Research Article ⬡

AN EAR FOR LANGUAGE

SENSITIVITY TO FAST AMPLITUDE RISE TIMES PREDICTS NOVEL VOCABULARY LEARNING

Marta Marecka* ⓘ

Institute of Psychology, Jagiellonian University

Tim Fosker ⓘ

School of Psychology, Queen's University Belfast

Jakub Szewczyk ⓘ

Institute of Psychology, Jagiellonian University and Department of Psychology, University of Illinois at Urbana–Champaign

Patrycja Kałamała ⓘ

Institute of Psychology, Jagiellonian University

Zofia Wodniecka ⓘ

Institute of Psychology, Jagiellonian University

ABSTRACT

This study tested whether individual sensitivity to an auditory perceptual cue called amplitude rise time (ART) facilitates novel word learning. Forty adult native speakers of Polish performed a perceptual task testing their sensitivity to ART, learned associations between nonwords and pictures

of common objects, and were subsequently tested on their knowledge with a picture recognition (PR) task. In the PR task participants heard each nonword, followed either by a congruent or incongruent picture, and had to assess if the picture matched the nonword. Word learning efficiency was measured by accuracy and reaction time on the PR task and modulation of the N300 ERP. As predicted, participants with greater sensitivity to ART showed better performance in PR suggesting that auditory sensitivity indeed facilitates learning of novel words. Contrary to expectations, the N300 was not modulated by sensitivity to ART suggesting that the behavioral and ERP measures reflect different underlying processes.

INTRODUCTION

What makes some people excel at learning words of a foreign language and others struggle with the acquisition of even basic vocabulary? Up to now, researchers identified several factors that characterize good learners of foreign languages. These include the so-called language aptitude (Carroll & Sapon, 1959), phonological short-term memory capacity (e.g., Baddeley et al., 1998), inhibitory control (Bartolotti et al., 2011), and musical abilities (Dittinger et al., 2016, 2017). However, to the best of our knowledge, no research to date has shown a relationship between any specific perceptual skill and learning novel words in healthy adults. In this article, we argue that sensitivity to an acoustic cue called amplitude rise time (ART) is relevant for vocabulary acquisition. This acoustic cue is considered important for speech segmentation in an individual's first language (Richardson et al., 2004; Thomson & Goswami, 2009). Given that better segmentation can lead to more efficient encoding of a novel word form (cf. Marecka et al., 2018 for an overview), high sensitivity to ART should make word learning much easier.

AUDITORY SKILLS AND NOVEL WORD LEARNING

Linguistically talented people are thought to have "an ear for language." However, there have been few studies on auditory skills and novel word learning to support this claim. Perhaps the most prominent line of research on this topic concerns the relationship between word learning and the ability to segment speech into linguistic units such as phonemes, rhymes, and syllables. To measure the ability to segment speech, researchers typically use tasks involving the extraction of a phonological unit from a word (i.e., saying the word cat without /k/), pointing to the unit that is shared by two words (e.g., the same phoneme or rhyme), or dividing words into phonemes or syllables (e.g., Farnia & Geva, 2011; Hu & Schuele, 2005; Marecka et al., 2018). Second language (L2) learners and children learning their first language (L1) who score higher in such tasks, tend to have larger vocabulary sizes and learn novel words more efficiently in auditory learning tasks (L1; Bowey, 1996, 2001; Farnia & Geva, 2011; Hu, 2003, 2008; Hu & Schuele, 2005; Marecka et al., 2018; Metsala, 1999).

Recently, Marecka et al. (2018) proposed that two different segmentation mechanisms might be at play, depending on whether learners process an already familiar language or a completely new language. It is implicit in the theory that these segmentation mechanisms facilitate auditory word learning—they help with processing the speech signal in the novel word and subsequently acquiring its form. One of the mechanisms—*phonological mapping*—operates when processing a known language, whereas the other—*universal segmentation*—operates when processing a novel language (Marecka et al., 2018).

Phonological mapping relies on the existing sublexical phonological representations, that is, representations of syllables, segmental categories (segments or phonemes), and larger sequences of segments in the listener's memory (a.k.a. speech chunks or ngrams; see Jones, 2016; Szewczyk et al., 2018). Within this mechanism learners segment words by pattern matching: they compare the acoustic form they hear against known phonological patterns. As a result, they decompose the stream of speech into a sequence of known phonological units that can then be encoded in the memory (see the EPAM-VOC and CLASSIC models: Jones, 2016). The better the quality and the larger the inventory of sublexical phonological representations in the learner's memory, the more efficient the mapping and the better the encoding of the word forms (Jones, 2016). This hypothesis is confirmed by experiments showing that children remember novel words better, if those words contain speech chunks they know (Storkel, 2001, 2003). Moreover, computational models show that well-established representations of speech chunks lead to more efficient encoding of word forms in memory (Jones & Witherstone, 2011).

In contrast to phonological mapping mechanism, the universal segmentation mechanism does not require the knowledge of any phonological representations—in this process listeners use language-universal acoustic cues to divide speech into smaller parts such as words, syllables, and subsyllabic elements like phonemes (see Endress & Hauser, 2010). Therefore, the universal segmentation mechanism is especially useful for processing speech in a language in which the learners have few phonological representations. It helps less proficient learners acquire novel word forms by allowing them to divide the speech stream into manageable units that can then be further processed and encoded (Marecka et al., 2018).

Summing up, research indicates that the ability to segment speech helps learning (auditorily) novel word forms. Two segmentation mechanisms might be involved in the process: *phonological mapping*, which depends on the individual's knowledge of phonological structures and thus is related to the experience with the phonology of the language, and *universal segmentation*, which depends on the detection of acoustic cues and thus is likely to be related to individual differences in perceptual skills. In this article we focus on the second mechanism, and in particular on one acoustic cue that might be related to it—sensitivity to temporal amplitude changes, and more specifically ART. We test if greater sensitivity to ART (understood as the ability to discriminate between different ARTs) helps individuals learn new words. In the following sections, we define ART and focus on the studies supporting the notion that temporal amplitude changes, and ARTs specifically, provide cues to universal segmentation and thus facilitate word learning.

Sensitivity to ART as a Predictor of Universal Segmentation and of Word Learning

The speech stream presumably contains a number of acoustic cues that could be relevant to universal speech segmentation, but most of the current research focuses on the cues related to how the amplitude (intensity, perceived as loudness of the sound) changes over time. One cue that receives particular attention is ART. ART is the time a sound takes to reach its maximum amplitude. ART is short when a sound's amplitude rises very rapidly (i.e., the sound gets loud very fast) and long when its amplitude rises more gradually. Tasks used to assess individual sensitivity typically involve deciding if two sounds have the same or different ART (e.g., Hämäläinen et al., 2005), or picking a sound with a different ART (the

odd-one-out) from a series of three (e.g., Surányi et al., 2008). The differences in ART between the sounds presented to the participants vary from very large to very small. If participants give correct answers for trials where the differences in ART between the sounds are very small, they have greater sensitivity to ART than participants who can distinguish only between sounds with larger ART differences.

In this study we test the hypothesis that greater sensitivity to ART (i.e., the ability to perceive smaller differences in ART between two sounds) facilitates word learning in a completely foreign language. This hypothesis rests on two premises: (a) greater sensitivity to ART leads to more efficient universal speech segmentation; and (b) more efficient universal speech segmentation leads to faster and more accurate novel word learning. In the following text, we discuss the data supporting each premise and the nascent evidence for the hypothesis derived from them.

Greater Sensitivity to ART Discrimination Leads to More Efficient Universal Speech Segmentation

Accurate perception of ARTs is thought to help speech segmentation and processing. Discriminating between sounds with relatively slow ART (slow ART discrimination) is thought to facilitate segmentation of speech into syllables (Goswami, 2011; Goswami et al., 2002; Richardson et al., 2004). Discriminating between sounds with relatively fast ART (fast ART discrimination) facilitates segmentation of speech into onset, rhymes, and phonemes (e.g., Goswami, 2011; Goswami et al., 2002; Hämäläinen et al., 2005; McAnally & Stein, 1996; Richardson et al., 2004).

Support for the relationship between sensitivity to ART and speech segmentation comes from studies on developmental dyslexia. Dyslexic individuals typically have poor segmentation skills and they also have problems discriminating sounds with different ARTs (e.g., Goswami et al., 2002; Hämäläinen et al., 2005; McAnally & Stein, 1996; Muneaux et al., 2004; Richardson et al., 2004; Surányi et al., 2008, although see Georgiou et al., 2010 for different results). Furthermore, these two abilities (speech segmentation and ART) are directly correlated in this population (Hämäläinen et al., 2005; Richardson et al., 2004; Surányi et al., 2008; Thomson & Goswami, 2009). Some studies show that individuals with dyslexia have problems with slow ART discrimination (Goswami et al., 2002; Hämäläinen et al., 2005), while other studies suggest that it is fast changes, that is, fast ART detection that should be problematic (McAnally & Stein, 1996).

