8-2018

Incorporating Memory Processes in the Study of Early Language Acquisition

Ferhat Karaman
University of Tennessee, fkaraman@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Ferhat Karaman entitled "Incorporating Memory Processes in the Study of Early Language Acquisition." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

Jessica S. F. Hay, Major Professor

We have read this dissertation and recommend its acceptance:

Aaron T. Buss, Devin M. Casenhiser, Gregory D. Reynolds, Shannon M. Ross-Sheehy

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Incorporating Memory Processes in the Study of Early Language Acquisition

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Ferhat Karaman
August 2018

Acknowledgments

I owe a great many thanks all the people who help me throughout my years of graduate school and especially during the writing process of this dissertation. First, I would like to express my sincere gratitude to my advisor, Jessica Hay, for her guidance as well as her continuous patience, support and immense knowledge. I would also like to thank my dissertation committee members: Aaron Buss, Devin Casenhiser, Greg Reynolds, and Shannon Ross-Sheehy for their insightful comments and encouragement despite their busy schedule.

I also ought to thank my fellow labmates, Ryan Cannistraci and Sara Parvenezadeh Esfahani and also our past and current lab managers, Dora Moore, Rebecca Bauer and Daniela Santos Oliveira. I would like to thank the participating families and the past and current members of the Infant Language and Perceptual Learning Lab. I would like to express a special thanks to our research assistants, Ryan Minchley and Maddox Myers who patiently coded my dissertation data. I also would like to thank Turkish Ministry of National Education for their financial support granted for my education. I also would like to thank Turkish Education Attaché, Dr. Hüseyin Tutar and the past and current officers of Office of Education Attaché in Turkish Consulate General in Houston for being so helpful and supportive.

I would like to thank my parents; whose love and guidance are with me in whatever I pursue. Last but not least, I would like to express heartfelt thanks to my love, my friend, my wife, Nuray for her love and never-ending support and my beautiful daughter Ipek for always cheering me up.

Abstract

Critical to the learning of any language is the learning of the words in that language. Therefore, an extensive amount of research in language development has examined how infants learn the words of their language so rapidly. In particular, research on statistical learning has suggested that sequential statistics may play a vital role in the discovery of candidate words, that become available to be mapped to meaning. One important limitation of this previous research is the lack of attention given to the memory processes involved in statistical word learning. Thus, the current set of experiments examine the availability of statistically defined words as object labels after a delay. To examine whether statistics found in speech supports infants' memory for label-object associations, in Experiment 1, 22- to 24-month-old infants were presented with 12 Italian sentences that contained 2 high transitional probability words (HTP) and 2 low transitional probability words (LTP). Ten-minute after familiarization, using a Looking-While-Listening procedure (Fernald et al., 2008), infants were trained and tested on 2 HTP and 2 LTP label-object associations. Results revealed that infants were able to learn HTP but not LTP words, suggesting that HTP words make better labels for objects after a minimal delay. Experiment 2 examined infants' memory for meaning representations that are statistically defined or not. Stimuli and procedure were identical to that of Experiment 1, except that the 10-minute delay was implemented after the referent training phase instead of after the familiarization phase. Infants in Experiment 2 were able to remember both HTP and LTP words when tested following a 10-min delay. Together, the findings suggest that statistical learning facilitates future word learning.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Language learning is one of the hallmark accomplishments of human development. Critical to the learning of any language is the learning of the words in that language. Typically-developing children go from being non-verbal, to producing their first word by the first year of life, to then saying thousands of words by age six. The tremendous speed and efficiency of this transformation have led researchers to search for the remarkable, early-available, abilities that support language development.

In order to build a lexicon, young learners must solve multiple challenging tasks, including, word segmentation – the process of determining where one word stops and the next starts – and word-object mapping – determining which sounds in the environment refer to which objects. In both tasks, young learners have to figure out how to deal with significant ambiguity in the information available in their natural language environment. For example, unlike in written language, the majority of spoken language, even in the form of infant-directed speech, appears to be a continuous speech stream with no reliable acoustic cues that demarcate word boundaries (Cole & Jakimic, 1980). Therefore, breaking up the continuous speech correctly into separate word units is a nontrivial challenge. Similarly, the contexts in which infants learn words often contain multiple candidate labels and multiple object referents. Despite the apparent complexity of these tasks, infants quickly learn what sound combinations form words in their language during the first years of life.

While some theorists argue that children are accomplished at learning language because they possess innate semantic and syntactic primitives (e.g., Chomsky, 1959; Pinker, 1984), others argue that children come to learn their language because they are equipped with general cognitive skills such as intention reading and pattern finding (e.g., Tomasello, 2009). On this account, it is important to explore how infants learn about words and build structure from information gleaned from the speech input. Statistical learning has been proposed as a way infants can track information in the linguistic signal (Saffran, Aslin, & Newport, 1996). Contrary to what Chomsky claims, Saffran and colleagues (Saffran et al., 1996) claimed that infants take advantage of existing general learning capacity that are not domain-specific to discover the structure of human language. A growing body of evidence, especially over the past two decades has confirmed that infants have powerful and robust computational abilities that may allow them to segment word-like acoustic units in both artificial language (Saffran et al., 1996) and natural language (Pelucchi, Hay, & Saffran, 2009a, 2009b) materials. They achieve this feat partly through tracking transitional probabilities that highlight word boundaries (Aslin, Saffran, & Newport, 1998; Saffran et al., 1996). Specifically, researchers have shown that infants as young as 6-8 months can track the transitional probability (hereafter TP) between syllable sequences in fluent speech (Saffran et al., 1996, Thiessen & Saffran, 2003). More importantly, infants use these computational abilities to generate potential candidate words, available for linking to meaning (Graf Estes, Evans, Alibali, & Saffran, 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011).

Despite numerous demonstrations of statistical learning, we know very little about whether and by what means infants' memories for statistical regularities persist and impact future word learning experiences. Critically, statistical learning has been typically tested in the seconds *immediately* after familiarization with an unfamiliar/novel language. However, the process supporting long-term memory unfolds over minutes, hours, and days. To our knowledge, with the exception of our own recent work (Karaman & Hay, 2018), there exists almost no work exploring long-term retention of statistically defined words. This creates a critical gap in knowledge because we do not know how experience with statistical regularities in the input translates into a long-term memory and supports future word learning.

To better understand the relationship between statistical learning, memory, and early word learning, the current set of studies explores whether experience with statistical regularities in natural language supports subsequent word learning following a delay. In this dissertation, I have taken a two-pronged approach to answer this question. The first experiment investigates whether statistically defined words will be treated as object labels after a short 10-min delay. The second experiment investigates whether the meanings of statistically defined words are better remembered than label-object associations where label goodness was not supported by strong sequential statistics. In both Experiment 1 and 2, I used a modified version of a statistical learning + label learning task (Graf Estes et at., 2007) that consists of four phases: familiarization (statistical learning phase), referent training (label-learning), testing (using the Looking-While-Listening procedure), and a 10-min delay period. Critically, while in Experiment 1 the 10-min delay period was inserted between the familiarization and referent training phases, in Experiment 2 the 10-min delay period was inserted between the referent training and testing phases. This manipulation allowed us to examine our main question from two different perspectives: memory for statistics and memory for meanings.

Additionally, both experiments examined the relationship between performance in the word learning task and vocabulary size in order to understand whether infants' expressive vocabulary size is predictive of early word-processing skills. By using a combined methodology of word segmentation and word learning, retention interval, and vocabulary measures, this dissertation aims to shed light on the contribution of statistical learning to an important real-world problem facing infants – remembering words.

Chapter II: Literature Review

In this chapter, I will set the stage for the work presented in this dissertation by providing an overview of what we know about infant statistical learning and the role of statistics in word segmentation and early word learning. I will also provide an overview of research on infant memory, the role of memory in early language learning and, finally, my current research investigating the link between memory and statistical learning.

Statistical Word Segmentation

A fundamental problem that infants face during early language acquisition is discovering the sound sequences that make words in their language. Adding to the complexity of language acquisition is the nature of the speech stream; unlike printed text, there are no clear-cut pauses between words. Thus, in order to identify potential words infants must use information in the speech signal to determine where words start and end.

A substantial literature has demonstrated that natural languages are rife with regularities. Infants are remarkably good at detecting many of these regularities, which is turn guides word segmentation. For example, as infants gain experience with their language, they begin to focus on salient prosodic patterns and make use of this information to build assumptions about the words in their language (Gleitman & Wanner, 1982; Morgan & Demuth, 1996). Prosody refers to the intonation (e.g., declarative versus question sentence types), word stress (e.g., récord (noun) versus recórd (verb)), and rhythm of a language (e.g., English is stress timed; Turkish is syllable timed), and this type of prosodic information often demarcates possible word boundaries and linguistic units. And indeed, infants can employ many of these types of language specific cues to extract words from continuous speech (e.g., Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999). Furthermore, at various ages infants can use native language phonotactic regularities (e.g., Mattys & Jusczyk, 2001; Friederici & Wessels, 1993), vowel harmony (Mintz, Walker, Kidd, & Welday, 2018), and allophonic variation (e.g., Jusczyk, Hohne, & Bauman, 1999) to segment the speech stream (see Saffran & Kirkham, 2017 for a recent review). However, all of the regularities mentioned above are language-specific, and so they cannot kick start word segmentation in young learners who are not familiar with the relevant features of their language. Given that using these language specific regularities in the service of word segmentation requires adequate experience with the ambient language, how do infants solve the word segmentation problem before they have learned the structure and the sound patterns of their native language?

Another source of information that highlights word boundaries is the statistical information which is present in all languages. While there are many different forms of statistical regularities available in languages, transitional probability (TP) is probably the one of the most frequently studied statistical word boundary cues. TP computes relations unidirectionally, but probably based on the idea that the speech unfolds over time, descriptions of TP tend to focus on forward-going TP (e.g., the probability that one syllable will follow another syllable). Forward-TP can be calculated with this equation:

$$\text{Forward TP} = P(Y \mid X) = \frac{frequency\ (XY)}{frequency\ (X)}$$

A seminal study by Saffran, Aslin, and Newport (1996) demonstrated that indeed 8-month-old infants can use TP information to extract syllable sequences from an artificial speech stream. In their study, infants listened to with 2 minutes of a made-up language, and were then tested on what they had learned from the language using a Headturn Preference Procedure

(Jusczyk & Aslin, 1995). TPs between the syllables were sole indicators of word boundaries, such that TPs within words were 1.0 and TPs across word boundaries were .33. In Experiment 1, after brief exposure with the artificial language, infants showed a novelty preference, listening to *nonwords* (novel syllables sequences, '*tilado'*, where the syllables had never occurred together in the corpus) longer than *words* (i.e., high probability sequences, '*golabu',* where the syllables always co-occurred). In Experiment 2, infants again showed a novelty preference, listening to *part-words* (low probability syllable sequences that spanned word boundaries, e.g., *bupado* from [*golabu*][*padoti*]) longer than *words*. These findings indicate that infants can rapidly detect TPs between syllables in artificial speech input.

This striking learning capability (later called *statistical learning*), involving no explicit instruction, feedback, or reinforcement has attracted wide attention especially in the field of language development and has been confirmed in a variety of modalities (e.g., auditory, visual and tactile: Conway & Christiansen, 2005), domains (e.g., music: Saffran, Johnson, Aslin, & Newport, 1999; vision: Fiser & Aslin, 2002; Kirkham, Slemmer, Johnson, 2002; Marcovitch & Lewkowicz, 2009), and species (e.g., rats: Toro & Trobalon, 2005; zebra-finches: Chen, & ten Cate, 2015; bengalise finches: Takahasi 2010; cotton-top tamarin monkeys: Hauser, Newport, & Aslin, 2001; Saffran, Hauser, Seibel, Kapfhamer, Tsao, & Cushman, 2008).

Since the artificial languages used in the early statistical learning studies were monotone, synthesized, and pause- and intonation-free, they lacked the complexity and richness found in natural languages. To overcome this ecological validity problem, Pelucchi and colleagues (Pelucchi, Hay, & Saffran, 2009a, b) increased the complexity in the materials to try to more closely approximate the natural languages infants confronted in their lives. So, for example, instead of presenting babies with artificial language materials, Pelucchi, Hay, & Saffran (2009a) created a natural Italian language where they manipulated the TP between syllables in four target words. Two of the target words had a high TP (HTP; TP = 1.0), as the syllables that made up the words did not appear anywhere else is the corpus. The TP was reduced to .33 in the other two words (low TP, LTP) by inserting extra exemplars of the first syllable throughout the language. Eight-month-old infants first listened this unfamiliar natural Italian corpus for ~ 2 minutes. Then, they were tested on their ability to discriminate high TP (HTP, TP=1.0) from low TP (LTP, TP=.33) words using the Headturn Preference Procedure. Infants showed familiarity preference – looking on HTP word trials longer than on LTP word trials. These results suggest even in these natural language materials, infants can successfully track TP information.

Although statistical learning studies tend to focus on forward TPs, we know that backward TPs are also prevalent in natural languages (e.g., the probability that one syllable was preceded by another syllable). Backward TP can be calculated with this equation:

$$\text{Backward TP} = P(X \mid Y) = \frac{frequency\ (XY)}{frequency\ (Y)}$$

For example, Pelucchi, Hay, & Saffran (2009b) showed that 8-month-olds can also track backward TPs. They created an Italian corpus in which the target words were distinguished solely by their backward TPs, controlling for their forward TPs. While the backward TPs of HTP words were 1.0 and the backward TP of LTP words were .33, the forward TPs of both word types were 1.0. When infants were tested on a word segmentation task, infants again exhibited significantly longer looking times to the HTP words than the LTP words. Taken together with the results of the Pelucchi et al. (2009a), these results suggest that infants are not only tracking the forward TPs but also the backward TPs in fluent speech.

Indeed, a corpus analysis of English infant-directed speech revealed that forward and backward TP are equally informative word boundary cues (Swingley, 1999). Swingley (1999) also suggested that mutual information (e.g., mutual probability of syllables within words) (see also Charniak, 1993) may also highlight words in fluent speech. Mutual information can be calculated with the following equation:

$$\text{Mutual Information} = log_2[P(XY)/P(A)P(B)] = \frac{frequency\ (XY)}{frequency\ (X)\ .\ frequency\ (Y)}$$

Statistical learning may also be facilitated by other cues present in natural languages. For example, a corpus analysis by Brent & Siskind (2001) demonstrated that the number of times a child hears a particular word in isolation is a significant predictor of whether the child knows and uses a given word (see also Fernald and Morikawa, 1993). Indeed, infants hear isolated words such as *mommy* and *daddy* very frequently, and these words are often some of the first words that they learn to produce (Ladd, 1997). To examine whether the presence of isolated words support statistical learning, Lew-Williams, Pelucchi, and Saffran (2011) familiarized 8- to 10-month-old English-learning infants with either only an abbreviated version of the fluent Italian speech stream or a mixture of the shorter Italian speech stream with interspersed isolated HTP and LTP words. They found that only infants who heard the combination of shortened Italian corpus with the isolated target words succeed at differentiating the HTP words from the LTP words at test, suggesting that isolated words may work in concert with sequential statistics in fluent speech to facilitate statistical learning. Further, we recently showed that the presence of isolated words also appears to preferentially strengthen infants' long-term memory for HTP words (Karaman & Hay, 2018). This study is described more fully below, in the section on memory.

A large number of statistical learning studies have focused on *word segmentation.* There is no doubt that infants can track statistical regularities available in speech (and non-speech) input and can make use of this familiarity when discriminating probable and improbable sequences. While that is a very important finding in itself, this discrimination measures do not explicitly tell us whether statistical learning supports word segmentation. Also, from these earlier findings of word segmentation studies, we did not know, if infants are pulling out actual candidate words from the speech stream or whether sequences with strong co-occurrence statistics are just easier to process. For example, when an infant listens to artificial speech stream containing the probable sequence 'timay', and then discriminates 'timay' from improbable sequence 'kuga' during testing, what is the nature of their representations of the word "timay"? Is it a potential word, available to be mapped to meaning? Or is it a familiar sound sequence that is easier to process but does not have a lexical status? The following section reviews the studies that have examined whether statistical learning supports subsequent word learning.

Statistical Word Learning

Extracting words from fluent speech is just one of the key challenges infants face over the course of language acquisition. While infants start to discover words of their language during the first year, the number of words in their lexicons significantly increases during the subsequent year. Acquiring a new word requires linking a sound representation with a meaning representation. In many cases, infants may need to first segment words from fluent speech before they can appropriately form associations between words and their referents in the world (Graf Estes et al., 2007; Hay et al., 2011; for examples of work exploring simultaneous segmentation and mapping see Cunillera, Laine, Càmara, Rodríguez-Fornells, 2010; François, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017; Shukla, White, and Aslin, 2011).

Statistical learning experience may reveal plausible candidate words that are readily available to be linked to meaning. However, to the best of my knowledge, there are only two infant studies (Graf Estes et al., 2007, Hay et al., 2011) and one adult study (Mirman et al., 2008) that have directly tested this potential link. In one study, Graf Estes and colleagues (Graf Estes et al., 2007) combined a statistical word segmentation paradigm (Saffran et al., 1996) with a modified version of a word learning paradigm (i.e., the Switch Paradigm) developed by Werker and colleagues (Werker, Cohen, Lloyd, Stager, & Casasola, 1998) to directly test whether 17-month-old infants treat sequences from fluent speech as candidate labels. Infants were exposed to an artificial sound sequence in which TP was the only cue word boundaries. Infants then entered a habituation-based word-learning phase, in which sound sequences from the speech stream were used to label novel objects. Labels were either words (TP = 1.0), partwords (TP = .5), or nonwords (TP = 0). Immediately after the habituation phase, infants were tested using the Switch task (e.g., Werker et al., 1998) to determine whether they had successfully learned the trained label-object pairs. There were two different types of test trials. On the Same trials – the original label-object combinations from habituation were maintained (e.g., object A combined with label A). However, on the Switch trials – the original label-object combinations were flipped (e.g., object A combined with label B). The logic behind the Switch task is that if the initial label-object combinations were learned, infants should look longer to Switch relative to Same trials, because Switch trials violate the initial associations. Results revealed that while infants looked longer to the Switch trials when the labels were words in the made-up language, they did not differ in their looking times to Switch and Same trials when the labels were non-word or partwords in the made-up language. The results of this study suggest that sequences with the strong co-occurrence statistics (i.e., words) make better object labels that those with weak internal TPs. In this way, the statistical learning experience affected subsequent word learning. Mirman and colleagues (Mirman et al., 2008) performed a similar experiment with adults using a modified version of label-object association task. Adults were able to learn all label-object associations, but associations were learned more quickly when the object labels were sequences with high TP (i.e., words).

However, the results of these adult and infant statistical word learning studies cannot conclusively show that learners use statistical information when learning the words of their language. Both of these studies used a simplified artificial language material. Nevertheless, the results support the hypothesis that infants can make use of statistical information available in speech to extract candidate words, available for subsequent mapping to meanings

To examine whether Graf Estes and colleagues' conclusions can be scaled up to natural language learning, Hay and colleagues (Hay et al., 2011) used speech from an unfamiliar natural language instead of an artificial language. Using the same combination of methods that was successfully used by Graf Estes and colleagues (Graf Estes et al., 2007), 17-month-old infants were exposed to an Italian speech stream and were then tested on their ability to map different words from the speech stream with novel objects. Like in previous work from this group (Pelluchi et al., 2009a, b), the corpus had four target words: two words had high TP (HTP; TP=1.0) as the syllables that made up the words did not occur anywhere else in the corpus, and two had low forward and low backward TP (LTP; TP =.33) because both the first and second syllables occurred many more times throughout the corpus. Results showed that infants readily learned the label-object associations when the labels were HTP syllable sequences (HTP, TP =

1.0) in both the forward and backward direction, but they failed when the labels were LTP syllable sequences (LTP, TP =.33) in both forward and backward directions. Taken together, the findings of these two infant studies indicate that the cohesive statistical structure of the HTP sequences made the words learnable as labels, suggesting that prior statistical segmentation opportunity facilitates subsequent learning of word-object associations.

