A Special Role of Syllables, But Not Vowels or Consonants, for Nonadjacent Dependency Learning

Ivonne Weyers[1,2] and Jutta L. Mueller[1,2]

Abstract

■ Successful language processing entails tracking (morpho) syntactic relationships between distant units of speech, so-called nonadjacent dependencies (NADs). Many cues to such dependency relations have been identified, yet the linguistic elements encoding them have received little attention. In the present investigation, we tested whether and how these elements, here syllables, consonants, and vowels, affect behavioral learning success as well as learning-related changes in neural activity in relation to item-specific NAD learning. In a set of two EEG studies with adults, we compared learning under conditions where either all segment types (Experiment 1) or only one segment type (Experiment 2) was informative. The collected behavioral and ERP data indicate that, when all three segment types are available, participants mainly rely on the syllable for NAD learning. With only one segment type available for learning, adults also perform most successfully with syllable-based dependencies. Although we find no evidence for successful learning across vowels in Experiment 2, dependencies between consonants seem to be identified at least passively at the phonetic-feature level. Together, these results suggest that successful item-specific NAD learning may depend on the availability of syllabic information. Furthermore, they highlight consonants' distinctive power to support lexical processes. Although syllables show a clear facilitatory function for NAD learning, the underlying mechanisms of this advantage require further research. ■

INTRODUCTION

Processing dependencies between temporally distant units of speech (e.g., "*he* sing*s*" or "*The girl* the boy kissed *ran away*.") is essential to human language. The hierarchical structure of language requires tracking the relationships between units, such as words or phrases, beyond the directly adjacent environment and across variable numbers of intervening elements. By studying the cognitive processes involved in the detection and learning of these so-called nonadjacent dependencies (NADs), we can learn about some of the most basic mechanisms supporting language processing and acquisition. Although it is known that NADs can be learned in principle, and many external cues have been identified that guide learning (Wilson et al., 2018), comparatively little research has explored the role of the speech sounds themselves that carry the dependency. The specific acoustic features of the input are particularly relevant during early acquisition of NADs because, initially, phonetic surface-level forms have to be identified as re-occurring patterns in the input. The resulting early, item-specific representations form the basis for the later development of more abstract, categorical relations (Mueller, ten Cate, & Toro, 2020; Culbertson, Koulaguina, Gonzalez-Gomez, Legendre, & Nazzi, 2016). In other words, the ability to detect and recognize dependencies between phonetic elements in linguistic input

can be understood as a precursor or an initial "bootstrapping" process that paves the way for the acquisition of higher-level syntactic rules. This study aims to compare learning-related changes in neural activity during NAD learning and processing and their dependence on the segmental level at which they are encoded, namely, syllables, consonants, and vowels.

In this line of research, artificial grammar learning (AGL) paradigms have proven a useful means to isolate surface-level structural processing from semantic meaning and syntactic function, while also controlling for effects of previous language learning (e.g., Frost & Monaghan, 2016; Newport & Aslin, 2004; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Reber, 1967). Gómez (2002), for example, used sequences of nonword triplets (*pel* kicey *rud*, *vot* wadim *jic*) in which the first (A) and third (B) word encoded a simple AXB NAD across a variable middle word (X). After passive auditory exposure to these strings, both adult and infant participants showed behavioral evidence of learning by successfully discriminating between consistent (*pel* kicey *rud*) and inconsistent (*pel* kicey jic) exemplars. Since then, many studies have confirmed that adults (Frost & Monaghan, 2016; Vuong, Meyer, & Christiansen, 2016; Mueller, Friederici, & Männel, 2012; van den Bos, Christiansen, & Misyak, 2012; Citron, Oberecker, Friederici, & Mueller, 2011; Mueller, Oberecker, & Friederici, 2009; Peña, Bonatti, Nespor, & Mehler, 2002), infants (Marchetto & Bonatti, 2015; Mueller et al., 2012; Gómez & Maye, 2005), and even nonhuman primates

[1]University of Vienna, [2]University of Osnabrück

(Malassis, Rey, & Fagot, 2018; Milne et al., 2016) are able to learn such arbitrary AXB NAD relations from speech (for a review, see Mueller, Milne, & Männel, 2018).

A variety of NADs and their learnability have been investigated, differing, for example, in the type of relationship between dependent units (e.g., repeated elements [AXA] or item-specific dependencies [AXB]) and the complexity of the structure that is encoded (e.g., simple AXB relationships or crossed dependencies $[A_1A_2A_3B_1B_2B_3]$; for a review, see Wilson et al., 2018). Many studies have further focused on identifying circumstances or cues that facilitate (or hinder) the learning of NADs. They are learned better, for instance, when the variability of intervening X elements is high (Gómez & Maye, 2005; Onnis, Monaghan, Christiansen, & Chater, 2004; Gómez, 2002), when the dependent elements appear in edge positions (Endress, Nespor, & Mehler, 2009), when they are highlighted by pauses (Mueller, Bahlmann, & Friederici, 2008; Peña et al., 2002) or other prosodic cues (Grama, Kerkhoff, & Wijnen, 2016; Mueller, Bahlmann, & Friederici, 2010), or when they are perceptually similar (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004). Few studies, however, have systematically investigated the role of the specific linguistic elements encoding the NAD and their impact on learning success and processing. In all of the studies cited above, the encoding elements were artificial monosyllabic or multisyllabic units. Yet, whether NADs are coded by those units or rather by segments forming those units, that is, consonants and vowels, is not known. Thus, in this study, we ask whether syllables, consonants, and vowels are equally suitable computational units for NAD learning. Before we turn to this study aiming to answer these questions, we briefly review the relevant previous literature on the role of linguistic segments in speech and particularly in NAD learning. To this aim, we consider both behavioral and neurophysiological experiments as both may provide complementary information about the nature of the involved cognitive processes.

The general notion that syllables are linguistic units relevant for both language comprehension and production is not new (e.g., Bertoncini & Mehler, 1981; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Hooper, 1972). Word production models largely agree that the syllable plays a role in the speech production process and merely disagree on when syllabic information is made available (e.g., Schiller & Costa, 2006; Levelt, Roelofs, & Meyer, 1999; Dell, 1986, 1988; Shattuck-Hufnagel, 1983). In Levelt's model of speech production (Levelt, 1989), for instance, it is presumed that for frequently used syllables of a given language, independent representations and motor programs are stored in the so-called mental syllabary, the activation of which allows for effective phonetic encoding and fluent articulation (e.g., Cholin, Dell, & Levelt, 2011; Cholin, 2008; Levelt et al., 1999; Levelt, 1992). Such articulatory motor programs at the syllable level have received support from both modeling (Guenther, 2016; Guenther, Ghosh, & Tourville, 2006) and experimental

work (Ziegler, Aichert, & Staiger, 2010; Cholin, Levelt, & Schiller, 2006; Carreiras & Perea, 2004). More recently, studies using EEG have shown that cortical activity recorded during continuous speech perception tracks linguistic structure at different levels, including syllable and word boundaries (Batterink, 2020; Choi, Batterink, Black, Paller, & Werker, 2020; Poeppel & Assaneo, 2020; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Ding & Simon, 2014); in fact, Giraud and Poeppel (2012) showed that syllabic structure is tracked already in primary auditory cortex, and a number of studies have related the precision of syllable tracking to language skills, particularly reading (Goswami, 2011; Abrams, Nicol, Zecker, & Kraus, 2009), which further supports the relevance of the syllabic processing level. Similarly, a recent computational approach highlights a possible role of (acoustic) syllables even in prelinguistic perception and speech sequencing (Räsänen, Doyle, & Frank, 2018).

Because of its apparent role as a basic perceptual and production unit, the syllable has been a natural target unit for AGL studies concerned with NAD learning (e.g., Mueller et al., 2012; de Diego-Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007; Endress & Bonatti, 2007; Peña et al., 2002). Mueller et al. (2012), for instance, used a classic oddball design and auditorily presented adult listeners with a segmented stream of syllable sequences encoding an item-specific AXB NAD (*fi*kato, *le*robu), which was interspersed with few deviant items in which the final syllable violated the AXB dependency (*fi*web<u>u</u>, *le*kot<u>o</u>). While participants performed a target detection task, their EEG response was recorded. For those participants who showed behavioral evidence of learning, deviant detection was indexed by an N2/P3 complex in the ERPs. De Diego-Balaguer et al. (2007) found similar ERP effects, across both learners and nonlearners, using comparable items (*nu*lade vs. *de*lanu) and design.

With regard to the smaller segmental level, there is evidence that consonants and vowels are not mere superimposed linguistic categories we use to classify speech sounds but actually constitute separable classes also at the neural level (Caramazza, Chialant, Capasso, & Miceli, 2000; Boatman, Hall, Goldstein, Lesser, & Gordon, 1997), which are processed by distinct neural mechanisms (Carreiras, Dunabeitia, & Molinaro, 2009; Carreiras, Gillon-Dowens, Vergara, & Perea, 2009; Carreiras & Price, 2008; Carreiras, Vergara, & Perea, 2007). Carreiras and Price (2008) presented participants with written words in which either consonants (e.g., PRI<u>VAM</u>ERA) or vowels (PRIM<u>EV</u>A<u>R</u>A) were transposed. Participants had to either read the words aloud or perform a lexical decision task while MRI brain scans were acquired. Whereas vowel changes induced increased relative activation in the STS during reading out loud, consonant changes exhibited increased activation in the right middle frontal cortex in the lexical decision task. The authors concluded that vowel changes placed additional demands on areas relevant for prosodic processing—possibly because of

self-monitoring processes engaged during production. Consonant changes, on the other hand, additionally taxed inhibitory control mechanisms during lexical decision, indicating more difficulties with lexico-semantic processing.

Consonants and vowels have been compared extensively concerning their functional role in word segmentation (Nazzi, Poltrock, & Von Holzen, 2016; Toro, Nespor, Mehler, & Bonatti, 2008; Mehler, Peña, Nespor, & Bonatti, 2006; Bonatti, Peña, Nespor, & Mehler, 2005; Newport & Aslin, 2004; Peña et al., 2002) and word identification/lexical selection (Delle Luche et al., 2014; Havy, Serres, & Nazzi, 2014; Carreiras, Dunabeitia, et al., 2009; Carreiras, Gillon-Dowens, et al., 2009; New, Araújo, & Nazzi, 2008; Cutler, Sebastian-Galles, Soler-Vilageliu, & Van Ooijen, 2000). When asked to reconstruct a word from a nonword (*kebra*) by changing a single phoneme, for instance, adult participants prefer to make a vowel change (*cobra*) rather than a consonant change (*zebra*; Sharp, Scott, Cutler, & Wise, 2005; Cutler et al., 2000; Van Ooijen, 1996). This observation holds even cross-linguistically, both for languages like Spanish with a larger number of consonants than vowels and for languages with a relatively equal consonant–vowel ratio, such as Dutch (Cutler et al., 2000). Similarly, adults can exploit co-occurrence statistics (transitional probabilities) between consonants to segment a continuous speech stream into word-like units (Nazzi et al., 2016; Toro, Nespor, et al., 2008; Mehler et al., 2006; Bonatti et al., 2005). Equivalent transitional probabilities between vowels can only be exploited for this purpose under highly redundant conditions, and in a direct comparison with equal distributional information across both segments, adults preferentially extract words based on consonant rather than vowel frames (Bonatti et al., 2005; Newport & Aslin, 2004).

This asymmetry is addressed by the "consonant–vowel (CV) hypothesis," which proposes that consonants and vowels assume at least partially distinct functions in linguistic processing: Consonants primarily encode lexical information, whereas vowels carry sentence prosody and thereby supply information about syntactic constituency and sentence structure (Nespor, Peña, & Mehler, 2003). Although there is abundant evidence for the former assumption (see above), evidence for the latter remains scarce. A possible structural role of vowels has mainly been tested with the help of AGLs encoding item-independent repetition rules (e.g., fef*u*f*u*, ku*f*ef*e*). These are deemed good examples of structural learning for two reasons: They require generalization of a regularity (ABB/ABA) beyond specific items (e.g., lum*ifi* vs. lum*ifa*), and vowel repetitions specifically can be conceptualized as an extreme case of vowel harmony, a phonetic assimilation operation that provides cues to morphosyntactic constituency in some languages (e.g., Turkish, Hungarian, Finnish). There is tentative evidence[1] that adults learn such reduplication rules better when they are encoded by vowels rather than consonants (Monte-Ordoño & Toro, 2017a; Toro, Nespor,

et al., 2008; Toro, Shukla, Nespor, & Endress, 2008), although participants in these studies were speakers of Catalan–Spanish (Monte-Ordoño & Toro, 2017a) and Italian (Toro, Nespor, et al., 2008; Toro, Shukla, et al., 2008), languages that do not typically harmonize.

