Sensitive periods in language development: Do children outperform adults on auditory word-form segmentation?

Authors	Smalle,E.; Bogaerts,L.
Published in	Cortex
DOI	10.1016/j.cortex.2024.07.001
Publication Date	2024
Document Version	publishersversion
Link	https://research.tilburguniversity.edu/en/publications/ac318911-e924-421f-8e5b-0a74e5ec48ab
Citation	Smalle, E & Bogaerts, L 2024, 'Sensitive periods in language development : Do children outperform adults on auditory word-form segmentation?', Cortex, vol. 179, pp. 35-49. https://doi.org/10.1016/j.cortex.2024.07.001
Download Date	
Rights	

CORTEX 179 (2024) 35–49

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Sensitive periods in language development: Do children outperform adults on auditory word-form segmentation?

Eleonore H.M. Smalle [a,b,*] *and Louisa Bogaerts* [a]

[a] *Department of Experimental Psychology, Ghent University, Belgium*
[b] *Department of Developmental Psychology, Tilburg University, the Netherlands*

ARTICLE INFO

Article history:
Received 16 May 2024
Reviewed 20 June 2024
Revised 1 July 2024
Accepted 1 July 2024
Action editor Valentina Borghesani
Published online 27 July 2024

Keywords:
Language development
Sensitive periods
Auditory statistical learning
Word segmentation
Implicit and explicit memory

ABSTRACT

Children are more successful language learners than adults, yet the nature and cause of this phenomenon are still not well understood. Auditory statistical learning from speech has been a prominent focus of research in the field of language development because it is regarded as a fundamental learning mechanism underlying word segmentation in early language acquisition. However, a handful of studies that investigated developmental trajectories for auditory statistical learning found no clear child advantages. The degree to which the learning task measures explicit rather than implicit mechanisms might obscure a potential advantage for younger learners, as suggested by recent findings. In the present study, we compared children aged 7–12 years and young adults on an adapted version of the task that disentangles explicit and implicit contributions to learning. They were exposed to a continuous stream of speech sounds comprising four repeating trisyllabic pseudowords. Learning of the hidden words was tested (a) online through a target-detection task and (b) offline via a forced-choice word recognition test that included a memory judgement procedure. Both measures revealed comparable learning abilities. However, children's performance on the recognition task showed evidence for both explicit and implicit word knowledge while adults appeared primarily sensitive to explicit memory. Since implicit memory is more stable in time than explicit memory, we suggest that future work should focus more on developmental differences in the nature of the memory that is formed, rather than the strength of learning, when trying to understand child advantages in language acquisition.

---

[*] *Corresponding author*. Department of Developmental Psychology, Tilburg University, Prof. Cobbenhagenlaan 225, Office 303, 5037 DB Tilburg, the Netherlands.

E-mail address: e.h.m.smalle@tilburguniversity.edu (E.H.M. Smalle).

https://doi.org/10.1016/j.cortex.2024.07.001

1. Introduction

A constant observation in the study of language acquisition is that learning abilities decline with age (Johnson & Newport, 1989; Lenneberg, 1967; Newport, 2018; Werker & Hensch, 2015). Children before age twelve achieve native-like fluency in first or second languages with seemingly less effort than adults, particularly for formal sequential properties of language such as phonology (e.g., Newport, Bavelier, & Neville, 2001). This suggests that there is a time-restricted window in childhood with a maximal opportunity for acquiring language, also referred to as a critical period (e.g., Thiessen, Girard, & Erickson, 2016). Today, most researchers favor the term *sensitive period*, given that one does not ‘lose’ the preparedness for acquiring language outside the developmental period; although learning happens with more effort and ends in less ultimate attainment at an adult age (Hakuta, Bialystok, & Wiley, 2003). Moreover, the age of the learner doesn't uniformly influence all aspects of language. For example, the acquisition of phonology is more restricted by a learner's age than the acquisition of meaning, and some formal properties of grammar such as basic word order are unrestricted (e.g., Newport et al., 2001).

Developmental advantages for language learning remain a puzzle, especially if one considers that it is adults who excel over children on most other measures of cognitive performance (Craik & Bialystok, 2006). In 2005, the journal Science put forward the sensitive period question as one of the most fundamental but still unresolved paradoxes in human science (Kennedy & Norman, 2005). To this day, the nature and causes of early advantages in language learning are still not well understood (Hartshorne, 2022).

1.1. *Statistical learning and early language acquisition*

One of the first challenges in language development is learning the forms of words, or phonotactics, by listening to speech (Swingley, 2009). When listening to an unfamiliar language spoken by a native speaker of that language, segmenting novel continuous speech into meaningful word units is not an easy task. Speech signals often do not contain reliable acoustic cues to word boundaries (Klatt, 1979), which can challenge early word-segmentation processes and, unavoidably, later vocabulary acquisition (Gaskell & Ellis, 2009; Swingley, 2009). Despite this challenge, most infants listening to continuous speech learn to understand a significant number of novel words within the first year of life (Bergelson & Swingley, 2012; Swingley, 2009; Werker & Hensch, 2015). Such a fast trajectory of learning is unlikely to be explained by instruction alone. In 1996, Saffran and colleagues demonstrated that 8-month-old infants can segment unfamiliar, artificial speech into word units after only 2 min of passive listening (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). They do this by unconsciously relying on statistical cues that are present in the speech signal (See also, Pelucchi, Hay, & Saffran, 2009, for similar observations when listening to natural speech). Transitional probabilities between sounds are higher within words than across words and thus can help define where a (novel) word begins and ends. This early process for word segmentation from continuous speech is referred to as speech-based statistical learning. It also supports later stages of language acquisition, such as syntax development and the mapping of sound to meaning (Estes, Evans, Alibali, & Saffran, 2007; Romberg & Saffran, 2010).

1.2. *The developmental trajectory of speech-based statistical learning*

The ability to track statistical regularities in speech, and discover novel word boundaries, has also been demonstrated in newborns, indicating that it is a very early developed learning mechanism in the brain (Fló et al., 2019; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). Older children and (older) adults, too, appear capable of tracking statistical regularities in speech (Frost, Armstrong, & Christiansen, 2019; Palmer, Hutson, & Mattys, 2018; Raviv & Arnon, 2018; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). These findings suggest that the mechanism(s) underlying the learning of statistical structure in speech presumably remain continuously available across the human life span (see also, Thiessen et al., 2016). This does not necessarily mean that learning is age-invariant (Forest, Schlichting, Duncan, & Finn, 2023). For instance, discussion arises on whether statistical learning in adults still operates solely implicitly (Christiansen, 2019; Rebuschat & Monaghan, 2019). We argue here that the study of whether and how speech-based statistical learning changes over life span can provide insights into long-standing questions like the sensitive period hypothesis for language acquisition (Forest, Schlichting, et al., 2023; Thiessen et al., 2016).

One of the reasons why we could expect age-related differences in nature of statistical language learning and, thus, language acquisition, is that children have a less developed cognitive system than adults and thus may acquire better sensitivity to the hidden statistical structures via spontaneous implicit learning processes which are more stable in time as opposed to explicit learning processes (Liu, Forest, Duncan, & Finn, 2023). Recent work showed that higher developed cognitive abilities such as attentional control and declarative memory *constrain* speech-based statistical learning abilities in adults (Smalle & Möttönen, 2023). For instance, the depletion of cognitive resources in adults via brain stimulation or via the induction of cognitive fatigue enhances their abilities for unconsciously extracting phonotactic regularities or word boundaries in artificial speech (Smalle et al., 2021, 2022; Smalle, Panouilleres, Szmalec, & Möttönen, 2017). Henceforth, a less developed cognitive system could make children better implicit statistical learners, benefitting their language acquisition relative to adults. Detecting statistical dependencies benefit from bottom-up and model-free processing whereby correlations between elements can be discovered from the raw signal (e.g., Finn, Lee, Kraus, & Hudson Kam, 2014; Lum et al., 2023; see also Janacsek et al., 2012). Such an explanation accords well with the less-is-more hypothesis that attributes developmental advantages in early language acquisition, such as phonology or grammar, to limited attention and memory capacities (Newport, 1990).

Only a handful of studies directly looked into developmental trajectories for speech-based statistical learning abilities and found no clear age-differences in support of a sensitive period for language development (i.e., no learning advantages for infants over adults, Choi, Batterink, Black, Paller, & Werker, 2020; for younger children over older children, Raviv & Arnon, 2018; Shufaniya & Arnon, 2018; Finn, Kharitonova, Holtby, & Sheridan, 2019; or for children over adults, Saffran et al., 1997; Moreau, Joanisse, Mulgrew, & Batterink, 2022). In most of these studies, statistical learning was measured in its typical way by using a post-learning word recognition task (e.g., Raviv & Arnon, 2018; Saffran et al., 1997; Shufaniya & Arnon, 2018). In a typical statistical-learning experiment, participants are repeatedly exposed to patterned stimuli such as recurrent syllable triplets in a continuous speech stream. They are asked to simply listen to the sounds, the hidden pattern repetitions are not announced, and thus participants are exposed to the language structure without intention to learn (i.e., incidental learning) (e.g., Saffran et al., 1997). Learning performance is then assessed post-exposure by means of a two-alternative forced-choice recognition task in which responses to previously encountered triplets (‘words’) versus foils (non-words, a sequence of syllables that did not occur together in the exposure) are measured. Participants indicate via preference methods such as head-turning (in infants) or by indicating their choice (in older children and adults), which of the two syllable sequences sounds more familiar. Above-chance accuracy on this task is taken as an indication of learning.