However, through the second year of life, infants' vocabulary size grows, they process language more quickly, and they learn new words more easily. Further, as they get older, infants become increasingly specialized in the types of sound sequences that they will accept as labels for novel objects. A recent set of studies conducted in our lab (Hay, Shoaib, Wang, Moore, Lohman, & Lany, 2017) tested whether older infants begin to rely less on sequential statistics during word learning. In the first experiment (henceforth referred to as Baseline Study 1), 22- to 24-month-olds were first presented with an Italian speech stream (a new recording of the corpus from Hay et al, 2011, Experiment 3) that had two embedded HTP words and two embedded LTP words and then infants were trained and tested on four novel-object pairings (two HTP and two LTP label-object pairings), using a Looking-While-Listening (LWL) procedure (Fernald, Zangl, & Marchman, 2008), which permits fine-grained analyses of word recognition. Like in the Hay et al., (2011), infants successfully learned the HTP words. However, surprisingly, they also successfully learned the LTP words. There are several possible explanations for the successful learning of LTP words. First, since the infants were a few months older (22- to 24-month olds) than those of Hay et al. (2011) (17-month-olds), it is possible that these more experienced word learners may not have been impacted by the internal co-occurrence statistics of the labels – both labels types had equivalent referential status. A second possibility, is that different factors are driving learning of the HTP vs LTP words. Indeed, vocabulary size differentially predicted word learning in the HTP versus LTP conditions (see below for further discussion of the relationship between vocabulary size and novel word learning), suggesting that different learning processes may have been at play. HTP words may be learned because of their high internal co-occurrence patterns, but LTP words may be learned because their syllables were highly frequent in the corpus. In the second experiment[1] (henceforth referred to as Baseline Study 2), Hay and colleagues (2017) examined whether syllable frequency may have been driving successful learning of LTP words. A different set of test words were created by maintaining syllable frequency from the corpus while violating the co-occurrence statistics of the HTP and LTP object labels (e.g., casa/bici → caci/bisa, TP = 0 for both modified LTP and modified HTP object labels). The authors reasoned that if high TP was driving mapping of the HTP words, then infants should fail to map these modified HTP words where the TP was 0 and the syllables were only heard minimally (i.e.,18 times) throughout the corpus. Conversely, if high syllable frequency was driving mapping of the LTP words, infants should continue to map the modified LTP words where the syllables were heard 54 times each in the corpus. Indeed, this is what was found; infants successfully learned the modified LTP words but failed to show evidence of learning the modified HTP words. Together, the findings from Baseline studies 1 and 2 suggest that, as infants become increasingly more proficient in their native language, they are able to simultaneously take advantage of co-occurrence statistics (i.e., transitional probabilities) and distributional statistics (i.e., syllable frequency) during early word learning.

Individual differences data exploring the relationship between vocabulary size and infants' accuracy on the word learning task also support the findings that HTP and LTP words are mapped for different reasons; while the correlation between vocabulary size and word

---

[1] Since Hay et al. (2017) study provided baseline for my dissertation, Experiment 1 and 2 of this study are hereafter respectively referred to as Baseline 1 and 2.

learning for the HTP and modified HTP words was not significant, vocabulary size and word learning for LTP and modified LTP words were positively correlated. These findings suggest that by 22- to 24-months of age, infants, regardless of vocabulary size, are able to map HTP words to meaning. Conversely, only infants with high vocabulary size were able to map the words with high syllable frequency (i.e., LTP and modified-LTP words) to novel objects. Similarly, a recent study by Shoaib, Wang, Hay, and Lany (in press) also demonstrated that there is an interaction between statistical word learning and vocabulary size. In a similar study, although, overall, infants failed to map both HTP and LTP labels onto novel objects, infants with smaller vocabularies successfully learned the HTP words but not LTP words. Similar to the Hay et al., (2017)'s findings, Lany and colleagues also demonstrated that infants with larger vocabularies were more likely to learn LTP words. It is important to note that the infants in the Shoaib et al. study were slightly younger, and unlike in Baseline studies 1 and 2, they were provided with minimal referential support (see General Discussion for additional comparisons between studies).

Another line of work by Graf Estes, Edwards, & Saffran (2011) investigated whether a different form of sequential statistics – phonotactic probability (e.g., the likelihood of a sound sequence occurs in a given position of a word from a given language) – influence infants' word learning. Eighteen-month-olds were trained on a novel word-learning task using the LWL procedure. Either they heard the novel labels that conformed to attested English phonotactic patterns (*dref* and *sloob*) or the same segments, reordered to create sequences that are unattested in English (*dlef* and *sroob).* Only the infants trained with the legal sequences as labels learned the label-object pairings, as indexed by the accuracy of their visual fixations during testing. Furthermore, they found a correlation between infants' performance and their expressive vocabulary size. While infants with smaller vocabulary size were relatively unaffected by English phonotactics, infants with larger vocabularies showed significant differences when learning phonotactically legal and illegal words – they learned only the phonotactically legal sequences – supporting the view that word learning, and phonological knowledge interact early in language acquisition. Taken together, these word learning studies suggest that statistics (e.g., TPs, phonotactic probabilities, frequencies) available in the ambient language impact early word learning.

The studies described so far have focused on identifying the process involved in resolving the word learning problem in one moment in time, but in real-world learning environments children must contend with significant ambiguity across several moments in time in order to learn words. Thus, in order to learn words, infants must create several hypotheses at a time, encode and store those hypotheses in memory, and compare them with competitive hypotheses across different learning experiences to find the best one. The cross-situational statistical learning studies of Smith and colleagues have demonstrated that both children and adults can track statistical co-occurrence of words and referents (Smith & Yu, 2007, 2008; Yurovsky, Fricker, Yu, & Smith, 2014) but not without memory constraints (Vlach & Johnson, 2013). For example, Smith and Yu (2008) demonstrated that infants can solve the word learning problem by tracking co-occurrence statistics between label and referents over time with multiple exposures. Twelve and 14-month-olds were randomly presented with a series of ambiguous naming events, containing 2 novel referents and 2 novel labels. Across the learning trials, the same word co-occurred with one object. Thus, it was not possible to map the labels onto referents in a single trial. However, by comparing co-occurrences across trials, infants could discover label-referent pairs. Results showed that both 12- and 14-month-olds succeeded in the task, suggesting that infants can track label-referent co-occurrences across trials. This seminal study shed light on the role of cross-situational statistical word learning in lexical development (see Smith, Suanda, & Yu, 2014 for a recent review). However, one recent study by Trueswell, Medina, Hafri, and Gleitman (2014) found that adult learners do not store all possible referents that co-occur with a label in a naming event. Instead, they use a propose-but-verify strategy in which they store and remember a single referent at a time and verify it against alternative referents in new learning situations.

Further, real-world language learning environments, are typically filled with multiple candidate labels and multiple object referents that create another challenge for infants. For instance, Vouloumanos and Werker (2009) investigated how sequential statistics between object labels and referents help infants overcome this challenge. They found that 18-month-old infants successfully mapped object labels onto their referents when the label-referent pairs have a perfect (1.0) and high co-occurrence statistics (i.e., 0.8) but fail to map label-referent pairs when the co-occurring statistics was much lower (i.e., 0.2). The infants also failed to map labels to referents when the same label co-occurred with more than one referent, despite occurring with high probability with one of them (i.e., 0.8) and low probability with the other (i.e., 0.2). This study suggests that the strength of co-occurrence between labels and referents may be an important statistical cue for word learning.

As reviewed above, infants and young children can use their statistical tracking ability in the service of word learning but obviously, tracking and using statistical regularities imposes a significant memory demand on the developing brain. However, how brief word learning experience translates into a memory trace remains poorly understood. Here, I suggest that memory processes are crucial for early language development and should be incorporated in studies of early language acquisition. To that end, in the next section, after briefly reviewing the infant memory development literature, I discuss what we know about infants' memory for words and describe the scant literature on the longevity of statistical language learning.

Memory Processing in Infancy

For years, researchers believed that during the first years of life infants were unable to form memories, because they lacked the capacity to encode information (Mandler, 1998; Nelson, 1990; Piaget, 1952; Pillemer & White, 198). However, the development of various non-verbal tasks (e.g., the deferred imitation paradigm, the mobile kicking paradigm, and the high amplitude sucking procedure) allowed researchers to challenge this assumption (Rovee-Collier, 1999; Rovee & Fagen, 1976; Meltzoff, 1985; 1988). A considerable amount of developmental research on memory using these behavioral measures has made it clear that infants can and do form memories of events: they can remember different kinds of information over a substantial period of time (Rovee-Collier & Fagen, 1976; Rovee-Collier, 1997; Greco, Rovee-Collier, Hayne, Griesler, & Earley, 1986; Rovee-Collier, Hartshorn, & DiRubbo, 1999). However, it is important to note that although infants can retain memories long after exposure in these procedures (Rovee-Collier, 1997), these tasks (especially the deferred imitation paradigm and mobile kicking paradigm) include motor movements and reinforcement during encoding that likely recruit learning systems with different characteristics than the ones underlying statistical learning.

Further, early memory studies show that infants' long-term memory increases with age. As they get older, infants habituate more quickly and efficiently and also remember more information across longer period of times (Hartshorn et al., 1998; Vander Linde, Morrongiello & Rovee-Collier, 1985; Greco, Hayne & Rovee-Collier, 1990; Hill, Borovsky & Rovee-Collier, 1988; Herbert & Hayne, 2000). In addition to the behavioral research, studies from developmental neuroscience that utilize electrophysiology and neuroimaging methodologies (Nyberg & Cabeza, 2000) and studies from behavioral neuroscience that utilize animal models (Nakashiba et al., 2008; Squire 1992) have informed us about how memory systems and brain structures that are associated with memory, change over the course of development (see Bauer, 2004, 2006; see also Gómez, 2017 for a recent review). Indeed, distinct learning systems with different properties of memory develop at different rates – while cortical learning systems are available early in infancy, hippocampal learning systems that are governed by rapid synaptic consolidation and slow system consolidation, matures significantly between 18 and 24 months (Olson & Newcombe, 2014).

Studies from behavioral and developmental neuroscience have also revealed that memory arises from different systems (working memory, short-term memory, and long-term memory), and sub-systems (declarative/explicit memory and non-declarative/implicit memory). Also, each subsystem comprises at least four different sub-processes such as encoding, consolidation, storage and retrieval of the learned material. Encoding refers to the first step in the process of creating a new memory (see Bauer, 2004, 2006, for reviews). Although most statistical learning studies test the encoding of statistical regularities, the term 'encoding' is not usually used in these studies. For example, infants may encode statistical regularities in speech (Saffran, et al., 1996). While encoding is an important aspect of memory, there are other sub-processes involved in forming a memory representation. After encoding, the encoded information must be also consolidated (Davis, Di Bietta, Macdonald, & Gaskell, 2009). The last sub-process is called retrieval, which is the process of getting information out of memory. These memory processes are particularly important for language learning. Supporting evidence comes from neuroimaging studies showing that brain areas consist of specialized memory systems, which seem to be also involved in language learning, particularly during the tracking of statistical information (Schapiro, Gregory, Landau, & Turk-Browne, 2014, Schapiro, Kustner, & Turk-Browne, 2012; Schapiro & Turk-Browne, 2015). For example, the striatum, the medial temporal lobe, and the hippocampus have been observed to be active during almost all types of statistical learning tasks (e.g., word segmentation, word learning, and cross-situational statistical learning tasks) (Berens, 2016; Berens, Horst, & Bird, 2018; Durrant, Cairney, & Lewis, 2012). Additionally, activation in the superior temporal gyrus, the inferior temporal gyrus, and the left inferior temporal gyrus has been reported to be involved in the segmentation of statistically defined words (Karuza, Newport, Aslin, Starling, Tivarus, & Bavelier, 2013; McNealy, Mazziotta, & Dapretto, 2006; Abla & Okanoya, 2008). Thus, integrating memory processes in language acquisition research is absolutely essential to understanding how memory and language interact.

Infants' Memory for Words

Although memory processes play an integral role in building a mental lexicon, surprisingly little research has directly explored the role of memory in word learning. Instead, researchers tend to test infants immediately following training and build theories and derive conclusion based their findings. Before addressing retention of statistical learning in infants, it is important to ask what we can draw from work done on the relationship between language and memory outside of statistical learning. There have been only a few studies in the literature that have directly examined infant's long-term memory for familiar words (Jusczyk and Hohne, 1997; Houston & Jusczyk, 2003). For example, Jusczyk and Hohne (1997) examined infants' memory for familiar words using the Headturn Preference Procedure. Infants were repeatedly presented with three tape-recorded children's stories, 10 times each, over a two-week period. When tested on target words two-weeks later, infants who had listened to children's stories looked significantly longer to the words that occurred more frequently in the stories (i.e., story words) than the words did not appear in the stories (i.e., foils). To ensure that story words were not listened to longer just because they were more interesting than foils, another group of infants who did not listen to the stories were tested. Results showed that these infants listened equally to story words and foils. This study suggests that with sufficient experience, infants can remember the sound pattern of words even after a 2-week delay. In a similar study, Houston and Jusczyk (2003) familiarized 7.5-month-olds with highly frequent isolated English words (i.e., 'feet' and 'bike' or 'cup' and 'dog') for 30 seconds each. They found that a day later, infants looked longer to the sentences comprising these familiarized words than to sentences with non-familiarized words, suggesting that 7.5-month-old infants seem to retain the sound patterns of words in their memory after a 24-hour delay (see Wojcik, 2013 for a recent review).

Taken together, these studies indicate that sufficient experience with the sound properties of words maybe one factor driving long-term memory in infants. Given that in these studies words and foils had different frequencies of occurrence, it is likely that successful retention at test may have been driven by this frequency difference. It is also possible that because the words used in the studies were real English words (such as cup, and dog), infants at this age might be already familiar with these words (Bergelson & Swingley, 2012). Thus, these studies did not provide conclusive evidence about infants' memory for recently segmented novel words.

Another line of evidence about infants' memory for words comes from grammar learning studies. There is converging evidence that infants can maintain simplistic grammatical regularities in their memory over short (e.g., 5 minutes) and long (e.g., 4 to 24 hr) delays. For instance, a seminal study by Gomez and Gerken (1999) showed that infants can remember and generalize grammatical regularities after a minimal delay. Twelve-month-old infants were presented with an artificial speech stream generated by a finite-state grammar. After 2 minutes of familiarization and a short 5-minute play break, infants were successful at discriminating novel grammatical test sequences from ungrammatical ones. Importantly, infants were also successful at generalizing the learned grammatical patterns after the delay.

Gómez and colleagues (Gómez, Bootzin, & Nadel, 2006; Hupbach, Gómez, Bootzin, & Nadel, 2009) have also conducted a set of studies to examine if sleep promotes memory for grammatical patterns. In one study, Gómez and colleagues presented infants with an artificial speech stream with non-adjacent dependencies - a conditional probability between two elements interleaved by at least one additional elements (AxB). Infants succeed at remembering and generalizing the nonadjacent dependencies after 4-hour delay (Gómez et al., 2006). Similarly, another study showed that only infants who napped within the 4 hours delay period remembered abstract grammatical regularities after a 24-hour delay (Hupbach et al., 2009). A recent study by Horváth, Myers, Foster, and Plunkett (2015) found that while 16-month-old infants in both wake and napping conditions did not differ in their immediate performance on a word-object association task, only infants who took a nap within 2 hours of training showed memory for word-object associations when tested 2 hours later. They also found a positive correlation between expressive vocabulary size and performance of infants in the nap group, suggesting that sleep and consolidation are more efficient if there are more representations stored in the memory (see Axelsson, Williams, & Horst, 2016 for recent review on the role of sleep on retention and generalization of words).

There is also evidence that infants can track some statistical relations between words (e.g. the serial order of words within a clause) and remember them over time. For example, infants appear to remember sequential order information between words (Benavides & Mehler, 2015;

Gulya, Mandel, Nelson, & Jusczyk, 1996) when tested following both short (e.g., 2 minute) and long (e.g., 24 hour) delays.

Although the studies described above have demonstrated that infants and children can remember words minutes or weeks after brief familiarization and sleep promotes the consolidation of newly learned rudimentary grammatical patterns, the word learning literature also provides evidence that the memory of newly learned words decays even after short delays (e.g., Bion, Borovsky, & Fernald, 2014; Horst & Samuelson, 2008; Vlach & Sandhofer, 2012; Werchan & Gómez, 2014). For example, Bion and colleagues (Bion et al., 2014) demonstrated that 18- and 24-month-olds were unable to remember recently learned label-object mappings after a short 5-min delay. Similarly, a recent study showed that when 24-month-olds were tested immediately after word-object training they showed evidence of successful learning, however, they performed more poorly when they were test on the label-object mappings after a 5-min delay (Horst and Samuelson, 2008). In a similar vein, Vlach and Sandhofer (2012) tested 3-year-old children's and adults' ability to remember fast-mapped words immediately, after a 1-week delay and after a 1-month delay. The results showed that both children and adults could not remember recently learned label-object mappings as time goes on. However, it is important to note that although forgetting might be detrimental for word learning, a recent study by Werchan & Gómez (2014) showed that forgetting due to wakefulness might be an important factor for promoting generalization of word learning in children.

Horst and Samuleson (2008) and Vlach and Sandhofer (2012) also examined how encoding conditions affect memory for newly learned words. Horst and Samuelson (2008) showed that if the words were labeled ostensively, 24-month-olds (but not 18-month-olds) can remember newly learned words after 5-min delay. Similarly, Vlach and Sandhofer (2012) found that 3-year-olds successfully remember novel words when they were provided additional memory support during the training phase. Together, these studies suggest that providing supporting cues makes memory less vulnerable to forgetting by creating more robust memory representations.

Retention of Statistically Defined Words

Some of the first studies on the retention of sequential statistics come from studies with adults (visual: Arciuli & Simpson, 2012; Kim, Seitz, Freenstra, & Shams, 2009 and auditory: Durrant, Taylor, Cairney, & Lewis, 2011). Adults exhibit equal retention of visually presented shape triples immediately after statistical learning experience and 24 hours later as measured on implicit (Kim et al., 2009) and explicit tests (Arciuli et al., 2012). In a separate study, discrimination of statistically predictable versus unpredictable tone sequences improved after a 24-hour delay (Durrant, et al., 2011). A recent visual statistical learning study also showed that adults remembered sequences even after 1-year delay and the acquired statistical knowledge was resistant to interference (Kóbor, Janacsek, Takács, & Nemeth, 2017).

Although retention of statistical information over a 24 h period is robust in adults, recent research, including some of our own, has shown initially weak memory representations for statistically defined words in young infants (Karaman, & Hay, 2018; Simon et al., 2017). We conducted a set of experiments to examined infants' ability to encode statistically defined words extracted from natural language and remember them after a 10-minute delay (Karaman, & Hay, 2018). Across four experiments, 8-month-old infants were first exposed to Italian speech stream that was comprised of 2 HTP words (TP = 1.0) and 2 LTP words (TP = 0.33). When tested 10 minutes after familiarization, infants failed to discriminate HTP and LTP words, suggesting that memory for TP information likely fades over time. These findings are consistent with cross-situational learning studies (Horst et al. 2008; Vlach et al., 2012) and the findings of Simon et al. (2017) that reported a weak memory for statistical regularities in 6.5-month-olds.

Why did memories for statistical regularities rapidly decay? Brief experience with either an unfamiliar natural language or an artificial language may not be adequate to support robust encoding of statistical regularities. In the light of previous findings (Lew Williams et al., 2011), we hypothesized that having an additional experience with both isolated HTP and LTP words may bolster infants' memory for statistics (Karaman & Hay, 2018). To test our hypothesis, immediately after hearing the Italian sentences, infants were presented with the HTP and the LTP words in isolation using either fixed-trial procedure or an infant-controlled procedure. Importantly, additional presentations of isolated words may have increased the TP of LTP words and hence reduced the TP difference between the HTP and LTP words. Nevertheless, across two separate experiments infants successfully discriminated HTP and LTP words after a 10-minute delay. Like in the previous studies that used the same Italian corpus (Pelucchi et al., 2009), infants again showed a familiarity preference, as evidenced by longer looking to the HTP words than to the LTP words. Together our results suggest that although infants' initially encoded memory representations for statistically defined words were not robust, hearing the words in isolation helped infants built more reliable memories for words with strong TP versus words with weak TP.