Item-specific NADs between vowels and consonants have only scarcely been researched. Newport and Aslin (2004) reported successful segmentation of a continuous syllable stream into trisyllabic words based on transitional probabilities between nonadjacent consonant (*p_g_t_*) and vowel (*_a_u_e*) frames. The authors concede, however, that the dependencies they employed may not exactly qualify as "nonadjacent" at the segmental level, as the entire consonant/vowel frame always remained fixed and the middle segment did not vary (i.e., *pxxxtx*). Specifically, if the assumed statistical learning mechanism operated on separate representations of consonant and vowel tiers, or if the segments were simply grouped together because of their perceptual similarity, the given dependencies would actually exist between adjacent units (Newport & Aslin, 2004).

This Study

Our aim in this study was to compare syllables, consonants, and vowels as carriers of item-specific NADs. We focused on item-specific dependencies of the type AXB, for example, as used by Gómez (2002). These lend themselves well to study NADs in natural language, because particularly at the local, morphosyntactic level, such NADs often exist between specific units (e.g., she *is* runn*ing*) whose phonetic surface-level form and arbitrary relationship need to be learned.

The summarized previous studies have shown that auditorily presented NADs at the syllable level can be learned by adults. From these results, it remains unclear, however, whether the relevant learning mechanism operates on the syllable level or whether NAD learning is possibly biased or guided by the lower segmental level. We addressed this question in Experiment 1, using an experimental design with alternating learning and test phases. In the learning phases, participants were exclusively exposed to a syllable-based NAD. In the test phases, they were also tested for the consonant- and vowel-based dependencies inherent in this syllable-based NAD. If participants also showed signs of discrimination for either of these, this would suggest a special role for segments smaller than the syllable in the learning of syllable-based NADs.

On the basis of the available evidence from studies comparing consonants and vowels, it seems that redundancies or repetition regularities are learned better across vowels, whereas dependencies between nonrepetitive, distinctive features are learned better across consonants. These previous studies have so far only compared the role of consonants and vowels in segmentation and/or repetition detection tasks. None of them have evaluated their role in NAD learning specifically and asked whether, in

principle, item-specific dependencies can be learned across these segments. We focused on this second question in Experiment 2, using an oddball paradigm to expose three separate groups of adults to input from which only one type of dependency (syllable/consonant/vowel-based) could be learned.

During both experiments, we recorded participants' EEG. The relatively low number of participants showing behavioral evidence of learning across previous studies (e.g., Mueller et al., 2012, 10/46 learners; de Diego-Balaguer et al., 2007, 8/16 learners) underscores the importance of an additional measure such as EEG. EEG can provide important insights into online processing even in the absence of offline learning success and into possible qualitative differences in the neural processes that support learning across these segments.

In line with the cited studies, we expected to find high accuracy rates along with ERP evidence of learning for syllable NADs in both experiments. With regard to the two segmental conditions/groups, hypotheses are more difficult to formulate. One could tentatively expect that item-based dependencies, which rely on the identification, association, and storage of specific segmental units, are learned better if they are encoded by consonants, as these appear to have larger distinctive power in the context of word identification and lexical selection. An advantage for vowels has mainly been postulated in the learning of repetitions. Although we do not employ repetitions, we purposefully chose phonetically similar segments for our stimulus material. It is thus conceivable that perceptual similarity in the vowel condition/group serves as a similar (albeit weaker) cue to the dependency relationship. As both of these options are conceivable, we did not have any specific hypotheses with regard to performance or neurophysiological responses in the two segmental conditions/groups.

EXPERIMENT 1

Methods

Participants

The experiment was approved by the ethics committee of the University of Osnabrück and adheres to the guidelines of the Declaration of Helsinki (2013). Participants were recruited from the University of Osnabrück, gave written informed consent before participating in the experiment, and received either course credit or payment as compensation for their participation. All tested participants were native speakers of German, were right-handed, and had normal hearing, normal or corrected-to-normal vision, and no history of neurological conditions. On the basis of the average number of participants in similar previous studies (e.g., Citron et al., 2011; Mueller et al., 2009; Mueller, Bahlmann, et al., 2008), we aimed for a minimum of 25 participants entering the final analysis. Five of the 34 tested participants had to be excluded from analysis

because of technical difficulties or high artifact rate in the EEG. The remaining 29 participants (2 men, 27 women) were between 18 and 29 years old (mean = 21.62 years, SD = 2.6 years).

Stimuli and Procedure

The stimulus material consisted of trisyllabic sequences of individually recorded CV syllables spoken by a trained female speaker. Recordings of similar length and pitch were selected from several recorded exemplars, digitized (44.1 kHz/16-bit sampling rate, mono), normalized to the same sound intensity, and cut to the same length (380 msec). The two syllable frames *bi* X *pe* and *go* X *ku* served as standards of an AXB-type NAD. Within items, syllables were separated by 50-msec pauses, and items were separated by 700-msec interstimulus pauses (Mueller, Bahlmann, et al., 2008; Peña et al., 2002). In an attempt to boost learning of this pairwise association, several cues known to aid NAD learning were integrated: The phonemes coding the dependency were selected for their perceptual similarity, that is, the consonants differed only in voicing (b–p, g–k) and the vowels were both either rounded back (o–u) or unrounded front (i–e) vowels (Creel et al., 2004; Newport & Aslin, 2004); the relevant units A and B were placed in edge positions (Endress et al., 2009); variability of the middle element X was high with 24 different syllables (*la*, *ma*, *na*, *ra*, *sa*, *ta*, *dä*, *nä*, *rä*, *sä*, *tä*, *wä*, *dö*, *lö*, *mö*, *sö*, *tö*, *wö*, *dü*, *lü*, *mü*, *nü*, *rü*, *wü*; Gómez, 2002); and attention to stimuli was required given the active design (Mueller et al., 2012).

Eight learning phases alternated with eight test phases. In each learning phase, the same 48 correct syllable items (see Table 1 for examples) were repeated twice in pseudo-random order while a fixation cross was shown on the screen (see Figure 1). Participants were instructed to listen attentively and detect a regularity inherent in the input, based on which they would have to make a grammaticality judgment in the test phases. During the test phases, items were presented individually,[2] and after a short delay of 900 msec, a response cue appeared on the screen (see Figure 1). No feedback was provided. The test items comprised correct (e.g., bidape) and incorrect (e.g., bida<u>ku</u>) exemplars of the syllable dependency as well as vowel-based and consonant-based variants of the AXB dependency. In correct exemplars of this segmental-level dependency, either the vowel dependency (e.g., k*owabu*) or the consonant dependency (e.g., *gewako*) remained constant compared to the learning phase syllable NADs. Incorrect exemplars of these lower-level NADs violated the respective dependency on either the final consonant (e.g., gewa<u>pi</u>) or the final vowel (e.g., kowa<u>ge</u>).[3] In contrast to Newport and Aslin's (2004) technically adjacent segmental dependencies ($p\_g\_t\_/\_a\_u\_e$), our segmental-level dependencies were truly nonadjacent. Eight separate middle syllables (*da*, *wa*, *lä*, *mä*, *nö*, *rö*, *sü*, *tü*) were used for the test phase items. Each test phase comprised 24

Table 1. Correct and Incorrect Exemplars by Condition, With Violations Underlined

Learning Phases	Test Phases					
bilape	Syllables		Consonants		Vowels	
bidäpe	correct	incorrect	correct	incorrect	correct	incorrect
bidöpe	bidape	bidaku	budapi	budako	pidage	pidabu
bimüpe	biläpe	biläku	buläpi	buläko	piläge	piläbu
gomaku	biröpe	biröku	buröpi	buröko	piröge	piräbu
gowäku	gowaku	gowape	gewako	gewapi	kowabu	kowage
gosöku	gomäku	gomäpe	gemäko	gemäpi	komäbu	komäge
gorüku	gotüku	gotüpe	getüko	getüpi	kotübu	kotüge

test items, holding correct and incorrect exemplars of the three conditions. We did not control for an exactly equal distribution of items per condition in each test phase of the four item lists created but restricted the number of items per segmental condition in each learning phase to a minimum of four and a maximum of 12.[4]

During the experiment, participants were seated in a chair at a distance of 100 cm from a computer screen while the stimuli were played via loudspeakers.

Data Acquisition and Preprocessing

The continuous EEG was recorded from a 64 Ag/AgCl electrode cap (TMSI B.V.; International 10–20 system of electrode placement), using a TMSi 72 Refa amplifier system and the TMSi Polybench recording software. The data were recorded with an implicit average online reference of all electrodes. The ground electrode was placed on the left collar bone; two additional single electrodes were placed on both temples, as well as one placed above and one below the left eye that recorded the horizontal and vertical EOG. Impedances of all electrodes were kept below 5 kΩ, and the data were sampled at a rate of 512 Hz with no hardware filters (except for antialiasing) in place.

The EEG data were processed offline with MATLAB (Version R2017a, The MathWorks Inc., 2010) and the EEGLAB open source toolbox (Version 14.1.1b; Delorme & Makeig, 2004). Before averaging, the continuous data were rereferenced to average mastoids, detrended, and filtered with two separate digital windowed-sinc finite impulse response filters (window type: Kaiser), one high-pass filter (−6 dB half-amplitude cutoff, 0.1-Hz cutoff frequency, filter order: 9274), and a low-pass filter (−6 dB, 30 Hz, 188), to remove slow drifts and line noise. For ERP averaging, epochs from −100 to 1000 msec after the onset of the final syllable (or vowel in case of the vowel condition) were cut out. Independent component analysis (ICA) was performed on the individual participant data to remove eye movement artifacts.[5] Trials containing any remaining artifacts were selected using a semiautomatic procedure coupled with visual inspection and excluded from further analysis. A baseline correction (−100 to 0 msec) was applied to the cleaned data, which were then averaged by participant for each experimental condition. The average number of epochs per participant entering the final analysis amounted to 28.86 (SD = 3.03) correct
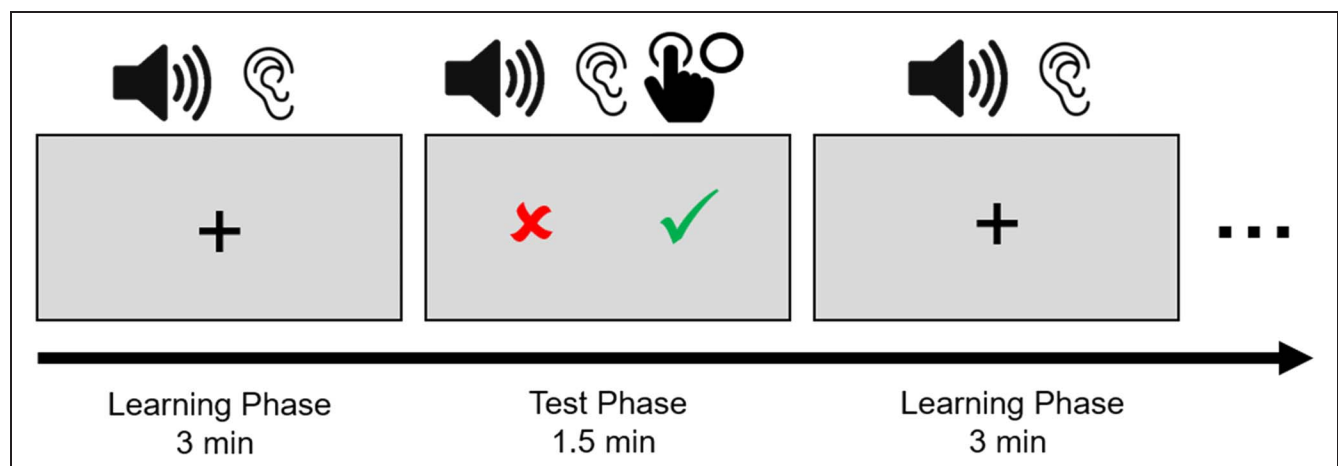

Figure 1. Experimental procedure in Experiment 1.

and 28.21 (SD = 3.65) incorrect syllable items, 28.55 (SD = 3.42) correct and 28.69 (SD = 3.17) incorrect consonant items, and 28.76 (SD = 3.86) correct and 29.17 (SD = 2.88) incorrect vowel items and did not significantly differ within or between conditions.

