S0093-934X(02)00542-4

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162. https://doi.org/10.1016/j.bandl.2011.07.001

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295. https://doi.org/10.1023/A:1010443001646

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196. https://doi.org/10.1038/5757

Tavano, A. (2017). Internal rhythms to sentences [Video]. Retrieved from https://www.aesthetics.mpg.de/institut/veranstaltungen/vergangene-veranstaltungen/neural-oscillations-in-speech-and-language-processing-berlin-2017.html

Uddén, J., de Jesus Dias Martins, M., Zuidema, W., & Tecumseh Fitch, W. (2019). Hierarchical Structure in Sequence Processing: How to Measure It and Determine Its Neural Implementation. *Topics in Cognitive Science*, 1-15. https://doi.org/10.1111/tops.12442

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition*, *18*(4), 380-393. https://doi.org/10.3758/BF03197127

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, *10*, 1-8. https://doi.org/10.3389/fnhum.2016.00274