Obviously with this limited set of studies on the retention of statistical learning we cannot conclusively claim that infants retain the statistics in the service language acquisition. Nevertheless, these studies raise important questions about whether memory for statistical regularities is prerequisite for further processing (e.g., label-object mapping) that occurs within minutes of initial segmentation. We know that in order to successfully acquire a new word, not only must infants pull words out of the speech stream, but they also need to learn how these words map onto objects and concepts in their environment. In addition to accomplishing these tasks, building a lexicon also requires infants to remember what they have encoded. Remembering words is very crucial for building a vocabulary because objects being talked about at any given time may not be in infants' immediate environment. This is especially true given that infants do not acquire all of the words in their lexicon solely in the context of adult naming contexts where the labels and objects are directly linked (e.g., "put that *xylophone* in the toy basket"). Infants can also learn new words by monitoring others' conversations (Akhtar, 2005; Akhtar, Jipson, & Callanan, 2001; Shneidman, Shimpi, Sootsman-Buresh, Knight-Schwartz, & Woodward, 2009) and thus it might be advantageous for infants to store the newly learned words into long-term memory and remember them over time and in a variety of different contexts.

Despite the importance of memory processes in forming a stable vocabulary, we know little about infants' long-term memory for statistically defined sound sequences because statistical learning and long-term memory in infancy have traditionally been studied separately. In particular, no research has examined the availability of statistically defined words in future word learning environments. This creates a critical gap in knowledge because we do not know to what extent infants may take advantage of their experience with sequential statistics in real speech. Given these factors, it is informative to ask how a memory trace acquired through statistical learning impacts subsequent word learning. Such knowledge is crucial for situating our demonstration proofs of learning (what early language learning researchers have measured in hundreds of labs) within real-world constraints imposed by the infants' developing brain. Inadvertently mischaracterizing early learning by infants' performance on immediate test as a proxy for what they actually remember will limit the applicability of statistical learning as a theory of early language acquisition. Thus, demonstrating that infants have the ability to encode the statistics of sound sequences in real speech into their long-term memory and remember them over time in the service of word learning will support the importance of statistical learning in early language acquisition.

Chapter III: The Current Study

To summarize, the current study aims to examine whether statistical learning found in natural language supports subsequent word learning following a delay. The main hypothesis is that statistically coherent sound sequences such as HTP words will make better object labels following a short 10-min delay than those with weaker internal co-occurrence statistics (i.e., LTP words). In this dissertation, I take two different approaches to address the relationship between statistical learning, word learning, and memory: memory for statistics and memory for meanings. Experiment 1 examines whether infants' memory for transitional probability between syllables during initial segmentation affects their word-object mappings after a 10-minute delay. If infants indeed exploit TPs in the service of discovering candidate words in fluent speech, then it is plausible to assume that the output of TP computations (i.e., high TP words) might be stored in long-term memory for future word learning. To test this hypothesis, infants were familiarized with an Italian corpus similar to the Karaman & Hay (2018). However, in contrast to Karaman & Hay (2018), infants entered a label-object association task instead of a word segmentation task following a 10-minute delay. In Experiment 2, I take a memory for meaning approach to test the hypothesis that statistically defined word meanings are better remembered than object label associations where label goodness was not supported by co-occurrence statistics. To test this hypothesis, as in the Experiment 1, infants were first familiarized with the same Italian corpus. However, unlike Experiment 1, infants were trained on word-object associations immediately following familiarization. Ten-minutes following training with word-object associations, infants were tested on their memory for these newly formed label-object associations.

Specifically, in both Experiment 1 and 2, the primary dependent measure was mean accuracy – the mean proportion of time spent looking to the target object following label onset divided by the total looking time. We calculated the mean accuracy for each participant on both HTP and LTP object label trials during a critical window that began 300 ms following label onset and ended 1700 ms later (ie., at 2000 ms after label onset). In the current work, I also examine infants' reaction times (latency to orient to target object from the distractor object) during the 300-2000 ms critical window, as infants' reaction times are thought to reflect underlying processing abilities (Fernald et al., 1998) and are often correlated with subsequent language outcomes (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006).

To collect accuracy and RT data, in both Experiment 1 and 2, I used a modified version of label-learning task, which uses a Looking-While-Listening procedure to test the learning of the label-object associations. This procedure has been successfully used by numerous researchers, including to test minimal pair label learning (Yoshida, Fennell, Swingley, & Werker, 2009), the effect of phonotactic probability on word-learning (Graf Estes, Edwards, & Saffran, 2011), category learning (Lany & Saffran, 2010), links between infant processing speed and measures of language proficiency (e.g., Marchman & Fernald, 2008), and in recent work from our lab, presented here as Baseline conditions 1 and 2. We used this methodology to test infants' ability to remember label-object associations because accuracy and reaction time data are more sensitive measures of the strength of label-object associations than are data derived from the Switch paradigm (e.g., Yoshida et al, 2009). Also, this methodology enables us to teach infants 4 label-object pairings, and thus we are able to employ a within subjects design.

Importantly, infants in the present study (e.g., 22- to 24-month-old infants) are older than those in the Karaman & Hay (2018) study (e.g., 8-month-old infants). We know that the duration of retention and memory capacity increases with age. Significant maturation of brain areas implicated in long-term memory including the dentate gyrus of the hippocampus (Olson & Newcombe, 2014; Schapiro, Turk-Browne, Botvinick, & Norman, 2017) and the prefrontal cortex occurs during the second year of life and may lead to better memories in older infants (see Gómez, 2017 for a review). Thus, given that the brain areas associated with long-term memory are more mature older in infants, we predicted that they might encode the TP information and word-object associations in their long-term memory more robustly and remember them after a 10-minute delay.

Additionally, in the current study, I also explored the role of vocabulary size on infants' performance in our word learning task. Prior studies have discovered that vocabulary size and cognitive abilities may be interconnected (Marchman & Fernald, 2008). Further, some studies have found correlations between vocabulary size, word learning, and memory abilities (e.g., Bion, Borovsky, & Fernald, 2013; Houston-Price, Caloghiris, & Raviglione, 2010; Lany & Saffran, 2011; Mills, Plunkett, Prat, & Schafer, 2005; Werker, Fennell, Corcoran, & Stager, 2002). Further, with increasing age and language experience, older infants tend to have larger lexicons. Thus, we predicted that as children acquire more words, they should encode, store, and remember the information more robustly.

Experiment 1

Imagine a child hearing a word (e.g., *doggy*) in a fluent speech at home. The child may not associate the word *doggy* with its referent because the referent may not be in child's immediate environment. When the dog comes in to the room the caregiver may to teach the child that the word 'doggy' and the four-legged furry animal go together by pointing at the animal while labeling it ("That's a *doggy*! Look at the *doggy*!"). If the child can remember hearing the word *doggy* previously, then word learning may be facilitated.

To examine infants' memory for statistically defined words, Experiment 1 uses a three-stage task, combining methods from the word segmentation and word learning literatures plus 10-minute retention interval. Twenty-eight 22- to 24-month-old infants first listened to a naturally produced Italian corpus. After a 10-minute delay, they entered a referent training phase, followed by a test phase.

Research Questions and Hypotheses

Question 1: Can infants retain statistical regularities available in speech in their memory and use this information to learn label-object mappings after a delay?

Aim 1: To examine whether statistics (i.e., TP and syllable frequency information) available in speech support retention of HTP and LTP object labels.

Predictions

1) Differential Memory for HTP and LTP object labels

   If TP information is represented more robustly that syllable frequency information, infants will show increased accuracy and decreased reaction times on HTP naming trials, as compared to LTP trials following a 10-minute delay.

2) Similar memory for both HTP and LTP object labels

   If both TP information and syllable frequency information are resilient to decay, we expect both HTP and LTP words to function as object labels following a 10-minute delay. If infants fail to learn both HTP and LTP mappings, this would suggest that neither HTP or LTP words would be mapped to meaning following a 10-minute delay.

Question 2: Is there a relationship between infants' vocabulary size and retention of HTP and LTP object labels?

Aim 2: To examine whether the expressive vocabulary size as measured by MCDI is predictive of retention of label-object associations.

Predictions:

1) The relationship between vocabulary size and accuracy

   There would be a positive correlation between accuracy and vocabulary size with larger vocabularies would showing greater accuracy.

2) The relationship between vocabulary size and RT

   There would be a negative correlation between reaction time and vocabulary size with larger vocabularies would find the target objects more quickly.

Method

Participants

Twenty-eight 22- to 24-month-old infants (M_{age}=23.24 months, range = 22.59 - 24, 12 females, 16 males) participated in Experiment 1. Parents indicated that their children were born full-term with no hearing or vision impairments. Twenty-two to 24-month-olds were chosen because we knew from the preliminary research done in our lab that infants within this age range can successfully map statistically defined words onto novel objects when trained and tested immediately after familiarization. The Child Development Research Group's database was used to recruit the participants. All parents signed consent forms. Participants received either book or t-shirt for their participation. Required approvals were obtained from Institutional Review Board at the University of Tennessee. Data from 19 additional infants were not included in the analysis due to the following reasons: fussiness, including whimpering and/or continuous crying leading to a failure to complete familiarization (2), training phases (3) or at least 6 of the 12 HTP or 12 LTP test trials (7), not paying attention as reflected by failure to orient to the TV screen (3), parental interference (e.g., giving pacifier to the infant during the experiment) (2), and experimental error (2). The attrition rate is slightly higher than or comparable to prior studies on word learning that used more pared-down tasks [Graf Estes et al., 2007 (17-mo, n=28, $n_{excluded} = 13$); Hay et al., 2011 (17-mo, n=40, $n_{excluded} = 15$); Wojcik & Saffran, 2015 (27-mo, n=24, $n_{excluded} = 13$); Lany et al., 2018 (20-mo, n=37, $n_{excluded} = 17$). High attrition rates might be due to our three-stage task, combining methods from the word segmentation and word learning literatures.

Stimuli

Auditory Stimuli

The language used during familiarization phase (i.e., word segmentation task) consisted of 12 Italian sentences (see the Appendix for the list of sentences) taken from Hay et al., (2011, Experiment 3). These grammatically accurate and semantically meaningful sentences were produced in an infant-directed manner by a native Italian who was blind to the purpose of the experiment. All sentences were intensity normalized to ~ 65 dB_{SPL}. A counterbalanced language, where HTP and LTP words were switched, was created to control for arbitrary label preferences. The familiarization language was presented 3 times for a total duration of 2 min and 30 s.

Four trochaic (i.e., strong-weak stress pattern) Italian bisyllabic target words: *bici*, *casa*, *fuga*, and *melo* (English translations respectively: *bike*, *house*, *escape*, and *apple tree*) were inserted in the corpus. Target words were comprised of phonetically and phonotactically permissible sequences in English (i.e., they contained sound sequences that occur in English), although their realization may have sounded non-native to the infants. Table 1 shows the phonotactic probability of both familiar and novel target words. I used an online phonotactic probability calculator to obtain each value (Vitevitch & Luce, 2004). Also, all target words followed a strong/weak (trochaic) stress pattern characteristic of English bisyllabic words (Cutler & Carter, 1987). Although English and Italian share a stress pattern in bisyllabic words, there are significant phonotactic, allophonic, and rhythmic variations across the two languages. Thus, as a whole, the languages that we used are likely to have sounded very novel to our monolingual English-learning infant participants.

Two HTP and 2 LTP words were presented 6 times in the corpus (18 times across the 3 presentations of the corpus), but importantly they differed in their internal transitional probabilities. In one of the counterbalanced languages, the syllables of the target words *fuga* and *melo* (i.e., *fu*, *ga*, *me*, and *lo*) appeared only in the words *fuga* and *melo*, and never appeared anywhere else in the language and thus, the TPs of these words were 1.0 (HTP words). Conversely, both the first and second syllables of two other words, *bici* and *casa* (i.e., *bi, ci, ca, sa*) appeared 12 additional times throughout the corpus (36 additional times across the 3 presentations of the corpus), and thus, the TPs of these words were .33 (LTP words). For example, to lower the backward and forward TPs of the LTP word '*bici',* 12 additional occurrences of '*bi'* in the stressed position and 12 additional occurrences of '*ci'* in the unstressed position were embedded within the Italian corpus. While the other counterbalanced language had the same structure, the HTP and the LTP words were flipped.

During referent training task, novel object labels (*bici, casa, fuga,* and *melo*) and familiar object labels (*baby*, *doggie, shoe*, and *book*,) were presented in isolation and also embedded in common naming carrier phrases (e.g., Bici! "See the bici! It's a bici! Bici!). At test, novel and familiar object labels were also embedded in carrier phrases (e.g., "Find the casa! Casa! Do you see it?", or "Where is the fuga? Fuga! Do you like it?". The length (500 ms) and intensity (~ 65 dB_{SPL}) of the test words were matched to ensure that acoustic differences between words did not affect infants' ability to map them to meaning.

Visual Stimuli

During familiarization phase, we used a silent cartoon video (Winnie-the Pooh) to attract infants' attention. Visual stimuli used during referent training and test trials consisted of colorful images of novel and familiar objects. Size and brightness of images were matched. To maintain infants' attention throughout the study, four different visual stimuli were used (*baby, doggie, shoe,* and *book*). Table 2 shows the images of familiar and novel objects with their paired labels.

To help capture infants' attention, in each referent training trial, a single object image moved across a white, rectangle-shaped box (~ 8"x 6") that appeared on either the bottom right corner or the bottom left corner of the screen, while the object was labeled ostensively. The movement of the object and timing of the naming was not tied to each other. On each training trial, the infant saw the image for 500 ms prior to the onset of the carrier phrase and for 1 s after the offset of the speech stimuli.

However, on each test trials, the target and the distractor object were shown at the same time on the screen for 500 ms prior to the onset of the carrier phrase and for 2 s prior to the onset of the target label in order to provide the infant with enough time to look at both images. The images stayed on the screen for 500 ms after the offset of the speech stimuli.

Apparatus

Both the word segmentation and the label-object learning phases were conducted in a sound-proof booth. The interior walls of the testing booth were covered with black curtains. To provide the most interesting visual for infants, we used low level-light in the booth. All visual stimuli were presented on a 106 cm Panasonic flat screen TV screen with a resolution of 1,024 x 768 pixels per inch and a refresh rate of 60 Hz. TV screen was located approximately 90 cm away from the infant's face. A hidden video camera below the TV screen recorded and relayed the infant's looking behavior to the experimenter in the adjacent room.

The speech stimuli were presented via two hidden loudspeakers, located behind the television screen, played at approximately 65 dB. The experiment was run from an adjacent control room using a MATLAB-based program (WISP) via a PC computer. The video of the infant's eye gaze with the experiment information (subject number, experiment condition, training and test phase trial numbers) and a timestamp was saved to the software program iMovies on an Apple© MacMini desktop computer in the control room.

Procedure

Before entering the booth, all experimental procedures were explained to the parent who then signed the consent form. Infants were seated on the parent's lap. To avoid any possible biases, the experimenter was blind to the auditory stimuli presented, and the parent listened to music during the experiment via Sennheiser studio monitoring headphones.

Each experiment consisted of four phases: familiarization (see Figure 1), 10-minute delay, referent training (see Figure 2) and test (see Figure 3). Figure 4 shows the overview of the experimental design at Experiment 1. While infants were listening to the speech stimuli during the familiarization, they were also presented with an unrelated silent video of the same duration to maintain their attention. After, infants were familiarized with the corpus, we implemented a 10-minute break where infants played with toys in the play area, while the parents completed the demographic information questionnaire and the MacArthur-Bates Communicative Developmental Inventory (MCDI).

Following the 10-minute delay, the parent and the infant returned to the booth for the referent training and test phases. When the infant looked at the attention-getter video (e.g., spinning pinwheel, the referent training phase started. During the referent-training phase, infants were presented with four novel word-object pairs (i.e., 2 HTP and 2 LTP object labels paired with novel objects) and 4 familiar word-object pairs. On each trial, a single moving object was presented on the TV screen while its corresponding label was presented. Each training trial started with an English phrase (e.g. "See the" or "Look at the") followed by 2 repetitions of the either a familiar target word or a novel target word. The training trials were randomized by block for a total of 20 referent training trials (4 familiar, 16 novel).

Finally, infants' label-object associations were tested using an LWL procedure (Fernald et al., 2008). On each test trial, infants were simultaneously presented with two stationary objects side by side on the TV screen. In order to ensure that, on any given trial, infants were equally as likely to have learned the label for both objects, objects were yoked. Thus, on HTP trials, the two objects that had been paired with the HTP labels appeared on the screen together, with one functioning as the target and the other the distractor. Similarly, on LTP trials, the two objects that had been paired with the LTP labels appeared on the screen. On each familiar label-object trial, objects were yoked based on their animacy (e.g., shoe-book, baby-doggie), and were presented side by side on the TV screen. In order to correctly code infants' eye gaze and shifting, objects were placed at in the bottom left and right corners of the TV screen with approximately 60 cm between them. Five hundred milliseconds after the objects appeared on the screen, infants were presented with an English carrier phrase (e.g. "Where's the" or "Find the"), followed by either the familiar target object or the novel target object (e.g., HTP and LTP objects). The onset of the target word always occurred at exactly 2 seconds after the beginning of the trial. Additional repetition of the isolated target word was presented at 1.5 seconds after the first target word onset. Five hundred milliseconds later infants heard another phrase (e.g. "Do you like it?" or "Do you see it?") and then the trial ended. Figure 5 shows the timeline for the 8-second long test trial.

To accustom infants with the format of the LWL task, the test phase started with 2 trials of objects and labels that are highly familiar to the infants of this age (e.g., doggie, baby, book, shoe). After these familiar word trials, trial type was counterbalanced in quasi-random testing orders for a total of 33 testing trials (8 familiar word trials, 12 HTP and 12 LTP target word trials, and an attention-getting whoopee trial). The whoopee trial, that consisted of 2 stimulating videos and a fun phrase (e.g. "Good job! You're doing great!"), was presented halfway through the testing phase to maintain interest in the task. Importantly, there were four pseudo-randomized testing orders: each label was tested on the left and right side of the TV screen an equal number of times and no labels occurred twice in succession. With the 10-minute delay period, the entire experiment lasted about 20 minutes.

Vocabulary Measures

We collected the McArthur-Bates Communicative Developmental Index (MCDI; Fenson et al., 2006) of expressive vocabulary data for each infant. The expressive vocabulary indicates the number of words the infant says. During the 10-minute delay period, the parent completed the infant short form (Level II, Form A, for 16-30-month-olds) that contained a 100-word vocabulary production checklist. The vocabulary scores for the infants in Experiment 1 ranged from 5 to 98[2]. The age and gender normed vocabulary percentiles ranged from 2 to 99. Further, each infants' raw expressive vocabulary scores were converted to age- and gender-normed vocabulary percentiles using normative tables provided by Fenson, Pethick, Renda (2000). This

[2] One of the female participants' vocabulary data could not be located in our vocabulary database (n =27).

parental report allowed us to investigate significant correlations between performance on our tasks and early language skills.

Coding

Coding Software and Coders

Videos of the infants' eye gaze were coded offline frame-by-frame by a trained research assistant with standardized coding protocols using iCoder which is a custom-made software developed by Anne Fernald's Language Learning lab at Stanford University (Fernald et al., 2008). We used eye gaze data to calculate accuracy within the critical window (300 ms to 2000 ms after word onset) and reaction time (i.e., time it takes infants to shift their eye gaze from the distracter to the target following word onset).

Pre-Screening of Trials

In order to exclude trials that should not be coded and save time for coders, each experimental session was prescreened using the iCoder software (Fernald et al., 2008). Importantly, during prescreening, the coder was blind to the side of the target presentation and the trial type. However, in order to identify if the infant or the caregiver talked during the video, the coder had access to the sound from the test booth.