Importantly, the relationship between ART discrimination and phonological segmentation in dyslexic individuals has been found in many language communities. ART discrimination is a consistent predictor of segmentation skills among English, Spanish, Chinese (Goswami et al., 2011), Hungarian (Surányi et al., 2008), Finnish (Hämäläinen et al., 2005), and French individuals with dyslexia (Muneaux et al., 2004).[1] This indicates that ART is a language-universal cue to speech segmentation—a very likely cue for universal segmentation.

More Efficient Universal Speech Segmentation Leads to Faster and More Accurate Novel Word Learning

As already mentioned, individuals who score lower on tasks involving segmenting speech into phonemes and syllables, learn words slower and less accurately (Bowey, 1996, 2001;

Farnia & Geva, 2011; Hu, 2003, 2008; Hu & Schuele, 2005; Marecka et al., 2018; Metsala, 1999). In the studies cited previously the segmentation involved in this relationship could be either universal segmentation or phonological mapping. However, studies looking at the relationship between musical training, speech segmentation abilities, and word learning suggest an involvement of specifically universal segmentation in word learning (Dittinger et al., 2016, 2017; François & Schön, 2011; François et al., 2013). These studies show that musically trained children (Dittinger et al., 2017) and adults (Dittinger et al., 2016) learn words of a completely foreign language better than those without musical training. Following a word learning task, participants with musical training were more accurate in assessing the meaning of the newly learned words. While there is no reason why musicians should have a richer inventory of phonological sequences, it is likely that they have superior sensitivity to acoustic cues, including those involved in segmenting speech.

In sum, the existing evidence shows a link between ART sensitivity and better segmentation of speech, as well as between better segmentation of speech and better word learning. However, it is not clear if sensitivity to ART contributes to novel word learning. Some indirect evidence has been provided by studies with dyslexic individuals, who typically have poor ART discrimination skills and experience difficulties learning novel words (Alt et al., 2017; Kalashnikova & Burnham, 2016; Kwok & Ellis, 2014). Children with specific language impairment (SLI)—a disorder that is often characterized by lower vocabulary scores—have been found to be less sensitive to ART (Cumming et al., 2015). Moreover, correlations were reported between sensitivity to ART and vocabulary size (Corriveau et al., 2007) as well as between sensitivity to ART and vocabulary learning in a paired-associates laboratory task (Thomson & Goswami, 2009). However, these studies included analyses on heterogonous groups, composed of both typically developing individuals and those with dyslexia or SLI. Therefore, it is unclear whether the relationship between ART and vocabulary size holds for a typically developed adult population.

CURRENT STUDY

In the current study, we directly addressed the hypothesis that a greater sensitivity to ART is related to more accurate and faster auditory word learning, especially for words that are phonologically novel, that is, have foreign phonological structure and accent. This has never been studied in typical adults.

To assess sensitivity to ART we used odd-one-out discrimination tasks, which estimated participants' thresholds for discriminating fast and slow ART separately. To assess the ability to learn new words, we used a paired associate learning paradigm (see de Groot, 2011 for an overview), in which participants were asked to remember novel words (nonwords) presented auditorily and paired with pictures of known objects. We additionally statistically controlled for other individual differences that, as suggested by earlier research, can independently explain word learning, namely phonological short-term memory as measured with digit span (e.g., Baddeley et al., 1998), and the resistance to interference (the ability to ignore interfering stimuli; see Friedman & Miyake, 2004) as measured by the Simon task (Bartolotti et al., 2011; Rey-Mermet & Gade, 2018). This second skill might be important specifically when participants learn new names of known objects that already have their names in L1, as was the case in the current study. Finally, we explored whether the universal segmentation mechanism is especially relevant for words that are phonologically novel.

992 *Marta Marecka, Tim Fosker, Jakub Szewczyk, Patrycja Kałamała, and Zofia Wodniecka*

Thus, we varied the nonwords' phonological similarity to words in the participants' L1. We manipulated both the nonwords' phonological structure (by using less or more typical phonological sequences to construct them) and the accent with which they were pronounced.

Manipulating both accent and structure of the word separately was done in an attempt to systematize the term "phonologically foreign." Several studies explore how different factors influence the learning or processing of foreign words, but none define what "foreign" means from a phonological point of view. We argue that the words can be foreign in terms of phonological structure (the phonotactic probability of these words) and/or the acoustic features that can be ascribed to accent (the quality of vowels and consonants, as well as prosodic patterns). These two aspects of "foreignness" are strikingly different. While accent is connected more with surface processing of the word (the processing of acoustic features), processing structure might be more linguistic— possibly depending on the knowledge of the sublexical representations (see e.g., Jones, 2016). Based on the current literature, we did not have a specific hypothesis as to whether learning words with foreign accent or structure would be more associated with ART discrimination. Foreign structure can result in foreign patterns of amplitude rises within the utterance, so the learning of such words might be facilitated by better ART discrimination. However, foreign accent could result in greater reliance on universal segmentation because participants might have problems with identifying known phonemes and syllables within foreign-accented words. In either case, making the distinction between accent and structure is important for understanding the potential relationship between ART and word learning.

To measure the accuracy and quality of representations in learning the paired associates, we used a picture recognition (PR) task in which participants were auditorily presented with a nonword followed by a congruent or incongruent picture, and were asked to indicate whether the picture matched the nonword. Apart from collecting behavioral data (accuracy and response latencies in the PR task), we also measured event-related potentials (ERPs) time-locked to the presentation of the pictures. ERPs may be particularly sensitive to the early stages of word acquisition, showing effects of learning even after a single exposure to a new word (Borovsky et al., 2010, 2012; McLaughlin et al., 2004; Osterhout et al., 2006, 2008) and can reveal differences in word learning by less and more skilled learners (Balass et al., 2010; Perfetti et al., 2005).

If greater sensitivity to ART leads to more better word learning, we should find that participants who are more sensitive to ART will be more accurate and faster on the PR task. This effect should be particularly pronounced in novel words that have a less familiar phonological structure and accent and thus require participants to use their universal segmentation mechanism.

With respect to the ERPs, we focused on the N300 ERP component. The N300 is a fronto-central component occurring 250–400 ms after the stimulus of interest. It is a member of a larger family of ERP components (the N3 complex). These components reflect the activity in occipitotemporal cortex related to the cognition and decision making connected with visual objects (Schendan, 2019). For example, higher (more negative) N300s are typically observed in response to pictures that are semantically incongruent (e.g., a picture of a women putting a checkerboard into the oven) or semantically unrelated to the preceding stimuli (e.g., Ganis & Kutas, 2003; Mazerolle et al., 2007; Mudrik et al., 2010; West & Holcomb, 2002), compared to congruent or related pictures. The N300 is

https://doi.org/10.1017/S0272263120000157 Published online by Cambridge University Press

also more negative to visually presented objects that are presented from an atypical perspective or are impoverished, so they match reference objects in memory to a lesser extent (Schendan & Ganis, 2015; Schendan & Stern, 2007).

In the present study, if a participant learns the nonword–picture association, the presentation of the nonword should activate memory representations of the object linked with this nonword. When the participant is then shown a picture consistent with the nonword, this picture should reduce the negativity in the N300 amplitude (make it more positive) compared to when the participant did not learn the association or when the participant is shown an incongruent picture. Building upon this logic, we propose that the better the participants learn the association, the more preactivated the concept should be and hence the more reduced the N300 amplitude in response to the congruent pictures. N300 amplitude on the congruent pictures should show how strongly participants activate the meaning upon hearing the word and thus could be considered an index of how well they learned the association between the form and meaning. For pictures incongruent with the newly acquired nonwords, the N300 amplitude should not be reduced no matter how well participants learned the nonword–picture associations because the concept expressed by the incongruent picture will not be activated by the preceding nonword. Thus, the difference between the N300 amplitude for congruent and incongruent pictures (the magnitude of the N300 effect) will provide an index of how strongly participants activated the association to the new word. As such, we consider the modulation of the N300 an additional index of paired associates learning, complementary to the behavioral data.

METHOD

PARTICIPANTS

Forty-four native speakers of Polish took part in the experiment. We rejected data for four participants who knew Russian because half of the stimuli in the experiment were produced with a Russian accent, which was supposed to be unfamiliar for the learners. We included 40 participants (26 female) in the final analyses. The participants were on average 23.3 years old ($SD = 2.9$, range: 19.0–30.10) and had 15 years of formal education ($SD = 2.42$, range: 12–25). All participants completed a background questionnaire, where they needed to assess their proficiency on all the known languages on a 9-point Likert scale for four components: reading comprehension, listening comprehension, speaking, and writing. We averaged the scores on these four components for each language. If the average score was above 6, we assumed that the participant knew the language on upper-intermediate or a higher level. Twenty-two participants (55% of the sample) reached this level of proficiency for at least one additional language, but none for Russian. A shortened version of Raven's Advanced Progressive Matrices test (sRAPM) was used to measure participants' fluid intelligence (see Marzecová et al., 2013 for a detailed description and rationale). On average, participants scored 13.18 points out of 18 ($SD = 2.65$, range: 7–17). All participants passed a hearing screen, in which they were played six pure tones at 30 dB HL. Each tone (250 Hz, 500 Hz, 1,000 Hz, 2,000 Hz, 4,000 Hz, and 8,000 Hz) was tested twice in the left and the right ear separately. No participant reported any language impairments or neurological disorders and all had normal or corrected-to-normal vision. The participants were recruited using a job-hunting Internet portal or through an

experimental recruitment system at Jagiellonian University, Kraków. They were paid 40 zł (about 12 dollars) for their participation and signed an informed consent form prior to the experiment.