Data Analysis

The behavioral data were analyzed using RStudio (R Core Team, 2020) and the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). A generalized linear mixed-effects model (GLMM) including a binomial link function was fitted to investigate the effects of condition (syllable, vowel, consonant), phase (1–8), and list (1–4), as well as the interaction between condition and phase, on the response accuracy data. The predictor phase was interval scaled and centered to facilitate model fitting. The factor list was included to confirm that the distribution of items between lists did not affect learning. Participants and items were included in the model as intercept-only random effects. Standard treatment contrasts were implemented with the consonant condition and list A as reference levels of the respective predictors. In total, 5568 data points (29 participants, eight test phases with 24 items each) were entered into the model. To test whether, across the whole group, participants' response accuracy rates in the three conditions exceeded chance level, separate intercept models were fitted to the accuracy data of each condition. Each of these models comprised 1856 data points. *p* Values for fixed effects were calculated via Wald tests (standard for *glmer* in lme4).

Participants were then split into learners (response accuracy ≥ 64%, indicated by a binomial test for chance response, *p* = .033) and nonlearners based on their behavioral performance in each condition. Because of the fact that the vowel and consonant dependencies were inherent in the syllable condition items, participants who learn either of these two dependencies should also be able to correctly evaluate the syllable condition items. Participants were thus classified as "syllable" learners, "vowel + syllable" learners, or "consonant + syllable" learners.[6] We further tested whether the latter two groups already performed above chance level in both conditions in the first test phase. If so, this might be indicative of segment-based NAD learning; if not, it may indicate sequential learning effects. To this end, we fit additional intercept models to the accuracy data of the learner groups, separately for the first and second test phases. For the vowel + syllable learners, the respective models included 216 data points, whereas 72 data points were entered into the consonant + syllable learner models.

The FieldTrip toolbox for EEG/MEG analysis (Oostenveld, Fries, Maris, & Schoffelen, 2011) was used for statistical analyses of the EEG data. Separate nonparametric cluster-based permutation tests using dependent samples *t* tests were run for each segmental condition, comparing ERP responses to correct and incorrect items.

All electrodes, except for the EOG and reference electrodes, were included in the analysis. Because previous literature does not provide specific expectations as to the timing and location of possible effects, the ERP analysis was exploratory. Only the latency range of 0–50 msec was excluded from analysis to increase power (Groppe, Urbach, & Kutas, 2011), because we were not interested in any auditory brain stem or primary auditory cortex responses in this very early time window (for a review, see Pratt, 2012; Picton, 2011). The initial sample-specific test statistic threshold was set to 0.05. The minimum number of neighboring channels to be included in a cluster was set to two, and neighboring channels were identified with a spatial neighborhood template by use of the triangulation method. For the cluster statistic permutation test, we employed the maximum sum approach and set the alpha level of the permutation test to .05 (distributed over both tails) and the number of draws from the permutation distribution to 2000 (Monte Carlo sampling).

Results

Behavioral Results

Accuracy rates were at 80.8% (SD = 14.4%) in the syllable condition, 59.4% (SD = 16.1%) in the vowel condition, and 47.9% (SD = 14.4%) in the consonant condition. Figure 2 further illustrates the distribution of the participant averages. In the consonant condition, the median is at 50% accuracy and the low dispersion and range of the data (except for a few outliers) suggest little deviation from chance level. The median response accuracy in the vowel condition is higher with 57.8%, and both dispersion and range are greater than for the consonants, but the two boxes still overlap slightly and the lower quartile range
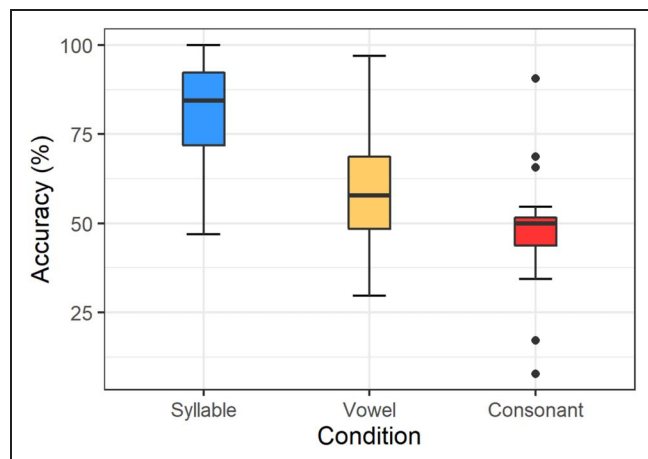

Figure 2. Average response accuracy separated by condition in Experiment 1. The length of each box represents the interquartile range (IQR), limited by the upper and lower quartiles; horizontal lines mark the median; whiskers illustrate data outside the IQR (maximum: 1.5 × IQR); and outlier participants are indicated by dots (any data points outside 1.5 × IQR).

of the vowel plot includes chance level. The accuracy plot of the syllable condition prominently differs from those of the other two conditions. Although the interquartile ranges are similar to the vowel condition, there is no overlap with either of the other two conditions as median accuracy lies at 84.4%.

Figure 3 depicts the development of average response accuracies by condition across test phases. It is clear that, in the consonant condition, accuracy remained at chance level (black dashed line) throughout the entire experiment. Although accuracy seems to increase slightly across phases in the vowel condition, it remained below the significantly above-chance threshold (64%) up to the penultimate phase. For the syllable condition, average performance already exceeded chance in the first test phase and improved almost continuously from thereon, suggesting a clear learning effect. The previously specified model showed a significant interaction between phase and condition for both the syllable and vowel conditions, as well as a significant syllable effect, when compared against the consonant condition (see Table 2). The factor list was nonsignificant. To further investigate the encountered interactions post hoc, separate GLMMs were fitted for each condition with phase as a predictor (not centered this time). The parameter list was excluded, but participants and items were again entered as random effects on the intercept. After correcting for multiple comparisons via the Bonferroni method ($p < .008$), both the intercept ($\beta_0 = 2.02$, $SE = 0.31$, $z = 6.44$, $p < .0001$) and the fixed effect phase ($\beta_1 = 0.61$, $SE = 0.07$, $z = 8.40$, $p < .0001$) were significant for the syllable condition. Phase was also significant for the vowel condition ($\beta_1 = 0.19$, $SE = 0.06$, $z = 3.34$, $p < .001$; but not the intercept [$\beta_0 = 0.62$, $SE = 0.38$, $z = 1.64$, $p = .10$]), suggesting a phase effect on response accuracy for both the syllable and vowel conditions. No significant effects were found for the consonant condition ($\beta_0 = -0.12$, $SE = 0.40$, $z = -0.31$, $p = .76$; $\beta_1 = -0.03$, $SE = 0.06$, $z = -0.59$, $p = .55$). The additionally fitted intercept models for each segmental condition revealed that the estimated intercept
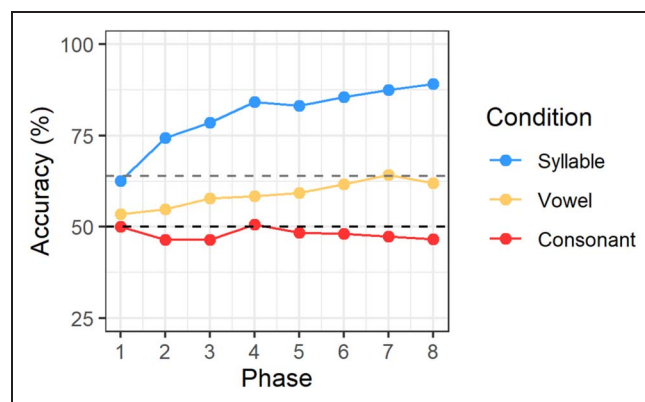
Table 2. Results of the GLMM Examining the Effects of Condition, Phase, and List on Behavioral Response Accuracy

	FE	*SE*	*z Value*	*p*
Syllable	**1.817**	**.390**	**4.655**	**<.001**
Vowel	0.635	.388	1.637	.102
Phase	−0.027	.053	−0.517	.605
List B	0.161	.225	0.715	.475
List C	0.378	.226	1.677	.094
List D	0.236	.232	1.015	.310
Syllable × Phase	**0.548**	**.086**	**6.373**	**<.001**
Vowel × Phase	**0.195**	**.075**	**2.593**	**<.01**

Bold print indicates significant effects ($p < .05$). FE = fixed effect estimates; *SE* = standard error.

was significantly different from zero only for the syllable condition ($\beta_0 = 1.89$, $SE = 0.29$, $z = 6.43$, $p < .0001$), suggesting above-chance performance in this condition (vowels: $\beta_0 = 0.61$, $SE = 0.38$, $z = 1.64$, $p = .10$; consonants: $\beta_0 = -0.12$, $SE = 0.40$, $z = -0.30$, $p = .76$).

Through the categorization of participants into learners and nonlearners, we identified 25 people who successfully learned the syllable dependency, nine of whom also performed well in the vowel condition and three who qualified as learners in both the consonant and syllable conditions. Four participants were nonlearners in all of the conditions. The models fitted to the vowel + syllable, consonant + syllable, and syllable learner data to investigate above-chance performance in the first and second test phases could not be fitted as initially described because of issues with singularity. We therefore reduced the models' complexity and fitted simple generalized linear models instead, omitting the random effects terms for participants and items (Bates, Kliegl, Vasishth, & Baayen, 2015). Bonferroni-corrected ($p < .008$) results for the estimated intercepts showed that the response accuracies in the syllable condition were already above chance in the first phase for both syllable ($n = 25$; $\beta_0 = 0.50$, $SE = 0.16$, $z = 3.09$, $p < .002$) and vowel + syllable ($n = 9$; $\beta_0 = 0.78$, $SE = 0.26$, $z = 2.98$, $p < .003$) learners. The latter group's response accuracy for vowels, however, only surpassed chance in the second test phase ($\beta_0 = 0.74$, $SE = 0.26$, $z = 2.85$, $p < .004$). For the sake of completeness, intercept models for the consonant + syllable learners were fitted as well, although the group size was admittedly very small ($n = 3$). The intercept models revealed chance-level responses in the syllable condition in the first test phase ($\beta_0 = -5.23$e-17, $SE = 4.71$e-01, $z = 0$, $p > .008$), which rose significantly above chance only in the second test phase ($\beta_0 = 1.39$, $SE = 0.50$, $z = 2.77$, $p < .006$). Performance for the consonant items remained at chance in both phases (Phase 1: $\beta_0 = -1.04$, $SE = 0.48$, $z = -2.19$, $p > .008$; Phase 2: $\beta_0 = 1.20$, $SE = 0.47$, $z = 2.59$, $p > .008$).

Figure 3. Average response accuracies per test phase (1–8) separated by condition in Experiment 1. The black dashed line indicates chance level; the gray dashed line marks significantly above-chance level (64%).

ERP Results

The cluster-based permutation test comparing the averaged ERP responses to correct and incorrect syllable condition items revealed a significant grammaticality effect ($p < .05$), corresponding to two clusters. The first cluster spanned a time window of approximately[7] 190–470 msec and was observed as a negativity in the incorrect condition. Inspection of both the ERP and topographical plots revealed, however, that the test seemed to have grouped two separate effects into one based on the parameters specified above (specifically the minimum number of neighboring channels being set to two) and a spatial overlap between the two clusters. In particular, two separate peaks were clearly visible at frontal electrodes (Figure 4A, i).[8] The first effect began frontally and developed into a broadly distributed effect including centro-parietal electrodes (Figure 4A, iii), peaking around 230–240 msec. The second effect began with a centro-parietal focus (hence the overlap), peaking between 370 and 380 msec (Figure 4A, iv), but developed into a right-lateralized effect with a broad frontal-to-parietal distribution. The second cluster began at around 830 msec, lasted until the end of the epoch, and was observed as a negativity with a fronto-central, right-lateralized distribution (Figure 4A, v).