There were 4 main trial exclusion criteria: 1) noise (e.g. if infant or caregiver was talking at target word onset) during the critical window 2) if the infant was not interested in the trial (i.e., not looking at either the target or the distractor for 15 or more consecutive frames, or 500 ms, during the critical window) 3) if the infant did not look at either target or the distractor object prior to target word onset 4) eyes not visible (e.g., if the both eyes were not visible, the trial was excluded but if the coder could see at least one eye, the trial was kept). Due to the reasons mentioned above, approximately 8% of the testing trials were excluded from the analysis (53 out of 672 total novel word test trials were excluded). Seven participants were excluded because we could not keep enough trials from their session.

Coding the Test

During coding, the coder did not have access to the sound from the booth and was also blind to the side of the target object presentation and the trial type (HTP vs LTP vs Familiar). The videos were coded from the coder's perspective. Each entry was displayed in four different columns in iCoder: trial number, trial status (on/off), response (right, left, off, away), and timecode. The coder indicated the changes in the infant's visual fixations on each frame with 4 possible eye gaze responses: 1) left (e.g., when the infant was looking at the object on the left), 2) right (e.g., when the infant was looking at the object on the right), 3) off (e.g., when the infant began to shift their eye-gaze off of one of the objects), and 4) away (e.g., when the infant was not looking at either the target or the distractor objects).

After each video was coded on iCoder, another custom-made software (DataWiz) which was also developed by Anne Fernald and colleagues (Fernald et al., 2008), was used to gather the collate the data. The DataWiz software allowed us to export an excel-formatted spreadsheet of group data (iChart) that was used for data analysis. The iChart data was used for summarizing and plotting the data in R software (R Core Team, 2017).

Reliability Coding

Approximately 25 % of the data (n=8) were coded by a second coder to check the intercoder reliability and 99.12% frame agreement and 99.83% shift agreement were obtained.

Dependent Measures

Infants' looking behavior in response to novel (HTP and LTP words) and familiar words were assessed using accuracy and reaction time (RT) measures. If the infants were looking at the distractor object at the onset of the testing trial, for a correct response, the infants should immediately shift their gaze from distractor to the target object when they hear the target word. However, if the infants were already looking at the target object at the onset of the test trial, for a correct response, the infants should not shift their gaze and should continue looking at the target object. Different studies have chosen the time windows for the analyses in different ways. In the current study, based on the ages of the infants and the complexity of the stimuli and the procedures, 300-2000 ms post-naming time window was chosen to analyze the data (Fernald, Perfors, & Marchman, 2006; Fernald et al., 2008). We excluded the first 300 ms following label onset from the analysis window to account for the time taken to initiate an eye-movement in response to hearing the target word. The critical window ended at 2000 ms because infants' looking behavior after 2000 ms may not be tied to auditory stimuli presented (Fernald et al., 2008; Swingley & Aslin, 2000; Swingley, Pinto, & Fernald, 1999).

Mean Accuracy

Accuracy or the proportion of time spent looking to the target object represents the reliability of infants' looking to the target object during the 300-2000 ms critical window. DataWiz software calculated the mean accuracy for each infant as the mean proportion of time spent looking to the target object divided by the total looking time (e.g., the mean proportion of time looking to the target object or to the distracter object). In accuracy analysis, all codable trials were included regardless of initial looking location (either target object or distractor object) at target word onset. Away trials at the word onset were also included in the accuracy analyses.

Reaction Time

Reaction time (RT) indicates the time taken to initiate a shift to the target object from the distractor object within the 300-2000 ms critical window. Differently from the accuracy analysis, only the trials on which infants were looking at the distractor object at the target word onset were included in the RT analysis. Thus, the target initial trials and away trials were excluded from the RT analysis.

Results

Familiar Words

Mean Accuracy

Univariate analysis of variance (ANOVA) showed that there were no significant main effects of age, F (1,14) = .54, p =.48, gender F (1,14) = .53, p =.48, counterbalanced language, F (1,14) = .19, p =.66, and order of presentation, F (1,14) = .09, p =.77, on accuracy. Thus, all the data were collapsed across these variables in subsequent accuracy analysis of familiar words.

To validate that infants had no preference before the word onset, I first examined infants' mean proportion of fixations on test trials during the baseline window (-2000 to 0 ms from familiar word onset). A one-sample t-test revealed no difference from chance level (50%) for the familiar words (M = 48%, SD = 7%), t (27) = -1.61, p = .12 (All t-tests are two-tailed, and effect sizes reported for significant t-tests are Cohen's d). However, the mean accuracy for the familiar words (M = 60 %, SD = 8.5%) within critical window (300-2000 ms) was significantly above chance level, t (27) = 6.23, p < .001, d = .85. In addition, planned comparison were performed to better understand the difference in baseline and critical windows. A paired sample t-test showed a significant increase in looking to the target object from the baseline window to the critical window, t (27) = -6.24, p < .001, d = 1.19 (see Figure 6 and Figure 7).

In addition to the profile plot, I have also created an *onset-contingent* plot (*OC*-plot) that tracks separately the time course of infants' looking patterns for target- and distractor-initial familiar word test trials (see Figure 8). This type of plot helps us to understand differences in how infants' gaze shifts relative to where they were looking at label onset. At the beginning of a test trial, infants have no way of knowing which object will be labeled, so they may be looking at the target or distractor object at the onset of the target word. On distractor-initial test trails, the infants should rapidly shift their eye-gaze away from the distractor object to the target object after the onset of the target word. However, on target-initial test trials, the infants should not shift their eye-gaze but stay on the target object after the onset of target word. The response pattern in Figure 8 clearly shows that infants showed considerably more shifts from the familiar distractor objects to the familiar target objects after the onset of the target word. The difference in target-initial and the distractor-initial trial trajectories suggests that 22-24-month-olds recognized our familiar words.

Reaction Time

Infants' latency to shift their eye-gaze from the distractor object to the target object can only be measured for the trials on which infants were initially looking at the distractor object at target word onset. As infants do not know which object will be labeled prior to the onset of the label, they were equally likely to be looking at distractor and target objects, thus this necessarily limits the number of trials that can be included in the RT analyses. Any participant that had less than 3 RT trials were excluded from the RT analyses[3]. When applying these criteria, 6 infants from our final sample were excluded, leaving 22 infants.

Univariate analysis of variance (ANOVA) showed that there were no significant effects of age, F (1,10) = 2.45, p =.15, gender F (1,10) = 2.36, p =.16, counterbalanced language, F (1,10) = .32, p =.59, and order of presentation, F (1,10) = 3.68, p =.08, on reaction times. Thus,

[3] Some studies using the similar procedures and the materials (e.g., Pomper & Saffran, 2018) used different exclusion criteria – excluding the infants that had less than 2 RT trials in RT analyses. In our study, if the subject number is too small after applying our ≤3 RT trial criteria, we also analyzed the RT data for the infants that had at least 2 RT trials.

all the data were collapsed across these variables in subsequent reaction times analysis of familiar words. Infants' average RT to orient to familiar target object from the distractor object was 931.46 ms (*SD* = 213.685).

Novel Words

Mean Accuracy

I first performed a repeated-measures analysis of variance (ANOVA) with object label type (HTP vs LTP) as a within-subject factor and age, gender, counterbalanced language, word order as between-subject factors. Repeated-measures ANOVA revealed that there were no main effects of age, $F(1,14) = .02$, $p = .89$, gender, $F(1,14) = 4.04$, $p = .064$, counterbalanced language, $F(1,14) = .004$, $p = .95$, and order of presentation, $F(1,14) = .29$, $p = .59$, on accuracy. Thus, all the data were collapsed across these variables in subsequent accuracy analysis of novel words (e.g., HTP and LTP words) (see Table 3).

To validate that infants had no object preferences before target word onset for any of the matched pairs (e.g., novel target word vs. novel distractor word), I first examined infants' mean accuracy for both HTP and LTP test trials during the baseline window (-2000 to 0 ms from novel target word onset). One-sample t-tests revealed no difference from chance level (50%) for either the HTP words (M = 50%, SD = 7%), $t(27) = .34$, $p = .74$, or the LTP words (M = 49%, SD = 7%), $t(27) = -.55$, $p = .12$. Also, a paired sample t-test showed that the mean accuracy for the HTP and LTP test trials during baseline window did not differ from each other, $t(27) = 63$, $p = .53$.

However, the mean accuracy during the critical window (300-2000 ms) was significantly above chance for HTP, $t(27) = 4.41$, $p < .001$, $d = 1.20$, but not for LTP object labels, $t(27) = -.21$, $p = .83$, suggesting that infants learned HTP object labels but not LTP object labels. Further, infants were significantly more accurate on HTP (*M* = 59%, *SD* = 11%) than LTP trials (*M* = 49%, *SD* = 13%), $t(27) = 3.97$ $p<.001$, $d = .76$, suggesting that HTP words made better object labels than LTP words following the 10-minute delay (see Figures 9 and 10).

In addition, planned comparison were performed to better understand the difference in infants' mean accuracy during the baseline and the critical windows. A paired sample t-test showed that the mean accuracy for HTP words during the baseline window and the critical window were significantly different from each other, $t\ (27) = -3.17$, $p < .01$, $d = .61$, indicating that infants' mean proportion of fixations for HTP words were higher in the critical window than in the baseline window. However, the mean proportion of fixations to the LTP object on test trials averaged across the baseline window and the critical window were not different from each other, $t\ (27) = -.09$, $p = .93$ (see Figure 11).

In addition, I created an *OC*- plot (see Figure 12) to examine infants' looking patterns for the target- and distractor-initial test trials on HTP and LTP words. It appears that on HTP object label trials, the infants showed more switches from the distractor object to the target object after the onset of the target word than from the target object to the distractor object. This suggests that the infants were able to map the HTP labels onto their referent objects. However, on LTP object label trials, we do not see such separation between distractor- and target-initial trials, supporting the finding that infants failed to learn LTP object labels.

Reaction Time

When applying our ≤ 3 RT criteria, 17 infants from our final sample were excluded, leaving 11 infants. I first performed a repeated-measures analysis of variance (ANOVA) with object label type (HTP vs LTP) as a within-subject factor and age, gender, counterbalanced language, word order as between-subject factors. Repeated measures ANOVA revealed that there were no main effects of age, *F* (1,2) = .03, *p* =.89, gender, *F* (1,2) = .001, *p* =.98, counterbalanced language, *F* (1,2) = .29, *p* =.64, and order of presentation, *F* (1,2) = .005, *p* =.95, on RT. Thus, all the data were collapsed across these variables in subsequent RT analysis of novel words (e.g., HTP and LTP words).

A paired sample t-test revealed that there was no RT difference for HTP (*M* = 991.61 ms, *SD* = 123.31 ms) and LTP (*M* = 1039.74 ms, *SD* = 293.38 ms) words, *t*(10) = -.45, *p* = .66, suggesting that infants' time to orient to the target object from distractor object on HTP and LTP trials were not different from each other.

Correlations Between Accuracy and RT

Familiar Words

Since RTs could only be computed for a subset of trials, for our RT analyses of familiar words, I only included infants who had at least 3 RT testing trials (N = 22). When performing Pearson's correlations, no significant correlation was revealed between accuracy and RT for familiar words, *r*(22) = - .20, *p* =.38. When I include only infants who had at least 2 usable RTs (N = 27), I again found no significant correlation between accuracy and RT for familiar words, *r*(27) = - .18, *p* = .36 (see Figure 13).

Novel Words

Infants who had at least 3 usable RTs in either HTP or LTP testing trials were included in the correlation analyses. When applying these criteria to testing trials, infants who had at least 3 RTs in HTP (N = 13) and LTP testing trials (N = 19) were included. When performing Pearson's correlations, I found no correlation between infants' accuracy and RTs for either HTP and LTP trials (see Figure 11). However, when I include only infants who had at least 2 usable RTs (N = 23), I found a marginally significant negative correlation between infants' accuracy and RTs for

LTP testing trials, $r(23) = -.366$, $p = .086$ (see Figure 11), suggesting that infants who had higher accuracy on LTP words were faster to orient to target LTP object from distractor object (see Figure 13).

Correlations Between Word Processing and Vocabulary

Since infants were expected to learn novel words, I predicted that infants' expressive vocabulary size and also their age-normed vocabulary percentiles will be positively correlated with mean accuracy for familiar and novel words (HTP and LTP words) and negatively correlated with reaction time.

Familiar Words

To examine whether the speech processing of familiar words was related to infants' vocabulary size, I performed Pearson's correlations. Individual differences in the size of infants' expressive vocabulary size did not predict their accuracy and RTs in learning of familiar words. Neither infants' raw expressive vocabulary scores nor their age-normed vocabulary percentiles were correlated with either accuracy (see Figure 14) or RTs (see Figure 15) for familiar words.

Novel Words

Infants' expressive vocabulary scores and their aged-normed vocabulary percentiles were not correlated with their accuracy on both HTP and LTP words. I also looked at the correlations between expressive vocabulary size and RTs and aged-normed vocabulary percentiles and RTs for HTP and LTP words. I found that infants' age-normed percentile on measures of expressive vocabulary size were significantly correlated with their RTs for HTP words, $r(11) = -.644$, $p = .02$, suggesting that infants with high vocabulary percentile were faster than infants with low vocabulary percentile on their RTs to orient to the target HTP objects from the distractor objects (see Figure 14 and 15).

Discussion

Familiar Words

In order to orient the infants to the format of the label-object association task, we used four English target words (e.g., shoe, book, baby, doggie). I predicted that infants' mean proportion of fixations during the 300-2000 ms critical window should be above chance level. When I analyzed infants' performance on familiar words, as I predicted, their accuracy for familiar words were significantly above chance. Since 22- to 24-month-old infants should already be familiar with our English target words, these findings are not surprising.

To better understand familiar word processing, I also analyzed at the RT measures. I expected that infants would show faster RT to familiar words compared to novel words because the processing demand for familiar words are minimal. However, the pattern of result was different than I expected – there was no difference between infants' average RT to orient to familiar target object (M = 931.46, SD = 213.69) and either novel HTP (M = 991.61, SD = 123.31) and LTP object labels (M = 1039.74, SD = 293.38).

Although the primary purpose of using familiar words was to familiarize infants with the nature of the label-object association task, I was also interested in to see how familiar words are processed and whether infants' performance on familiar words are predicted by their vocabulary development. Results showed that infants with larger vocabularies showed a tendency towards recognizing the familiar words better than the infants with smaller vocabularies, $r(27) = 282$, p=.15. This result is consistent with the previous findings that have demonstrated a relation between vocabulary development and accuracy comprehension of familiar words (Fernald, Perfors, & Marchman, 2006; Grieco-Calub et al., 2009). However, I found no correlations between accuracy and RT between vocabulary size and RT for the familiar words. Given that previous studies have found a relation between vocabulary development and the processing of familiar words (Fernald, Perfors, & Marchman, 2006; Grieco-Calub et al., 2009), the lack of relation in the current experiment is surprising. I think that this unexpected pattern of results in RT analyses might be due to a lack of power. This is discussed further in General Discussion.

Novel Words

I first analyzed infants' accuracy (mean proportion of fixations during the 300-2000 ms critical window) on HTP and LTP object label trials. Infants were significantly above chance on HTP object label trials, but the same infants did not perform significantly above chance on LTP object label trials, suggesting that they were able to remember HTP but not LTP object labels 10 minutes after familiarization with the Italian corpus. We also found that infants were significantly more accurate on HTP than LTP trials, suggesting that HTP words make better object labels than LTP words following a 10-minute delay.These results are consistent with previous findings showing that statistically coherent sequences (i.e., words, HTP words) make better object label than less predictive sound sequences (i.e., partwords, LTP words) when 17-month-old infants were trained and tested immediately after familiarization with either artificial (e.g., Graf Estes, et al., 2007) or natural languages (e.g. Hay, et al., 2011). Taken together with the findings of Baseline Study 1 (Hay et al., 2017), Experiment 1 suggests that while the representations of TP information are maintained in long-term memory and remain available to support word-object associations, memory representations of syllable frequency information may decay more quickly.

In addition to accuracy measures, I also analyzed the RT measures for novel words. In Experiment 1, infants' RT to orient to HTP (*M* = 991.6, *SD* = 123.31) and LTP object labels (*M* = 1039.74, *SD* = 293.38) from the distractor objects were not significantly different from each other. Further, I found no significant correlation between infants' accuracy and RTs on both HTP and LTP object label trials. However, when I include only the infants who had at least 2 usable RT trials (n= 23) instead of 3 RT trials (n=13), I found a marginally significant correlation between accuracy and RT (p = .086), suggesting that infants who had higher accuracy on the LTP words were faster to orient to target LTP objects from the distractor objects.

In Experiment 1, correlation analyses for the novel words revealed no correlation between vocabulary and novel word accuracy on either HTP and LTP object label trials. Thus, infants' vocabulary size did not predict their performance on HTP (p = .77) or LTP words (p = .29). Results from the HTP words are consistent with the results from Baseline study 1 and 2 (immediate tests), suggesting no correlation between vocabulary size and accuracy on HTP words. However, infants' vocabulary size and their accuracy on LTP words were significantly correlated in immediate tests reported from Baseline studies 1 and 2.

Together with the previous findings, the results of the Experiment 1 demonstrate that once extracted from the Italian corpus HTP syllable sequences appear to be maintained in memory across a10-minute delay and serve as potential candidate object labels that are available to be linked to meaning. However, we do not know if the meaning representations that are supported by the statistics (i.e, HTP words) will be also maintained in memory better than the meaning representations that are not supported by statistics 10 minutes after training on the label-object associations. Can infants still remember the HTP words better than LTP words when tested following a 10-minute delay, if they were trained on the label-object associations immediately after familiarization? By testing a new sample of 22- to 24-month old infants with a 10-minute delay between referent training and test, Experiment 2 aims to shed light on infants' memory for meaning representations that are statistically defined or not.

Experiment 2

Imagine a child hearing a word (e.g., *doggy*) in a fluent speech and then mom subsequently tries to teach her child that the word 'doggy' and the four-legged furry animal go together by pointing at the animal while labeling it (e.g., "That's a *doggy*! Look at the *doggy*!"). Learning this one meaning association is an important step to learning what *doggy* means, but the child must move beyond that specific learning moment because the referent might go out of the child's sight and may not be available for future mappings. Will the child remember the meaning of the word in future word learning context when the referent is next available?

To answer this question, a new group of 22- to 24-month-olds were first presented with an Italian corpus. Immediately after familiarization, infants were trained on word-object associations. Following a 10-minute delay, infants were tested on their ability to remember the label-object association.

Research Questions and Hypotheses

Question 1: Are statistically defined word meanings better remembered than label object associations where label goodness was not supported by sequential statistics (i.e, but TP)?

Aim 1: To examine whether HTP word meanings will be better remembered than LTP word meanings 10 minutes after they were trained on label-object associations.

Predictions

1) Differential Memory for HTP and LTP object labels

   If object labels with strong sequential statistics (i.e., HTP words) facilitate the formation of more robust label-object associations than do labels with less strong sequential statistics (i.e., LTP words), infant should show increased accuracy and decreased reaction times to look HTP object labels, as compared to LTP object labels 10 minutes after they were trained on label-object associations.

2) Similar memory for both HTP and LTP object labels

   If the robustness of label-object associations are independent of the statistical structure of the labels, once they are formed, we expect infants to demonstrate similar memory for both HTP and LTP object labels 10 minutes after they were trained on label-object associations.

Question 2: Is there a relationship between infants' vocabulary size and retention of HTP and LTP word meanings?

Aim 2: To examine whether the expressive vocabulary size as measured by MCDI is predictive of retention of label-object associations.

Predictions:

1) The relationship between vocabulary size and accuracy

   There would be a positive correlation between accuracy and vocabulary size with larger vocabularies would showing greater accuracy.

2) The relationship between vocabulary size and RT

   There would be a negative correlation between reaction time and vocabulary size with larger vocabularies would find the target objects more quickly.