PREDICTOR MEASURES

Sensitivity to ART

Sensitivity to ART was tested using two 3-interval 2-alternative forced choice tasks (3I-2AFC) separately for fast ART and slow ART. Both tasks were administered on a laptop using a custom program coded in Real Basic. In each task participants saw three cartoon owls, which produced tones. On each trial, either the first or the third owl produced a tone that was different from the others in ART. Participants had to choose which of these two tones (the first or the last one) was the odd-one-out. Feedback on the participant's accuracy was provided automatically after each trial.

This task used an adaptive (staircase) procedure to establish an individual's discrimination threshold, that is, the smallest possible change in ART a participant can detect. We used a combined 1-up 2-down and 1-up 3-down staircase. Under the 1-up 2-down staircase, every time a participant made two consecutive correct responses (i.e., correctly identified the odd-one-out), the task became harder (i.e., the difference between the tones was reduced) and every time they got one trial incorrect the discrimination became easier. After two reversals (a change in direction from getting easier to getting harder or vice versa) had been reached, the staircase changed. Under the 1-up 3-down staircase, every time a participant made three consecutive correct responses, the task became harder (i.e., the difference between the tones was reduced) and every time they got one trial incorrect the discrimination became easier. The initial step size of 58 ms was halved after the fourth and sixth reversals to enable the staircase to converge more closely on the participant's threshold. The tasks ended after 40 trials. The discrimination threshold for each condition/task was computed as the average of the last three reversals of the staircase procedure.

Each task contained sinusoidal tones of 450 ms duration with a fixed frequency of 500 Hz. All tones were played out at a peak intensity of 75 dB SPL. The tones in the two tasks were manipulated in the following way:

(a) Task testing sensitivity to fast ART
    The two standard tones always had a 15 ms linear ART, 385 ms steady rate, and a 50 ms linear fall time. The target tone (odd-one-out) varied in linear ART from the standard tones. At the start, the ART of the target tone was 300 ms (i.e., the procedure started with the maximal difference of 285 ms ART) and it got shorter in subsequent trials.

(b) Task testing sensitivity to slow ART
    The two standard tones always had a 300 ms linear ART, 385 ms steady rate, and a 50 ms linear fall time. The target tone (odd-one-out) varied in linear ART from the standard tones. In the initial step, the ART of the tone was 15 ms and it got longer in subsequent trials.

Simon Task

The task was based on paradigms employed by Paap and Sawi (2014) and Bialystok et al. (2004) who used it to measure resistance to interference in bilinguals. The task was presented in the DMDX software (Forster & Forster, 2003).

A trial consisted of a 500 ms blank interval, followed by a fixation cross for 300 ms and then a letter "P" or "Q." Participants had to press the letter they saw as quickly as possible. The timeout was 1,000 ms. The letters remained on the screen until the timeout. Throughout the task, participant's left hand rested on the "Q" key and the right hand rested on the "P" key on the keyboard. Participants first did a noncritical block (practice session), where the letters appeared 2.3° above or below from the central fixation. This block contained 80 trials. Then they did one critical block, where the letters were presented either 3.9° to the left or to the right of the central fixation. In the congruent condition of the critical block, localization of the letter on the screen (left or right) matched localization of response button on the response keyboard (left or right) ("P" appeared on the right side of the screen or "Q" appeared on the left side of the screen). In the incongruent condition, the letter on the screen did not match the location of the response button on the keyboard ("Q" appeared on the right side of the screen or "P" appeared on the left side of the screen). There were 160 trials in the critical block, 80 congruent and 80 incongruent, presented in a random order. The Simon effect was calculated as the difference in mean reaction times for correct responses between incongruent and congruent condition in the critical block.

Forward Digit Span

To measure individual differences in short-term memory, we used a forward digit span task. This was a pen and article task, taken from the Polish version of the Wechsler Intelligence Scale for adults (Brzeziński et al., 2004) and administered as per the standard procedure. The participants were instructed that they would hear strings of digits and would be asked to repeat those digits in the order of presentation. The experimenter read each string aloud once at a pace of approximately one digit per second and asked the participant to repeat the string immediately after. The participant could get either one point for each string—if all digits in a string were recalled correctly—or zero points. The task always began with two strings of three digits, then two strings of four digits and so on up until nine digits (overall 14 trials, two per each span level). The participants got feedback only for the first two strings (these trials are also included in the final score). The task was terminated if the participant repeated two strings of the same length incorrectly.

OUTCOME MEASURE: WORD LEARNING

Stimuli

Novel Words
Twelve novel words (nonwords) were created for the purpose of the task. First, on the basis of frequency data from the Polish corpus SUBTLEX-PL (Mandera et al., 2015), we

algorithmically created a set of 80 bisyllabic nonwords that were phonotactically possible, but maximally differed in terms of phonotactic probability, as measured with ngram frequency. Ngram frequency is a mean (log) frequency of all ngrams, that is, phoneme sequences (that is bigrams—two-phoneme sequences, trigrams—three-phoneme sequences, and so forth, up to the length of the stimulus minus 1) in a given nonword, weighted for the length of the ngram.[2] A previous study has shown that this metric was the superior over other indices of phonotactic probability (see Szewczyk et al., 2018).

From the nonwords created by the algorithm, we chose 12 that had a high ngram frequency (4.66–6.45) and 12 that had a low ngram frequency (2.66–3.88). All the chosen nonwords had a CCVCV (consonant-consonant-vowel-consonant-vowel) or CVCCV structure and were stressed on the penultimate syllable. The nonwords were then recorded by a native speaker of Polish and tested for wordlikeness by 20 native speakers of Polish, who had to indicate on a 5-point Likert scale to what extent a particular nonword sounded like a typical Polish word, with 1 meaning "this could never be a Polish word" and 5 meaning "this could very well be a Polish word." On the basis of the test, from the high ngram frequency nonwords, we chose six that were assessed as most wordlike ($M = 4.03$, $SD = 0.25$). These were the native structure items. We also chose six nonwords from the low ngram frequency set that had lowest wordlikeness ($M = 2.82$, $SD = 0.59$), which became nonnative structure items. Each of the nonwords selected started with a different phoneme. All items are presented in Appendix A.

All items were then recorded in two versions by a bilingual speaker of Polish and Russian. The speaker recorded each of the 12 items with a Russian accent and, during a separate session, with a Polish accent. Thus, we created two accent versions of the stimuli: native accent (Polish) and nonnative accent (Russian). Each nonword was recorded three times with each accent. All three sets were interchangeably used in the experiment to provide some variance in the acoustic realization of the nonwords.

Pictures

From the SUBTLEX-PL corpus, we chose 12 concrete high-frequency nouns. Next, we searched in Google Image search engine for black-and-white line drawing depicting the nouns. For each concept, we chose three pictures depicting it. To check whether the pictures tend to be named unambiguously, we conducted a pilot study on 17 participants. The participants received a questionnaire, in which they were asked to name all 36 pictures. The accuracy rate on the test was 99.18%, which shows that the names of the pictures were unambiguous. All pictures were adjusted to be perceptually similar in size. The images were displayed in the center of the screen. Using three versions of the pictures portraying each object was aimed at forcing participants to associate the nonwords with concepts, rather than specific depictions of these concepts.

Stimuli Lists

We created four stimuli lists. In each of them we had three nonwords with native accent and native structure, three nonwords with native accent and nonnative structure, three words with nonnative accent and native structure, and three words with nonnative accent and nonnative structure. Because we recorded each of the nonwords in a native accent version and in a nonnative accent, we counterbalanced the accents in the
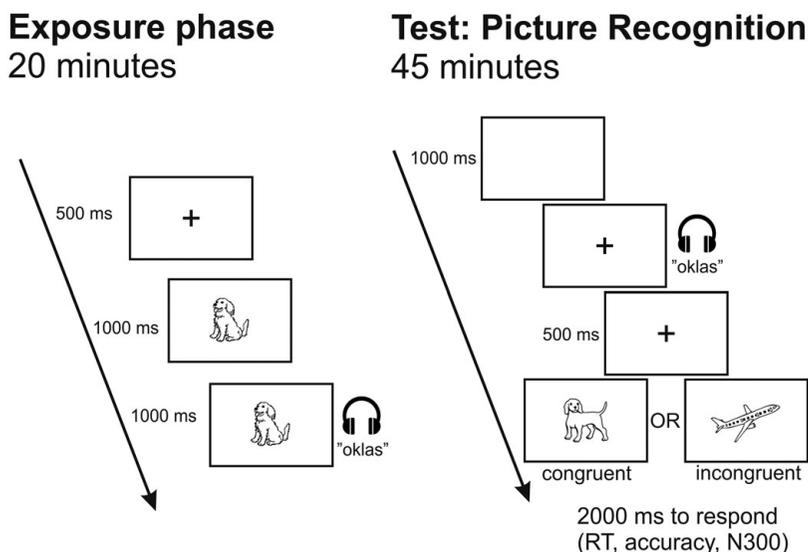
FIGURE 1. The word learning task. The left panel represents an exemplar trial of the exposure phase. The panel on the right represents the trial structure of the test phase (Picture Recognition Task), which followed the exposure phase.

nonwords across the lists. That means that the nonwords produced with native accent in two of the lists, was produced with nonnative accent in the other two and vice versa. For each list we randomly assigned nonwords to concepts depicted in the pictures, so that for each of the four lists the associations between nonwords and pictures were different.