Although responses on a post-exposure word recognition test can yield significant insights into auditory statistical learning abilities, it has several limitations when aiming for meaningful comparisons across age. First, word recognition reflects the outcome of learning but not the *time course* of the learning, this is, the processing of the underlying statistical structures. Second, and perhaps most important, the forced-choice recognition task requires participants to overtly reflect on what they have learned and thus relies at least partly on explicit decision making and memory retrieval abilities, which are not yet fully developed in children (Arnon, 2020; Kidd et al., 2020). It is often wrongly assumed that performance on the offline word recognition task reflects implicit learning performance simply because the learning during exposure happened without explicit instruction. Recent evidence by Batterink and colleagues demonstrated that, even without the intention to learn, offline recognition performance reflects explicit rather than implicit knowledge of the novel word-forms (Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015). This was demonstrated by adding a memory judgement procedure to the task. After each word recognition response, participants were asked to indicate their level of confidence in remembering the word from the exposure. If word recognition reflects solely explicit knowledge, accuracy is expected to be highest when experiencing detailed recollection and not better than chance when guessing. On the other hand, if participants still perform above chance when they indicate they are guessing, knowledge is inferred to also partly reflect implicit knowledge (i.e., for some items they are not consciously aware of what they have learned; Dienes & Berry, 1997).[1] In a couple of studies with adult participants, Batterink and colleagues showed that task performance is above chance when remembering the triplet (i.e., when confidence is high) but at chance when not remembering the triplet (i.e., when confidence is low), indicating that their performance on this task primarily reflects explicit knowledge (Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015).

Today, little is known about what type of word knowledge, or memory, children develop during exposure with structured auditory syllables. A recent study comparing performance on the reflection-based two-alternative forced-choice recognition task with performance on a more processing-based measure of statistical learning (i.e., via serial recall of sequences embedded during the exposure) showed that, in young children, the latter is more sensitive and more reliable (Kidd et al., 2020). This suggests that studies with children that operationalized learning solely as recognition test performance—which is indeed what most of the published studies with children did—might have underestimated children's word-form learning abilities. In addition, even if children and adults do not differ in terms of statistical processing, they might nevertheless differ in the mechanism(s) they recruit to support the learning or the knowledge that they gain, such that children form more implicit memory representations of the regularities. This is an important question to ask because implicit memory is more robust and long-lived than explicit memory which fades quickly in the absence of recall (Baars, 1995; Dienes & Berry, 1997; see also Liu et al., 2023, for direct evidence in the statistical word-segmentation task). Children's use of implicit rather than explicit memory could therefore help explain why children eventually attain better language-learning outcomes than adults (see also, Smalle & Möttönen, 2023).

Recent work started to characterize other measures of auditory statistical learning and directly compared these between children and adults. For instance, Choi et al. (2020) studied when preverbal infants, as compared to adults, become sensitive to the statistical cues in the speech signal by relying on an electroencephalogram (EEG) index of neural entrainment to the words in the speech signal (Batterink, 2017). They found that infants and adults follow similar learning trajectories. Very recently, Moreau et al. (2022) compared English-speaking children aged 8 to 12 with adults on this EEG-index and found slightly stronger entrainment to the word structures in adults than in children – potentially reflecting better attention or auditory processes. They additionally implemented other offline measures of learning such as word familiarity rating and target detection, for which no significant age differences were observed. In a supplementary file, they present comparative data on memory judgements during the typical word recognition task. At each word recognition trial, participants were asked to indicate which of

---

[1] The complete absence of a correlation between memory and accuracy would indicate the absence of explicit learning. This is also referred to as the zero-correlation criterion (Dienes & Berry, 1997).

two words sound more familiar to the exposure and their level of confidence for their choice, i.e., high confidence (remembering the word), medium confidence (familiarizing with the word without clear memory) or low confidence (guessing). Interestingly, they reported (in their supplementary) that while both children and adults performed equally above chance on confident trials, only the children also performed above chance on unconfident trials. This pattern of results indicates that children formed implicit knowledge of the words, suggesting they rely more heavily on additional implicit learning mechanisms than adults.

1.3. *Present study*

In the present study, we aimed to directly compare children and adults on a speech-based auditory statistical learning task, also taking into account the above-mentioned methodological concerns and recent findings by Moreau et al. (2022). *First*, we attempted a child-adult comparison on *online* word-segmentation abilities. This was accomplished behaviorally by implementing a target detection task during the speech exposure. Whereas in Moreau et al. (2022), this task was implemented indirectly, i.e. after the exposure, we aimed to implement it directly during the exposure, i.e. while listening to the repeating syllable sequences for the first time. In a target detection task, participants are required to press a button whenever they hear a given sound in the stream. The target sounds are either at a predictable position in the speech stream (i.e., they are the second or third syllable of an embedded triplet in the stream), or at an unpredictable position (i.e., the first syllable). If participants learn the underlying word structure in the speech signal, they should become increasingly faster in detecting targets at predictable positions than at unpredictable positions (hereafter, RT facilitation effect). Previous studies implementing the target detection task (albeit offline) showed that the RT facilitation effect did not correlate with scores on the typical forced-choice recognition task (Batterink, Reber, Neville, & Paller, 2015; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015), suggesting that performance on the target detection task reflects a different, processing-based memory process (Christiansen, 2019). In support of a child advantage for language learning, we predict here that (1) children show an enhanced RT facilitation effect compared with adults, and (2) perhaps show this effect already at an earlier time point during exposure than adults. Alternatively, and in line with recent findings using other measures for online statistical learning (i.e., EEG-based neural entrainment), we predict no differences (e.g., Choi et al., 2020) or even a stronger learning effect in adults (e.g., Moreau et al., 2022). The latter prediction accords with observations in natural language learning environments where it is often seen that in both instructional and immersion contexts, language acquisition sets off more quickly with increasing age, presumably because of a stronger reliance on conscious strategies (Huang, 2016). *Second*, we aimed to compare children and adults on the typical offline, i.e., post-exposure, forced-choice word recognition task but, similarly to Moreau et al. (2022), implement a memory judgement procedure to dissociate between implicit and explicit memory. In support of a child advantage for auditory statistical learning that is specific to implicit memory, we predict that (3) children show enhanced (i.e., above-chance) word recognition compared to adults when confidence is low (i.e., when guessing). We predict no child advantage, or even better performance in adults, for the explicitly remembered words (i.e., when confidence is high).

2. Method

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. *Participants*

Based on the smallest effect size for a group difference on statistical word recognition with memory judgement, as seen in Smalle et al. (2022), we estimated a required sample size of 36 participants per group (independent samples *t*-test, $d = .6$, alpha $= .05$, power $= .80$, calculated using G*Power: Faul, Erdfelder, Lang, & Buchner, 2007). The size of a group-dependent learning effect for the online target detection task is unknown. We decided to recruit as many children between 6 and 12 years old[2] as possible within the foreseen period and match the sample size for the adults. We report the estimated group effect sizes on the online target detection task and the precision given our actual sample size in the result section. In total, 100 participants (50 participants per group) completed the experiment but 12 participants (6 in each age group) were excluded from the analyses because of equipment error resulting in missing data ($n = 1$), failure to follow task instructions ($n = 2$), because they reported having a developmental learning disorder ($n = 3$) or mother tongue other than Dutch ($n = 1$), or because they scored 1.5SD below the norm of their age group on the Dutch Peabody Picture Vocabulary Test, version 3 ($n = 6$). This latter is a test of prior vocabulary knowledge and known to affect statistical learning abilities in artificial speech (e.g., Stärk, Kidd, & Frost, 2022). Altogether, 44 primary school-aged children and 44 young adults were included in the final dataset. All 88 included participants had Dutch as their native language. Descriptive statistics of the remaining characteristics are provided in Table 1. Children were recruited by contacting local primary schools and organizations. Adults were recruited via social media and/or student recruitment platforms at the university.

---

[2] This wide age range for the children group allowed for a larger sample size and was also theoretically motivated as most estimates for a sensitive age in language learning range between 2 and 12 years of age (e.g., Bialystok & Hakuta, 1994; Singleton, 2006). Six years was chosen as the minimum age because of practical reasons (i.e., avoiding difficulties in understanding and keeping attention to the tasks). Twelve years was chosen as maximum age because higher cognitive abilities (including explicit retrieval abilities) expand rapidly after this age (Craik & Bialystok, 2006). Although the study was not designed for this purpose, the wide age range in the child group allowed for exploratory analyses with developmental age as continuous predictor. No effects of age within the child group were found on the learning measures (supplementary materials, S2).