The equivalent comparisons in the consonant and vowel conditions did not yield any significant differences. The syllable learners ($n = 25$) showed a similar significant grammaticality effect ($p < .05$) as reported for the whole group, except that here, three clusters were identified. The first cluster began around 150 msec and was observed to last until approximately 520 msec with a broad distribution. The second cluster was observed between 530 msec and roughly 800 msec as a negativity with a fronto-central, right-lateralized distribution. The third cluster, also visible as a negativity, occurred between approximately 830 and 1000 msec, also with a right-lateralized fronto-central focus (see Figure 4). Cluster-based permutation tests performed on the averaged data of the successful vowel ($n = 9$) and consonant ($n = 3$) learners' averaged data were nonsignificant.

Discussion

In Experiment 1, we investigated adults' learning of auditorily presented NADs from segmented streams of trisyllabic AXB items. Whereas the syllable condition tested for successful learning of the syllable dependency as presented in the learning phase, the vowel and consonant conditions aimed at assessing whether the basis of this learned association lay with either of these smaller segmental units. In other words, we tested whether, given the availability of all three segments, participants would memorize entire syllable frames or build representations based on vowels or consonants, respectively.

The behavioral data showed that participants were by far most successful at distinguishing correct and incorrect syllable items at test. We thereby replicated previous findings (Mueller et al., 2012; de Diego-Balaguer et al., 2007; Endress & Bonatti, 2007; Peña et al., 2002). Neither of the other two smaller segments seem to have been particularly accessible for NAD learning. If at all, correct and incorrect exemplars of the vowel-based NAD were successfully differentiated offline by a larger number of learners ($n = 9$) than consonant-based NADs ($n = 3$).[9] However, the comparison of these learners' accuracy rates for syllables and vowels in the first two test phases showed a sequential learning pattern: Although their accuracy rates for syllable items already exceeded chance in the first phase, they only performed above chance level starting from the second test phase in the vowel condition. The specific design employed here possibly invited strategic evaluation of (early) test phase exemplars, resulting in the (later) application of a learned regularity that was not built exclusively on learning phase input. The data available from the few consonant + syllable learners ($n = 3$) were less clear but tentatively suggested simultaneous onset of above-chance performance in both conditions.

The EEG data supported the conclusion that participants likely built a syllable-based and not a consonant- or vowel-based representation, because the only significant ERP effects were found in the syllable condition. In the latter, we encountered a broadly distributed negativity followed by another late negativity with a fronto-central focus in response to incorrect items. We interpreted the first negative shift as two separate effects, namely, an N200 with a broad distribution followed by an N400-like effect with a typical centro-parietal topography. This combination and distribution of effects has typically been found in auditory speech processing studies investigating semantic violations within the sentence context (Van Den Brink, Brown, & Hagoort, 2001; Hagoort & Brown, 2000; Connolly, Phillips, & Forbes, 1995; Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990; McCallum, Farmer, & Pocock, 1984). Van Den Brink et al. (2001) specifically investigated the differentiation of the two effects by comparing participants' ERP responses to sentences with semantically incongruous (target: *penseel*/brush) but phonetically congruous final words (*De schilder kleurde de details in met een klein pensioen*/The painter colored the details with a small pension) to those elicited by semantically and phonetically incongruous sentence-final words (*De schilder kleurde de details in met een klein doolhof*/The painter colored the details with a small labyrinth). Whereas the N400 appeared in response to words at odds with the semantic sentence context, the additional N200 was elicited whenever the target word also constituted a mismatch with the expected word on the phonological level. As a result, the N200 was interpreted as an index of phonological processing that interacted with semantic
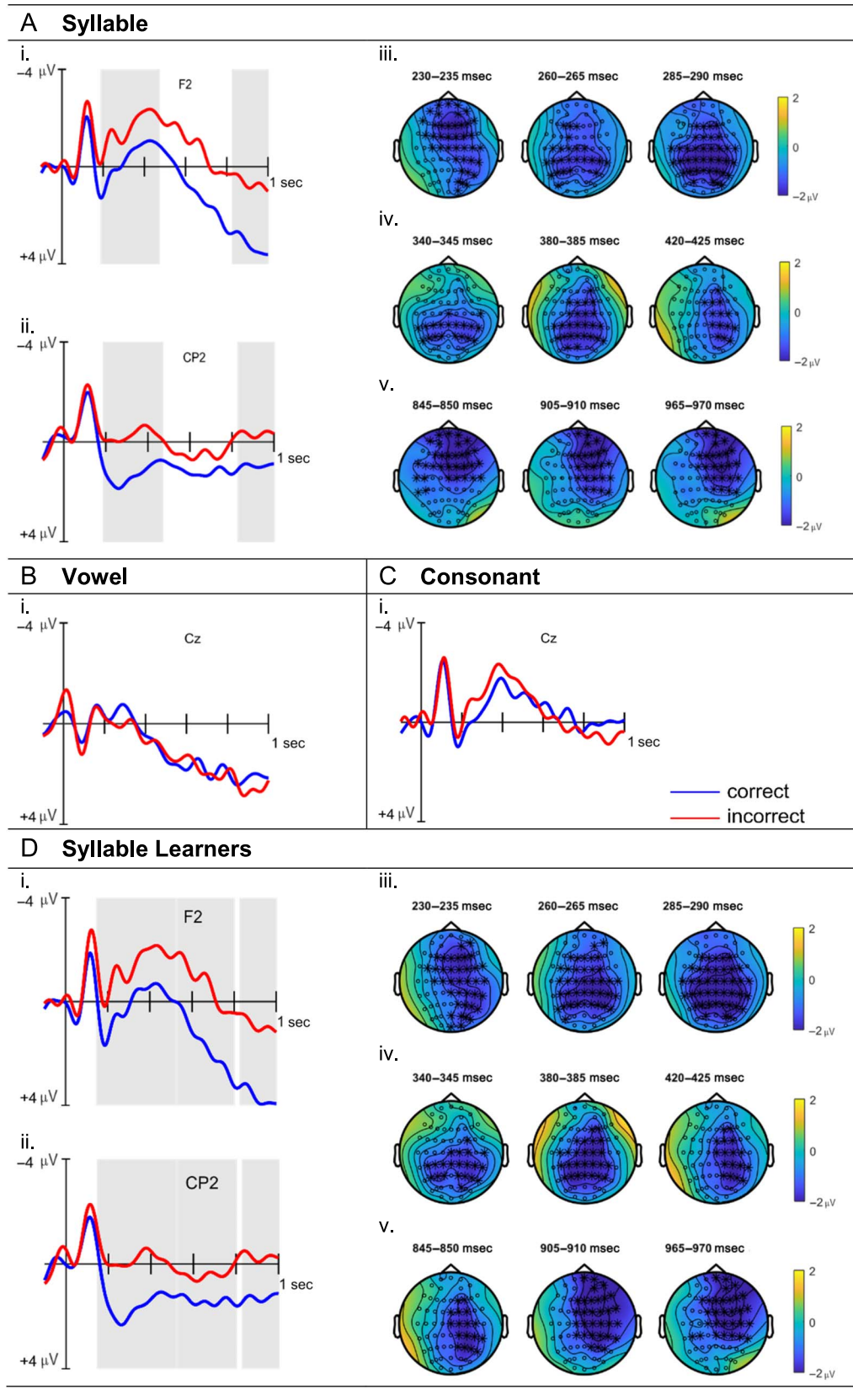
Figure 4. ERP waveforms[10] of data averaged across all participants for the syllable (A.i–A.ii), vowel (B.i), and consonant (C.i) conditions at representative electrodes, and waveforms of data averaged across syllable learners ($n = 25$) only (D.i–D.ii). Significant differences between correct and incorrect conditions are shaded in gray. Topographical plots of the syllable condition (A.iii–A.v/D.iii–D.v) show 5-msec averaged time windows that best illustrate distribution and time course of the respective effects; significant electrodes contributing to the effects are marked with asterisks.

context effects in the lexical selection process (Van Den Brink et al., 2001).

More recently and in the context of AGL tasks, the N200 has been established more generally as an attention-dependent marker of novelty detection or sequence matching (for a review, see Folstein & Van Petten, 2008). Mueller et al. (2012), for instance, found an N200 effect in response to incorrect nonadjacent syllable combinations in their previously described oddball experiment, which employed stimulus material very similar to the present input. Thus, it is likely that, in our experiment, the N200 reflects two aspects: First, because the effect is attention dependent, it shows participants' attentional focus on the stimulus material and specifically the relevant final unit in the syllable condition; and second, it suggests auditory discrimination processes that identify the final syllable of the incorrect AXB sequences as a mismatch with the previously acquired phonemic template of the NAD. No such effect was visible for the vowel and consonant conditions, which is likely because here both correct and incorrect test items included an unexpected phonemic mismatch with the learning phase items (see Table 1).

The second negative effect, which we identified as an N400-like component, is also in line with previous research. Although the N400 component is typically associated with lexical and semantic processing (e.g., Lau, Phillips, & Poeppel, 2008), similar N400-like effects have also been reported in a number of AGL studies (Citron et al., 2011; Mueller et al., 2009; Mueller, Girgsdies, & Friederici, 2008; de Diego-Balaguer et al., 2007; Cunillera, Toro, Sebastián-Gallés, & Rodríguez-Fornells, 2006; Sanders, Newport, & Neville, 2002). These studies have shown that the effect does not depend on the availability of semantic meaning but is also sensitive to presemantic levels of processing, specifically lexical access (i.e., the identification of familiar word forms). In word segmentation tasks, for example, nonwords elicited an N400-like effect, which is explained by lexical search processes failing to match them with previously established lexical items (de Diego-Balaguer et al., 2007; Cunillera et al., 2006; Sanders et al., 2002). A set of NAD learning studies (Citron et al., 2011; Mueller et al., 2009) in which German native speakers successfully learned a morphosyntactic dependency from mere exposure to Italian sentences also reported such a lexically interpreted N400-like component in response to violations of this NAD. On the basis of this evidence, we assume that our participants built a representation of the AXB syllable dependency. When the final syllable failed to match it, difficulties with lexical access resulted in the N400-like response. This interpretation also aligns with recent accounts that more generally assume surprisal (e.g., Kuperberg, 2016; Frank, Otten, Galli, & Vigliocco, 2015) or prediction error (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2019; Rabovsky, Hansen, & McClelland, 2018) as the basis for the N400. Under this view, participants in our experiment experienced surprisal, that is, a

low match between their probabilistic prediction and the bottom–up input, upon encountering the final syllable of an incorrect syllable item, inciting them to update their internal model.

Interestingly, however, we did not find a late positivity, contrary to what has been reported in combination with the N400-like response in some of the cited studies (Citron et al., 2011; Mueller et al., 2009; Mueller, Bahlmann, et al., 2008; de Diego-Balaguer et al., 2007). Mueller, Bahlmann et al. (2008) found such a negativity–positivity complex in response to deviant items (e.g., *tile* puwo moku) after exposure to a segmented, rule-based stream of bisyllabic nonwords encoding an item-specific AXB dependency (e.g., *tile* puwo *semi*). The authors interpreted the positivity as indicating controlled structural processes, akin to a P600 effect, which has been established as a marker of sentence-level syntactic (and semantic) integration difficulty (Friederici, 2011). Recently, it has become subject of considerable debate whether the P600 is actually specific to language or constitutes a more general marker of incongruency detection for complex structured sequences (e.g., Christiansen, Conway, & Onnis, 2012; Coulson, King, & Kutas, 1998) similar or even identical to the P300 (e.g., Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014; Bornkessel-Schlesewsky et al., 2011). Although this discussion is outside the scope of this article, an aspect that is relevant to the present investigation is the notion that the process underlying the P600 is one of (predictive) structured sequence processing, be it domain general or domain specific (cf. Christiansen et al., 2012, for a similar argument). If one assumed that item-specific NAD learning in general was a somewhat structural processing task, one might have expected a P600-like positivity also in the context of this study. There are several reasons this might not have been the case: First, our stimulus material might simply not have induced such structural processing operations. Because Mueller et al.'s (2008; and Citron et al.'s [2011]) stimulus material consisted of larger units encoding the dependency, it is conceivable that their mere size was decisive in triggering rather structural (syntactic-like) processing strategies, whereas our smaller units more closely resembled words and warranted lexically based operations.