Procedure

The word learning task was presented using the DMDX software. It involved an exposure phase and a test phase. The procedure is graphically represented in Figure 1 and described in more detail in the following text.

Exposure Phase

In the exposure phase, the participants were familiarized with all the stimuli. They were told that they would have to learn 12 new words and their association to objects, and that later they would be tested on how well they learned the nonwords.

The exposure phase consisted of four blocks of 108 trials each. Each trial took 2,500 ms. The trial structure is presented in Figure 1 on the left. The participant saw a fixation cross for 500 ms, followed by the presentation of a picture for 1,000 ms. After those 1,000 ms the participant heard the nonword associated with the picture through the headphones —the picture stayed on the screen for the remaining 1,000 ms. We used three pictures of the object and three different recordings of each nonword. This was done to ensure that the

participants did not learn the association between only one particular version of the recording and a particular picture, but that they generalized over three acoustically different versions of the word and built the link between this word and a conceptual representation of the object portrayed using the three different pictures. We paired all the versions of the nonword with all the pictures of the object, which gave nine different combinations for each picture–nonword association. We presented all these nine combinations once per block for all 12 nonword-object associations, yielding 108 trials. The order of trials was fully randomized within each block.

There were breaks between each of the four blocks. The whole exposure phase took around 20 minutes.

Testing Phase: Picture Recognition Task

The task consisted of three blocks of 264 trials. Each trial began with 1,000 ms of blank screen, immediately followed by a fixation cross and the auditory presentation of a nonword, which occurred at the same time (see Figure 1 on the right). After the offset of the word, the fixation cross stayed on the screen for 500 more ms and then a picture was presented on the screen. The picture displayed was either the one that was associated with the presented novel word (congruent) or another picture that was associated with one of the other 11 novel words (incongruent). The participant had to decide if the picture matched or did not match the word heard. The timeout for the reaction was 2,000 ms. There was no feedback during the PR task. In the task both EEG data and behavioral data were collected.

Within each block, each novel word occurred in 11 congruent trials (with different versions of recordings and pictures) and in 11 incongruent trials. Each incongruent trial used a picture associated with a different novel word. The order of trials was randomized within the block. The blocks were divided by short breaks during which the participants were encouraged to rest.

PROCEDURE

The testing session took about 3 hours with a short break in the middle. First, the participants were informed about the aim of the experiment and they underwent a hearing screen and a shortened version of sRAPM, which served as screening measures (see details described in the "Participants" section). Then they performed the digit span, sensitivity to ART tasks, and the Simon task. Following the behavioral tests, they were prepared for the EEG testing and underwent the word learning procedure. First, they were exposed to the novel word names and pictures for 20 minutes and then were tested with a PR task. At end of the session, participants completed the background questionnaire and were paid for participation.

EEG ACQUISITION

The EEG was recorded at 256 Hz from 32 Ag/AgCl scalp electrodes, positioned at the standard 10–20 locations, mounted in an elastic cap, using the BioSemi ActiveTwo system. Electrodes were initially referenced online to the Common Mode Sense electrode

located at C1 and re-referenced off-line to linked mastoids. Horizontal and vertical electrooculogram was recorded bipolarly using electrodes placed at the outer canthus of each eye (horizontal), as well as below and above a participant's right eye (vertical). The EEG signal was filtered off-line with an IIR band-pass filter (0.05–25 Hz frequency range; low cutoff slope: 24 dB/oct; high cutoff slope: 12 dB/oct).

We used BrainVision Analyzer software to analyze the ERPs in the PR task. From the EEG recordings we extracted epochs of 150 ms before and 800 ms after the onset of the picture. Both correct and incorrect trials were analyzed. The epochs were baseline corrected using the 150-ms prestimulus time window. We removed ocular and other stationary artifacts using independent components analysis (ICA; Delorme et al., 2007; Jung et al., 2000). We then inspected all trials manually to remove any remaining artifacts (9.3% of congruent and 8.9% of incongruent trials in total).

On the basis of a previous study, which used a similar paradigm (presentation of word and then a picture—Mazerolle et al., 2007), we investigated mean voltage amplitudes of the N300 on the frontocentral electrodes (Fz, FC1, FC2, Cz) within the 250–350 ms time window.

DATA ANALYSIS

We created three linear mixed-effects models, to establish the effects of ART on the outcome variables taken from the PR Task: RTs, accuracy, and N300 amplitude.[3] The models for RT and N300 were mixed effects linear regressions, calculated using lmer function in lme4 package (Bates et al., 2015b). The model for accuracy was a mixed-effects binary logistic regression, calculated using the lme4 package. For all three models, we used the Satterthwaite approximation for p values, as implemented in the lmerTest package (Kuznetsova et al., 2015). In all models, we started with the maximal structure of random effects. If the model did not converge, we first removed correlations between random effects, and—in the next step—the random effects with the smallest unique variance, following the recommendation by Bates et al. (2015a). All predictor variables in all the models were standardized. For the RT model, only the correct responses were taken. The RTs in the model were log-transformed.

For both the RTs and accuracy models, the participant-related fixed effects entered into the model were forward digit span, the mean Simon effect, and the estimates of fast ART and slow ART thresholds. The item-related fixed effects were accent (native vs. nonnative), structure (native, nonnative), and congruence (congruent, incongruent). We also entered interactions between accent and ART thresholds and interactions between structure and ART thresholds.

In the N300 model, we used the same fixed effects as in the preceding analyses, with the addition of the interactions of congruence and ART thresholds, accent and ART thresholds, structure and ART thresholds, and the three-way interaction of accent, congruence, and ART thresholds, as well as the three-way interactions of structure, congruence, and ART thresholds. In this model, the primary effects of interest were interactions with congruence because our index of learning was the reduction of N300 in the congruent trials, as compared to the incongruent trials (see also the last paragraph of the "Introduction" section).

The R scripts containing the specification of all models are available at https://osf.io/3fxgc/. All graphs were created in R using the ggplot2 package (Wickham et al., 2018).

TABLE 1.   Modeling the predictors of the picture recognition accuracy: Random effects

Random variable	Name	Variance	SD
Participant	*Intercept*	0.46	0.68
Participant	Accent	0.04	0.21
Participant	Congruence	0.15	0.38
Item	*Intercept*	0.01	0.10
Item	slow ART	0.01	0.12
Item	Digit span	0.01	0.09

TABLE 2.   Modeling the predictors of the picture recognition accuracy: Fixed effects

	Estimate	Upper CI	Lower CI	SE	z value	p
Intercept	**3.31**	**3.08**	**3.55**	**0.12**	**27.96**	**<0.001*****
Simon effect	**−0.27**	**−0.50**	**−0.04**	**0.12**	**−2.31**	**0.021***
Digit span	**0.26**	**0.00**	**0.52**	**0.13**	**1.98**	**0.047***
Accent	**−0.19**	**−0.35**	**−0.03**	**0.08**	**−2.29**	**0.022***
Structure	0.06	−0.08	0.20	0.07	0.91	0.364
Congruence	**0.18**	**0.01**	**0.36**	**0.09**	**2.07**	**0.038***
fast ART	**−0.40**	**−0.66**	**−0.15**	**0.13**	**−3.14**	**0.002****
slow ART	0.01	−0.08	0.45	0.13	1.39	0.163
Accent × structure	0.20	−0.07	0.47	0.14	1.42	0.156
Accent × fast ART	−0.04	−0.19	0.11	0.08	−0.52	0.603
Structure × fast ART	−0.06	−0.18	0.05	0.06	−1.07	0.284
Accent × slow ART	−0.08	−0.26	0.10	0.09	−0.89	0.376
Structure × slow ART	0.04	−0.11	0.19	0.08	0.52	0.603

Significant effects are bolded (* $p < .05$, ** $p < .01$, *** $p < .001$).