Table 1 – Participant characteristics.

	Children (N = 47, 32 female)		Adults (N = 47, 37 female)		Comparison		
	Mean (SD)	Range	Mean (SD)	Range	*t*	*p*	*Cohen's d*
Age in years	8.9 (1.0)	7.3–11.4	23.2 (2.5)	17.9–26.5	35.5	<.001	7.5
Vocabulary[a]	102.8 (9.5)	86–126	107.5 (7.0)	96 -123	2.7	< .01	.56
Socio-economic status[b]	3.0 (.9)	1–4	3.0 (.8)	1–4	<1	*ns*	.03

[a] Standardized scores on the Peabody Picture Vocabulary Test (PPVT-III, Dutch version) to assess prior vocabulary knowledge.

[b] Highest completed educational level of mother is taken as measure of family's socio-economic status: 1 = Lower education, 2 = High School, 3 = Bachelor's degree, 4 = Master's degree or higher. Seven responses on SES are missing in the child group.

All participants were tested individually at school or via a visit at home. Written informed consent was obtained prior to the start of the experiment. Parental consent was additionally obtained for the children. After participation, the children received financial compensation. The adult participants either received financial compensation or a course credit. The research was approved by the Ethical commission of the Faculty of Psychology and Educational Sciences at Ghent University.

2.2. *Stimuli, design and procedure*

All participants were tested on a Dell latitude E5520 laptop with 15.6-inch screen and 1366 × 768 resolution. They listened to the sounds via a noise-cancelling headphone. Participants started with the target detection task during which they were exposed to the repeating syllable triplets. Forty-four participants (23 children and 21 adults) had a break of ~10 min during which they completed the PPVT-III vocabulary test. The remaining participants received no break in between exposure and test. These participants completed the vocabulary test at the start or end of the experiment. This was not intended but the result of two experimenters collecting the data independently. Nevertheless, the interesting prediction can be made that the short delay is necessary for learning effects and a group difference to emerge. The experiment ended with a forced-choice recognition task with memory judgement. The experimental procedure for the main tasks is visualized in Fig. 1.

2.2.1. *Target detection task*

Twelve unique consonants were selected from the Dutch alphabet (i.e., C, D, F, G, K, L, M, P, Q, R, S, V) and structured into four trisyllabic novel word-forms that are pronounceable in Dutch, namely “eLQuPe” (/ϵɫkype:/), “VeeGee-eS” (/ve:ɣe:es/), “eMeRDee” (/ϵmϵrde:/), and “CeeKaa-eF” (/se:ka:ϵf/). We chose distinguishable (syllabic) sounds from the alphabet to assure that the (young) children could easily detect the target sounds

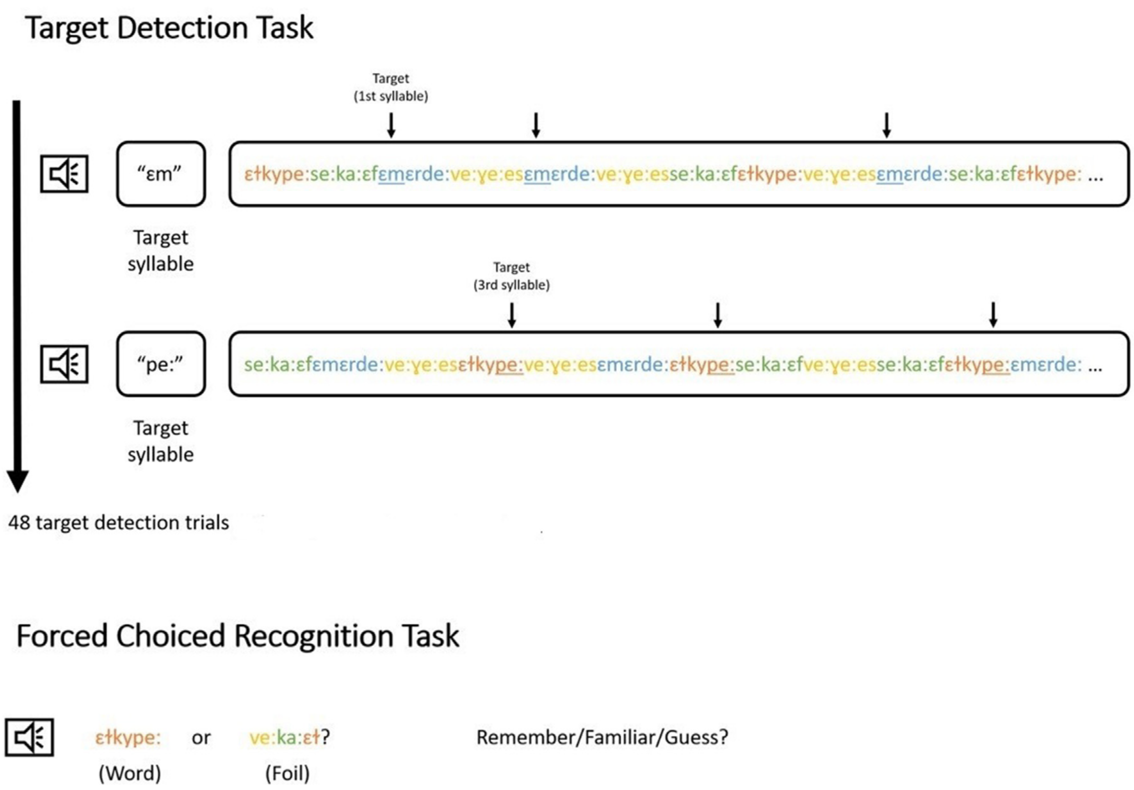

Fig. 1 – Experimental procedure for the main tasks: target detection and forced-choice recognition.

in the stream. The resulting words contained mixed consonant-vowel (CV) and vowel-consonant (VC) syllable structures: CV.CV.VC (2), VC.CV.CV (1), VC.VC.CV (1). None of the words had existing Dutch neighbours, checked via WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004). The twelve syllable sounds were created individually using audio recordings of a female native-Dutch speaker with a neutral vocal affect. The audio files were noise-cancelled and edited to have an exact duration of 300 ms using audacity® software.

At the beginning of each speech stream (i.e., the exposure trial), a target syllable was presented two times in its auditory format (e.g., /ϵł/), with an interstimulus interval of 1000 ms, while also depicted visually on the screen (e.g., “L”). The instruction was to detect the target syllable as accurately and as fast as possible in a subsequently presented speech stream. After a short break of 1.5s during which the word ‘START’ was presented, the speech stream was presented at a rate of 300 ms per syllable, in which the four embedded word triplets were repeated six times in a pseudorandom order. The restriction was that the same word was not repeated twice in a row and that the stream did not begin or end with the word triplet that includes the target. There were reliable transitional probabilities (TP) that signaled the word boundaries. For instance, considering all syllable streams, L [ϵł] was always followed by Q [ky] (i.e., a TP of 100%) while P [pe:] could be equally followed by C [se:], V [ve:] or M [ϵm] (i.e., a TP of 33%). The participants were not made aware of the underlying probability structure. They were made aware that multiple examples of the targets were hidden though they were not told in advance how many examples (namely, that there were six).

Across four blocks (with a self-determined break in between) the exposure trial procedure was performed 12 times so that each syllable along the four triplets were represented as target. In total, participants were thus exposed to 48 speech streams that each consisted of 72 structured syllables (or 6 word repetitions). Hence, in total, target detection could be performed 288 times.

The order of the target detection trials and subsequent streams was pseudo-randomized such that three trial order versions were created and contra-balanced across participants in the two groups upon entry to the experiment.

2.2.2. *Word recognition task*

Subsequent to the target detection task, participants completed a two-alternative forced-choice word recognition task. Four foil triplets were created from the same list of twelve unique sounds that were used to create the trisyllabic structured word-forms at exposure: “RSP” (/ϵrϵspe:/); “CGM” (/se:ɣe:ϵm/), “QDF” (/kyde:ϵf/) and “VKF” (/ve:ka:ϵł/). The syllables making up these foils never followed each other in the speech stream, not even across word boundaries, and the three syllables always came from different word-forms. The foils matched with the structured word-forms on syllable structure, i.e. CV.CV.VC (3) and VC.VC.CV (1), and on summated bigram frequency which represents how often each pair of syllables in a triplet occurs in the Dutch language ($p = .90$). None of the foils had existing Dutch neighbours. These matching variables were checked via WordGen (Duyck et al., 2004).

At each trial, a fixation cross appeared while the auditory presentation of the word triplet and a foil word were presented, separated by an inter-stimulus interval of 1500 ms. The task was to indicate which of the two triplets sounded most familiar to what they previously heard during the target detection task. They were then asked to judge on their memory of the chosen word (‘I remember it’, ‘It sounds familiar but I have no clear memory of it’, ‘I guessed’). The next trial started 1500 ms after the participant entered his or her response. The triplets were presented at the same rate as during the exposure, i.e. at a rate of 300 ms per syllable (~3.3 Hz) (e.g., Batterink & Paller, 2017; Smalle et al., 2022). Each of the four target words and four non-words (foils) were paired exhaustively to a total of 16 trials. In half of the trials, a target word was followed by a non-word while in the other half a non-word was followed by a target word. The order of the 16 trials was randomly shuffled at the start of the experiment.