Second, and apart from functional differences in processing dependent on the type of input, the experimental design might have played a role. Citron et al. (2011) reported the positivity only for a design with a single prolonged learning phase, but not for a design with alternating learning and test phases (here, they only report the N400-like effect), as we used in Experiment 1. De Diego-Balaguer et al. (2007) further found the negativity–positivity complex only when deviant items were inserted into an oddball-like stream, but not when presented in isolation (here, only an N400 modulation was reported). These differences in results between research designs might be related to the finding that both the P600 (e.g., Hahne & Friederici, 1999) and the P300 (e.g.,

Duncan-Johnson & Donchin, 1982; Duncan-Johnson & Donchin, 1977) are sensitive to the conditional probability of occurrence of a target (or violation). As such, these positivities might appear primarily in paradigms that specifically manipulate the frequency of occurrence of target items and render them highly unexpected.

Furthermore, second-language learners initially show N400 effects in response to violations of morphosyntactic rules at early stages of learning, which may later develop into a more native-like N400/P600 complex (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Mueller et al., 2009; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006). From the available evidence, it is impossible to determine, however, whether this difference in effects actually suggests functional differences in how item-specific NADs of different sizes are processed, whether they are attributable mainly to the specific experimental designs used, or whether they reflect learners' relative "proficiency."

The late negative effect in the syllable condition is less tangible than the two previously classified effects, at least from what is typically seen in AGL designs. Because of the effect's late appearance, relatively long duration, and fronto-central topography, two possible candidates come to mind. First, the reorienting negativity, which reflects the process of reorienting one's attention from an unexpected or unpredicted distractor back toward task-relevant information, including its retrieval from working memory (Bendixen, Schröger, Ritter, & Winkler, 2012; Escera & Corral, 2007; Munka & Berti, 2006; Berti, Roeber, & Schröger, 2004; Escera, Alho, Schröger, & Winkler, 2000; Schröger & Wolff, 1998). Typically, however, the reorienting negativity has only been reported when the deviant is employed as a behavioral distractor that is to be ignored (e.g., tonal changes in a visual task) and not when the distractor is part of the target stimulus set. In the present experimental design, however, it is conceivable that the incorrect syllable items acted as distractors. After extended exposure to correct AXB sequences (~3 min in each learning phase), an incorrect syllable item, although task relevant, was highly unexpected and required participants to reallocate attention to the maintenance of the correct representation.

An alternative but related explanation comes from the working memory literature, where not only the retention of tones (Lefebvre et al., 2013) or simple acoustic features such as pitch (Guimond et al., 2011) or timbre (Nolden et al., 2013) but also the retention of verbal information (Ruchkin et al., 1997; Lang, Starr, Lang, Lindinger, & Deecke, 1992) in auditory working memory have elicited sustained anterior negative waves during the retention period. These fronto-central slow waves have typically been interpreted as reflecting active maintenance of stimulus information in working memory, a process that includes both the sustained (re)activation of stimulus representations and their rehearsal (for a review, see Kaiser, 2015). It is possible that, in the present experiment,

encountering an unexpected stimulus in the test phase initiated such maintenance and rehearsal processes to avoid "contamination" of the established correct target representation.

The fact that we did not find any ERP effects across the whole group to suggest differentiation of correct and incorrect vowel or consonant items, respectively, indicates that the given input did not induce a learning process that is intuitively or preferentially based on either of the two segments. The small number of participants who successfully distinguished items in these conditions did not provide an appropriate signal-to-noise ratio to obtain any significant ERP effects, rendering such a comparison uninformative. These results suggest that, with all three segments available, the syllable was the most relevant unit for NAD learning. The sequential learning pattern for syllables and vowels and the absence of ERP effects in the vowel condition further suggest that participants likely built a syllable-based representation, which only a subgroup of learners was able to flexibly access and apply also to the segment-based test items.

EXPERIMENT 2

In Experiment 2, we employed a between-participant design to test whether adults are capable of learning NADs under conditions where only one type of segmental information, namely, syllables, vowels, or consonants, respectively, is available as a learning cue. To prevent the previously explained possibly confounding effects of strategic learning from test items, we decided to abolish the learning phase/test phase design in favor of an oddball paradigm. The oddball task was followed by a forced-choice grammaticality judgment task (GJT) akin to a test phase in the previously used design. We additionally administered a small debriefing questionnaire after the experiment, asking participants if they had identified a regularity in the input and, if so, if they could spell it out. Most importantly, they were asked to indicate when during the experiment, approximately, they thought they had identified the regularity, with "test phase" being one of the possible answers. Such retrospective evaluations of the learning process have been shown to yield valuable information not only on what was learned but also to which degree the learned representation was consciously accessible (Rebuschat, 2013).

Methods

Participants

The same circumstances and requirements as in Experiment 1 were applied. On the basis of similar previous studies (e.g., Monte-Ordoño & Toro, 2017a; de Diego-Balaguer et al., 2007), we aimed for a minimum of 20 participants entering the final analysis per group. Note that the smaller number of participants compared with

Experiment 1 was justified given that, in the between-participant oddball design, participants were exposed to an overall larger number of deviant items. Seventy-seven adults participated in Experiment 2, 10 of whom had to be excluded because of technical difficulties or high artifact rate in the EEG recording. The remaining 67 participants (15 men, 52 women) were between 18 and 33 years old (M = 21.8 years, SD = 3.10 years). Participants were randomly assigned to one of three experimental groups (SYL: n = 23, CON: n = 22, VOW: n = 22), which did not significantly differ with regard to age or sex.

Stimuli

For the second experiment, a number of additional syllables were used, which had already been recorded for Experiment 1. Four hundred forty-eight stimuli were used for each experimental group, with a distribution of 384 standard trisyllabic sequences interspersed with 64 deviants (~14%). In the syllable group, standard and deviant items were identical to the correct and incorrect syllable condition items from Experiment 1, except for a different set of 32 middle syllables (/d, l, m, n, r, s, t, w/ each combined with /a, $ä$, $ö$, $ü$/). In the vowel group, standard sequences were defined as the fixed vowel combinations xi X xe and xo X xu, whereas all other slots in the CVCVCV structure were filled equally often with the consonants /b, g, k, l, m, r, s, w/ (and the vowels /a, $ä$, $ö$, $ü$/ in the middle syllable) in a nonrepetitive manner. The consonant group items were built correspondingly, filling the open positions with /d, l, m, s/ and /a, e, i, o, u, $ä$, $ö$, $ü$/ (see Figure 5).

Procedure

Four different item lists were created per group, with the sequence of items pseudoranzomized according to the number of constraints. The first 16 items of each list consisted of standards for familiarization. Each deviant item was preceded by a minimum of four and a maximum of eight standards (see Figure 5). To focus participants' attention on the input, they were instructed to attentively listen to the input, determine a regularity inherent in it, and perform a target detection task

(Mueller et al., 2012). Throughout the continuous stimulus presentation, they would press a button whenever they detected an item deviating from the regularity. The oddball task was followed by a forced-choice GJT comprising 64 items (half standards and half deviants). Test items were presented individually and, after a 900-msec delay, had to be evaluated as adhering to or deviating from the regularity previously identified. Stimulus presentation and all other external conditions were the same as in Experiment 1. After the experiment, participants received the previously mentioned debriefing questionnaire, asking whether they thought they had learned a regularity ("Yes," "No," "Not sure") and, if "Yes" or "Not sure," when they had learned it ("Beginning," "Middle," "End," "Test Phase") and if they could spell it out.

Data Acquisition and Preprocessing

The same software, parameters, and preprocessing steps as in Experiment 1 were used, except that, for ERP averaging, epochs from −100 to 800 msec after onset of the final syllable (or vowel in case of the vowel condition) were cut out from the oddball task. Data epochs were cut shorter than in Experiment 1 to avoid contamination of the signal by the button press during exposure. Standard items immediately after a deviant were excluded from analysis to avoid effects of refamiliarization. The average number of epochs per participant entering the final analysis in each group was as follows: 311.22 (SD = 16.14) standards and 62.04 (SD = 3.57) deviants in the syllable group, 310.95 (SD = 7.56) standards and 61.77 (SD = 1.85) deviants in the consonant group, and 308.09 (SD = 9.40) standards and 60.59 (SD = 2.65) deviants in the vowel group. Differences in the number of standard/deviant items between groups were not significant.

Data Analysis

Similar to Experiment 1, accuracy scores were calculated for each participant for the GJT items. For the responses during the oddball phase, d' was additionally calculated as an index of sensitivity. To identify learners and non-learners, responses in the target detection task (d' > 1) and/or accuracy in the GJT (response accuracy ≥ 64%,

SYLLABLES	**go** X **ku**	**bi** X **pe**	goläku	bidape	gotöku	goläpe	bimüpe	gosöku	bidaku
CONSONANTS	**g**x X **k**x	**b**x X **p**x	gelokä	boläpi	gusakü	gidopu	bodupe	gäsekö	bömeki
VOWELS	x**o** X x**u**	x**i** X x**e**	sigäle	miläwe	lorasu	rowäke	sibawe	monöru	loküge
			S	S	S	D	S	S	D

Figure 5. Exemplary series of standard (S) and deviant (D) stimuli by group.

indicated by a binomial test for chance response, $p = .033$) were taken into account. Because we were mainly interested in whether participants had learned the respective dependency they had been exposed to, visible in above-chance performance in the GJT, we fitted generalized linear mixed-effects intercept models including a binomial link function to the accuracy data of each group, with participants and items as random effects. Furthermore, 1472 (syllable condition) and 1408 (consonant and vowel condition) data points were entered into these models, respectively. p Values for fixed effects were calculated via Wald tests (standard for *glmer* in lme4). With regard to the EEG data, separate cluster-based permutation tests with the same parameters stated above were run for each group comparing responses to standard and deviant items in the oddball task.

Table 3. Results of the GLMMs Separately Investigating Divergence from Chance-Level Response in Each Group

	FE	SE	z Value	p
		Syllable		
Intercept	**1.037**	**.368**	**2.814**	**<.005**
		Vowel		
Intercept	0.254	.148	1.715	.086
		Consonant		
Intercept	0.167	.117	1.432	.152

Bold print indicates significant effects ($p < .05$). FE = fixed effect estimates; *SE* = standard error.

Results

Behavioral Results

Average response accuracy in the GJT was at 63.6% ($SD = 21.3\%$) in the syllable group, at 54.9% ($SD = 10.8\%$) in the vowel group, and at 53.6% ($SD = 10.7\%$) in the consonant group. The boxplots in Figure 6 provide further descriptive evidence for largely chance-level responses in both the vowel and consonant groups: Both medians are close to 50% accuracy, and range and dispersion of the data are low. Only one outlier per condition is visible, each with an accuracy rate close to ceiling. Whereas the lower quartile range of the syllable condition overlaps with the other two boxes and the median is similar (56.3%), range and dispersion of the accuracy rates are greater in this
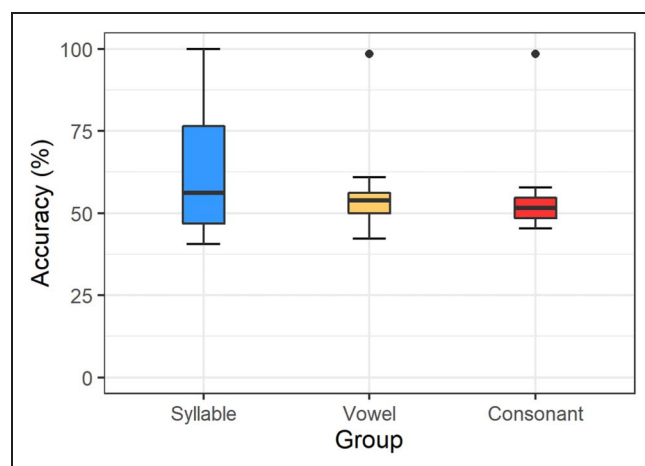
group and the data are visibly skewed toward the upper percentages.