RESULTS

PICTURE RECOGNITION TASK: ACCURACY DATA

Mean accuracy for the PR was 94.9% (range: 75.25%–99.24%; $SD = 0.05\%$). The binary logistic regression model for the accuracy in PR is presented in Tables 1 (random effects) and 2 (fixed effects). Lower fast ART thresholds, smaller Simon effect and larger digit span were related to greater accuracy in the PR task. The participants were more accurate for items pronounced with native rather than nonnative accent and for incongruent rather than congruent trials. Figures 2 and 3 show the accuracy in behavioral data as a function of sensitivity to fast ART. Figure 2 shows raw data with the LOESS line fitted, while Figure 3 shows the predictions of the model. It is worth mentioning that while fast ART and slow ART were moderately correlated ($r = 0.43$), the variance inflation factor (VIF) diagnostics for this and all the other models reported here had appropriately low collinearity (VIF < 1.5)

PICTURE RECOGNITION TASK: RT BEHAVIORAL DATA

Mean RT for the PR was 679.42 ms (range: 445.64–1,109.41; $SD = 151.12$). The estimates of the random and fixed effects for the analysis of picture-naming latencies
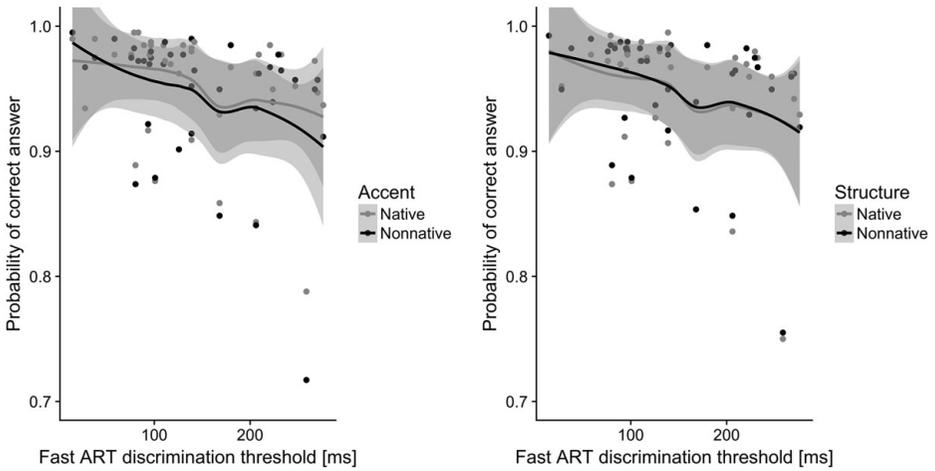
FIGURE 2.   Accuracy on the picture recognition task as a function of sensitivity to fast ART. The panels show raw data along with the LOESS line fitted. The x-axis shows fast ART discrimination thresholds, i.e., the smallest perceived differences in the Amplitude Rise Times measured in ms (a lower value is better). The y-axis shows the log-odds ratio of giving a correct response as predicted by the model. The first panel shows the interaction with accent and the second one shows the interaction with structure.
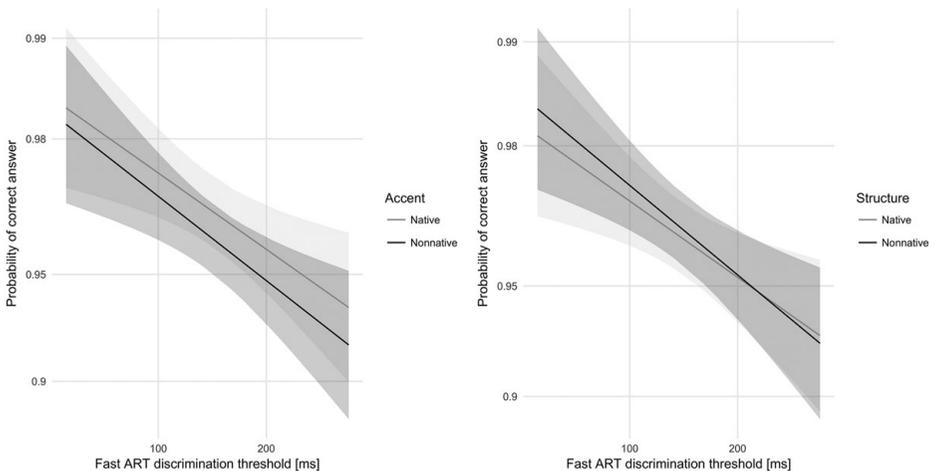

FIGURE 3.   Accuracy on the picture recognition task as a function of sensitivity to fast ART. The panels show the regression lines taken from the model along with 95% confidence intervals, marked with ribbons. The x-axis shows fast ART discrimination thresholds, i.e., the smallest perceived differences in the Amplitude Rise Times measured in ms (a lower value is better). The y-axis shows the log-odds ratio of giving a correct response as predicted by the model. The first panel shows the interaction with accent and the second one shows the interaction with structure.

TABLE 3.   Modeling the predictors of the picture recognition RTs: Random effects

Random variable	Name	Variance	SD
Participant	*Intercept*	0.03	0.18
Participant	Accent	0.00	0.05
Participant	Structure	0.00	0.03
Participant	Congruence	0.00	0.04
Participant	Accent × structure	0.00	0.04
Item	*Intercept*	0.00	0.03
Item	fast ART	0.00	0.01
Item	slow ART	0.00	0.01
Item	Simon effect	0.00	0.02
Item	Digit span	0.00	0.02
Residual		0.11	0.34

TABLE 4.   Modeling the predictors of the picture recognition RTs: Fixed effects

	Estimate	SE	*df*	*t* value	*p*
Intercept	**6.44**	**0.03**	**36.24**	**217.21**	**<2e−16*****
Simon effect	**0.07**	**0.03**	**34.41**	**2.20**	**0.035***
Digit span	**−0.08**	**0.03**	**33.96**	**−2.52**	**0.017***
Accent	**0.03**	**0.02**	**29.94**	**2.05**	**0.049***
Structure	−0.02	0.01	23.34	−1.48	0.151
Congruence	**0.12**	**0.01**	**37.27**	**15.29**	**<2e−16*****
fast ART	**0.07**	**0.03**	**33.52**	**2.04**	**0.049***
slow ART	−0.01	0.03	33.25	−0.21	0.837
Accent × structure	−0.01	0.03	21.24	−0.32	0.751
Accent × fast ART	0.01	0.01	39.34	0.99	0.327
Structure × fast ART	−0.01	0.01	28.65	−0.80	0.432
Accent × slow ART	0.00	0.01	35.57	−0.02	0.983
Structure × slow ART	0.00	0.01	29.27	0.25	0.801

Significant effects are bolded (* $p < .05$, ** $p < .01$, *** $p < .001$).

are presented in Tables 3 and 4. The results show that participants arrived at the correct decision faster if they had (a) a larger forward digit span, (b) smaller Simon effect, and (c) greater sensitivity (i.e., lower threshold) to fast ART. Incongruent trials elicited longer RTs than congruent trials and items pronounced with a nonnative accent elicited longer RTs than those pronounced with native accent. The collinearity in the model was appropriately low (VIF below 1.5). Figures 4 and 5 show the RT values as a function of sensitivity to fast ART. Figure 4 shows raw data with the LOESS line fitted, while Figure 5 shows the predictions of the model.

PICTURE RECOGNITION TASK: EEG DATA

We observed a negativity between 250 and 350 ms and peaking at around 300 ms, for the incongruent, compared to the congruent trials, which we identified as the N300 component (see Figure 6). Figure 7 presents the difference waves showing the effect of
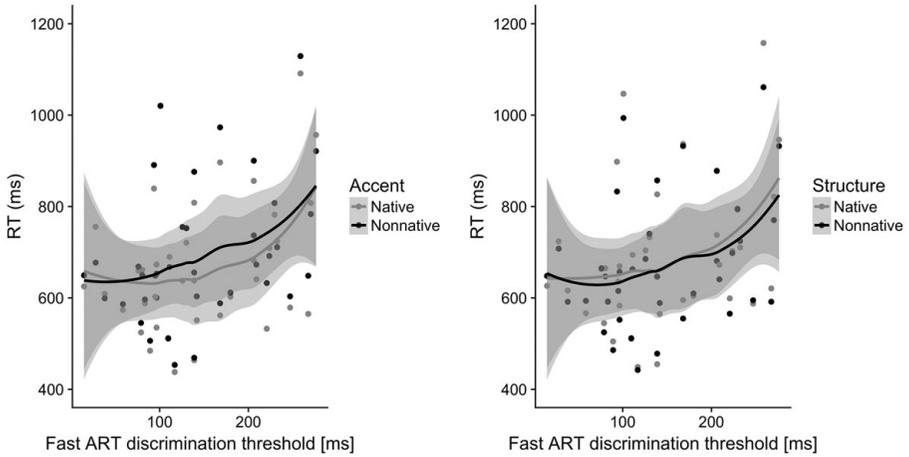
FIGURE 4. The RTs on the picture recognition task as a function of sensitivity to fast ART. The panels show raw data along with the LOESS line fitted. The x-axis shows fast ART discrimination thresholds, i.e., the smallest perceived differences in the Amplitude Rise Times measured in ms (a lower value is better). The y-axis shows the RT in ms. The first panel shows the interaction with accent and the second one shows the interaction with structure.
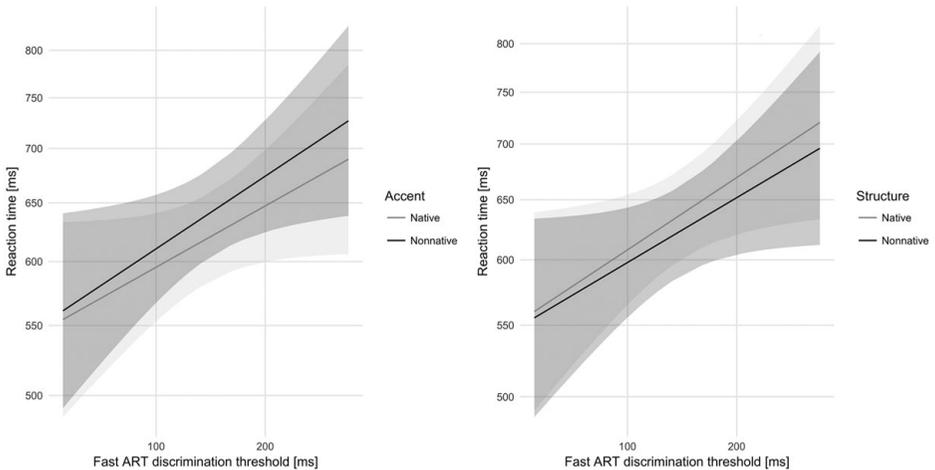