2.3. *Statistical analyses*

We used an alpha level of .05 for all statistical tests. Participant characteristics were compared across groups using two-tailed independent *t*-tests and Cohen's d effect sizes with Hedges correction. Welch *t*-tests were used when the two samples had unequal variances. The target detection data and the word recognition data were analyzed using linear mixed effects modeling with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and the afex package (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2015) in R (R Core Team, 2021). We always strived for models including a maximal random effects structure for participant and item justified by the design; in case of convergences issues (e.g., singular fits), we refitted the maximal model by first removing correlations among random slopes, after which the highest order random slopes with the least estimated variance were removed (Singmann & Kellen, 2019). If necessary, also the item factor (i.e., word) was removed. For the target detection data, responses that occurred within 1500 ms after the target were considered as hits and log-transformed before the start of the analyses.[3] The model included a fixed effect for Triplet Position with user-defined contrasts. A first contrast coded the comparison between the unpredictable first position and the average of the predictable second and third position (as proposed by Siegelman, Bogaerts, Kronenfeld, & Frost, 2018), referred to as the RT facilitation effect. A second contrast coded the comparison between the two predictable positions, not expected to differ (i.e., Siegelman et al., 2018). Other fixed effects were Exposure Trial (1–48), Group (effects coded) and all interactions. The implemented control variables were participant's SES and Vocabulary Score, as well as the position of the target within the exposure stream (i.e., Stream Position: 4 to 68). The latter variable was included because the tendency to respond faster to targets later in the stream may otherwise

---

[3] This is important to control for baseline reaction time differences, because otherwise a child advantage in the syllable position effect might be explained by slower response times allowing for more gains across syllable positions (Siegelman et al., 2018).

Table 2 – Means and SDs for raw RTs and log-transformed RTs for syllables in first, second and third positions.

	Triplet Position 1	Triplet Position 2	Triplet Position 3
Children			
Raw RT (SD)	647 (195)	572 (220)	631 (229)
Log-transformed	6.33 (.41)	6.16 (.48)	6.28 (.51)
Adults			
Raw RT (SD)	510 (100)	451 (138)	462 (117)
Log-transformed	6.18 (.19)	6.02 (.32)	6.05 (.32)

be confounded with the effect of triplet position (Himberger, Finn, & Honey, 2019). The final converging model was *Log(RT) ~ Triplet Position x Exposure Trial x Group + Stream Position + Vocabulary Score + SES + (1+Triplet Position | Subject) + (1 | Word)*. For the word recognition data, the outcome variable was accuracy (binomial: 1 = correct, 0 = incorrect). The factors were Group (effects coded, adults set to −1), Memory Judgement (polynomial linear contrast) and their interaction. Participant's SES and vocabulary score were again added as control variables. The final converging model was *Accuracy ~ Memory Judgement x Group + Vocabulary Score + SES + (1 | Subject)*. To check whether the accidental delay after the exposure affected word recognition, we ran a second model which included Experimenter as separate factor. All *p* values in the linear mixed effects analyses were derived using Satterthwaite's method with the anova function provided in the afex package. Pairwise contrasts were calculated using the phia package. We used Bonferroni correction for multiple comparisons. Cohen's *f* effect sizes and their precision are calculated using the sjstats package (Levine & Hullett, 2006). We additionally performed one-sample *t*-tests and computed Cohen's *d* effect sizes with Hedges correction to test for above-chance performance in the memory judgement task. To check whether groups differ on overall proportion of metamemory responses, supplementary repeated measure ANOVA's are performed (Supplementary S1). Study materials, data files and scripts for analysis are available via an open science repository (https://osf.io/u6qk9/). No part of the study procedures and analyses were pre-registered prior to the research being conducted.

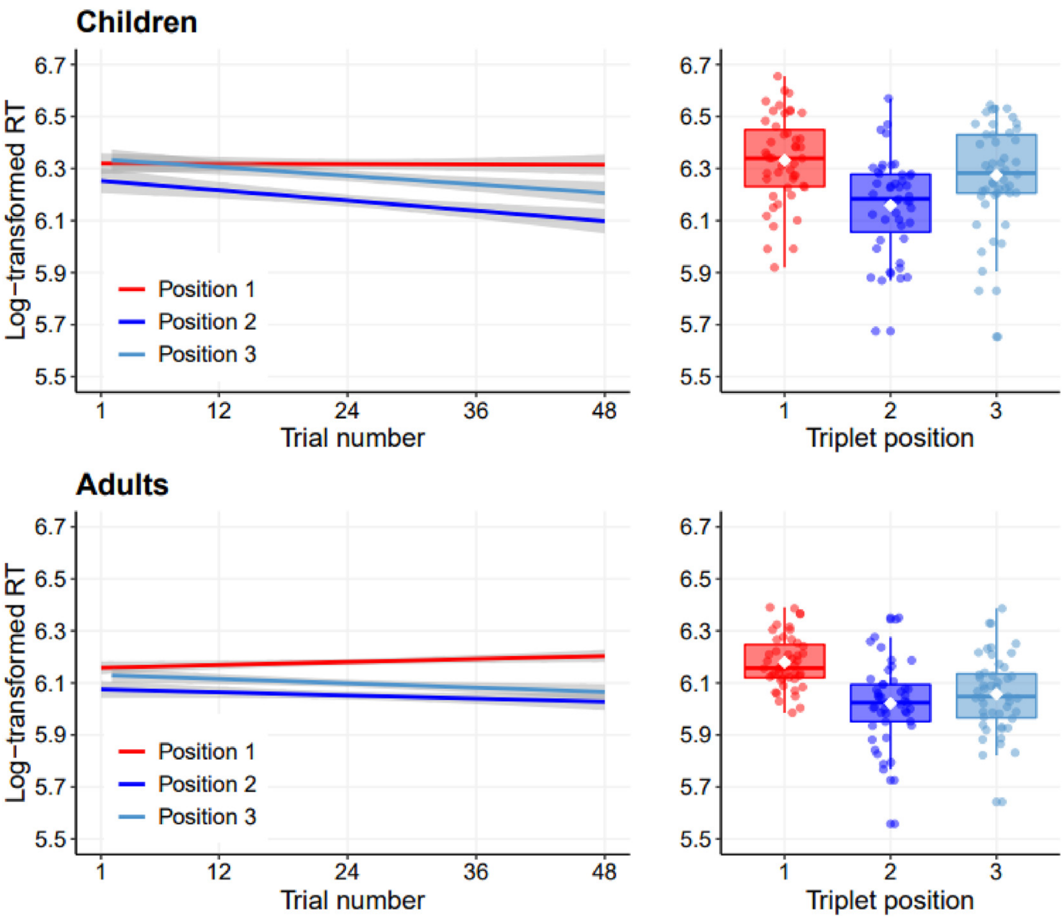

Fig. 2 – Reaction time data as a function of syllable triplet position (first, second or third syllable in the word) and group (children versus adults). Left panels present Log-transformed RT as a linear function of position and time (i.e., trial 1 to 48). The shaded area around the lines shows the 95% confidence intervals. Right panels show boxplots per triplet position (averaging across the entire exposure phase). White diamonds show the means and midlines represent medians. Box limits indicate the 25th and 75th percentiles and whiskers extend to minimum and maximum value, with the exception of outliers. Dots depict individual participants. Please note that syllable position 3 was never presented as a target in the first trial.

3. Results

3.1. *Online statistical learning: target detection*

On average, adults detected 260 of the 288 hidden target syllables (i.e., a hit rate of 90% $\pm$ 6$_{SD)}$ while children detected 185 target syllables (i.e., a hit rate of 64% $\pm$ 14$_{SD;}$ independent Welch $t_{63.6} = 11.7$, $p < .001$, $d = 2.5$). Table 2 presents the means and standard deviations (SD) for raw and log-transformed response times for detected targets (hits) as a function of the position of the target within a triplet and across the age group. Fig. 2 plots the log-transformed RTs for the three syllable positions, as well as their change throughout the exposure. The model estimate output is presented in Table 3. Type III Anova tests reveal a significant main effect of Triplet Position ($F_{2,\ 78.8} = 79.1$, $p < .001$, Cohen's $f = 1.42$, 95%CI[1.10, 1.73]) that interacts with Group ($F_{2,\ 78.6} = 4.9$, $p = .01$, Cohen's $f = .35$, 95% CI[.09, .57]. Estimate results in Table 3 (see also Fig. 2) show evidence for a RT facilitation effect; RTs are faster for targets at the second and third predictable positions than for targets at the first unpredictable position. This effect increases with exposure and does not differ between groups (i.e., no significant two- and three-way interaction). In contrast, while adults show the expected pattern of the predictable positions (i.e., positions 2 and 3) not differing from one another (Cohen $d = -.07$) children unexpectedly show faster RTs on second syllable positions relative to third syllable positions (Cohen $d = -.22$).