The fitted intercept models revealed that the estimated intercept was significantly different from zero only for the syllable group, whereas the other two models did not provide evidence for above-chance performance in the vowel and consonant groups (see Table 3). We identified seven learners in the syllable group (detection rate: 48.2%, $SD = 32.9\%$; grammaticality judgment: 92.4%, $SD = 12.6\%$), one in the vowel group (detection rate: 85.9%; grammaticality judgment: 98.4%), and one in the consonant group (detection rate: 21.9%; grammaticality judgment: 98.4%). The debriefing questionnaire revealed that five of the seven syllable group learners were able to correctly spell out the syllable dependency, as were the individual learners in the vowel and consonant groups, respectively. The syllable learners and the vowel learner further believed to have learned the regularity "at the beginning" or "in the middle" of the experiment, whereas the consonant learner perceived detection to have happened only "at the end" of the experiment.

ERP Results

The comparison between the evoked responses of standard and deviant items in the syllable group indicated a significant difference ($p < .05$). The difference was manifested in a fronto-centrally distributed positivity, based on a cluster between approximately 335 and 440 msec (Figure 7A, i–ii).[10] The cluster-based test yielded no significant differences in the vowel group (Figure 7B, i) but exposed a significant difference between conditions in the consonant group ($p < .05$), corresponding to a fronto-central to centro-parietal, slightly left-lateralized positive cluster between around 200 and 290 msec (Figure 7C, i–ii).
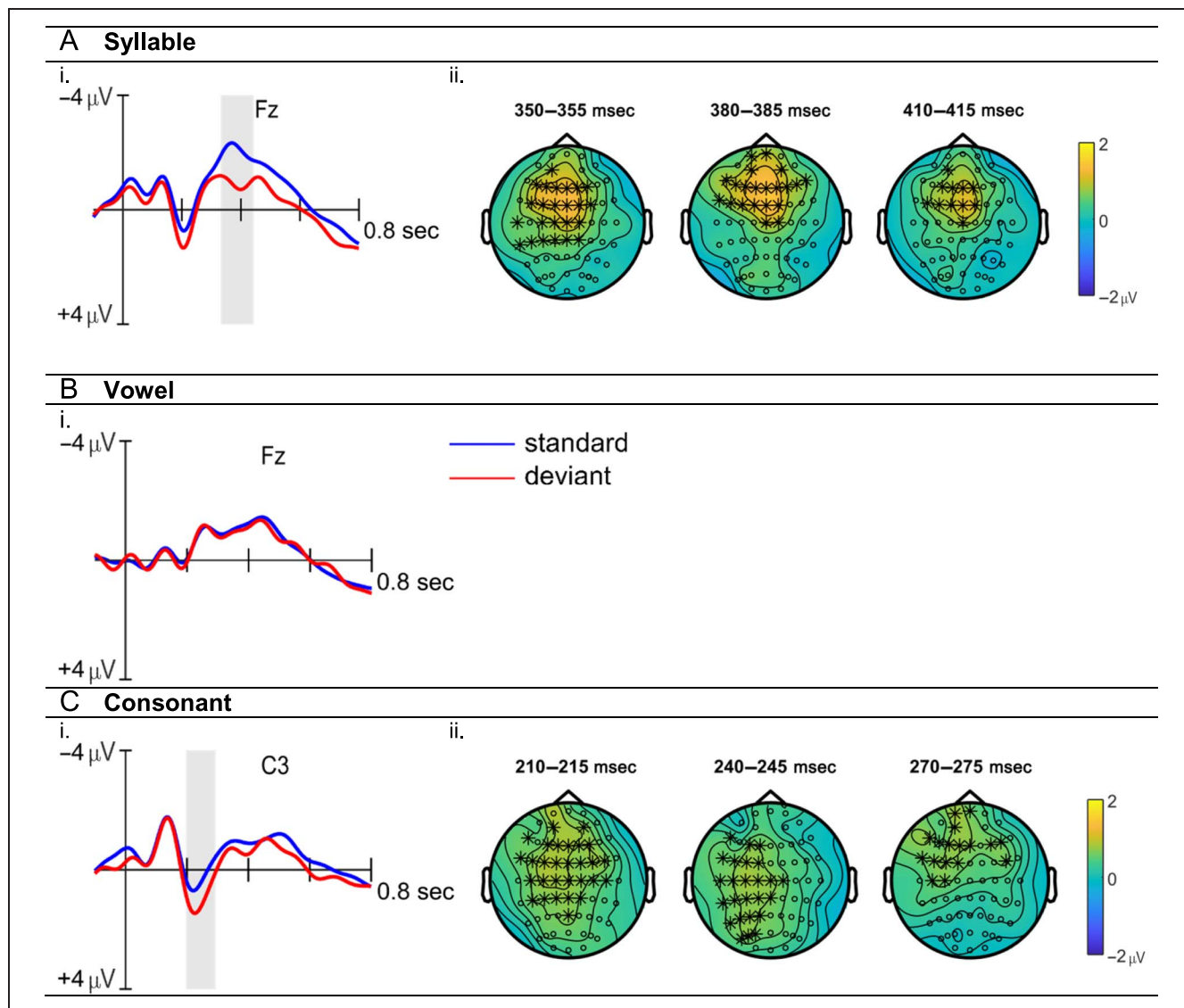

Figure 6. Average response accuracy separated by groups in Experiment 2. The length of each box represents the interquartile range (IQR), limited by the upper and lower quartiles; horizontal lines mark the median; whiskers illustrate data outside the IQR (maximum: 1.5 × IQR); and outlier participants are indicated by dots (any data points outside 1.5 × IQR).

Figure 7. ERP waveforms[11] of data averaged across all participants of the syllable (A.i), vowel (B.i), and consonant (C.i) groups at the representative electrodes Fz and C3. Significant differences between standard and deviant condition are shaded in gray. Topographical plots of the syllable (A.ii) and consonant (C.ii) groups show 5-msec averaged time windows that best illustrate distribution and time course of the respective effects; significant electrodes contributing to the effects are marked with asterisks.

DISCUSSION

Experiment 2 aimed at assessing adult listeners' ability to detect and evaluate NADs between specific syllables, vowels, or consonants online. In both behavioral and neurophysiological results, the syllable again clearly emerged as the unit from which adults discerned the dependency most successfully.[11] We thereby closely replicated Mueller et al.'s (2012) findings and extended them by showing that explicit instructions specifying which units are relevant to the regularity (the authors hinted at the order of the syllables) are not necessary for successful learning. The corresponding EEG data further highlighted the role of the syllable unit. We interpreted the encountered positivity in response to deviant syllable items as a P3 effect, which is typical in oddball designs and indicates

discrimination of the infrequently occurring deviant (target) from the frequently occurring standard (for a review, see Polich, 2007). It mirrored the effect Mueller et al. (2012) found and, similarly, the effect reported by Monte-Ordoño and Toro (2017b) in an oddball paradigm in response to stimuli that violated a syllable-based ABB repetition rule.

The P3 is often classified as P3a or P3b depending on its distribution. Whereas the former shows a frontal distribution and is suggestive of a stimulus-driven (unconscious) attention switch to previously unattended material or stimulus features (Escera & Corral, 2007; Debener, Kranczioch, Herrmann, & Engel, 2002; Escera, Alho, Winkler, & Näätänen, 1998), the latter is centro-parietally distributed and indexes task-relevant conscious attention to stimuli and subsequent working memory updating

processes (Ferdinand, Mecklinger, & Kray, 2008; Sergent, Baillet, & Dehaene, 2005; Wetter, Polich, & Murphy, 2004). The fact that we found a fronto-central distribution of the P3 suggests a mixture of P3a and P3b probably attributable to averaging across the entire participant group, of which only seven people consciously discriminated the items, resulting only in a slight shift of the effect toward central electrode sites. Nevertheless, the effect was significant across the entire syllable group, suggesting that even those who did not demonstrate behavioral evidence of learning possibly passively recognized the dependency violation.

When it comes to the smaller segmental level, we found no evidence, either behavioral or neurophysiological, that vowels are accessible for item-based NAD learning. Only a single participant performed well at test, and there were no significant ERP effects across the whole group that would suggest any kind of differentiation between standards and deviants. Because the experimental design prevented strategic learning from test items, these results support our suspicion that the behavioral learning effects seen in the vowel condition in Experiment 1 were largely provoked by the experimental design.

In the consonant group, we also only found a single participant who exhibited behavioral evidence of learning. Nonetheless, the ERPs computed for the entire group showed a significant P2 effect in response to deviants compared to standards—a component that has been found to signal selective attention to acoustic stimulus change and feature detection in a variety of auditory tasks (see e.g., Paulmann, Bleichner, & Kotz, 2013; Cunillera et al., 2006; Yingling & Nethercut, 1983). Cunillera et al. (2006), for example, encountered a P2 effect during segmentation of stressed words from a continuous stream when compared to the same unstressed word stream. Although this is evidence for an emergent dissociation, the lack of a concurrent behavioral response or P3(b) effect similar to the syllable group suggests that whichever consonant-based representation participants might have built, it was not accessible to be translated into an offline response. One might assume a less sophisticated NAD representation than in the syllable group, for instance, one that rested merely on phonological memory of the perceptual similarity of the paired consonants.

In any case, the finding confirmed the tentatively formulated expectation that the use of item-based dependencies between specific units might result in a slight advantage for consonants over vowels in NAD learning because of their previously shown prevalence in word identification and lexical selection (Delle Luche et al., 2014; Havy et al., 2014; Carreiras, Dunabeitia, et al., 2009; Carreiras, Gillon-Dowens, et al., 2009; New et al., 2008; Cutler et al., 2000), where item specificity is of importance. In other words, the fact that the dependency relationship between consonants was more salient to adults than that between vowels is suggestive of lexical processes having been at play during learning.

GENERAL DISCUSSION

The goal of this study was to compare syllables, consonants, and vowels as carriers of item-specific NADs and to identify possible differences between the three segments with regard to online processing and offline learning success for NADs. We used an artificial grammar consisting of trisyllabic sequences from which a dependency relationship between specific units had to be learned. In Experiment 1, we reported evidence that adults preferentially learn the syllable-based dependency when all three segments are available for NAD learning and likely build a syllable-based representation of the input. We found no evidence that would suggest that item-specific NAD learning was biased or guided by the smaller segmental level. From Experiment 1 alone, it remained unclear whether it is the syllable per se that receives a special role in NAD learning or whether adults simply focused on the largest informative unit. We further investigated this in Experiment 2, where three separate groups of participants were exposed to material containing either a syllable-, consonant-, or vowel-based dependency that could be learned. Again, the syllable clearly emerged as the unit from which the dependency was learned most successfully. When it was not available and the only units informative to the nonadjacency relationship were consonants or vowels, respectively, there was little evidence for successful learning. We therefore conclude that it was not relative informativeness that triggered syllable-based NAD learning in Experiment 1 but rather a more general (attentional) preference for the syllable unit over the two smaller segments.

A caveat to be considered in this regard is that, in both of our experiments, the syllables within the trisyllabic AXB units were separated by 50-msec pauses, possibly resulting in a perceptual bias toward the syllable unit. Recent oscillation-based approaches to natural speech processing have shown, however, that neuronal oscillations automatically track the dynamics of continuous speech input at the syllable rate even without such artificially inserted segmentation markers (Poeppel & Assaneo, 2020; Giraud & Poeppel, 2012). Although these oscillator models cannot explain how isolated subword units such as the ones used here are decoded, they do highlight a general possible perceptual advantage for syllables in competent language users. Whether this advantage is based on a signal-driven, bottom–up mechanism or whether it constitutes a learned, top–down process is subject of considerable debate and is beyond the scope of this article (for a discussion, see, e.g., Räsänen et al., 2018; Giraud & Poeppel, 2012).