FIGURE 5. The RTs on the picture recognition task as a function of sensitivity to fast ART. The panels show the regression lines taken from the model along with 95% confidence intervals, marked with ribbons. The x-axis shows fast ART discrimination thresholds, i.e., the smallest perceived differences in the Amplitude Rise Times measured in ms (a lower value is better). The y-axis shows the RT in ms. The first panel shows the interaction with accent and the second one shows the interaction with structure.
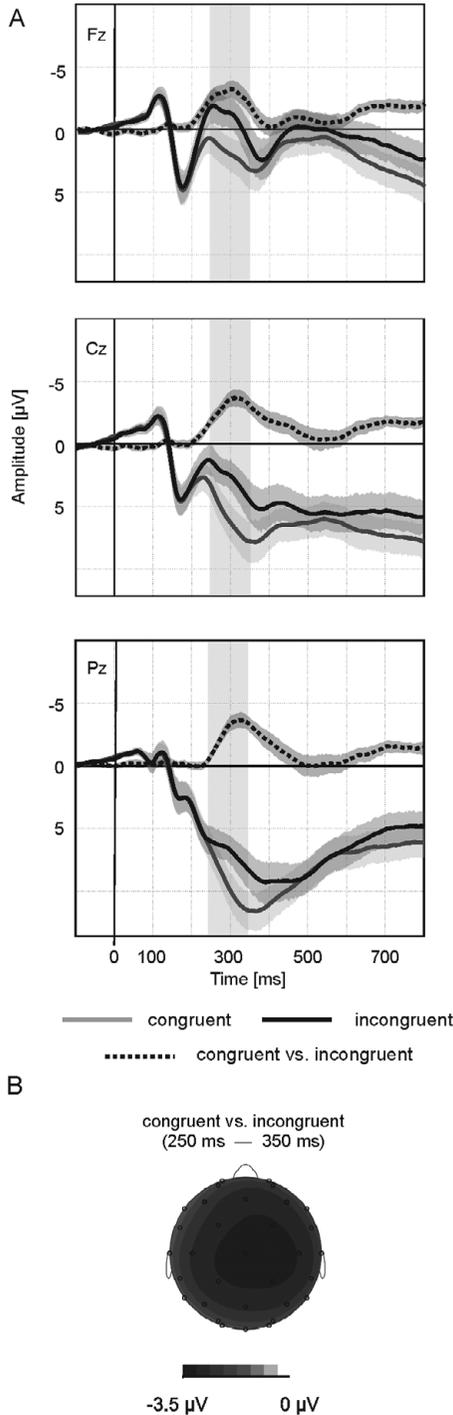
FIGURE 6. Stimulus-locked grand-averaged waveforms for congruent and incongruent trials at representative midline electrodes Fz, Cz, and Pz (top) with scalp potential difference maps for the N300 component (bottom). Confidence intervals are marked in gray. The shaded vertical stripe corresponds to the time-window of the N300 component.
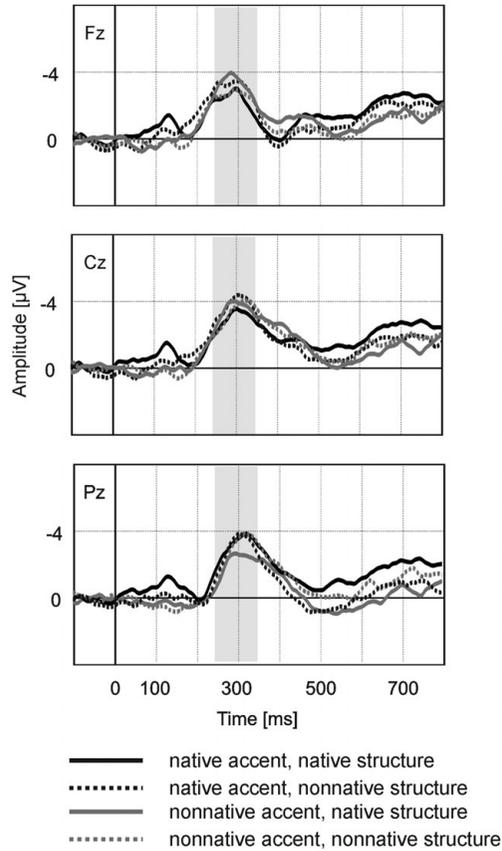
FIGURE 7. Comparison of the N300 effect (incongruent vs. congruent stimulus-locked grand-averaged waveform) across the different types of words at representative midline electrodes Fz, Cz, and Pz.

TABLE 5. Modeling the predictors of the N300 amplitude: Random effects

Random variable	Name	Variance	SD
Participant	Intercept	4.98	2.23
Participant	Accent	0.32	0.56
Participant	Structure	0.27	0.52
Participant	Congruence	1.88	1.37
Item	Intercept	0.08	0.28
Item	fast ART	0.02	0.16
Item	slow ART	0.06	0.23
Item	Simon effect	0.01	0.09
Item	Digit span	0.00	0.02
Item	Congruence × fast ART	0.23	0.48
Residual		66.99	8.18

TABLE 6.   Modeling the predictors of the N300 amplitude: Fixed effects

	Estimate	SE	df	t value	p
Intercept	**1.69**	**0.37**	**34.50**	**4.54**	**0.000*****
Simon effect	−0.26	0.38	33.15	−0.69	0.498
Digit span	**1.41**	**0.42**	**33.05**	**3.39**	**0.002****
Accent	−0.29	0.18	26.42	−1.65	0.110
Structure	0.18	0.17	25.80	1.04	0.310
Congruence	**−0.59**	**0.24**	**34.45**	**−2.42**	**0.021***
fast ART	−0.17	0.41	33.39	−0.41	0.684
slow ART	−0.78	0.42	33.80	−1.85	0.073
Congruence × fast ART	0.02	0.29	40.42	0.05	0.958
Congruence × slow ART	−0.33	0.27	34.48	−1.24	0.222
Accent × fast ART	−0.05	0.16	26.86	−0.29	0.771
Structure × fast ART	0.02	0.16	25.20	0.11	0.912
Accent × slow ART	0.14	0.18	28.48	0.82	0.419
Structure × slow ART	0.02	0.17	26.77	0.13	0.898
Accent × congruence × fast ART	−0.06	0.30	25.74	−0.21	0.836
Structure × congruence × fast ART	0.09	0.30	25.74	0.30	0.765
Accent × congruence × slow ART	0.23	0.22	25,670.00	1.08	0.282
Structure × congruence × slow ART	0.11	0.22	25,680.00	0.50	0.618

Significant effects are bolded (* $p < .05$, ** $p < .01$, *** $p < .001$).

congruence (incongruent–congruent) for the different word types—native in accent and structure; native in accent, but not structure; native in structure, but not accent; and nonnative in accent and structure. We analyzed the data for the fronto-central electrodes (Fz, FC1, FC2, Cz), in accordance with previous study using a similar paradigm (Mazerolle et al., 2007) so as to avoid double-dipping. However, it needs to be observed that the topography of our effect is more central than in the quoted study.

Tables 5 and 6 provide the random and fixed effects for model predicting the N300 amplitude in response to the pictures. The model met the no-collinearity assumption (VIF below 1.5).

Pictures congruent with the preceding nonword elicited the N300 component with a reduced amplitude, compared to pictures incongruent with the preceding nonword. Furthermore, participants with a longer digit span had a smaller (less negative) N300 amplitude across the board. Prior to the experiment we hypothesized that participants with better ART sensitivity (i.e., lower threshold) should have a smaller (i.e., more positive) N300 amplitude for the congruent pictures, compared to participants with lower ART sensitivity. Thus, we expected a positive and significant interaction of congruence and ART. This was, however, not the case: the effect of congruence was not modified by ART (see Figure 8).