Method

Participants

Twenty-eight 22- to 24-month-old infants (M_{age}= 22.92 months, range = 22.01-23.92, 12 females, 16 males) participated in Experiment 2. Participant eligibility criteria and recruitment procedures were identical to Experiment 1. Data from 12 additional infants were not included in the analysis due to the following reasons: fussiness, including whimpering and/or continuous crying leading to a failure to complete familiarization (2), and training phases (3) or at least 6 of the 12 HTP or 12 LTP test trials (5), not paying attention as reflected by failure to orient to the TV screen (1), and experimental error (1). Like in the Experiment 1, the attrition rate is again slightly higher than or comparable to prior studies on word learning. High attrition rates are likely due to have a relatively long procedure that included a three-stage task (familiarization, referent training, testing phases).

Stimuli

All auditory and visual stimuli were the same as those used in Experiment 1.

Apparatus

All apparatuses were the same as those used in Experiment 1.

Procedures

Procedures were similar to those of Experiment 1, except that a 10-minute delay was implemented immediately after the referent training phase instead of following familiarization with the Italian corpus (see Figure 16 for the overview of the experimental design at Experiment 2). Infants were first familiarized with the Italian corpus. Immediately after familiarization, infants were trained on the 4 label-object associations. Following a 10-minute delay, infants' label-object associations were tested using Looking-While Listening Procedure.

Vocabulary Measure

The vocabulary scores for the infants in Experiment 2 ranged from 10 to 82. The age and gender normed vocabulary percentiles ranged from 2 to 89.

Coding

The coding procedures were identical to those of Experiment 1. Experiment 2 was coded by the same coder using iCoder software (Fernald et al., 2008). The trial exclusion criteria were identical to those used in Experiment 1. Using these criteria, approximately 10 % of the testing trials were excluded from the analysis (71 out of 672 total novel word test trials were excluded). Five participants were excluded because we could not keep enough trials from their session. Approximately 10 % of the data (n=3) were coded a second coder to check the intercoder reliability and 98.8% frame agreement and 98.53% shift agreement were obtained.

Dependent Measures

As the in Experiment 1, the dependent measures were accuracy and the reaction time.

Results

Familiar Words

Mean Accuracy

Univariate analysis of variance (ANOVA) showed that there were no significant main effects of age, F (1,15) = .32, p =.58, gender F (1,15) = .1, p =.76, counterbalanced language, F (1,15) = .17, p =.69, and order of presentation, F (1,15) = .72, p =.41, on accuracy. Thus, all the data were collapsed across these variables in subsequent accuracy analysis of familiar words.

To validate that infants had no object preferences before the word onset, I first examined infants' mean proportion of fixations on test trials during the baseline window (-2000 to 0 ms from familiar word onset). A one-sample t-test revealed no difference from chance level (50%) for the familiar words (M = 48%, SD = 7%), t (27) = -1.18, p = .25. However, the mean accuracy for the familiar words (M = 62 %, SD = 10 %) within critical window (300-2000 ms) was significantly above chance level, t (27) = 6.37, $p < .001$, d = 1.2. In addition, planned comparison were performed to better understand the difference in baseline and critical windows. A paired sample t-test showed a significant increase in looking to the target object from the baseline window to the critical window, t (27) = -6.24, $p < .001$, $d = 1.21$ (see Figure 17 and Figure 18).

To better understand infants' looking patterns on familiar words, I also created an *OC*-plot (see Figure 19) that separately tracks looking behavior for the target- and distractor-initial test trials. It appears that the infants showed more switches from the distractor to the target object following word onset than from the target object to the distractor object, suggesting that the infants successfully recognized the familiar labels.

Reaction Time

Any participant that had less than 3 usable RTs were excluded from the RT analyses. When applying these criteria, 13 infants from our final sample were excluded, leaving 15 infants. Univariate analysis of variance (ANOVA) showed that there were no significant effects of age, F (1,5) = .44, p =.54, gender F (1,5) = .9, p =.39, counterbalanced language, F (1,5) = .76, p =.42, and order of presentation, F (1,5) = .39, p =.56, on reaction times. Thus, all the data were collapsed across these variables in subsequent reaction times analysis of familiar words. Infants' average RT to orient to familiar target object from the distractor object was 925.35 ms (SD = 215.10).

Novel Words

Mean Accuracy

I first performed a repeated-measures analysis of variance (ANOVA) with object label type (HTP vs LTP) as a within-subject factor and age, gender, counterbalanced language, word order as between-subject factors. Repeated-measures ANOVA revealed that there were no main effects of age, F (1,15) = 2.02, p =.18, gender, F (1,15) = 3.22, p =.09, counterbalanced language, F (1,15) = 2.43, p =.14, and order of presentation, F (1,15) = 1.83, p =.2, on accuracy. Thus, all the data were collapsed across these variables in subsequent accuracy analysis of novel words (e.g., HTP and LTP words) (see Table 4).

To validate that infants had no object preferences before target word onset for any of the matched pairs (e.g., novel target word vs. novel distractor word), I first examined infants' mean accuracy for both HTP and LTP test trials during the baseline window (-2000 to 0 ms from novel target word onset). One-sample t-tests revealed no difference from chance level (50%) for the HTP words (M = 49%, SD = 9%), t (27) = -.60, p = .55, d = .11, and the LTP words (M = 49%, SD = 6%), t (27) = -.66, p = .51. Also, a paired sample t-test showed that the mean accuracy for the HTP and LTP test trials during baseline window were did not differ from each other, t (27) = -.17, p = .86.

The mean accuracy during the critical window (300-2000 ms) was significantly above chance for both HTP, t (27) = 4.29, p < .001, d = .81, and LTP object labels, t (27) = 3.76, p <.01, d = .71, suggesting that infants remembered both HTP and LTP object labels. However, there was no significant difference between infants' accuracy on HTP (M = 59%, SD = 11%) and LTP test trials (M = 58%, SD = 11%), t(27) = .31 p =.75, (see Figures 20 and 21).

Additionally, I created an *OC*- plot (see Figure 22) to examine infants' looking patterns for the target- and distractor-initial test trials for both HTP and LTP word trials. It appears that on both HTP and LTP object label trials, the infants showed more switches from the distractor object to the target object at word onset than from the target object to the distractor object at word onset, indicating that the infants successfully remembered the link between both the HTP and LTP labels and their referents.

Reaction Time

Measures of RT included only trials in which the infant was looking at the distractor object at the onset of the target word. In addition, each participant needed to contribute at least 3 trials on both HTP and LTP object label trials to be included in the analysis. When applying our criteria, 19 infants from our final sample were excluded, leaving 9 infants.

I first performed a repeated-measures analysis of variance (ANOVA) with object label type (HTP vs LTP) as a within-subject factor and age, gender, counterbalanced language, word order as between-subject factors. Repeated measures ANOVA revealed that there were no main effects of age, F (1,2) = .005, p =.95, gender, F (1,2) = .68, p =.5, counterbalanced language, F (1,2) = .26, p =.66, and order of presentation, F (1,2) = 0.0, p =.99, on RT. Thus, all the data were collapsed across these variables in subsequent RT analysis of novel words (e.g., HTP and LTP words).

A paired sample t-test revealed that there was no RT difference for HTP (M = 760.99 ms, SD = 414.20 ms) and LTP (M = 654.19 ms, SD = 190.94 ms) words, $t(8)$ = 1.09, p = .31, suggesting that infants' time to orient to the target object from distractor object on HTP and LTP trials were not different from each other.

Correlations Between Accuracy and RT

Familiar Words

When I perform Pearson's correlation between accuracy and RT for familiar words, no correlation was revealed, $r(15)$ = - .15, p =.59. However, when we include only infants who had at least 2 usable RTs (N = 23), I found a significant negative correlation between accuracy and RT for familiar words, $r(23)$ = - .55, p <.01 (see Figure 23), indicating that infants who had higher accuracy on familiar words were faster to orient to familiar object from distractor object, which was expected because the more quickly the infants looked at to the target object, the longer they are able to explore the target object.

Novel Words

Infants who had at least 3 usable RTs in either HTP or LTP testing trials were included in the correlation analyses. When applying these criteria to testing trials, infants who had at least 3 RTs in HTP (N = 13) and LTP testing trials (N = 19) were included. When I perform Pearson's correlations, I found no correlation between infants' accuracy and RTs for either HTP and LTP words (see Figure 23).

Correlations Between Word Processing and Vocabulary

Familiar Words

To examine whether the speech processing of familiar words was related to infants' vocabulary size, I performed Pearson's correlations. Individual differences in the size of infants' expressive vocabulary size did not predict their accuracy and RTs in recognizing familiar words. Neither infants' expressive vocabulary scores nor their age-normed vocabulary percentiles were correlated with their accuracy (see Figure 24) or RTs (see Figure 25) for familiar words.

Novel Words

Neither infants' productive vocabulary scores nor their age-normed vocabulary percentiles were correlated with their accuracy (see Figure 24) or RTs (see Figure 25) on HTP and LTP words.

Discussion

Familiar Words

Like in the Experiment 1, accuracy for familiar words were significantly above chance. Different from the Experiment 1, infants in the Experiment 2 were tested 10 minutes after training on familiar word-object associations. Because 22- to 24-month-old infants were likely already familiar with our English target words (e.g., shoe, book, baby, doggie), regardless of the referent training, their accuracy on familiar word test trials, was predicted to be above chance. Although we did find above chance performance on infants' accuracy on familiar word recognition, they were much less good at the task than we predicted. While previous work has revealed mean accuracy in familiar word processing to be around 75 to 85 % correct (Grieco-Calub et al., 2009; Zangl, Klarman, Thal, Fernald, & Bates, 2005; Fernald, Pefors, Marchman, 2006; Robertson, 2014), our participants were considerably less accurate in both Experiments 1 (60%) and 2 (%62). For example, when Fernald and colleagues used the same LWL procedure and the similar familiar words (ball, shoe, baby, and doggie) to test familiar word recognition, they found a higher accuracy score in both 21-month-olds (%64) and 25-month-olds (%78) compared to the 22- to 24-month-olds tested in the current studies. Importantly, in our study, familiar words were produced by the same native Italian speaker. Previous research has clearly showed that accented speech affects both accuracy and speed of speech processing. Thus, the accuracy difference in the current study and the previous studies might be due to foreign accented speech.

Also, this might be due to low vocabulary scores reported by parents in our studies. Approximately 30% of our infants were below the 20^{th} percentile on their age- and gender-normed expressive vocabulary size. Although infants' above chance performance at test does not tell us about the role of training on test, their relatively low accuracy and vocabulary size compared the previous studies suggest that their accuracy performance might be supported by the training on label-object associations.

To better understand the familiar word processing, I also analyzed RT measures. Since the processing demand for familiar words are minimal, I expected that the infants' RT to orient to familiar words will be faster than their RT to orient to HTP or LTP words. However, I again found no significant difference between the speed of processing of familiar and novel words. Further, I examined the correlation between accuracy and RT and found a negative correlation between these two measures, indicating that infants who had higher accuracy on familiar words were faster to orient to familiar object from distractor object.

To examine the relationship between vocabulary size and familiar word processing, I performed Pearson's correlations between vocabulary size and accuracy and RTs for familiar words. Results revealed that unlike Experiment 1, I found that vocabulary size was not predictive of infants' accuracy and RTs for familiar words in Experiment 2.

Novel Words

In Experiment 2, infants were significantly above chance on both HTP and LTP object label trials, suggesting that they were able to remember both HTP and LTP object labels when tested following a 10-minute delay, if they were trained on the label-object associations immediately after familiarization. A first key finding from Experiment 2, however, is that counter to our hypothesis, there was no difference in infants' accuracy on HTP and LTP object labels. Thus, the hypothesis that HTP words made better object labels was not supported. These findings suggest that experience with label-object associations immediately after familiarization may strengthen initially weak memory representations of syllable frequency information in LTP syllable sequences. This is discussed further in General Discussion.

RT analyses revealed that infants' RT to orient to HTP and LTP object labels from the distractor objects were not significantly different from each other. Further, I found no significant correlation between infants' accuracy and RTs on either HTP or LTP object label trials. To inspect whether familiar word processing was linked to novel word processing, Pearson's correlations were carried out. Like in the Experiment 1, I again found a relation between infants' accuracy in familiar words and their accuracy in LTP words. However, accuracy in familiar words did not predicted their performance in HTP words.

Like in the Experiment 1, correlation analyses for the novel words revealed no correlation between vocabulary and accuracy on either HTP or LTP object label trials. Thus, infants' vocabulary size did not predict their performance on HTP (p = .59) or LTP word trials (p = .52).

Chapter IV: General Discussion

By combining word learning tasks with word segmentation tasks, studies of statistical learning have revealed a relationship between tracking statistics of sounds in fluent speech and learning how those sound combinations map onto meaning. These studies have typically tested statistical word learning immediately after familiarization with an artificial (Graf Estes et al., 2007) or a natural language (Hay et al., 2011; Hay et al., 2017). However, word learning is a much more complicated developmental task – one that can be grounded in other cognitive processes such as memory. Infants must find words in speech and map them onto referents, but they must also encode and retrieve the sound and meaning representations of words over time. Thus, how statistical learning supports future word learning remains unknown. The current set of studies was designed the take a first step toward understanding the availability of statistically defined words as object labels after a delay. Specifically, the dissertation had 4 main aims: 1) to examine whether infants' memories for statistical properties from the speech stream (i.e., Tp and syllable frequency information) support word learning after a 10-minute delay (Experiment 1), 2) to examine whether infants' memories for the meaning representations that are characterized by different degrees of sequential statistics, support word learning after a 10-minute delay (Experiment 2), 3) to examine familiar word processing to understand whether the processing of familiar and novel words are interrelated, and 4) to examine whether individual differences in vocabulary size relate to infants' performance in tasks that tap into memory for sound and meaning representations.

To address the first aim, 22- to 24-month old infants were first presented with an unfamiliar natural language – Italian – that has been successfully employed in previous studies (e.g., Hay et al., 2017). Ten-minutes following familiarization with the Italian speech, infants were trained and tested on the 2 HTP and 2 LTP words using the LWL procedure (Fernald et al., 2008). The results revealed that while infants successfully learned the HTP words as object labels, they failed the learned the LTP words. Further, infants were significantly more accurate on HTP words than on LTP words within the 300-2000 ms critical window[4], suggesting that HTP words serve as better labels for objects than the LTP words following a 10-minute delay. These results are consistent with the previous research showing that the sequences whose constituent syllables have stronger co-occurrence statistics (such as HTP words) were better labels for objects when the learning was tested immediately after exposure to an artificial (Graf Estes et al., 2007) or a natural language (Hay et al., 2011).

Did infants learn and remember HTP words better than LTP words because HTP words have strong referential (e.g., word-like) status or are HTP words generic sequences that are easier to encode, just like any other probable sequences? A recent study by Lany and colleagues (Lany et al., 2018) tested these questions and demonstrated that tracking TPs between syllable sequences results in representations of potential words. Although, as a group, 20-month-old infants failed to learn either HTP or LTP words, infants' performance was correlated with their vocabulary score. Only infants with smaller vocabularies successfully mapped the HTP words (but not the LTP words) to meaning, suggesting that as infants' vocabularies grow, they become less open to learn sequences that deviate from native language word forms. Essentially, Lany and colleagues (2018) argued that infants were able to track TP information in the speech stream, but for infants with larger vocabularies, the newly segmented HTP words had referential status

---

[4] The same pattern of result was found when I analyzed the late window (300-2700 ms). I analyzed infants' accuracy for both HTP and LTP words within late window (300-2700 ms) and again found an above-chance performance in HTP words, $t(27) = 4.14$, $p<.001$, but not in LTP words , $t(27) = 1.32$, $p>.05$. Also, infants were significantly more accurate on the HTP (M=.61, SD=.32) than the LTP words (M=.54, SD=.33), $t(27) = 3.13$, $p<.01$, suggesting that HTP words make better object labels than the LTP words.

but did not sound like good words. Infants with smaller vocabularies, who presumable knew less about sounds and words in their native language, were more open to the newly segmented HTP sequences being acceptable words. Based on these findings, I suggest that in Experiment 1, HTP words are better remembered than LTP words because HTP words have a stronger referential (word-like) status.

However, a recent set of experiments conducted in our lab (Hay et al., 2017) reported enhanced word learning performance in slightly older infants – infants were able to learn the both HTP and LTP words as object labels when they were trained and tested immediately following exposure to an Italian speech stream (Baseline study 1). There are several possible explanations for this pattern of results. First, it may be that older infants are better at learning words than younger infants and they might be relying less on the TP information. Second, it may be that infants are processing the HTP and LTP words differently. In order to examine these two possible explanations, Hay and colleagues (2017) created another language where they disrupted the TPs of the both HTP and LTP words (TP = 0 for both HTP and LTP words) while preserving the syllable frequency information in the corpus. When TPs were violated, infants were unable to learn the modified HTP words, suggesting that older infants still track and rely on TP information for word learning. However, when the statistics of the LTP words, whose syllables occurred 3 times as often in the corpus, were violated, infants still learned the modified LTP words. These findings suggest that older infants can also make use of syllable frequency information in the service of learning of label-object associations. Since infants' continued reliance on TP information impacted their subsequent word learning performance, the first explanation – that older infants may be learning the both types of words just by associating labels and objects presented in isolation during the training phase – is rather unlikely. Hay and colleagues (2017) also found a different relationship between vocabulary size and infants' performance on HTP and LTP words, further suggesting that infants are mapping HTP and LTP words to meaning using different processess. Although it is admittedly speculative interpretation, based on the findings of Hay et al. (2017), it is worth suggesting that the HTP and LTP words are mapped to meaning for different reasons: the HTP words may be learned because their constituent syllables have stronger co-occurrence statistics and thus have strong referential status, whereas the LTP words may be learned because their syllables are more frequent in the Italian corpus. Thus, the sheer familiarity with the syllables in the LTP words may have driven a more associative learning process between the familiar syllables in the LTP words and the objects.

The results of the Experiment 1 are particularly interesting considering that in Hay et al.'s (2017) study, infants were able to learn the both HTP and LTP words as object labels when they were trained and tested immediately following familiarization with the same Italian speech stream. Why did infants show a similar learning for both the HTP and LTP words immediately after familiarization, while they show preferential learning for the HTP words relative to the LTP words 10 minutes after familiarization? The HTP words seems to be readily mapped onto meanings either immediately or 10-minute after familiarization with the Italian corpus. However, even though the LTP words' highly frequent syllables facilitated the learning of the LTP words at immediate test, syllable frequency did not have such facilitation effect on learning of the LTP words after the delay. The differential memory for the HTP and the LTP words as object labels highlights the role of prior exposure to the Italian speech stream because in order to show this kind of learning pattern, infants would have had to encode the TP information in their memory. Given that infants were able to learn the LTP words at immediate label-object association test, it is likely that they also encoded the syllable frequency information in their memory. However, the encoded memory representation of syllable frequency information was not robust enough to support learning of LTP words after the 10-minute delay. Taken together with the Hay et al. (2017), these results suggest that TP information is more resilient to decay in memory than syllable frequency information. Hay and colleagues also found that infants' vocabulary size and their performance on LTP words were significantly correlated in both Baseline Study 1 and 2, suggesting that infants with larger vocabularies were more likely to learn LTP words at immediate test. This individual difference data also supports the interpretation that infants process HTP and LTP words differently: tracking TP versus tracking syllable frequency information.

To address the second aim of investigating the retention of meaning representations, Experiment 2 implemented a 10-minute delay between the referent training and test phases. Twenty-two to 24-month-old infants were first familiarized to the same Italian corpus followed by a referent training phase. After a delay of 10 minutes, infants were tested on their memory for the 2 HTP and 2 LTP label-object associations. Based on the findings from the previous studies and Experiment 1, we predicted that statistically defined word meanings may be better remembered than label object associations where label goodness was not supported by the TP information. However, we found that infants were able remember both HTP and LTP word meanings when tested following a 10-minute delay between label-object training and test. Further, infants' performance on the HTP and the LTP words were not statistically different from each other, suggesting that infants showed similar memory for HTP and LTP label-object associations 10 minutes after training with the label-object associations. This finding goes against our prediction that statistically defined word meanings will be better remembered. These findings are somewhat surprising given that infants in the Experiment 1 showed fragile memory for syllable frequency information when they were trained and tested 10 minutes after familiarization with the Italian corpus (see Figure 26 and Figure 27 for comparison of Baseline Study 1 and 2 of Hay et al., 2017 and Experiment 1 and 2 in the current study).