With regard to vowels as carriers of NADs in Experiment 2, we found no evidence for a "perceptual similarity effect" as hypothesized initially: It seems that nonrepeated but phonetically close vowels do not suffice to induce the learning advantage vowels have tentatively demonstrated over consonants in the context of studies

using repetition rules (Monte-Ordoño & Toro, 2017a; Toro, Nespor, et al., 2008; Toro, Shukla, et al., 2008). As such, vowels do not seem to be particularly accessible for the identification and memorization of item-specific dependency relations when they are the only informative segment. What is intriguing, however, is that at least a small group of participants in Experiment 1 was able to correctly evaluate the vowel test items by identifying them as subcomponents of the syllable dependency. That is, once a syllable-based representation of the NAD has been built, it seems that a matching vowel-based dependency is recognized more easily than a matching consonant-based dependency. The question remains whether this is because of mere perceptual differences in the two segments (i.e., relative saliency of vowels) or because of a top–down mechanism that identifies such "word"-internal regularities and preferentially operates across vowels. One might speculate that if working memory and specifically articulatory rehearsal processes play a role for such explicit, strategic consideration of the stimulus material, it might be easier to rehearse and connect individual vowels than consonants. Single vowels can be full syllables in some cases (e.g., in a word like "a" or an exclamation like "uh") and thus serve as independent production units, whereas consonants typically occur in conjunction with vowels (e.g., even during the production of individual consonants in spelling out loud).

When it comes to consonants, the ERP evidence for deviant processing (at least at an acoustic level) found in Experiment 2 further substantiates the superior role of consonants as "identifiers," as formulated, for instance, in the CV hypothesis (Nespor et al., 2003). Together with the lexically related ERP effects found for syllables in Experiment 1, these results additionally show that the processing of item-specific NADs seems to prompt lexical rather than (morphosyntactic-like) structure-related processes. The question remains whether the given ERP patterns were because of the "low proficiency" attained by the learners throughout the experiment, that is, similar to what Morgan-Short et al. (2010) and Osterhout et al. (2006) find for second-language learners at early stages of morphosyntactic learning, or whether they are attributable to the nature of the stimulus material. A potential methodological issue to be mentioned in this context is the fact that consonants always preceded vowels in our stimulus material. In combination with the mentioned pauses between units, this might have provided consonants with a positional advantage and increased their relative salience. To exclude this possibility, future investigations could directly compare CVXCV and VCXVC structures.

Future research could further investigate which features of the syllable specifically facilitate the extraction of dependencies in speech. Are specific acoustic properties of the syllable responsible, for example, its duration (providing sufficient time for working memory processes) or spectro-temporal complexity (providing a unique and rich

representation for memory), or is it its association with articulatory gestures (e.g., in the mental syllabary) that supports inner rehearsal processes during NAD learning?

Overall, it seems that, at the smaller segmental level, vowels are susceptible to conscious, strategic manipulations in working memory, whereas the processing of relationships between consonants instead induces implicit learning processes. Thus, we can conclude that, on the basis of the present evidence, syllables, consonants, and vowels clearly do not constitute equally suitable computational units for NAD learning. Indeed, we were able to show that syllables are by far the most accessible unit for the learning of such relatively local, item-specific dependencies.

Reprint requests should be sent to Ivonne Weyers, Department of Linguistics, University of Vienna, Sensengasse 3A, Vienna 1090, Austria, or via e-mail: ivonne.weyers@univie.ac.at.

Funding Information

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Notes

1. In all of the cited studies, the authors used only eight (four: Toro, Shukla, et al. [2008]) pairs of test items in the generalization test, which makes their average correct response rates tentative indicators of learning at best. The authors reported average response accuracies of 66.6% (SD = 11.2; Toro, Nespor, et al., 2008), 61.6% (SD = 12.9; Toro, Shukla, et al., 2008), and 61.18% (SD = 22.78; Monte-Ordoño & Toro, 2017a) for NAD generalizations across vowels, that is, as low as five hits out of eight trials. A simple binomial test reveals that the cumulative probability of $P(X \geq 5)$ = .363, suggesting there is a 36% probability that the average result was obtained by chance.

2. To test for learning, previous behavioral studies have typically used a two-alternative forced-choice task (e.g., Peña et al., 2002; Saffran, Newport, & Aslin, 1996). This task, in which two

items are presented to the participant sequentially, is problematic when simultaneously using EEG, however, because it additionally taxes working memory. Therefore, we chose a standard ERP violation paradigm, in which items are presented and evaluated individually.

3. Note that other similar EEG studies have sometimes used novel phonemes in their violation items (e.g., Monte-Ordoño & Toro, 2017a, 2017b) and subsequently compared ERP responses to phonemically very different sets of correct (e.g., fefuku) and incorrect (e.g., mimomi) exemplars. Because this can lead to confounding phoneme change effects, we ensured our stimulus design would allow comparing responses to the same phonemes in the relevant positions (e.g., pi and ko in the consonant condition; see Table 1).

4. Even an exactly equal distribution of items per segmental condition would likely result in a perceived imbalance of correct and incorrect items by the participant, for example, assuming they learn the vowel dependency and correctly evaluate the vowel and syllable items but reject all consonant condition items.

5. ICA was run on 1-Hz high-pass filtered (-6 dB, 1 Hz, 930) data to improve decomposition performance (Winkler, Debener, Muller, & Tangermann, 2015). The resulting ICA weights were then applied to the 0.1-Hz filtered data sets.

6. Theoretically even "vowel + consonant + syllable" learners, although this could arguably be evidence for a syllable-based representation and subsequent access to its segmental subcomponents rather than segment-specific functionalities.

7. Note that it is in the nature of the cluster-based permutation method that the reported cluster T-statistic does not allow any inferences with regard to the specific temporal or spatial distribution of the identified cluster(s), since there is no error rate control for the inclusion of individual sample statistics in a specific cluster. The spatiotemporal characteristics of the effects established by the test are, however, likely to be highly correlated with those of the true effect (Sassenhagen & Draschkow, 2019; Groppe et al., 2011; Maris & Oostenveld, 2007).

8. An additional 10-Hz low-pass filter (-6 dB, 10, 620) was applied to the averaged data exclusively for plotting to improve visibility.

9. In this context, it should be mentioned that a few participants actually performed below chance level in one of the two conditions (vowel or consonant; see Figure 1) but show high response accuracy in the respective other condition. Apparently, these participants ($n = 3$ for vowels, $n = 2$ for consonants) established (perceptual) classes of elements (Wilson et al., 2018), evaluating any combination of o/u and i/e (or b/p, g/k) as correct, regardless of the vowels' (or consonants') specific positions in the items. Therefore, they evaluated the incorrect consonant items bu X ko/ge X pi (vowel items pi X bu/ko X ge) as correct because the vowel pairings were intact but rejected their correct counterparts bu X pi/ge X ko (pi X ge/ko X bu for vowels), because these violated the pairings they had learned.

10. An additional 10-Hz low-pass filter (-6 dB, 10, 620) was applied to the averaged data exclusively for plotting of the ERP graphs to improve readability.

11. On the basis of the information provided in the debriefing questionnaire after Experiment 2, it seems that at least some participants (SYL: $n = 5$, CON: $n = 6$, VOW: $n = 2$) were distracted by the use of umlaut in the middle positions irrelevant for the NAD and pressed a button whenever no umlaut appeared in a stimulus. This is in line with the finding that, initially, adults primarily identify the distributional properties of an input and only extract the conditional regularities potentially at odds with them after extended exposure (Endress & Bonatti,

2016). Even so, the average detection rates of the remaining participants (consciously) undisturbed by the umlaut did not exceed chance in either the vowel or consonant group.

REFERENCES

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *Journal of Neuroscience*, *29*, 7686–7693. https://doi.org/10.1523/JNEUROSCI.5242-08.2009, PubMed: 19535580

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *ArXiv*. https://doi.org/10.48550/arXiv.1506.04967

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Batterink, L. (2020). Syllables in sync form a link: Neural phase-locking reflects word knowledge during language learning. *Journal of Cognitive Neuroscience*, *32*, 1735–1748. https://doi.org/10.1162/jocn_a_01581, PubMed: 32427066

Bendixen, A., Schröger, E., Ritter, W., & Winkler, I. (2012). Regularity extraction from non-adjacent sounds. *Frontiers in Psychology*, *3*, 143. https://doi.org/10.3389/fpsyg.2012.00143, PubMed: 22661959

Berti, S., Roeber, U., & Schröger, E. (2004). Bottom–up influences on working memory: Behavioral and electrophysiological distraction varies with distractor strength. *Experimental Psychology*, *51*, 249–257. https://doi.org/10.1027/1618-3169.51.4.249

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247–260. https://doi.org/10.1016/S0163-6383(81)80027-6

Boatman, D., Hall, C., Goldstein, M. H., Lesser, R., & Gordon, B. (1997). Neuroperceptual differences in consonant and vowel discrimination: As revealed by direct cortical electrical interference. *Cortex*, *33*, 83–98. https://doi.org/10.1016/S0010-9452(97)80006-8

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science*, *16*, 451–459. https://doi.org/10.1111/j.0956-7976.2005.01556.x, PubMed: 15943671

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., et al. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*, 133–152. https://doi.org/10.1016/j.bandl.2010.09.010, PubMed: 20970843

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*, 298. https://doi.org/10.3389/fpsyg.2019.00298, PubMed: 30846950

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*, 428–430. https://doi.org/10.1038/35000206, PubMed: 10667794

Carreiras, M., Dunabeitia, J. A., & Molinaro, N. (2009a). Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming. *Cerebral Cortex*, *19*, 2659–2670. https://doi.org/10.1093/cercor/bhp019, PubMed: 19273459

Carreiras, M., Gillon-Dowens, M., Vergara, M., & Perea, M. (2009b). Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter

paradigm. *Journal of Cognitive Neuroscience*, *21*, 275–288. https://doi.org/10.1162/jocn.2008.21023, PubMed: 18510451

Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, *90*, 393–400. https://doi.org/10.1016/j.bandl.2003.12.003, PubMed: 15172555

Carreiras, M., & Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex*, *18*, 1727–1735. https://doi.org/10.1093/cercor/bhm202, PubMed: 18234690

Carreiras, M., Vergara, M., & Perea, M. (2007). ERP correlates of transposed-letter similarity effects: Are consonants processed differently from vowels? *Neuroscience Letters*, *419*, 219–224. https://doi.org/10.1016/j.neulet.2007.04.053, PubMed: 17507160

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*, 1161–1173. https://doi.org/10.1177/0956797620933237, PubMed: 32865487

Cholin, J. (2008). The mental syllabary in speech production: An integration of different approaches and domains. *Aphasiology*, *22*, 1127–1141. https://doi.org/10.1080/02687030701820352

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 109–122. https://doi.org/10.1037/a0021322, PubMed: 21244111

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*, 205–235. https://doi.org/10.1016/j.cognition.2005.01.009, PubMed: 15939415

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *27*, 231–256. https://doi.org/10.1080/01690965.2011.606666, PubMed: 23678205

Citron, F. M. M., Oberecker, R., Friederici, A. D., & Mueller, J. L. (2011). Mass counts: ERP correlates of non-adjacent dependency learning under different exposure conditions. *Neuroscience Letters*, *487*, 282–286. https://doi.org/10.1016/j.neulet.2010.10.038, PubMed: 20971159

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256–266. https://doi.org/10.1162/jocn.1994.6.3.256, PubMed: 23964975

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *94*, 276–287. https://doi.org/10.1016/0013-4694(95)98479-R, PubMed: 7537200

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1–18. https://doi.org/10.1016/0093-934X, PubMed: 1643505

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*, 302–318. https://doi.org/10.1016/0093-934X(90)90016-A, PubMed: 2224497

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58. https://doi.org/10.1080/016909698386582

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119, PubMed: 15355140

Culbertson, J., Koulaguina, E., Gonzalez-Gomez, N., Legendre, G., & Nazzi, T. (2016). Developing knowledge of nonadjacent dependencies. *Developmental Psychology*, *52*, 2174–2183. https://doi.org/10.1037/dev0000246, PubMed: 27893252

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*, 168–178. https://doi.org/10.1016/j.brainres.2006.09.046, PubMed: 17064672

Cutler, A., Sebastian-Galles, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory and Cognition*, *28*, 746–755. https://doi.org/10.3758/BF03198409, PubMed: 10983448

Debener, S., Kranczioch, C., Herrmann, C. S., & Engel, A. K. (2002). Auditory novelty oddball allows reliable distinction of top–down and bottom–up processes of attention. *International Journal of Psychophysiology*, *46*, 77–84. https://doi.org/10.1016/S0167-8760(02)00072-7, PubMed: 12374648

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. https://doi.org/10.1037/0033-295X.93.3.283, PubMed: 3749399

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142. https://doi.org/10.1016/0749-596X(88)90070-8, PubMed: 3749399

Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazzi, T. (2014). Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1–15. https://doi.org/10.1016/j.jml.2013.12.001

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*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

de Diego-Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One*, *2*, e1175. https://doi.org/10.1371/journal.pone.0001175, PubMed: 18000546

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164. https://doi.org/10.1038/nn.4186, PubMed: 26642090

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. https://doi.org/10.3389/fnhum.2014.00311, PubMed: 24904354

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x, PubMed: 905483

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*, 1–52. https://doi.org/10.1016/0301-0511(82)90016-3

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*, 247–299. https://doi.org/10.1016/j.cognition.2006.09.010, PubMed: 17083927

Endress, A. D., & Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*, 19–35. https://doi.org/10.1002/wcs.1376, PubMed: 26683248

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, 13, 348–353. https://doi.org/10.1016/j.tics.2009.05.005, PubMed: 19647474

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, 5, 151–166. https://doi.org/10.1159/000013877, PubMed: 10859410

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, 10, 590–604. https://doi.org/10.1162/089892998562997, PubMed: 9802992

Escera, C., & Corral, M. J. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *Journal of Psychophysiology*, 21, 251–264. https://doi.org/10.1027/0269-8803.21.34.251, PubMed: 10859410

Ferdinand, N. K., Mecklinger, A., & Kray, J. (2008). Error and deviance processing in implicit and explicit sequence learning. *Journal of Cognitive Neuroscience*, 20, 629–642. https://doi.org/10.1162/jocn.2008.20046, PubMed: 18052785

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45, 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x, PubMed: 17850238

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006, PubMed: 25461915

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91, 1357–1392. https://doi.org/10.1152/physrev.00006.2011, PubMed: 22013214

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70–74. https://doi.org/10.1016/j.cognition.2015.11.010, PubMed: 26638049

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517. https://doi.org/10.1038/nn.3063, PubMed: 22426255

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431–436. https://doi.org/10.1111/1467-9280.00476, PubMed: 12219809

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183–206. https://doi.org/10.1207/s15327078in0702_4, PubMed: 33430549

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15, 3–10. https://doi.org/10.1016/j.tics.2010.10.001, PubMed: 21093350

Grama, I. C., Kerkhoff, A., & Wijnen, F. (2016). Gleaning structure from sound: The role of prosodic contrast in learning non-adjacent dependencies. *Journal of Psycholinguistic Research*, 45, 1427–1449. https://doi.org/10.1007/s10936-016-9412-8, PubMed: 26861215

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48, 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x, PubMed: 21895683

Guenther, F. H. (2016). Neural control of speech. In F. H. Guenther (Ed.), *Neural control of speech*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, 280–301.

https://doi.org/10.1016/j.bandl.2005.06.001, PubMed: 16040108

Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory. *Psychophysiology*, 48, 1500–1509. https://doi.org/10.1111/j.1469-8986.2011.01234.x, PubMed: 21824153

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, 38, 1518–1530. https://doi.org/10.1016/S0028-3932(00)00052-X

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, 194–205. https://doi.org/10.1162/089892999563328, PubMed: 10198134

Havy, M., Serres, J., & Nazzi, T. (2014). A consonant/vowel asymmetry in word-form processing: Evidence in childhood and in adulthood. *Language and Speech*, 57, 254–281. https://doi.org/10.1177/0023830913507693, PubMed: 25102609

Hooper, J. B. (1972). *The syllable in phonological theory*, *Language*, 48, 525. https://doi.org/10.2307/412031

Kaiser, J. (2015). Dynamics of auditory working memory. *Frontiers in Psychology*, 6, 613. https://doi.org/10.3389/fpsyg.2015.00613, PubMed: 26029146

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, 31, 602–616. https://doi.org/10.1080/23273798.2015.1130233, PubMed: 27570786

Lang, W., Starr, A., Lang, V., Lindinger, G., & Deecke, L. (1992). Cortical DC potential shifts accompanying auditory and visual short-term memory. *Electroencephalography and Clinical Neurophysiology*, 82, 285–295. https://doi.org/10.1016/0013-4694(92)90108-T

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933. https://doi.org/10.1038/nrn2532, PubMed: 19020511

Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., et al. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia*, 51, 2939–2952. https://doi.org/10.1016/j.neuropsychologia.2013.08.003, PubMed: 23938319

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press. https://doi.org/10.2307/1423219

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1–22. https://doi.org/10.1016/0010-0277(92)90038-J, PubMed: 1582153

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75. https://doi.org/10.1017/S0140525X99001776, PubMed: 11301520

Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. *Cognitive Science*, 42, 1677–1699. https://doi.org/10.1111/cogs.12617, PubMed: 29781135

Marchetto, E., & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, 42, 873–902. https://doi.org/10.1017/S0305000914000452, PubMed: 25300736

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80. https://doi.org/10.1126/science.283.5398.77, PubMed: 9872745

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024, PubMed: 17517438

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruites on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials*, *59*, 477–488. https://doi.org/10.1016/0168-5597(84)90006-6

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*, 298–305. https://doi.org/10.1016/S0022-5371(81)90450-3

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846–854. https://doi.org/10.1016/S0010-9452(08)70427-1

Milne, A. E., Mueller, J. L., Männel, C., Attaheri, A., Friederici, A. D., & Petkov, C. I. (2016). Evolutionary origins of non-adjacent sequence processing in primate brain potentials. *Scientific Reports*, *6*, 1–10. https://doi.org/10.1038/srep36259, PubMed: 27827366

Monte-Ordoño, J., & Toro, J. M. (2017a). Different ERP profiles for learning rules over consonants and vowels. *Neuropsychologia*, *97*, 104–111. https://doi.org/10.1016/j.neuropsychologia.2017.02.014, PubMed: 28232218

Monte-Ordoño, J., & Toro, J. M. (2017b). Early positivity signals changes in an abstract linguistic pattern. *PLoS One*, *12*, e0180727. https://doi.org/10.1371/journal.pone.0180727, PubMed: 28678863

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*, 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x, PubMed: 21359123

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008a). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*, *20*, 892–905. https://doi.org/10.1162/jocn.2008.20511, PubMed: 18201121

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science*, *34*, 338–349. https://doi.org/10.1111/j.1551-6709.2009.01093.x, PubMed: 21564216

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences, U.S.A.*, *109*, 15953–15958. https://doi.org/10.1073/pnas.1204319109, PubMed: 23019379

Mueller, J. L., Girgsdies, S., & Friederici, A. D. (2008). The impact of semantic-free second-language training on ERPs during case processing. *Neuroscience Letters*, *443*, 77–81. https://doi.org/10.1016/j.neulet.2008.07.054, PubMed: 18201121

Mueller, J. L., Milne, A. E., & Männel, C. (2018). Non-adjacent auditory sequence learning across development and primate species. *Current Opinion in Behavioral Sciences*, *21*, 112–119. https://doi.org/10.1016/j.cobeha.2018.04.002

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure an ERP study in adult learners. *BMC Neuroscience*, *10*, 1–9. https://doi.org/10.1186/1471-2202-10-89, PubMed: 19640301

Mueller, J. L., ten Cate, C., & Toro, J. M. (2020). A comparative perspective on the role of acoustic cues in detecting language structure. *Topics in Cognitive Science*, *12*, 859–874. https://doi.org/10.1111/tops.12373, PubMed: 30033636

Munka, L., & Berti, S. (2006). Examining task-dependencies of different attentional processes as reflected in the P3a

and reorienting negativity components of the human event-related brain potential. *Neuroscience Letters*, *396*, 177–181. https://doi.org/10.1016/j.neulet.2005.11.035, PubMed: 16356637

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*, 291–296. https://doi.org/10.1177/0963721416655786

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*, 201–227. https://doi.org/10.1418/10879

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, *19*, 1223–1227. https://doi.org/10.1111/j.1467-9280.2008.02228.x, PubMed: 19121127

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2, PubMed: 14732409

Nolden, S., Bermudez, P., Alunni-Menichini, K., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2013). Electrophysiological correlates of the retention of tones differing in timbre in auditory short-term memory. *Neuropsychologia*, *51*, 2740–2746. https://doi.org/10.1016/j.neuropsychologia.2013.09.010, PubMed: 24036359

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*, 1047–1052.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869, PubMed: 21253357

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230. https://doi.org/10.1111/j.1467-9922.2006.00361.x

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://doi.org/10.3389/fpsyg.2013.00345, PubMed: 23801973

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607. https://doi.org/10.1126/science.1072901, PubMed: 12202684

Picton, T. W. (2011). *Human auditory evoked potentials*. Plural Publishing.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*, 322–334. https://doi.org/10.1038/s41583-020-0304-4, PubMed: 32376899

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019, PubMed: 17573239

Pratt, H. (2012). Sensory ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of ERP components* (pp. 89–114). Oxford University Press.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*, 693–705. https://doi.org/10.1038/s41562-018-0406-4, PubMed: 31346278

Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition*, *171*, 130–150. https://doi.org/10.1016/j.cognition.2017.11.003, PubMed: 29156241

R Core Team. (2020). *A language and environment for statistical computing* (Vol. 2). R Foundation for Statistical Computing https://www.r-project.org.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595–626. https://doi.org/10.1111/lang.12010

Ruchkin, D. S., Berndt, R. S., Johnson, R., Ritter, W., Grafman, J., & Canoune, H. L. (1997). Modality-specific processing streams in verbal working memory: Evidence from spatio-temporal patterns of brain activity. *Cognitive Brain Research*, *6*, 95–113. https://doi.org/10.1016/S0926-6410(97)00021-9, PubMed: 9450603

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621. https://doi.org/10.1006/jmla.1996.0032

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*, 700–703. https://doi.org/10.1038/nn873, PubMed: 12068301

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335. https://doi.org/10.1111/psyp.13335, PubMed: 25151545

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010, PubMed: 25151545

Schiller, N. O., & Costa, A. (2006). Different selection principles of freestanding and bound morphemes in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1201–1207. https://doi.org/10.1037/0278-7393.32.5.1201, PubMed: 16938057

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport*, *9*, 3355–3358. https://doi.org/10.1097/00001756-199810260-00003, PubMed: 9855279

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*, 1391–1400. https://doi.org/10.1038/nn1549, PubMed: 16158062

Sharp, D. J., Scott, S. K., Cutler, A., & Wise, R. J. S. (2005). Lexical retrieval constrained by sound structure: The role of the left inferior frontal gyrus. *Brain and Language*, *92*, 309–319. https://doi.org/10.1016/j.bandl.2004.07.002, PubMed: 15721963

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. *The Production of Speech*, 109–136. https://doi.org/10.1007/978-1-4613-8202-7_6

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, *19*, 137–144. https://doi.org/10.1111/j.1467-9280.2008.02059.x, PubMed: 18271861

Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics*, *70*, 1515–1525. https://doi.org/10.3758/PP.70.8.1515, PubMed: 19064494

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*, 507–520. https://doi.org/10.1016/j.jml.2012.07.008

Van Den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967–985. https://doi.org/10.1162/089892901753165872, PubMed: 11595099

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory and Cognition*, *24*, 573–583. https://doi.org/10.3758/BF03201084, PubMed: 8870528

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, *66*, 8–30. https://doi.org/10.1111/lang.12137

Wetter, S., Polich, J., & Murphy, C. (2004). Olfactory, auditory, and visual ERPs from single trials: No evidence for habituation. *International Journal of Psychophysiology*, *54*, 263–272. https://doi.org/10.1016/j.ijpsycho.2004.04.008, PubMed: 15331217

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., et al. (2020). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, *12*, 843–858. https://doi.org/ttps://doi.org/10.1111/tops.12381, PubMed: 32729673

Winkler, I., Debener, S., Muller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4101–4105. https://doi.org/10.1109/EMBC.2015.7319296, PubMed: 26737196

Yingling, C. D., & Nethercut, G. E. (1983). Evoked responses to frequency shifted tones: Tonotopic and contextual determinants. *International Journal of Neuroscience*, *22*, 107–118. https://doi.org/10.3109/00207459308987389, PubMed: 6668128

Ziegler, W., Aichert, I., & Staiger, A. (2010). Syllable- and rhythm-based approaches in the treatment of apraxia of speech. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, *20*, 59–66. https://doi.org/10.1044/nnsld20.3.59