Table 3 – Model output for the log-transformed target response times as a function of Triplet Position (1, 2 and 3), Exposure trial (1–48) and Age Group (Children versus Adults).

Variable	Estimate	Standard error	t-value	Pr (|t|)
(Intercept)	6.35	.18	35.3	< .001***
Triplet Position[a]	−.15	.13e-01	−10.9	< .001***
Triplet Position[b]	.37e-01	.59e-02	6.2	< .001***
Trial[c]	−.21e-02	.28e-03	−7.3	< .001***
Group[d]	.83e-01	.16e-01	5.2	< .001***
Triplet Position[a] × Trial	−.37e-02	.58e-03	−6.45	< .001***
Triplet Position[b] × Trial	−.48e-03	.35e-03	−1.35	.18
Triplet Position[a] × Group	.52e-02	.13e-01	<1	.70
Triplet Position[b] × Group	.18e-01	.59e-02	3.09	<.01**
Trial × Group	−.47e-03	.28e-03	−1.69	.091
Triplet Position[a] × Trial × Group	−.93e-04	.58e-03	<1	.87
Triplet Position[b] × Trial × Group	.41e-04	.35e-03	<1	.90
Stream Position	.26e-03	.19e-03	1.34	.18
SES	−.30e-01	.16e-01	−1.86	.07
PPVT	−.97e-03	.17e-02	<1	.57

[a] User defined contrast for the RT facilitation effect (i.e., unpredictable position 1 set to −.67 and predictable positions 2 and 3 each set to .33).

[b] User defined contrast for the comparison across predictable positions (i.e., second and third positions set to −1 and 1, respectively).

[c] Trial was coded as a numeric variable and centered.

[d] Sum contrast with children set to 1 and adults set to −1.

3.2. *Offline statistical learning: word recognition with memory judgement*

Participants performed above chance on the forced-choice word recognition task, indicating offline statistical learning of the triplets (Children: mean = 71.0, SE = .022, $t_{43} = 9.5$, $p < .001$, $d = 1.41$; Adults: mean = 67.3, SE = .027, $t_{43} = 6.4$, $p < .001$, $d = .95$). Children performed numerically higher than adults ($t_{86} = 1.06$, $p = .29$, $d = .22$).

The overall proportion of metamemory responses differs between groups (Memory × Group: $F(2,172) = 12.7$, $p < .001$; Supplementary Fig. S1a). Mean proportion of "remember" responses is 46.5% (range = 18.8%–100%) in children and 30.1% (range = 0%–100%) in adults ($t_{86} = 3.7$, $p < .001$). Mean proportion of "familiar without clear memory" responses is 33.1% (range = 0%–62.5%) in children and 49.6% (range = 0%–81.3%) in adults ($t_{86} = -4.6$, $p < .001$). Mean proportion of "guess responses" is 20.5% (range = 0%–50%) in children and 20.3% (range = 0%–75%) in adults ($t < 1$, $p = .96$). Seven adults never

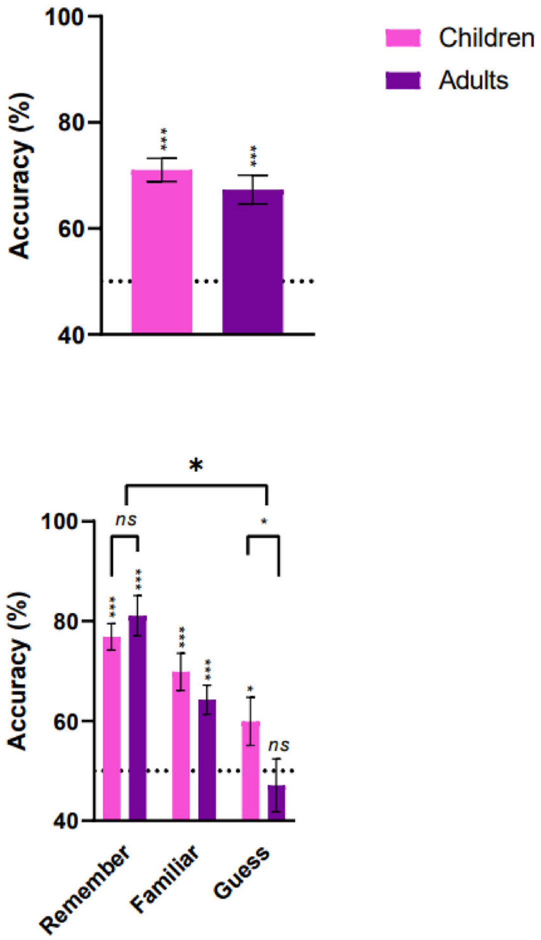

Fig. 3 – Performance on the word recognition task. Word recognition accuracy by group. Lower panel: Accuracy as a function of memory judgment. Error bars represent standard error of the mean. Chance performance on this task is 50%. Vertically aligned asterisks on top of the bars denote significance for above-chance performance. *<.001, **<.01, *<.05.**

reported remembering the triplet. Four children and three adults never guessed.

Fig. 3 presents the behavioral results on the word recognition task. Model output for accuracy (1 = correct, 0 = incorrect) as a function of Group and Memory Judgement reveals that performance increases with confidence (i.e., $\beta = -.92$, SE = .13, Z = −7.06; Memory Judgement: $X^2_{(2)} = 51.7$, $p < .001$). This effect appears stronger in adults than in children (i.e., $\beta = .32$, SE = .13, Z = 2.44; Memory Judgement × Group: $X^2_{(2)} = 6.87$, $p < .05$). While both children and adults show above chance performance for the remembered responses (with adults scoring numerically higher than children, $p_{\text{one-sided}} = .08$, $d = .39$) and for the familiar (without clear memory) responses (with children scoring numerically higher than adults, $p_{\text{one-sided}} = .09$, $d = .32$), only the children show above chance performance for the guess responses (see Fig. 3; children versus adults: $p_{one\text{-}sided} = .04$, $d = .51$).

Recall from the method section that half of the participants received a short delay after the exposure before completing the offline test. Subsequent exploratory analyses show that this delay influences the group effect above (i.e., Memory Judgement × Group × Delay: $\beta = .63$, SE = .31, Z = 2.06, $p < .05$). In the next analyses we split up the data according to whether a delay period was implemented or not. The behavioral results are presented in Fig. 4. Proportion of meta-memory responses across delay groups is presented in a supplementary file (S1; Fig. S1b).

With delay, a significant Memory by Group interaction emerges (i.e., $\beta = .82$, SE = .25, Z = −3.30; Memory Judgement × Group: $X^2_{(2)} = 11.3$, $p < .01$). Word recognition is better for adults than for children when confidence is high (i.e., remembered responses: $p_{one\text{-}tailed} < .01$, $d = 1.55$) but better for children than for adults when confidence is low (i.e., guessed responses: $p_{one\text{-}tailed} < .05$, $d = .78$). No significant interaction between Memory Judgement and Group emerges without the delay (i.e., $\beta = .18$, SE = .17, Z < 1).

3.3. *Relationship between online and offline learning measure*

Correlations were calculated between the online statistical learning score (i.e., RT facilitation effect) and the offline statistical learning score (i.e., 2AFC accuracy). As shown in Fig. 5, there is a significant correlation for the adults but not for the children.

4. Discussion

Children are more successful language learners than adults. However, today, the exact nature of the child advantage in language acquisition is still poorly understood. According to Newport's less is more hypothesis, the very limitations of the child's information processing abilities can provide the basis