Why did infants show enhanced retention for LTP word meanings when they were tested 10 minutes following training with the label-object associations? It is possible that when infants learn the label-object associations immediately after familiarization, they show successful retention for both HTP and LTP word meanings. These findings were also in line with the findings of Hay et al. (2017). Once extracted from the fluent speech stream HTP words appear to function as candidate labels that are mapped to meaning after a delay because of their strong word-like status. However, although they do not have strong lexical status, LTP words appear to be mapped to meaning because their syllables occurred frequently in corpus. However, once the LTP words are mapped to meaning, they appear to enjoy similar representational status as HTP word meanings. Additional support for this claim comes from our correlation analysis between familiar word processing and learning of HTP and LTP words in Experiment 2. While there was no correlation between infants' accuracy on familiar words and HTP words, the correlation between infants' accuracy on familiar words and LTP words were significant. Infants who were more accurate on familiar words were more likely to learn the LTP words. However, regardless of their performance at familiar words, infants were able to accurately map the HTP words onto their referents.

Another possible explanation is that having experience with the label-object associations immediately after familiarization may strengthen the initially weak memory representations of syllable frequency information in LTP words. Thus, weak retention of LTP words in Experiment

1 may just have resulted from insufficient encoding of syllable frequency information into memory. Support for this interpretation comes from previous studies suggesting that infants' and children's memory representations of novel label-object associations are very fragile and require additional memory support for later retrieval (Horst & Samuelson, 2008; Karaman & Hay, 2018; Vlach & Sandhofer, 2012). For example, Horst and Samuelson (2008) found that 2-year-olds may forget label-object associations at rapid rate. However, when the label-object associations presented via ostensive naming, infants were able to retain those associations after a 5-minute delay. Similarly, Vlach and Sandhofer (2012) demonstrated that 3-year old children's memory for fast-mapped words are very fragile but when the children were provided with different memory supports (e.g., saliency support: telling the children that the object was special; repetition support: labeling target objects repeatedly, generation support: asking the children to generate the target words), they were able to retain the fast-mapped words over long-delays. In a recent study, we also demonstrated that hearing target words (e.g., HTP and LTP words) in isolation reinforced 8-month old infants' memory for statistically defined words after a 10-minute delay (Karaman & Hay, 2018). Together, these studies suggest that providing additional cues to support infants' memory may have help them to encode the relevant words more robustly and efficiently. Thus, based on these prior findings, it is plausible to assume that additional experience with target words in both isolation and in the context of carrier phrases during referent training phase might have supported initially weak representations of the LTP words.

While this interpretation partly explains why infants learned the LTP words after the 10-minute delay, it is not clear why they showed similar memory for the HTP and the LTP words after the 10-minute delay. It is likely that hearing the target words in isolation and within the contexts of carrier phrases during the referent training phase might have functioned to make the

HTP and LTP words less distinguishable at test by reducing the TP between these two word types. During the referent training phase, each novel label-object association was randomly presented 4 times for a total of 16 novel label-object pairs across the training trials. In each training trial, infants heard each target word 2 times in isolation and twice within the carrier phrases (e.g., Bici! "See the bici! It's a bici! Bici!) while the associated object was on the TV screen. While presenting HTP object labels during the referent training trials did not change the overall TP of HTP words (e.g., TP remained 1.0), 16 (4 trials x 4 tokens) additional presentation of LTP words in isolation and the common naming carrier phrases increased the overall TP of the LTP words in both forward and backward direction [e.g., TP increased from .33 (=18/54) to .49 (=34/70)]. If indeed infants are continually updating TP information, the decreased TP difference [.67 (=1-.33) → .51 (=1-.49)] between the HTP and LTP words might have resulted in similar learning of these word types after the 10-minute delay. Even though we do not know how infants encode and retrieve the words with fine-graded statistics, it is likely that the TP of HTP (1.0) and LTP (.49) words is adequately dissimilar to lead to more reliable memories for the HTP words. However, this is not the pattern we see here. To the best of my knowledge, no studies have tested infants' sensitivity to graded TPs. Further research on the processing and retention of graded statistic is needed because with the current set of findings it is difficult to tease apart these possible interpretations.

To better understand the processing differences in the mapping of HTP and LTP words, in addition to the accuracy analysis, I also examined whether there is a difference in infants' speed of processing HTP and LTP words. I predicted that if infants encode TP information and label-object pairings into their long term-memory, following a delay they should show faster reaction times to initiate a shift in fixation to HTP as compared to LTP object labels. However, I found no meaningful difference in infants' RT latencies to orient to either target HTP or LTP objects from the corresponding distractor objects. Since the RT measure can only be calculated when the child is looking at the distractor object at label onset, the RT analyses included fewer participants (~12-18) than the accuracy analyses. Thus, the lack of meaningful differences found in Experiment 1 and 2 might be due to insufficient power. A priori power analysis revealed that at least 32 participants is necessary to obtain 80% power to find a significant difference at α level of .05. It is also likely that the lack of meaningful difference in RT measures might be due to increased task and memory demands. Greater number of familiar and novel training and test trials (8 familiar, 24 novel, and 1 whoopie trials, a total of 33 testing trials), and the 10-minute retention interval undeniably increased the cognitive demands.

Together, these findings provide evidence that TP information is more resilient to decay in memory over time than syllable frequency information. Further, although the mechanisms underlying the mapping of HTP and LTP words might be different, having experience with label-object associations immediately after familiarization may strengthen initially weak memory representations of syllable frequency information in LTP words.

Familiar Word Processing and Its Relation to Learning of Novel Words

The main purpose of having the familiar words in the current study was to familiarize infants with the nature of the word learning task. Thus, infants were trained on familiar label-object associations and either immediately (Experiment 1) or 10 minutes following training (Experiment 2), they were also tested on the same familiar word-object associations. Importantly, the presence of familiar labels and objects did not only orient infants to the structure of our test trials, but also served to help establish that the novel words should similarly be treated as labels.

Although the familiar label-object trials were basically 'the filler trials', they may help us to better understand the mechanisms underlying the mapping of HTP and LTP words. To examine whether infants' performance on familiar word trials were correlated with their performance on either HTP or LTP word trials, Pearson's correlations were performed. In Experiment 1, I found that infants' accuracy on familiar words and HTP words were not correlated (p = .24), whereas the same infants' accuracy on familiar words and LTP words were significantly correlated to each other, $r(28)$ =.466, p <.05[5]. Similarly, in Experiment 2, the correlation between infants' accuracy on familiar words and LTP words were marginally significant, $r(28)$ =.326, p = .09, but there was no correlation between infants' accuracy on familiar words and HTP words, (p = .66). These findings from both Experiment 1 and 2 suggest that only the infants who were more accurate on familiar words were also more accurate on LTP words. However, regardless of their performance on familiar word trials, infants were able to accurately map the HTP words onto their referents.

In both Experiments 1 and 2, in addition to looking at the correlations between infants' accuracy on familiar and novel word trials, I also performed Pearson's correlations between the RT latency to orient to the familiar objects and to the novel objects from label onset. Given that there were some significant correlations between infants' accuracy on familiar word trials and their accuracy on LTP word trials in both Experiment 1 and 2, I expected to obtain a *positive* correlation between infants' RT latency to orient to the familiar and LTP objects – infants with faster RTs to orient to the familiar objects should be also faster to the orient to the novel objects. However, we found a marginally significant *negative* correlation between infants' RT latency to

[5] Note that even in the significant correlations found in the current study, the effect was weak (correlation coefficient r ranged between .2 to .4 or -.2 to -.4) or moderate (correlation coefficient r ranged between .4 to .6 or -.4 to -.6). Thus, it is difficult to draw strong conclusions based in the correlations found.

orient to familiar objects and to the LTP objects, $r(19) = -.432$, $p = .06$ in Experiment 2, indicating that infants with faster RTs to orient to the familiar objects were slower to the orient to the LTP objects. It is possible that as infants spent more time on the distractor object, they might be more certain that the distractor object is not the referent for the LTP label. Thus, after they shifted their eye gaze, they might not to shift back to the distractor and stay longer on the target LTP object. However, in familiar word trials they did not need to stay longer on the distractor object to ensure that the label and the referent were mismatched. Since they already knew the familiar label-object pairings, they might quickly shift away from the distractor to the named target object. This *ensuring behavior* might also partially explain infants' enhanced accuracy/retention in the mappings of LTP words in Experiment 2. However, it appears that infants do not show the same *ensuring behavior* in the mappings of HTP words, suggesting that they may be treating the HTP words as the familiar words and so do not necessarily stay longer on the distractor object to learn the correct HTP label-object pairings. This interpretation, however, is highly speculative.

The Relationship Between Vocabulary Size and Learning of Familiar and Novel Words

Previous studies have shown a mixed pattern of results on the relationship between individual differences in vocabulary size and infants' and children's performance at learning of novel words – while some studies (Bion et al., 2013; Hay et al., 2017; Lany et al., 2018; Law & Edwards, 2015) reported significant correlations, others failed to find any relationship (Horst & Samuelson, 2008; Kucker & Samuelson, 2012; Pomper & Saffran, 2018)[6].

In one study, Bion and colleagues (Bion et al., 2013) used an LWL procedure (Fernald et al., 2008) to examine the relationship between the individual differences in vocabulary size and

---

[6] Note that, however, there are some methodological differences between these studies.

the ability to disambiguate novel objects (e.g. children's tendency to select to the novel object when they are presented with a novel object and a familiar object as they hear a novel object label) and remember them after a short 5-minute delay. They found that across three age groups (18-, 24-, and 30-month-olds), only the 24-month-old infants' expressive vocabulary size predicted their performance on immediate disambiguation trials. The authors speculate that the lack of relationship might be due to relatively low vocabulary level of 18-month-olds and the ceiling effect on the vocabulary measures of 30-month-olds. However, on retention trials, only the 30-month-old infants' accuracy was correlated with their vocabulary size, suggesting that by 30 months, infants' ability to remember the novel label-object associations were related to their vocabulary growth. Similarly, Hay and colleagues (2017) looked at the relationship between infants' vocabulary size and their performance in the mappings of HTP and LTP words. Correlation analyses revealed that there were no significant correlations between infants' vocabulary size and their performance on HTP words in both Baseline Study 1 and 2, suggesting that 22- to 24-month old infants, regardless of their vocabulary sizes, are able to learn HTP words in the immediate label-object associations task. However, infants' vocabulary size and their performance on LTP words were significantly correlated in both Baseline Study 1 and 2, suggesting that infants with larger vocabularies were more likely to learn LTP words at immediate test.

However, in the current set of studies, individual differences in infants' expressive vocabulary size did not predict their performance on familiar, HTP, or LTP word test trials in the delayed testing conditions. Given that prior research has provided inconsistent results linking vocabulary size and novel word learning performance, the lack of significant relations in the current set of studies is not surprising. Several possible reasons might account for why individual difference in vocabulary size did not reveal any meaningful patterns. First, as Bion et al. (2013) pointed out that before 30 months, infants' ability to remember the novel label-object associations may not be related to their vocabulary growth. Second, studies that report significant correlations have typically measured infants' expressive vocabulary size with the long form of the MCDI Words and Sentences (680 words), whereas in the current study, we used the short form of MCDI (100 words) to measure expressive vocabulary size. I speculate that using different measurement tool for expressive vocabulary size in the current study might hide the possible relationship between vocabulary size and infants' retention of novel-object associations. Note, however, that Hay and colleagues (2017) did find significant link between vocabulary size and immediate accuracy performance using the short form of MCDI. Further, even using the short MCDI form, there was a wide distribution of vocabulary sizes across our participants (ranged from 5 to 98). Even though it is challenging to interpret null findings, I hope that this discussion will help inform future research examining the relationship between vocabulary size and retention of statistically defined object labels.

While prior findings do not provide a clear picture on the role of vocabulary development on novel word learning, some studies have found robust relationships between vocabulary size and familiar word processing (Fernald, Perfors, & Marchman, 2006; Grieco-Calub et al., 2009). For instance, Fernald and colleagues (2006) used an LWL experimental design to study the link between processing efficiency and vocabulary growth in typically developing children from 12 to 25 months of age. They measured speed and accuracy of familiar word comprehension at 15, 18, and 25 months of age, and measured vocabulary using the MCDI at 12, 15, 18, and 25 months. They found that speed and accuracy measures at 25 months were highly correlated with vocabulary measures from 12 to 25 months, indicating that children who had larger vocabularies between 12 to 25 months were faster and more accurate in word recognition at 25 months. Despite these demonstrations, in the current study, I found no correlation between infants' vocabulary size and familiar word processing.

However, when we collapsed the data from familiar word trials of the current study with those from Hay et al. (2017), a different pattern of results emerges. Since everything (e.g., age range, stimuli, apparatus, procedures, and the critical window) was identical in the Hay et al. (2017) and Experiment 1 of the current study[7], data from Baseline Study 1(n=32), Baseline study 2 (n=32) and the Experiment 1 (n=27) were collapsed for an additional analysis. When I performed a Pearson's correlation, I found a significant positive correlation between vocabulary size and accuracy performance on familiar word trials, $r(91) = .204$, $p < .05$. These results suggest that infants' vocabulary size might be relate to their processing of familiar words.

Limitations and Future Directions

While the results of this dissertation work extend previous findings on statistical word learning and the retention of statistically defined words, there are some limitations that may be considered as suggestions for future research. One limitation of the current study is that the attrition rate is high. I think that high attrition rates are likely due to have a relatively long procedure that included a three-stage task (familiarization, referent training, testing phases). Although in the current study there was no difference between the vocabulary size of participants that were included in the analyses (Experiment 1, n=27: M_{Vocab}= 43.66, SD_{Vocab} = 22.92, Experiment 2, n=28: M_{Vocab}= 50.85, SD_{Vocab} = 24.55), and the participants that were excluded due

[7] Since differently from the Baseline study 1, Baseline study 2 and Experiment 1, infants in the Experiment 2 were tested 10-minutes after the training with the familiar words, the accuracy and vocabulary data from this study was not included in the correlation analysis. However, the pattern of results is unchanged if we also include the data from the Experiment 2, r (119) = .195, p <.05.

to fussiness (Experiment 1, n = 15: M_{Vocab}= 45.87, SD_{Vocab} = 22.26, Experiment 2, n=11: M_{Vocab}= 52.82, SD_{Vocab} = 26.64), it is also possible that that infants with higher vocabularies might be more likely to become bored than infants with smaller vocabularies.

Another limitation is that although the speech materials used in the current study were more ecologically valid than those used in some prior statistical learning studies (e.g., Graf Estes et al., 2007), it is still difficult to mimic natural settings infants experience in daily life. Further, our design lacked the social cues[8] such as pointing, eye contact, facial expressions, gestures, and joint attention that facilitate word processing and learning in natural settings. Thus, future research may systematically integrate the social cues in the experimental design to examine the role of these cues on the retention of statistical learning.

Even though using natural language as a speech material increase the ecological validity of the task, it also creates additional challenges for us. First, although English and Italian differ in many of their allophonic and phonotactic characteristics, and thus likely sounded quite unfamiliar to our English-speaking participants, the target words in the corpus shared the strong/weak stress patterning found in English bisyllabic words. It is well-established that infants are sensitive to the prosodic patterns their native language from a young age and can use stress cues to segment speech (Thiessen & Saffran, 2003, 2007; Johnson & Jusczyk, 2001). Thus, in the current set of experiments, it is not entirely possible to know how essential the stress patterns of the target words were to infants; ability to segment and map the target words. Thus, I do not assert that infants learned our novel words just by tracking the TP information alone. Future

[8] Note that in the current set of experiments only social cue used was the carrier phrases. During training and test, our target words were embedded within the carrier phrases (e.g., Look at the bici!)

research may consider replicating the current study with an iambic (weak/strong) language such as Farsi[9] to examine the learning and remembering of statistically defined iambic words.

Another limitation of the present study was using the highly predictive, deterministic TP information in HTP words (TP = 1.0) and contrasting them with much lower probability words (i.e., LTP words; TP = .33). In natural languages, the TP of words are likely to be lower than the TP used in the previous studies and the current study (Willits, Seidenberg, & Saffran, 2009). Importantly, in the current experiments, hearing additional tokens of the novel target words either in isolation or in carrier phrases during the referent training phase changed the TP of LTP words (.33 →.49) but the TP of HTP words did not change. If indeed infants are progressively updating their statistical computations and the memory representations of TP information, it is likely that increasing TP may increase the learnability and retention of LTP words. Further, increasing TP of LTP words would make HTP and LTP words less discriminable at test by reducing the TP difference between HTP and LTP words. Thus, future research may examine infants' sensitivity to graded statistics and also how infants retain the representations of words defined with graded statistics.

In the current study, we employed a 10-minute retention interval. According to the Craik and Lockhart's (1972) the levels of processing model, information can be consolidated from short-term memory to long-term memory within minutes. Also, we know that synaptic consolidation (a.k.a. late-phase long-term potentiation) which is a form of memory consolidation is achieved within minutes (Bramham & Messaoudi, 2005; Dudai, 2004). However, for more stable and long-lasting memories, memory traces should be transferred from hippocampus to the

---

[9] However, note that a recent set of studies conducted in our lab (Parvanezadeh Esfahani & Hay, in prep) found that 8-month-old infants are having difficulty to track statistics when they were exposed to an iambic natural language: Farsi.

cortex (e.g., system consolidation), which occurs within hours of learning (Dudai, 2004). Thus, by varying the delay between familiarization/training and test, future research may examine the availability of statistically defined words as object labels after longer delays. Further, we know that sleep is crucial for memory consolidation (e.g., Stickgold, 2005). Thus, future research may also investigate the role of sleep on the retention of statistically defined words as object labels.

Last but not least, based on the assumption that statistical segmentation and word-object mapping are dependent processes and often operate in a sequential manner, in the current set of experiments and in the previous studies (Hay et al., 2011; Hay et al., 2017; Graf Estes et al., 2007; Mirman et al., 2008) the word segmentation and the label-learning tasks were implemented sequentially. However, behavioral (Shukla et al., 2011), neuropyshiology (Cunillera et al., 2010; François et al., 2017) and modeling studies (Räsänen et al., 2015) have suggested that segmentation and mapping processes may also occur simultaneously. For example, a recent study by Räsänen and Rasilo (2015) proposed a computational model for joint construction of sound and meaning representations and the results from the model stimulations showed that sound and meaning representations might be constructed simultaneously. Similarly, Shukla and colleagues (Shukla et al., 2011) demonstrated that 6-month-old infants can simultaneously segment novel words from prosodically structured speech stream and map them onto objects. Thus, future research may examine the retention of statistically defined object labels that are simultaneously segmented and mapped.

Conclusion

Word learning is central to language development. Children learn thousands of words in only a few short years. Previous studies of statistical learning have revealed remarkable statistical tracking abilities in infancy, which may, at least partially, explain how infants solve the word learning puzzle quickly without apparent effort. With little exception such studies test word learning in the seconds immediately after familiarization. Thus, to fill in this critical gap in the literature, the current set of experiments examined the availability of statistically defined words as object labels following a delay.

When 22- to 24-month-old infants were tested 10 minutes following familiarization with an Italian corpus, they were able to remember the HTP, but not LTP object labels, suggesting that HTP words not only have stronger referential status, but that they are also better maintained in long term memory than are LTP words. These findings suggest that TP information is more resilient to decay in memory than syllable frequency information (Experiment 1). However, if the infants were trained on the label object associations immediately after familiarization, they were able to remember both HTP and LTP words when tested following a 10-minute delay (Experiment 2). Together, these findings suggest that although the mapping of HTP and LTP words might be driven by different underlying processes, having experience with label-object associations immediately after familiarization may strengthen initially weak memory representations of syllable frequency information in LTP sequences.