DISCUSSION

In this article, we asked whether increased sensitivity to basic auditory information is related to greater accuracy and better-quality word representations in language learning. We focused on sensitivity to a single auditory parameter, namely ART, and investigated
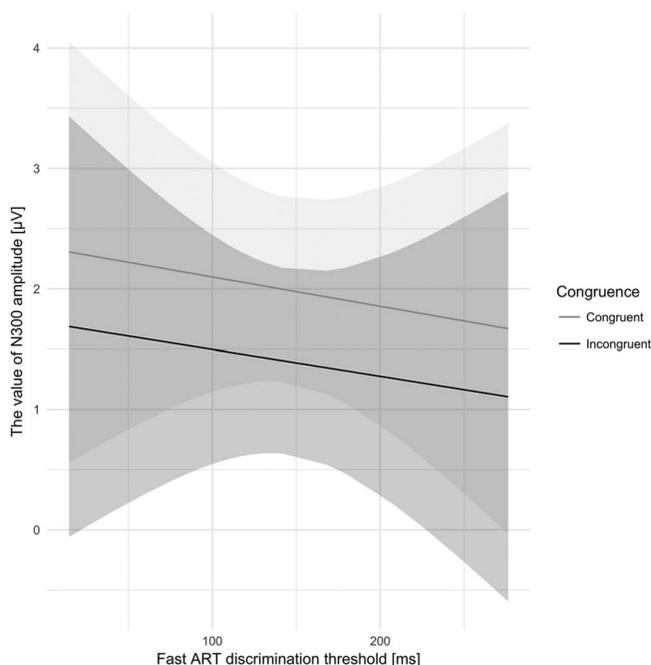
FIGURE 8.   The lack of interaction between the type of trial (congruent vs. incongruent) and sensitivity to fast ART in the picture recognition task. The lines indicate the values of the N300 amplitude (in microvolts) for congruent (gray) and incognruent (black) trials as a function of sensitivity to fast ART. The ribbons represent the 95% confidence intervals.

whether it predicts word learning, especially for words that have a foreign phonological structure and accent. We exposed participants to nonword–picture pairs and then tested them on their knowledge of the pairs with a PR task. We hypothesized that greater sensitivity to ART would predict better accuracy in word learning and a better quality of the representations of the learned words: participants with smaller thresholds for detecting differences in ART would show higher accuracy and shorter reaction times in the PR, as well as more reduced N300 in response to congruent pictures. We predicted that sensitivity to ART would particularly improve learning of novel words with unfamiliar phonological structure and accent, that is, those that require participants to use their universal segmentation mechanism. Additionally, we explored what rate of ART—slow (connected with segmentation into syllables) or fast (connected with segmentation into phonemes)—would be more strongly associated with word learning. There was no prior evidence allowing us to make strong predictions with regard to this question.

Our analyses showed that sensitivity to fast rather than slow ART correlated with improved word learning accuracy. In line with the predictions, sensitivity to fast ART predicted accuracy and reaction times on the PR task. However, contrary to our expectations, the sensitivity to ART was equally predictive for learning nonnative and native

nonwords. Moreover, the relationship between the sensitivity to ART and word learning was only visible in behavioral measures. The amplitude of the N300 component at the target picture was not affected by the individual sensitivity to ART.

SENSITIVITY TO ART IS CONNECTED WITH BETTER PERFORMANCE ON BEHAVIORAL INDICES OF WORD LEARNING

So far, the relationship between sensitivity to ART (or any other single auditory characteristic) and novel vocabulary learning hypothesized only in studies of dyslexic individuals and children with SLI (Alt et al., 2017; Cumming et al., 2015; Kalashnikova & Burnham, 2016; Kwok & Ellis, 2014). Furthermore, only two studies that tested this relationship directly, but in both, samples of typically developing participants were mixed in with dyslexic or SLI children (Corriveau et al., 2007; Thomson & Goswami, 2009). Our study is the first one to show that sensitivity to ART is important not only for reading and word processing in developmentally delayed children but also continues to be important for word learning in typically developed adults who are already proficient in at least one language.

WORD LEARNING TASK IS INFLUENCED BY SENSITIVITY TO FAST RATHER THAN SLOW ART

The performance on the PR task in our study was connected with sensitivity to fast, but not slow ARTs. Previous research suggested that sensitivity to slow amplitude cues such as slow ART might be connected with the ability to segment speech into syllables, while sensitivity to fast amplitude change, such as fast ART might be important for discriminating phonemes and phoneme boundaries (Goswami, 2011; Hämäläinen et al., 2005; McAnally & Stein, 1996). Our data suggest that in typically developing adult population, sensitivity to fast ART might be of greater importance for word learning than sensitivity to the slow ART.

THE SENSITIVITY TO FAST ART FACILITATES LEARNING BOTH NATIVE AND NONNATIVE WORDS

In this article we hypothesized that sensitivity to ART influences word learning by facilitating a universal segmentation mechanism. According to our theory (based on Marecka et al., 2018), this mechanism enables segmentation of the word form, particularly for words that have an unfamiliar phonological structure. We hypothesized that segmentation of words with familiar structure is done using a different segmentation mechanism—phonological mapping, which relies on sublexical phonological representations, that is, representations of syllables, phonemes, and ngrams (speech chunks). If this hypothesis were correct, we would see a greater effect of sensitivity to ART on learning words with nonnative rather than native accent or structure.

However, our results do not provide support for the hypothesis—sensitivity to fast ART predicted the acquisition of both familiar and unfamiliar novel word types to the same extent. There are two possible explanations for this finding. The first one is that ART

helps in word segmentation regardless of whether the word is native or nonnative. This would mean that there is a single segmentation mechanism that is sensitive to language-universal auditory cues such as ART.

The alternative explanation is that the theory of two segmentation mechanism is correct, but we do not see evidence for it in our data due to potential methodological problems. One of them is that our task did not force participants to focus on learning word forms in a greater detail. Because the nonwords acquired in the study were quite distinct, participants could perform the task successfully by making very coarse distinctions between these nonwords. Consequently, participants may have not learned the nonwords in very fine-grained detail. It is possible that if the participants were tested with a task that required a more detailed knowledge of the word form—for example, a production task or a task where participants needed to choose a correct form from a set of similar nonwords—the effect of nativeness would be present.

Whatever the explanation ultimately is, our study suggests that sensitivity to ART supports word learning at all stages of word acquisition, even when learning words with a well-known phonological structure.

ERP INDICES OF WORD LEARNING ARE NOT SENSITIVE TO THE EFFECT OF ART

While we found the effect of sensitivity to ART on the behavioral measures—that is, the overall word learning efficiency—we did not see it in our EEG data. The difference in the N300 amplitude evoked by the congruent and incongruent pictures was not modulated by participants' sensitivity to fast ART. There are three possible explanations of this result, however because there is no previous data on this topic, all of them are speculative at this point.

The first explanation is that in all learners, irrespective of their sensitivity to fast ART, presentation of the nonword fully activated the associated concept. In the study we assumed that size of N300 effects would depend on the strength of conceptual activation. However, present data suggest that presentation of a (non)word may activate the corresponding concept in a binary fashion (i.e., the concept is either fully activated or not activated) and there is no gradation in its strength that could correlate with fast ART. This would mean that, contrary to our initial assumptions, the N300 amplitude on the congruent pictures cannot be considered an index of conceptual activation.

The second explanation is that the N300 component can capture changes in word learning only at the very initial stages of acquisition and our measurement of the learning processes occurred rather late (after most of the participants had learned the words very well). With this account, we would expect to see more variability in the N300 for the congruent condition, if the participants were tested at the beginning of learning, when the link between the word form and meaning was still volatile. In light of this conclusion, it would be beneficial to conduct a further study that would use a much larger number of items (as to make the task more difficult) and to investigate the ERPs at the very beginning of the learning process.

The third explanation is that the N300 and the behavioral indices are to some extent dissociable and rely on different cognitive processes and sources of information. Hamm et al. (2002) suggested that N300 might reflect presemantic processing of pictures, related to the categorization of objects. The authors claim that this process is dissociable from semantic processing. Even though there is some evidence against that claim (Draschkow

et al., 2018), it is possible that the N300 captures an aspect of conceptual knowledge that does not fully reflect word learning efficiency or that cannot be graded based on individual differences measured behaviorally.

ADDITIONAL FINDINGS

In all our models we controlled for phonological STM (measured by digit span) and resistance to interference (measured by the Simon task). We found that both those measures influenced the behavioral results. Participants who had a longer digit span were more accurate and faster on the PR task. This confirms previous findings that better phonological STM tasks predict better performance on novel word learning tasks and greater vocabulary size in L1 and L2 learners (e.g., Farnia & Geva, 2011; Gathercole & Baddeley, 1989, 1990; Masoura & Gathercole, 1999; Papagno & Vallar, 1995; Service & Kohonen, 1995).