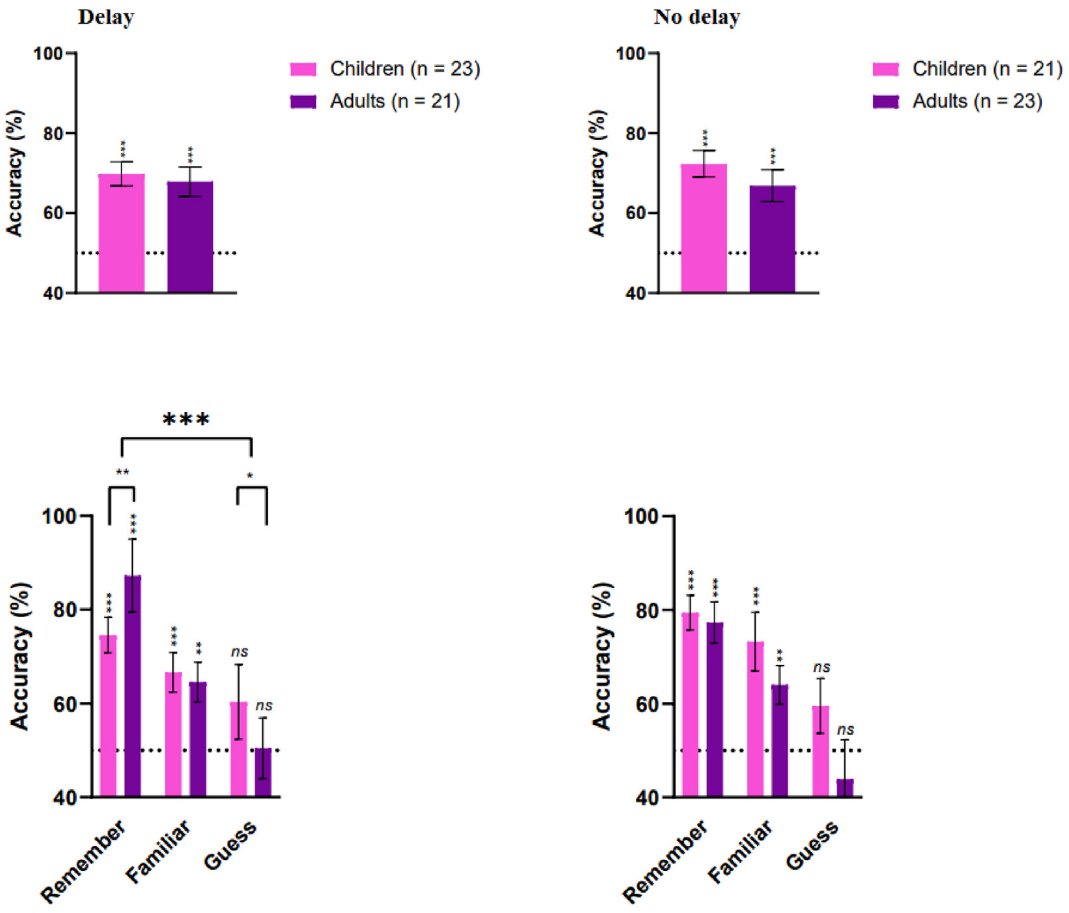

Fig. 4 – Performance on the word recognition task, split by delay. *<.001, **<.01, *<.05.**

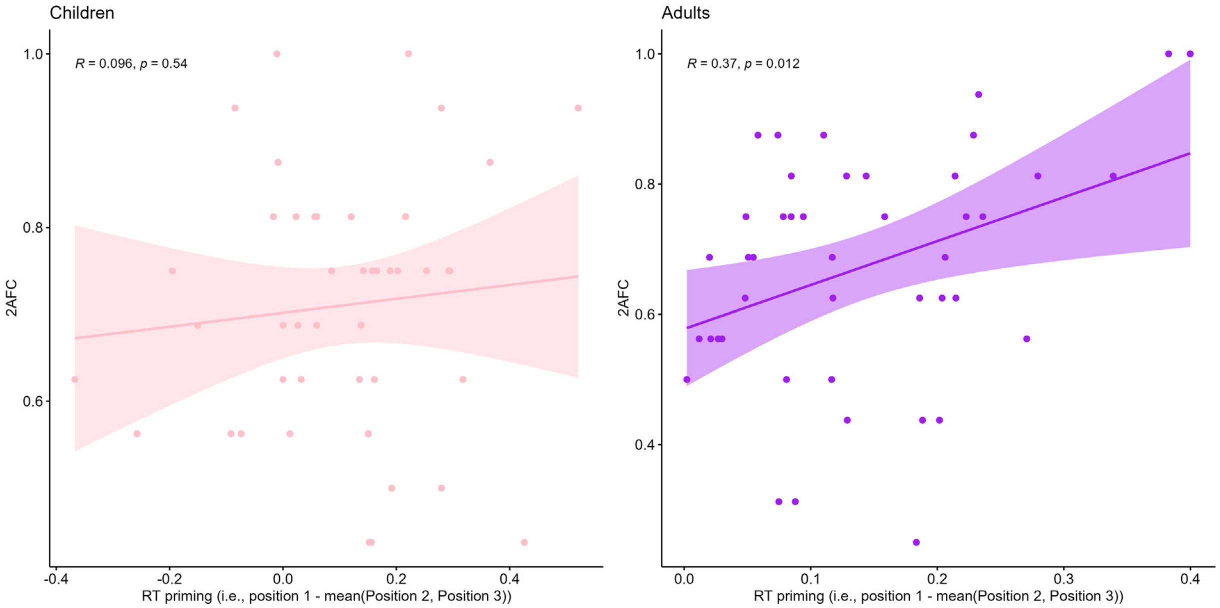

Fig. 5 – Correlations between the online and offline statistical learning score. Y-axis denote accuracy on the forced choice recognition task. X-axis denote the RT priming effect on the target detection task (i.e., Position 1 – mean (Position 2 & 3)). 95% confidence intervals are added.

on which their successful language acquisition occurs (Newport, 1990). According to her view, the more limited abilities of children in terms of attentional control and memory span provide an advantage for tasks, like language learning, that rely on componential analyses such as unconsciously extracting recurring word-forms in speech. However, the few studies that compared statistical learning directly across age did not find clear support for early advantages (e.g., Choi et al., 2020; Moreau et al., 2022; Raviv & Arnon, 2018; Saffran et al., 1997). This could be due to the fact that statistical learning is typically measured in the lab via tasks that recruit explicit, cognitive processes that are not directly relevant for the learning process and may not be well-developed in the child's brain (Arnon, 2020; Kidd et al., 2020). Consequently, this could obscure a potential learning advantage for the young child when compared to the older child or adult. Here, we investigated whether children are more sensitive learners when focusing on the implicit knowledge acquired during the learning task. We compared seven- to twelve-year old Dutch-speaking children and young adults on 1) an online, processing-based measure of novel word-form learning and 2) an offline, reflection-based measure of learning that dissociates implicit from explicit knowledge of the segmented speech via a guessing criterion.

First, we found that both children and adults show fairly comparable *online* learning abilities. Learning was measured via a target detection task that participants performed while listening to the structured speech sounds. In both groups, response times indicated that targets at the predictable second and third syllable positions of a repeating triplet in the stream were detected faster than targets at the initial syllable position, i.e. there was a RT facilitation effect. Surprisingly, children displayed slower response times to targets at the third position compared to response times to targets at the second syllable position. This effect is not typically seen in adult samples, for which response times gradually speed up for increasing syllable positions (e.g., Batterink & Paller, 2017; Franco et al., 2015) or response times on the second and third syllable position remain indifferent (Siegelman et al., 2018).

The unexpected syllable position effect pattern in our children could suggest that they relied on a different mechanism when listening to the speech sounds. While the statistical structure of our sequence was defined in terms of transitional probabilities, the underlying learning mechanism could be a memory-based chunking mechanism, i.e., the chunking of frequently co-occurring syllables into gradually higher units rather than computing transitional probabilities per se (see for instance, Siegelman, Bogaerts, Armstrong, & Frost, 2019). Overall, the chunking hypothesis suggests that repeated exposure to a stimulus set, for example, a stream of sounds, leads to the stimuli being represented in larger and larger chunks to free up space in memory, also referred to as the “now-or-never” bottleneck principle (Christiansen & Chater, 2016). Sensitivity to transitional probabilities could reflect such chunking process according to which adjacent syllables are grouped into chunk representations that receive activation every time the chunk is encountered during the exposure. Representations of groupings across word boundaries show less (re)activation because they are re-encountered less frequently; hence they will suffer in competition with representations of groupings within word boundaries (Erickson & Thiessen, 2015). Within our speculation, smaller attention and memory capacities may potentially force children to encode the syllable stream at a smaller grain size (e.g., bi-syllabic chunks) compared with adults (i.e., tri-syllabic chunks). It is possible that since the syllable stream in the present study unfolds one syllable at a time and the first syllable is the most surprising yet informative syllable (it signals there is something familiar coming), children prioritize attending the first two syllables of a triplet rather than the final two syllables in the target detection task.

The speculation on different chunking mechanisms during auditory statistical learning across development would be in line with recent work showing that newborns only retain the first two syllables of a statistically embedded triplet in speech (Fló, Benjamin, Palu, & Dehaene-Lambertz, 2022). It would also corroborate Newport's less is more hypothesis and previous work with the Hebb repetition learning paradigm in memory research (Smalle et al., 2016). In this work, it was shown that both children and adults become better in recalling repeating syllables sequences over non-repeating sequences as a function of exposure, also referred to as Hebb repetition learning. However, when the repeating sequences contained the same syllables as the non-repeating sequences (albeit in a different order), learning was strongly reduced in children but not in adults. This item-overlap effect was explained by assuming a different chunking mechanism in children than adults, with children focusing on smaller, two-syllable chunks than adults. Because of the item-overlap, anagrams of small chunks turned up more often than anagrams of large chunks, disrupting children's Hebb learning process. Recent work on (visual) statistical learning indeed suggests that children are more rigid learners of the transitions between items within a triplet than adults (i.e. they are less likely to falsely judge anagrams as previously seen during the exposure; Forest, Abolghasem, Finn & Schlichting, 2023). A different chunking mechanism in children versus adults was confirmed in a follow-up experiment of Smalle et al. (2016), wherein adults were encouraged to chunk the sequences into smaller, two-syllable items (via embedded pauses that directed their attention) and resulting in a similar item-overlap effect as in children. Future work implementing memory-based measures of statistical learning, such as statistically induced chunking recall (Isbilen, McCauley, Kidd, & Christiansen, 2020; Kidd et al., 2020) could bring important new insights in this underlying chunking mechanism and should be considered as an alternative method for studying child advantages in language learning. So far, we are not aware of any study making this comparison (but see previous work on verbal serial recall paradigms, for which child advantages in immediate serial recall and long-term Hebb learning of syllable sequences are found; Bishop, Barry, & Hardiman, 2012; Smalle et al., 2016, 2018).