This study provides the first piece of evidence that statistical learning experience supports the learning of word-object associations after a minimal delay. Knowledge of retention of statistical learning will shed light on the relevance of statistical learning to early language acquisition. This knowledge is also important for identifying deficits in statistical learning in atypical populations.

References

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia*, *46*(11), 2787-2795.

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199-209. doi: 10.1111/j.1467-7687.2005.00406.x

Akhtar, N., Jipson, J., & Callanan, M. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416-430. doi: 10.1111/1467-8624.00287

Arciuli J, Simpson I. 2012 Statistical learning is lasting and consistent over time. Neurosci. Lett. 517, 133 – 135. (doi:10.1016/j.neulet.2012.04.045)

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. doi: 10.1111/1467-9280.00063

Axelsson, E. L., Williams, S. E., & Horst, J. S. (2016). The Effect of Sleep on Children's Word Retention and Generalization. *Frontiers in psychology*, *7*, 1192.

Bauer, P. J. (2004). Getting explicit memory off the ground: Steps toward construction of a neuro-developmental account of changes in the first two years of life. *Developmental Review*, *24*, 347-373.

Bauer, P. J. (2006). Constructing a past in infancy: A neuro-developmental account. *Trends in Cognitive Sciences*, *10*(4), 175-181. doi: 10.1016/j.tics.2006.02.009

Benavides-Varela, S., & Mehler, J. (2015). Verbal positional memory in 7-month-olds. *Child development*, *86*(1), 209-223.

Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*(7), 1132-1136.

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.

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39-53.

Bramham, C. R., & Messaoudi, E. (2005). BDNF function in adult synaptic plasticity: the synaptic consolidation hypothesis. *Progress in neurobiology*, *76*(2), 99-125.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.

Charniak, E. (1993). *Statistical language learning.* Cam- bridge, MA: MIT Press.

Chen, J. and ten Cate, C. (2015) Zebra finches can use positional and transitional cues to distinguish vocal element strings. Behavioral Processes 117, 29–34.

Chomsky, N. (1959). A note on phrase structure grammars. *Information and control*, *2*(4), 393-395.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), *Perception and Production of Fluent Speech*, 133–163. *Hillsdale, NJ: Erlbaum*.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Cunillera, T., Laine, M., Càmara, E., & Rodríguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, *63*(3), 295-305.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3-4), 133-142.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. Journal of Cognitive Neuroscience, 21(4), 803-820.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51-86.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322-1331.

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2012). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, *23*(10), 2467-2478.

Fenson, L., Marchman, V., Thal, D. J., Dale, P. S., Reznick, S., & Bates, E. (2006). MacArthur-Bates Communicative Development Inventories (2nd Ed.). Baltimore, MD: Brookes Publishing Company.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95-116.

Fernald, A. (2008). Chapter Two Hearing, Listening, and Understanding: Auditory Development in Infancy. *Blackwell handbook of infant development*, 35.

Fernald, A., & Marchman, V. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83*, 203–222.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child development*, *64*(3), 637-656.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98.

Fernald, A., Pinto, J., Swingley, D., Weinberg, A., & McRoberts, G. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 72-75.

Fernald, A., Zangl, R., Portillo, A. L., Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In: Sekerina, I., Fernandez, E. M., Clahsen, H. (Eds.) Developmental psycholinguistics: On-line methods in children's language processing. Amsterdam: John Benjamins, 97-135.

Fiser J, Aslin RN. 2002. Statistical learning of new visual feature combinations by infants. *PNAS* 99(24):15822–26

François, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*, *98*, 56-67.

Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Attention, Perception, & Psychophysics*, *54*(3), 287-295. doi:10.3758/BF03205263

Gleitman, Lila R. & Eric Wanner. 1982. Language acquisition: the state of the art. In Eric Wanner & Lila Gleitman (eds.), *Language acquisition: The state of the art*, 3–48. Cambridge: Cambridge University Press.

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of Royal Society B*, *372*(1711), 20160054. doi: 10.1098/rstb.2016.0054

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, *17(8)*, 670–674. doi: 10.1111/j.1467-9280.2006.01764.x

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135. http://dx.doi.org/10.1016/S0010-0277(99)00003-7

Greco, C., Rovee-Collier, C., Hayne, H., Griesler, P., & Earley, L. (1986). Ontogeny of early event memory: I. Forgetting and retrieval by 2-and 3-month-olds. *Infant Behavior and Development*, *9*(4), 441-460.

Greco, C., Hayne, H., and Rovee-Collier, C. (1990). Roles of function, reminding, and variability in categorization by 3-month-old infants. The *Journal of Experimental Psychology: Learning, Memory & Cognition, 16*, 617–633.

Graf Estes, K., 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, 254–260.

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy*, *16*(2), 180-197.

Grieco-Calub, T. M., Saffran, J. R., & Litovsky, R. Y. (2009). Spoken word recognition in toddlers who use cochlear implants. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1390-1400.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64.

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Wondoloski, T. L., Klein, P., et al. (1998). The ontogeny of long-term memory over the first year-and-a-half of life. *Developmental Psychobiology, 32*, 69–89.

Hay, J. F., Shoaib, A., Wang, T., Moore, D., Lohman, J., & Lany, J. (November, 2017). Learning words in an unfamiliar language: The role of statistics and context. Paper presented at the *42nd Boston University Conference on Language Development*, Boston, MA.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93-106. doi: 10.1016/j.cogpsych.2011.06.002

Herbert, J., & Hayne, H. (2000). Memory retrieval by 18–30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology*, *36*(4), 473. doi: 10.1037/0012-1649.36.4.473

Hill, W., Borovsky, D., and Rovee-Collier, C. (1988). Continuities in infant memory development over the first half-year. *Developmental Psychobiology, 21*, 43–62.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1143. doi: 10.1037/0096-1523.29.6.1143

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125-150.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128-157. doi: 10.1080/15250000701795598

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of sleep research*, *24*(5), 503-509.

Hupbach, A., Gómez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science 12(6)*, 1007–1012. doi: 10.1111/j.1467-7687.2009.00837.x

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*(4), 548-567.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675-687. doi:10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science, 277*(5334), 1984-1986. doi: 10.1126/science.277.5334.1984

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Attention, Perception, & Psychophysics*, *61*(8), 1465-1476. doi: 10.3758/BF03213111

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3), 159-207. doi:

10.1006/cogp.1999.0716

Karaman, F., & Hay, J. F. (2018). The longevity of statistical learning: When infant memory decays, isolated words come to the rescue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 221.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and language*, *127*(1), 46-54.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit?. *Neuroscience letters*, *461*(2), 145-149.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Kóbor, A., Janacsek, K., Takács, Á., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific reports*, *7*(1), 760.

Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*, *17*(3), 295-323.

Ladd, R. (1997). Intonational phonology. London: Cambridge University Press.

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284-291.

Lany, J., & Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Developmental science*, *14*(5), 1207-1219.

Law, F., & Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, *11*(4), 331-355.

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323-1329.

Mandel, D. R., Nelson, D. G. K., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, *11*(2), 181-196.

Mandler, J. M. (1998). Representation. In D. Kuhn & R. Siegler (Eds.), *Cognition, perception, and language,* Vol. 2 of W. Damon (Series Ed.), *Handbook of child psychology.* (pp 255-308). New York: Wiley.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–16.

Marcovitch S, Lewkowicz DJ. 2009. Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. *Dev. Sci.* 12(6):1020–25

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121. doi: 10.1016/S0010-0277(00)00109-8

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *Journal of Neuroscience*, *26*(29), 7629-7639.

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen- and twenty-month-old infants. *Child Development, 56(1)*, 62–72. doi: 10.2307/1130174

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: long- term memory for novel acts and multiple stimuli. *Developmental Psychology 24(4)*, 470–476. doi: 10.1037/0012-1649.24.4.470

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development*, *20*(1), 19-31.

Mintz, T. H., Walker, R. L., Welday, A., & Kidd, C. (2018). Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition*, *171*, 95-107.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271-280.

Morgan, James L & Demuth, Katherine. (1996) Signal to Syntax: an overview. In Morgan, JL & Demuth, K. (eds.). Signal to Syntax: Bootstrapping from speech to grammar in early acquisition (pp. 1-22), Mahwah, New Jersey: Lawrence Erlbaum Associates.

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L., & Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science*, *319*(5867), 1260-1264. doi: 10.1126/science.1151120

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (Vol. 26, pp. 125-173). Academic Press.

Nyberg, L., & Cabeza, R. (2000). Brain imaging of memory. In E. Tulving & F.I.M. Craik (Eds.), The Oxford handbook of memory (pp. 501– 519). *New York: Oxford University Press.*

Olson, I. R., & Newcombe, N. S. (2014). The Wiley Handbook on the Development of Children's Memory.

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. doi: 10.1111/j.1467-8624.2009.01290.x

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244-247. doi: 10.1016/j.cognition.2009.07.011

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* (Vol. 8, No. 5, p. 18). New York: International Universities Press.

Pillemer, D. B., & White, S. H. (1989). Childhood events recalled by children and adults. In *Advances in child development and behavior* (Vol. 21, pp. 297-340). JAI.

Pinker, S. (1984). Language learnability and language development. Cambridge, MA: Harvard University Press. (a)

Pomper, R., & Saffran, J. R. (2018). Familiar Object Salience Affects Novel Word Learning. *Child development*.

Räsänen, O., & Rasilo, H. (2015). A joint model of word segmentation and meaning acquisition through cross-situational learning. *Psychological review*, *122*(4), 792.

Rovee-Collier, C. (1997). Dissociations in infant memory: rethinking the development of implicit and explicit memory. *Psychological Review*, *104*(3), 467.

Robertson, S. (2014). The effect of hearing loss on word processing of infant- and adult-directed speech. Department of Audiology and Speech Pathology, University of Tennessee Health Sciences Center, UT Memphis

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science*, *8*(3), 80-85. doi: 10.1111/1467-8721.00019

Rovee, C. K., & Fagen, J. W. (1976). Extended conditioning and 24-hour retention in infants. *Journal of Experimental Child Psychology*, *21*(1)*,* 1-11. doi: 10.1016/0022-0965(76)90052-7

Rovee-Collier, C., Hartshorn, K., and DiRubbo, M. (1999). Long-term maintenance of infant memory. *Developmental Psychobiology, 35*, 91–102.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants.

Science, *274*(5294), 1926-1928. doi: 10.1126/science.274.5294.1926

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479-500.

Saffran, J. R., & Kirkham, N. Z. (2017). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1).

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of cognitive neuroscience*, *26*(8), 1736-1747.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622-1627.

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain mapping: an encyclopedic reference*, *3*, 501-506.

Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. 2017 Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Phil. Trans. R. Soc. B 372, 20160049. (doi:10.1098/rstb.2016.0049)

Simon, K. N., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L., & Gómez, R. L. (2017). Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain and Language*, *167*, 3-12. doi: 10.1016/j.bandl.2016.05.002

Shneidman L, Sootsman-Buresh J, Shimpi P, Knight-Schwartz J, Woodward AL. (2009) Social experience, social attention, and word learning in an overhearing paradigm. Language Learning and Development.;5:266–281.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038-6043.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in cognitive sciences*, *18*(5), 251-258.

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word– referent mappings via cross-situational statistics. Cognition, 106, 1558–1568.

Squire. L.R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231. doi: 10.1037/0033-295X.99.2.195

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272.

Swingley, D. (1999). Conditional probability and word discovery: A corpus analysis of speech to infants. In *Proceedings of the 21st annual conference of the cognitive science society* (pp. 724-729). Erlbaum Mahwah, NJ.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147-166.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73-108.

Takahasi, M. et al. (2010) Statistical and prosodic cues for song segmentation learning by Bengalese finches (Lonchura striata var. domestica). Ethology 116, 481–489

Team, R. C. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, *39*(4), 706. doi: 10.1037/0012-1649.39.3.484.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3*(1), 73-100.

Tomasello, M. (2009). *Constructing a language*. Harvard university press.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Attention, Perception, & Psychophysics*, *67*(5), 867-875.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126-156.

Vander Linde, E., Morrongiello, B. A., and Rovee-Collier, C. (1985). Determinants of retention in 8-week-old infants. *Developmental Psychology, 21*, 601–613.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481-487.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382.

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*(4), 46. doi: 10.3389/fpsyg.2012.00046

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental psychology*, *45*(6), 1611.

Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child development*, *85*(2), 429-436.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental psychology*, *34*(6), 1289.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1-30.

Willits, J., Seidenberg, M., & Saffran, J. (2009, January). Verbs are LookING good in early language acquisition. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in Psychology*, *4*: 151. doi: 10.3389/fpsyg.2013.00151

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental science*, *12*(3), 412-418.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414-420.

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic bulletin & review*, *21*(1), 1-22.

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development*, *6*(2), 179-208.

Appendix

Language A: HTP = fuga & melo; LTP = casa & bici

1. Spesso Lisa capita in fuga nella casa dove giaci gracile e tesa. **2.** Se cadi con la bici prima del bivio del melo cavo ti do dieci bigoli e una biro. **3.** Gli amici della cavia Bida poggiano le bici in bilico presso il melo per difesa dalla biscia. **4.** Sovente carico la spesa nel vicinato dopo una fuga con la bici nuova. **5.** Carola si è esibita in una fuga verso il melo perché offesa dagli amici scortesi. **6.** Se vai a casa in bici ti debiliti ma cali e non sei più obesa. **7.** Dietro la casa del capo ho sprecato i ceci sotto al melo ombroso. **8.** Se cuci subito sulla divisa bigia il distintivo col melo vado in casa a dormire**.** **9.** Teresa si abitua alla fuga da casa con la vecchia bici senza luci posteriori. **10.** Taci sulla fuga di Marisa con il caro lattaio. **11.** Il bel melo sta tra la casa dei Greci e la chiesa arcana dove hai giocato con le bilie. **12.** I soci della ditta Musa si danno alla fuga con la bici della maglia rosa.

Language B: HTP = casa & bici; LTP = fuga & melo

1.Roméro fu coinvolto in una futile fuga in bici verso il profumo del mélo ombroso. **2.** Il collega di Paolo Fusi trovò la bici per la fuga presso la casa del molo. **3.** La maga tiene in casa almeno un fuco, uno squalo e una tartaruga del Nilo. **4.** Il fuco procede parallelo alla casa sulla riga tracciata dalla cometa. **5.** Il gattone Refuso medita sul mélo presso casa ascoltando una fuga di Verdi. **6.** Il fu Medo Rossi ruppe la braga nella bici il mese scorso durante la gara. **7.** Giga ogni mese paga con zelo l'affitto per la casa con il melo in fiore. **8.**Meco prega il cielo che ogni fuga da casa termini sotto melo ombroso. **9.**Il delfino beluga si dimena tutto solo nella fuga verso il Nilo azzurro. **10.**Un pezzo di filo si è infilato nella bici appoggiata al melo dietro la méscita. **11**.Vi fu un tempo in cui la bici in lega non temeva il gelo del rifugio della Futa. **12.** La strega del melo fu vista in fuga sulla bici con un chilo di rametti.

Table 1. Phonotactic probability of target words

	Probability for Phonemes					Probability for Biphones			
Target Word	**1st**	**2nd**	**3rd**	**4th**	**5th**	**1st**	**2nd**	**3rd**	**4th**
baby (IPA: /'beɪbi/, Klattese: /beIbi/)	.0512	.0292	.0350	.0179	.0404	.0017	.0000	.0006	.0008
doggie (IPA: /dɑgi/, Klattese: /dagi/)	.0518	.0605	.0179	.0432		.0023	.0007	.0005	
book (IPA: /'bʊk/, Klattese: /bUk/)	.0512	.0102	.0535			.0012	.0010		
shoe (IPA: /'ʃuː/, Klattese: /Su/)	.0097	.0221				.0002			
bici (IPA: /ˈbitʃi/, Klattese: /biCi/)	.0512	.0318	.0080	.0432		.0022	.0006	.0001	
casa (IPA: /ˈkaː.sa/, Klattese: /kasa/)	.0927	.0605	.0788	.0174		.0166	.0024	.0008	
fuga (IPA: /ˈfʊ.ga/, Klattese: /fUga/	.0466	.0102	.0179	.0174		.0007	.0002	.0003	
melo (IPA:/meː.loː/, Klattese:/melo/)	.0572	.0292	.0737	.0210		.0028	.0029	.0026	

Table 2. Familiar and Novel Object-Label Associations

Familiar Objects	Familiar Object Labels	Novel Objects	Novel Object Labels
	baby		bici
	doggie		casa
	book		fuga
	shoe		melo

Table 3. Analysis of Variance Source Table for Experiment 1.

Source	df	F	η^2
Repeated Measures			
(A) Word Type (HTP vs. LTP)	1	20.067*	.589
Between Groups			
(B) Age	1	.020	.001
(C) Gender	1	4.039	.223
(D) Language	1	.004	0
(E) Order	1	.295	.021
AXB (Word Type X Age)	1	.136	.01
AXC (Word Type X Gender)	1	2.078	.129
AXD (Word Type X Language)	1	2.014	.126
AXE (Word Type X Order)	1	.768	.052
Error	14		

* $p < .001$.

Table 4. Analysis of Variance Source Table for Experiment 2.

Source	df	F	η^2
Repeated Measures			
(A) Word Type (HTP vs. LTP)	1	.066	.004
Between Groups			
(B) Age	1	2.019	.119
(C) Gender	1	3.217	.177
(D) Language	1	2.431	.139
(E) Order	1	1.829	.109
AXB (Word Type X Age)	1	5.382*	.264
AXC (Word Type X Gender)	1	.012	.001
AXD (Word Type X Language)	1	3.335	.182
AXE (Word Type X Order)	1	.272	.018
Error	15		

* $p < .05$.

Spesso Lisa capita in fuga nella casa dove giaci gracile e tesa. Se cadi con la bici prima del bivio del melo cavo ti do dieci bigoli e una biro. Gli amici della cavia Bida poggiano le bici in bilico presso il melo per difesa dalla biscia. Sovente carico la spesa nel vicinato dopo una fuga con la bici nuova. Carola si è esibita in una fuga verso il melo perché offesa dagli amici scortesi. Se vai a casa in bici ti debiliti ma cali e non sei più obesa. Dietro la casa del capo ho sprecato i ceci sotto al melo ombroso. Se cuci subito sulla divisa bigia il distintivo col melo vado in casa a dormire**.** Teresa si abitua alla fuga da casa con la vecchia bici senza luci posteriori. Taci sulla fuga di Marisa con il caro lattaio. Il bel melo sta tra la casa dei Greci e la chiesa arcana dove hai giocato con le bilie. I soci della ditta Musa si danno alla fuga con la bici della maglia rosa.

Target Words
Each target word appeared 6 times
HTP words: fuga, melo
LTP words: bici, casa

Roméro fu coinvolto in una futile fuga in bici verso il profumo del mélo ombroso. Il collega di Paolo Fusi trovò la bici per la fuga presso la casa del molo.La maga tiene in casa almeno un fuco, uno squalo e una tartaruga del Nilo.Il fuco procede parallelo alla casa sulla riga tracciata dalla cometa. Il gattone Refuso medita sul mélo presso casa ascoltando una fuga di Verdi. Il fu Medo Rossi ruppe la braga nella bici il mese scorso durante la gara. Giga ogni mese paga con zelo l'affitto per la casa con il melo in fiore. Meco prega il cielo che ogni fuga da casa termini sotto melo ombroso.Il delfino beluga si dimena tutto solo nella fuga verso il Nilo azzurro. Un pezzo di filo si è infilato nella bici appoggiata al melo dietro la méscita. Vi fu un tempo in cui la bici in lega non temeva il gelo del rifugio della Futa. La strega del melo fu vista in fuga sulla bici cn un chilo di rametti.

Target Words
Each target word appeared 6 times
HTP words: bici, casa
LTP words: fuga, melo

Figure 1. Overview of familiarization phase.