Furthermore, participants who exhibited a smaller Simon effect (i.e., were more resistant to interference) were also more accurate and faster on the PR task. This suggests that learning words of a different language requires efficient ability to resist interference—a domain general aspect of cognitive control (e.g., Friedman & Miyake, 2004; Rey-Mermet & Gade, 2018). A possible explanation for this finding is that when learners acquire novel word forms for concepts that already have names in their L1, they need to ignore the interference from the L1 form. Our results contribute to research on elementary cognitive process influencing second language learning and to the field that investigates relation between bilingualism and cognitive control (e.g., Bartolotti et al., 2011; Blumenfeld & Marian, 2011; Teubner-Rhodes et al., 2016).

CONCLUSIONS

The presented research is the first study to that detecting fast ART can support vocabulary acquisition. In particular, we propose that sensitivity to ART allows for better segmentation of the word form, which translates to better encoding. Our finding paves the way for future studies exploring the elusive "ear for language"—the relationship between auditory skills and word learning. It offers a glimpse into the issue of language talent, by showing the possible sources of the difference between good and bad language learners. Finally, it suggests a possibility that training basic auditory skills such as sensitivity to ART can improve vocabulary learning in people with second language learning problems.

NOTES

[1]It should be noted, however, that this result was not found for the Greek children in the study by Georgiou et al. (2010).

[2]For example, for the sequence "#bask#," where # signifies a beginning or end of the word, the sequences would be: "#b," "as," "ba," "sk," "k#," "#ba," "bas," "ask," "sk#," "#bas," "bask," "ask#," "#bask," "bask#."

[3]It is theoretically possible that second language experience enhances ART discrimination, so before we ran the following analyses, we also calculated correlations between fast ART discrimination scores and L2 age of acquisition as well as between fast ART discrimination and L2 experience (calculated as age minus L2 age of acquisition). Neither factor explained significant variance in ART (fast ART and age of acquisition: $r = -0.13$, $p = 0.427$, $-95\%\ CI\ [-0.44, 0.20]$; fast ART and L2 experience: $r = -0.05$, $p = 0.744$, $-95\%\ CI\ [-0.37, 0.27]$). Thus, at least in this sample, there is no evidence for a relationship between second language experience and ART.

REFERENCES

Alt, M., Hogan, T., Green, S., Gray, S., Cabbage, K., & Cowan, N. (2017). Word learning deficits in children with dyslexia. *Journal of Speech, Language, and Hearing Research*, 60, 1012.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, 35, 126–140.

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, 2, 234.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. ArXiv:1506.04967 [Stat]. http://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19, 290–303.

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition*, 118, 245–257.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8, 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, 116, 289–296.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology*, 63, 44–78.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, 22, 441–469.

Brzeziński, J., Gaul, M., Hornowska, E., Jaworowska, A., Machowski, A., & Zakrzewska, M. (2004). *WAIS-R (PL)—Skala Inteligencji Wechslera dla Dorosłych—Wersja Zrewidowana*. Pracownia Testów Psychologicznych.

Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*. Psychological Corporation.

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research*, 50, 647–666.

Cumming, R., Wilson, A., Leong, V., Colling, L. J., & Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments. *Frontiers in Human Neuroscience*, 9, 200–221.

de Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. Psychology Press.

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34, 1443–1449.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, 28, 1584–1602.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience*, 11, 111.

Draschkow, D., Heikel, E., Võ, M. L. H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, 120, 9–17.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, 61, 177–199.

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, 32, 711–738.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116–124.

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21, 2357–2365.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*, 2038–2043.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*, 123–144.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200–213. http://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439–454.

Georgiou, G. K., Protopapas, A., Papadopoulos, T. C., Skaloumbakas, C., & Parilla, R. (2010). Auditory temporal processing and dyslexia in an orthographically consistent language. *Cortex*, *46*, 1330–1344.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*, 3–10.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, *99*, 10911–10916.

Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*, 325–337.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*, 1339–1350.

Hämäläinen, J., Leppänen, P. H. T., Torppa, M., Müller, K., & Lyytinen, H. (2005). Detection of sound rise time by adults with dyslexia. *Brain and Language*, *94*, 32–42.

Hu, C. -F. (2003). Phonological memory, phonological awareness, and foreign language word learning. *Language Learning*, *53*, 429–462.

Hu, C. -F. (2008). Rate of acquiring and processing L2 color words in relation to L1 phonological awareness. *The Modern Language Journal*, *92*, 39–52.

Hu, C. -F., & Schuele, C. M. (2005). Learning nonnative names: The effect of poor native phonological awareness. *Applied Psycholinguistics*, *26*, 343–362.

Jones, G. (2016). The influence of children's exposure to language from two to six years: The case of nonword repetition. *Cognition*, *153*, 79–88.

Jones, G., & Witherstone, H. L. (2011). Lexical and sublexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition*, *39*, 588–599.

Jung, T. -P., Makeig, S., Lee, T. -W., McKeown, M. J., Brown, G., Bell, A. J., & Sejnowski, T. J. (2000). Independent component analysis of biomedical signals. In *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation* (pp. 633–644).

Kalashnikova, M., & Burnham, D. (2016). Novel word learning, reading difficulties, and phonological processing skills. *Dyslexia*, *22*, 101–119.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest. R package version 2.0.

Kwok, R. K. W., & Ellis, A. W. (2014). Visual word learning in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*, 583.

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). Subtlex-pl: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, *47*, 471–483.

Marecka, M., Szewczyk, J., Jelec, A., Janiszewska, D., Rataj, K., & Dziubalska-Kołaczyk, K. (2018). Different phonological mechanisms facilitate vocabulary learning at early and late stages of language acquisition: Evidence from Polish 9-year-olds learning English. *Applied Psycholinguistics*, *39*, 1–35.

Marzecová, A., Asanowicz, D., Krivá, L., & Wodniecka, Z. (2013). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition*, *16*, 608–623.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology*, *34*, 383–388.

Mazerolle, E. L., D'Arcy, R. C. N., Marchand, Y., & Bolster, R. B. (2007). ERP assessment of functional status in the temporal lobe: Examining spatiotemporal correlates of object recognition. *International Journal of Psychophysiology*, *66*, 81–92.

McAnally, K. I., & Stein, J. F. (1996). Auditory temporal coding in dyslexia. *Proceedings of the Royal Society of London B: Biological Sciences*, *263*, 961–965.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*, 3–19.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*, 507–517.

Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: Evidence from French. *NeuroReport*, *15*, 1255–1259.

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.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., & Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*, 509–521.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, *5*, 1–15.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*, 98–107.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1281–1292.

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*, 1695–1716.

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, *10*, 215–233.

Schendan, H. E. (2019). Memory influences visual cognition across multiple functional states of interactive cortical dynamics. In K. *Federmeier (Ed.),* The psychology of learning and motivation (pp. 303–386). Academic Press.

Schendan, H. E., & Ganis, G. (2015). Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions. *Frontiers in Psychology*, *6*, 1289.

Schendan, H. E., & Stern, C. E. (2007). Where vision meets memory: Prefrontal–posterior networks for visual object constancy during categorization and recognition. *Cerebral Cortex*, *18*, 1695–1711.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*, 155–172.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*, 1321–1337.

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, *46*, 1312–1323.

Surányi, Z., Csépe, V., Richardson, U., Thomson, J. M., Honbolygó, F., & Goswami, U. (2008). Sensitivity to rhythmic parameters in dyslexic children: A comparison of Hungarian and English. *Reading and Writing*, *22*, 41–56.

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes. Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36.

Teubner-Rhodes, S. E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition*, *150*, 213–231.

Thomson, J. M., & Goswami, U. (2009). Learning novel phonological representations in developmental dyslexia: Associations with basic auditory processing of rise time and phonological awareness. *Reading and Writing*, *23*, 453–473.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Brain Research Cognitive Brain Research*, *13*, 363–375.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., & Woo, K. (2018). ggplot2. R package version 3.1.0

APPENDIX

LIST OF STIMULI

Native structure words

Word	IPA transcriptions	Mean wordlikeness (questionnaire)	Phonootactic probability (all.ngrams)	n-coltheart	Neighborhood density (pld20)
oklas	/ˈoklas/	3.95	4.66	2	1.90
wyska	/ˈviska/	4.37	5.33	6	1.70
jesty	/ˈjesti/	3.89	6.13	3	1.85
zarki	/ˈzarki/	3.74	4.89	8	1.60
tenty	/ˈtenti/	4.29	5.55	6	1.70
pszego	/ˈpʃego/	3.94	6.45	2	1.90

Nonnative structure words

Word	IPA transcriptions	Mean wordlikeness (questionnaire)	Phonootactic probability (all.ngrams)	n-coltheart	Neighborhood density (pld20)
czułfi	/ˈt͡ʃuwfi/	2.75	2.90	0	2.00
gaszna	/ˈgaʃna/	3.42	3.45	0	2.00
uchekt	/ˈuxekt/	2.15	3.55	0	2.50
rzołse	/ˈʒowse/	2.85	3.14	1	1.95
izbur	/ˈizbur/	2.20	3.52	1	1.95
cerdze	/ˈt͡serd͡ze/	3.56	3.88	1	1.95