Second, we observed that both children and adults show comparable performance on the offline recognition task, with children scoring only numerically higher than adults. The added memory judgement procedure, however, indicated that the nature of learning on the task, and thus the knowledge of the words, differed between the two groups. In line with previous work, adults' performance on the task reflected explicit word knowledge (e.g., Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015; Smalle et al., 2022). Children, in contrast, performed above chance on the recognition task also when confidence was low (i.e., during guessing reports) suggesting that their performance on the tasks reflected both explicit and implicit knowledge of the hidden words.

Very recently, Batterink and her team published similar word recognition findings in a supplementary file (Moreau et al., 2022). Together, our results for different language populations provide increasing evidence for the hypothesis that children might rely more heavily on implicit memory processes during speech-based auditory statistical learning. We believe that such knowledge might be an important finding in light of the sensitive age debate for language acquisition because implicit memory is assumed to be more stable and long-lived than explicit memory (see for instance, Ullman, 2004). The hypothesis of dissociated consolidation trajectories for implicit and explicit memory has recently been demonstrated in speech-based statistical learning using a similar memory judgement task (Liu et al., 2023). In their study with adults, implicit knowledge of the extracted word patterns strengthened over a 24-h period while explicit knowledge decayed. Moreover, implicit knowledge was unaffected by repeated testing, further corroborating the stability of implicit (over explicit) representations. Finally, recent cross-sectional work on perceptual-motor sequence learning suggests no age-related differences in the overall retention of statistical knowledge over a 24-h delay (e.g., Tóth-Fáber, Nemeth, & Janacsek, 2023). An interesting avenue for future follow-up work would be to look more closely into the long-term retention and stability of implicitly versus explicitly extracted patterns in children versus adults, and test whether there are any language-specific effects therein. Note that our results demonstrate that implementing even a very short delay period after the exposure already made the group difference in the word recognition task more obvious, particularly for what concerns the explicitly remembered representation of the novel words (though also the proportion of the meta-memory judgments differed across groups in the delay). This might suggest that a break is necessary for revealing explicit learning advantages in adults – yet future work is needed to test this hypothesis more directly.

4.1. *Study limitations and further considerations*

It is noteworthy that although children and adults differ in their likelihood to select "remember" or "familiar" responses, with children showing a higher proportion of "remember" responses and adults showing a higher proportion of "familiar" responses, they do not differ in their guessing behavior (see supplementary materials, S1). This is important because the "guess" responses are our primary measure indicating implicit memory of the words (as measured with the recognition task.) We think it is rather unlikely that our children were not able to do the memory judgement and selected their confidence responses randomly. If that would be the case, then performance on the word recognition task would have been unrelated to memory. This was not seen in the data with remembered responses being on average most accurate (and above chance), followed by "familiar" and "guess" responses. However, it may be well possible that children are less reliable overall in making memory judgements and therefore approach task differently than adults.

Two methodological choices in the current study also deserve additional consideration. Firstly, we used word recognition test trials contrasting real word patterns with non-word foil patterns rather than with part-word foil patterns. Both part-word and non-word foils have been used in the statistical learning literature. According to a test item-analysis in Siegelman, Bogaerts, and Frost (2017) contrasting words with nonword foils discriminates better between low and high statistical learners than contrasting words with partword foils. Future work could further investigate whether child advantages are also seen in other (perhaps more cognitively challenging) testing conditions, yet notably, the previously discussed study by Moreau et al. (2022) included both foil types which showed very similar results. Moreover, recent work on visual statistical learning suggests a strong developmental effect on testing the memory of part-words (Forest, Schlichting, et al., 2023). In their study, older children and adults were more likely than younger children to erroneously remember part-words (as well as group lures) as previously seen, arguably because they have better higher cognitive abilities and therefore tend to memorize overall group membership rather than the precise statistical dependencies of the items (Forest, Schlichting, et al., 2023). Secondly, because of the implementation of the target detection task, the speech stream during exposure was broken up into 48 shorter streams. The silences before and after each stream could therefore provide additional cues to word boundaries and make the performance on the word recognition task difficult to interpret as a pure measure of statistical learning. However, silences between words also occur in natural speech, such as at the beginning or end of a spoken sentence, making the combination of cues in our setup a more realistic approximation of the natural learning environment (see, Ibsilen & Christiansen, 2022). We believe that in addition to the findings of Moreau et al. (2022), our findings demonstrate important developmental differences in memory representations of the words (irrespective of the discussion of how they are learned, e.g., via additional silences or chunking mechanisms). We nevertheless call for future work to investigate the observed developmental differences also in an unconfounded version of the target detection task.

While our main results are based on group differences, we also considered inter-individual variability in learning performance within each group. An additional correlation analysis shows that the online RT facilitation effect is positively associated with performance on the offline word recognition task in adults only. Performance on the online target detection task likely reflects processing-based chunking mechanisms underlying statistical learning. Hence, a different correlation with the reflection-based measure in children versus adults suggests that different underlying processes operate during the exposure. Although speculative, we think that the positive correlation with performance on the recognition task in adults could reflect explicit learning strategies during exposure. Alternatively, as the reliability of a measure at the level of the individual determines the upper bound for the correlation that can be found (Bogaerts, Siegelman, Ben-Porat, & Frost, 2018; Siegelman et al., 2017), the lack of a significant correlation in children could be the result of lower reliability of the learning measures in child participants as opposed to adult participants (Arnon, 2020).

An important final question arises whether child advantages for implicit statistical learning could be a domain-general effect or is specific to speech-based statistical learning. It is not completely understood how statistical learning is implemented in different modalities (i.e., auditory, visual, or tactile) and domains (i.e., linguistic, or non-linguistic) (Bogaerts, Siegelman, Christiansen, & Frost, 2022). Statistical learning follows different developmental trajectories in visual, auditory, and motor-based forms of statistical learning (Arciuli & Simpson, 2011; Janacsek, Fiser, & Nemeth, 2012; Raviv & Arnon, 2018). Moreover, there exist no or only weak correlations between performance on distinct statistical learning tasks, both within a modality (e.g., using different syllables, Erickson, Kaschak, Thiessen, & Berry, 2016) and across modalities (e.g., visual and auditory SL; Siegelman & Frost, 2015). In 2015, Frost and colleagues proposed a theoretical framework to explain how SL can operate across different domains and modalities while also demonstrating specificity. More specifically, they argue that SL involves domain-general computational principles that operate on internal representations that are modality-dependent. Even shared computations for extracting regularities could be implemented by distinct neural networks in the brain, depending on the properties of the input signal that is processed (i.e., visual, auditory, or somatosensory). In line with this, we would speculate that the child advantage is specific to the perceptual encoding of and learning over speech-based stimuli. However, an alternative prediction of a domain-general learning difference could also be made as children are found to also excel adults on other forms of implicit sequence learning, such as perceptual-motor sequence learning in the serial-reaction time task (e.g., Janacsek et al., 2012; Juhasz, Nemeth, & Janacsek, 2019; Nemeth, Janacsek, & Fiser, 2013). Already in 2012, Janacsek and colleagues proposed that a developmental shift from model-free learning to model-based learning processes occurs around the onset of adolescence, emphasizing the intrinsic nature of learning differences regardless of modality (Janacsek et al., 2012). In this framework, the underdevelopment of learning systems that depend on prefrontal- and medial temporal lobe in children causes internal models to have little influence on the extraction of environmental regularities. Learning is then mainly implicit and driven by basal-ganglia-related subsystems. This shifts after adolescence, when internal models start to play a more important role. Recent work in both the language domain and perceptual-motor domain provide direct evidence for the competitive nature of above learning systems in the adult brain, by demonstrating improved implicit statistical learning abilities when lesioning the dorsolateral prefrontal cortex with non-invasive brain stimulation (Ambrus et al., 2020; Smalle et al., 2022). An important avenue for future research will be to investigate the domain-specific versus general nature of the observed developmental differences by directly comparing age-related sensitivities in different domains of (skill) learning.

5. Conclusion

In the present work, we investigated the potential of child advantages for implicit statistical learning in speech supporting sensitive age effects in language development.

Overall, children and adults showed comparable online learning abilities. However, additional memory judgements to the typical offline post-exposure word recognition test revealed that the resulting word knowledge was at least partly implicit in children. Taking into account the recent comparable findings from Moreau et al. (2022) with a different language population and experimental materials, we conclude that the human ability for auditory word segmentation is fairly similar in children and adults but that developmental changes can exist in the nature of the memory that is eventually formed. Additional work is needed to show whether similar dissociations in implicit and explicit word knowledge can be found when using other tests of (pure) statistical learning, and to determine what advantages the recruitment of implicit over explicit learning processes has on memory processes, such as the stabilization of the novel extracted words in light of a longer delay and its resistance to interference, and thus on ultimate language performance.

Open practices

The study in this article has earned Open Data and Open Materials for transparent practices. The data and materials studies are available at: https://osf.io/u6qk9/.

Ethics and integrity statement

Data and materials are available at https://osf.io/u6qk9/.

The research was funded by a grant from The Research Foundation – Flanders (FWO) to the first author. The Ethical Commission of the Faculty of Psychology and Educational Sciences at Ghent University (reference number 2020/81) approved the study. We state no conflict of interest.

CRediT authorship contribution statement

Eleonore HM. Smalle: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Louisa Bogaerts:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2024.07.001.

REFERENCES

Ambrus, G. G., Vékony, T., Janacsek, K., Trimborn, A. B., Kovács, G., & Nemeth, D. (2020). When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *Journal of Memory and Language, 114*, 104144.

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental science, 14*(3), 464–473.

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods, 52*(1), 68–81.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science, 9*(4), 321–324.

Baars, B. J. (1995). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. *American Scientist, 83*(1), 92–94.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme 4. *Journal of Statistical Software, 67*(1), 1–48.

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological science, 28*(7), 921–928.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 90*, 31–45.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Medicine and Life, 83*, 62–78.

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & Memory, 22*(11), 544–556.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258.

Bialystok, E., & Hakuta, K. (1994). *In other words: The psychology and science of second language acquisition*. New York: Basic Books.

Bishop, D. V., Barry, J. G., & Hardiman, M. J. (2012). Delayed retention of new word-forms is better in children than adults regardless of language ability: A factorial two-way study. *Plos One, 7*(5), Article e37326.

Bogaerts, L., Siegelman, N., Ben-Porat, T., & Frost, R. (2018). Is the Hebb repetition task a reliable measure of individual differences in sequence learning? *The Quarterly Journal of Experimental Psychology: QJEP, 71*(4), 892–905.

Bogaerts, L., Siegelman, N., Christiansen, R., & Frost, R. (2022). Is there such a thing as a "good statistical learner". *Trends in Cognitive Science, 26*(1), 25–37.

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*(9), 1161–1173.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in cognitive science, 11*(3), 468–481.

Christiansen, M., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*, E62. https://doi.org/10.1017/S0140525X1500031X

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in cognitive sciences, 10*(3), 131–138.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*(1), 3–23.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers, 36*(3), 488–499.

Erickson, L. C., Kaschak, M. P., Thiessen, E. D., & Berry, C. A. S. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra, 2*(1). https://doi.org/10.1525/COLLABRA.41

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological science, 18*(3), 254–260.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2019). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *Journal of Cognitive Neuroscience, 31*(1), 126–137.

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). Whenit hurts (and helps) to try: The role of effort in language learning. *Plos One, 9*(7), Article e101806.

Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports, 12*(1), 4391.

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental science, 22*(4), Article e12802.

Forest, T. A., Abolghasem, Z., Finn, A. S., & Schlichting, M. L. (2023). Memories of structured input become increasingly distorted across development. *Child Development, 94*, e279–e295.

Forest, T. A., Schlichting, M. L., Duncan, K. D., & Finn, A. S. (2023). Changes in statistical learning across development. *Nat Rev Psychol, 2*, 205–219.

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. *Experimental psychology, 62*(5), 346.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128.

Gaskell, M. G., & Ellis, A. W. (2009). In *Word learning and lexical development across the lifespan* (Vol. 364, pp. 3607–3615). The Royal Society.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological science, 14*(1), 31–38.

Hartshorne, J. K. (2022). When do children lose the language instinct? A critical review of the critical periods literature. *Annual Review of Linguistics, 8*(1), 143–151.

Himberger, K., Finn, A., & Honey, C. J. (2019). *Reconsidering the automaticity of visual statistical learning*. https://doi.org/10.31234/osf.io/r659w. Preprint.

Huang, B. H. (2016). A synthesis of empirical research on the linguistic outcomes of early foreign language instruction. *International Journal of Multilingualism, 13*(3), 257–273.

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis into 25 years of research. *Cognitive Science, 46*(9), Article e13198.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive science, 44*(7), Article e12848.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental science, 15*(4), 496–505.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology, 21*(1), 60–99.

Juhasz, D., Nemeth, D., & Janacsek, K. (2019). Is there more room to improve? The lifespan trajectory of procedural learning and its relationship to the between-and within-group differences in average response times. *Plos One, 14*(7), Article e0215116.

Kennedy, D., & Norman, C. (2005). What don't we know? *Science, 309*(5731), 75-75.

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology, 200*, Article 104964.

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of phonetics, 7*(3), 279–312.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice, 2*(12), 59–67.

Levine, T. R., & Hullett, C. R. (2006). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612–625.

Liu, H., Forest, T. A., Duncan, K., & Finn, A. S. (2023). What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition, 26*, Article 105439.

Lum, J. A., Byrne, L. K., Barhoun, P., Hyde, C., Hill, A. T., Enticott, P. G., et al. (2023). Resting state electroencephalography power correlates with individual differences in implicit sequence learning. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.16059

Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience, 57*, Article 101154.

Nemeth, D., Janacsek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in computational neuroscience, 7*, 147.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science, 14*(1), 11–28.

Newport, E. L. (2018). Is there a critical period for L1 but not L2? *Bilingualism: Language and Cognition, 21*(5), 928–929.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehlter* (pp. 481–502). Cambridge, MA: MIT Press.

Palmer, S. D., Hutson, J., & Mattys, S. L. (2018). Statistical learning for speech segmentation: Age-related changes and underlying mechanisms. *Psychology and aging, 33*(7), 1035.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child development, 80*(3), 674–685.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental science, 21*(4), Article e12593.

R Core Team. (2021). *R: A language and environment for statistical ## computing*. Vienna, Austria: R Foundation for Statistical Computing. ## URL https://www.R-project.org/.

Rebuschat, P., & Monaghan, P. (2019). Editors' introduction: Aligning implicit learning and statistical learning: Two approaches, one phenomenon. *Topics in cognitive science, 11*(3), 459–467.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906–914.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science, 8*(2), 101–105.

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive science, 42*(8), 3100–3115.

Siegelman, N., Bogaerts, L., Armstrong, B. C., & Frost, R. (2019). What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition, 192*, Article 104002.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*, 418–432.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science,* 42(Suppl 3), 692–727.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Medicine and Life, 81*, 105–120.

Singleton, D. (2006). The critical period hypothesis: Some problems. *Interlingüística*, (17), 48–56.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). *afex: Analysis of factorial experiments*. R package version 0.13–145.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology* (pp. 4–31). Routledge.

Smalle, E. H., Bogaerts, L., Simonis, M., Duyck, W., Page, M. P., Edwards, M. G., et al. (2016). Can chunk size differences explain developmental changes in lexical learning? *Frontiers in psychology, 6*, 1925.

Smalle, E. H., Daikoku, T., Szmalec, A., Duyck, W., & Möttönen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. *Proceedings of the National Academy of Sciences, 119*(2), Article e2026011119.

Smalle, E. H., & Möttönen, R. (2023). Cognitive development as a piece of the language learning puzzle. *Cognitive Science, 47*(5), Article e13296.

Smalle, E. H., Muylle, M., Duyck, W., & Szmalec, A. (2021). Less is more: Depleting cognitive resources enhances language learning abilities in adults. *Journal of Experimental Psychology: General, 150*(12), 2423.

Smalle, E. H., Page, M. P., Duyck, W., Edwards, M., & Szmalec, A. (2018). Children retain implicitly learned phonological sequences better than adults: A longitudinal study. *Developmental science, 21*(5), Article e12634.

Smalle, E. H., Panouilleres, M., Szmalec, A., & Möttönen, R. (2017). Language learning in the adult brain: Disrupting the dorsolateral prefrontal cortex facilitates word-form learning. *Scientific reports, 7*(1), 1–9.

Stärk, K., Kidd, E., & Frost, R. L. A. (2022). The effect of children's prior knowledge and language abilities on their statistical learning. *Applied Psycholinguistics*, 1–27.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3617–3632.

Tóth-Fáber, E., Nemeth, D., & Janacsek, K. (2023). Lifespan developmental invariance in memory consolidation: Evidence from procedural memory. *PNAS nexus, 2*(3), Article pgad037.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*(1), 1–8.

Thiessen, E. D., Girard, S., & Erickson, L. C. (2016). Statistical learning and the critical period: How a continuous learning mechanism can give rise to discontinuous learning. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*(4), 276–288.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1–2), 231–270.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual review of psychology, 66*(1), 173–196.