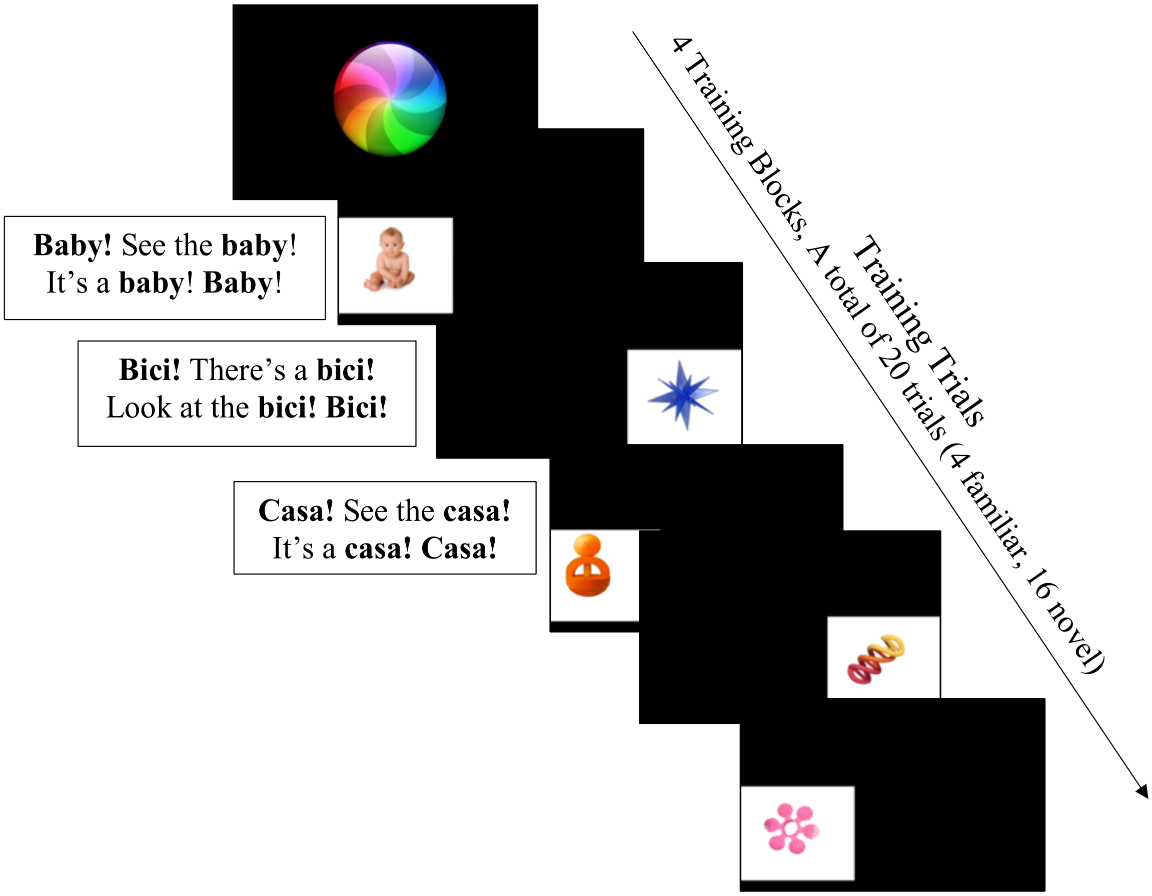

Figure 2. Overview of referent training phase. During the training trials the objects, presented in isolation on either left or right side of the screen, moved within the white box in various patterns while the object was being labeled.

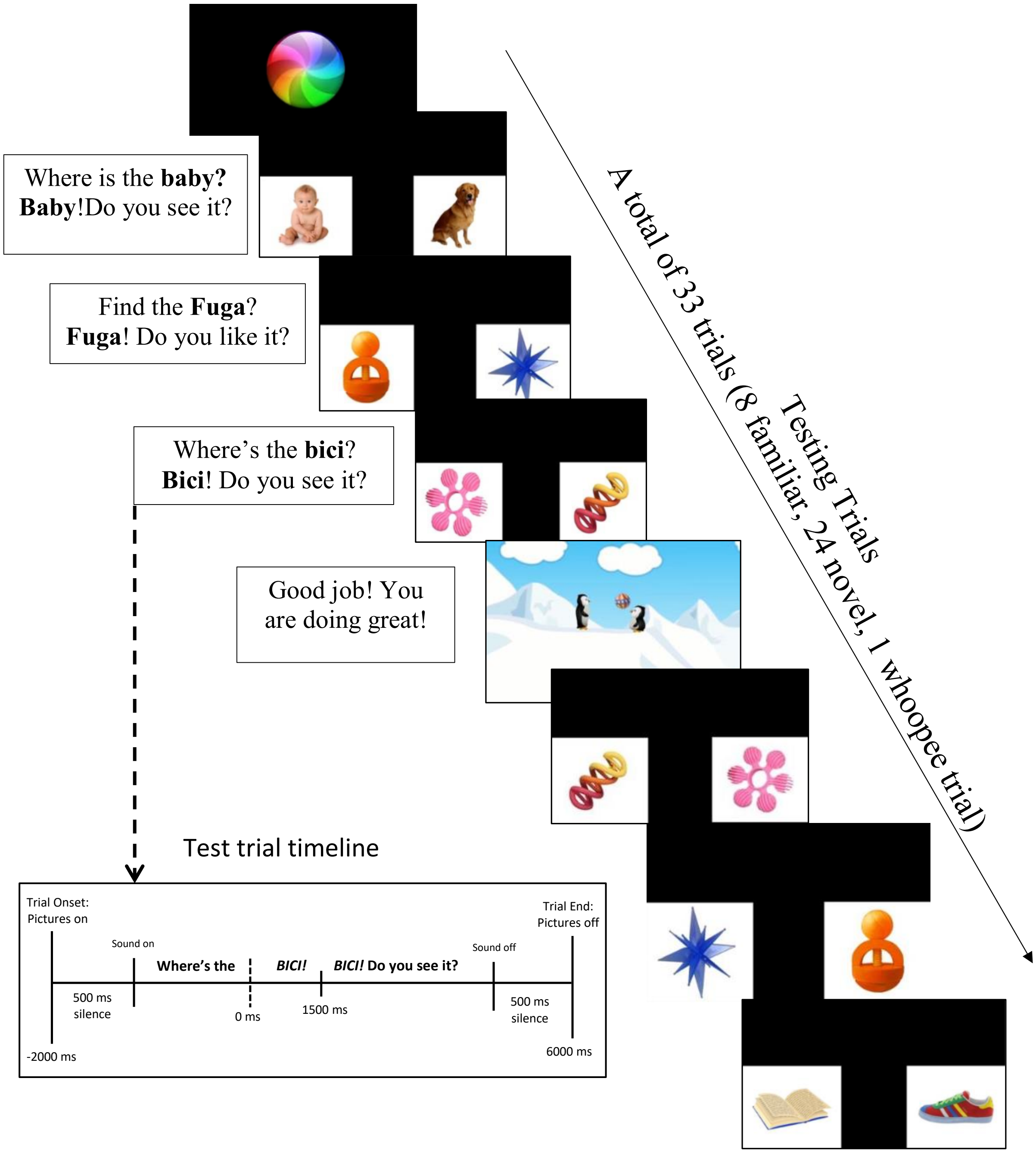

Figure 3. Overview of testing phase.

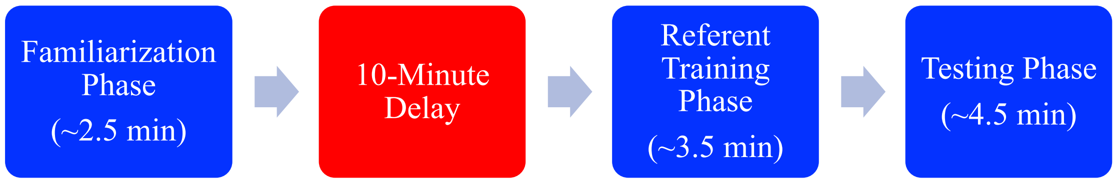

Figure 4. Overview of the experimental design at Experiment 1.

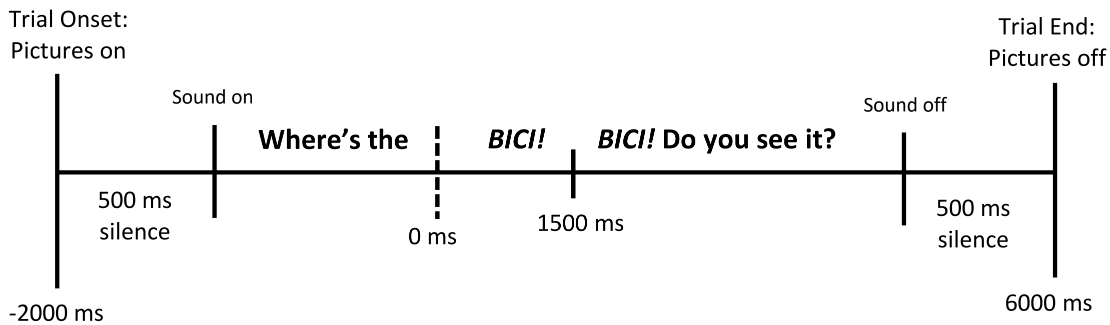

Figure 5. Schematic timeline for the test trials.

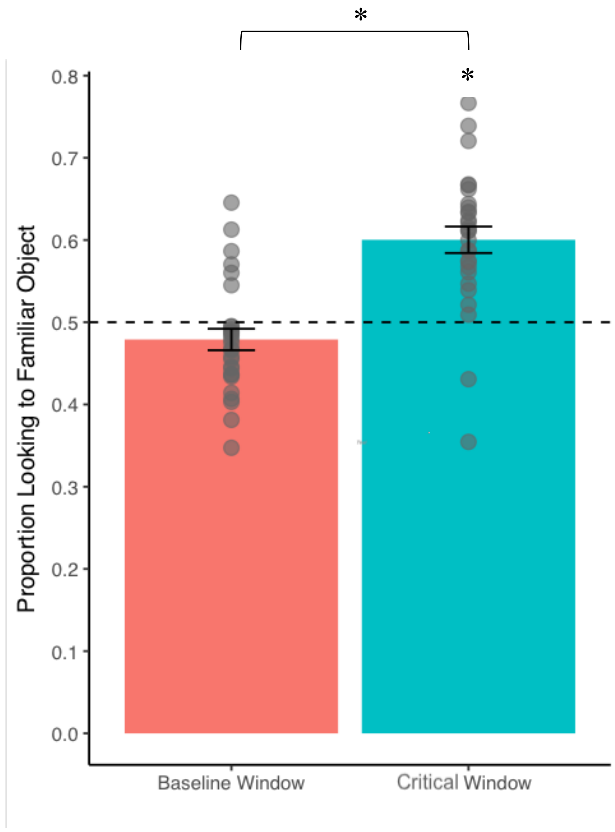

Figure 6. Mean proportion of time looking to the target object on familiar word test trials averaged across the baseline window (-2000 to 0 ms from familiar word onset) and the critical window (300-2000 ms) in Experiment 1. Data points represent the proportion for each infant averaged across trials. Error bars represent ± 1 *SE*. The dashed horizontal line at 0.5 marks equal looking to the target and distractor objects. * p <.001.

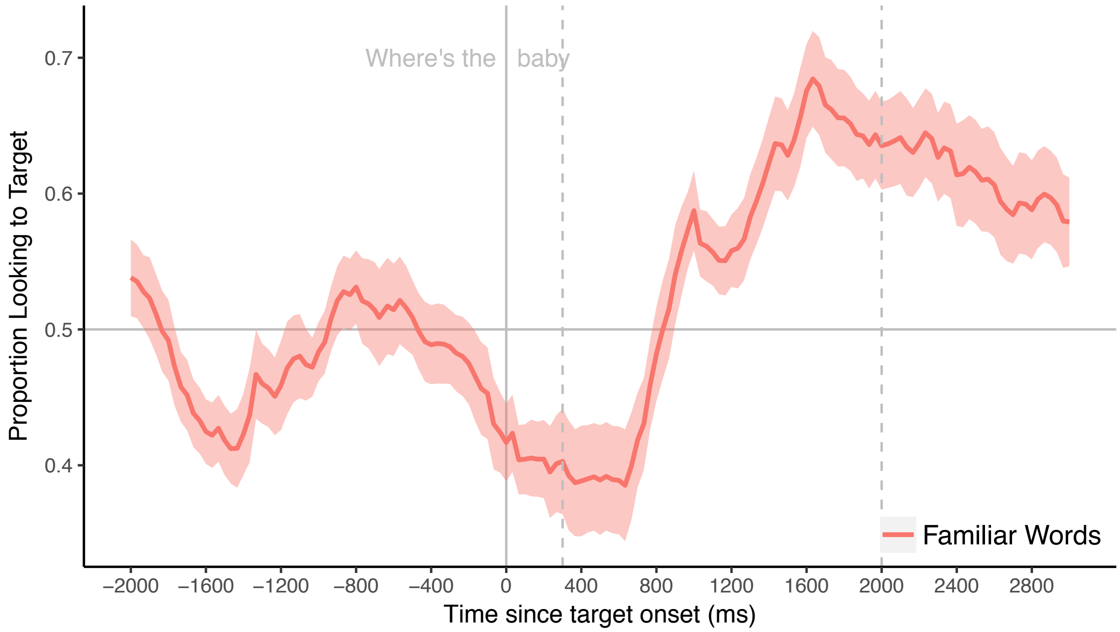

Figure 7. Mean proportion of looking time to the target object on familiar word test trials at each frame (33 ms interval) as a function of time (since the onset of the object label) in Experiment 1. The red line represents the proportion of fixations to the target object in 33 ms increments averaged across infants. The ribbon around the line indicates ± 1 SE. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the target familiar word. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

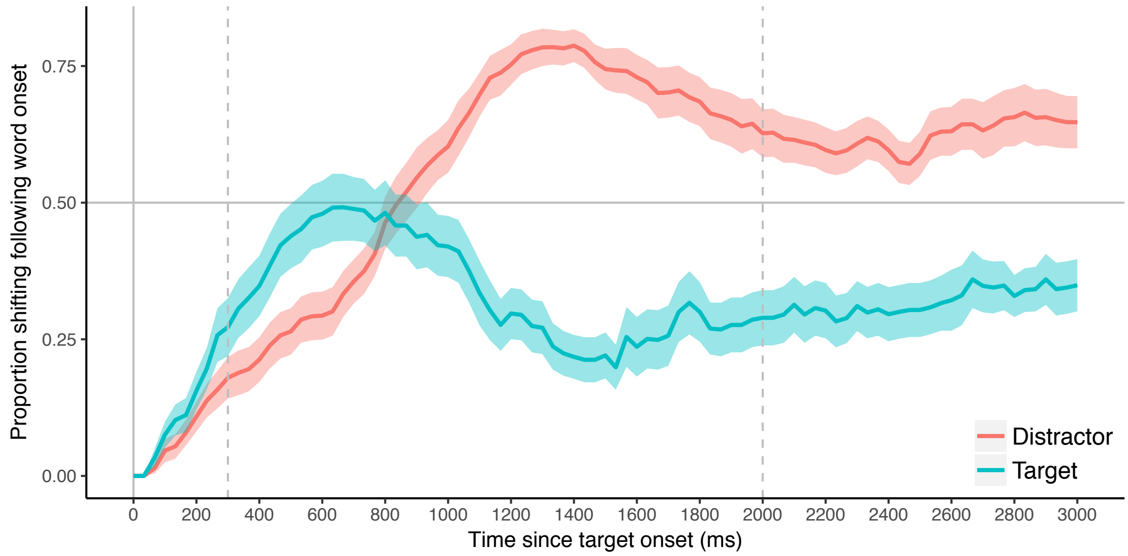

Figure 8. Onset-contingent (OC) plot: Proportion of shifting following familiar word onset on target initial trials and on distractor initial trials in Experiment 1. The ribbon around the line indicates ± 1 SE. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the familiar label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

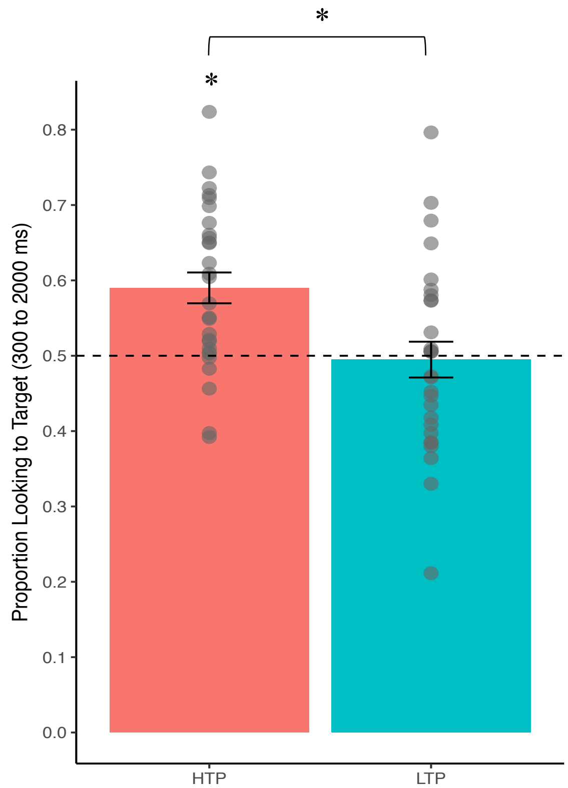

Figure 9. Mean proportion of looking time to target objects on HTP and LTP test trials averaged across the critical window (300-2000 ms) in Experiment 1. Data points represent the proportion for each infant averaged across trials. Error bars represent ± 1 *SE*. The dashed horizontal line at 0.5 marks equal looking to the target and distractor objects. * p <.001.

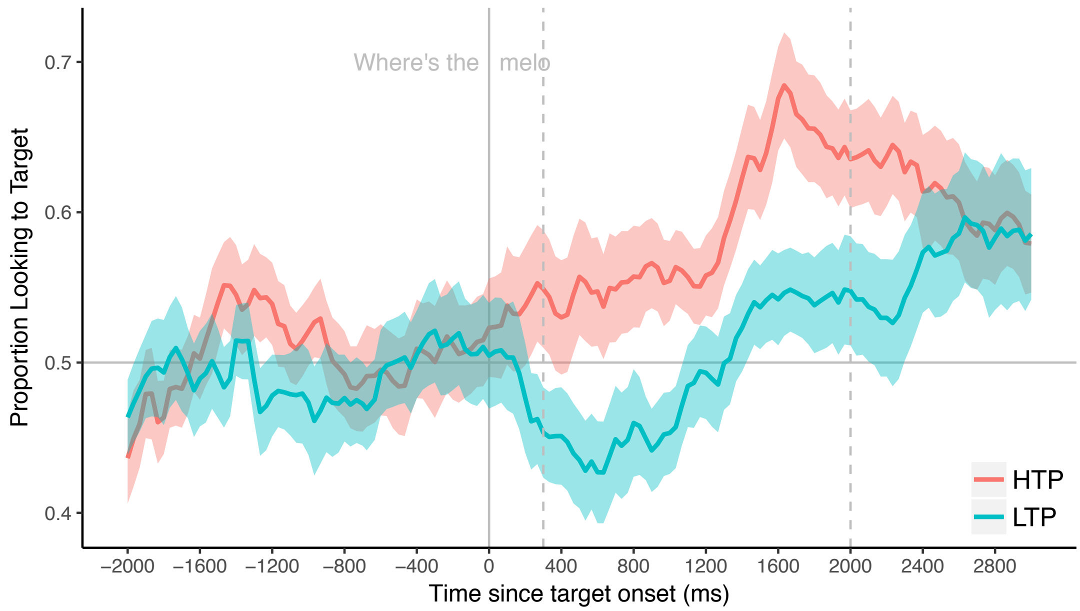

Figure 10. Mean proportion of time looking to the target objects on HTP and LTP test trials at each frame (33 ms interval) as a function of time (since the label onset) in Experiment 1. The red line represents the proportion of time looking to the target HTP object and green line represents the proportion of time looking to the target LTP object in 33 ms increments averaged across infants. The ribbon around the lines indicates ± 1 SE. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

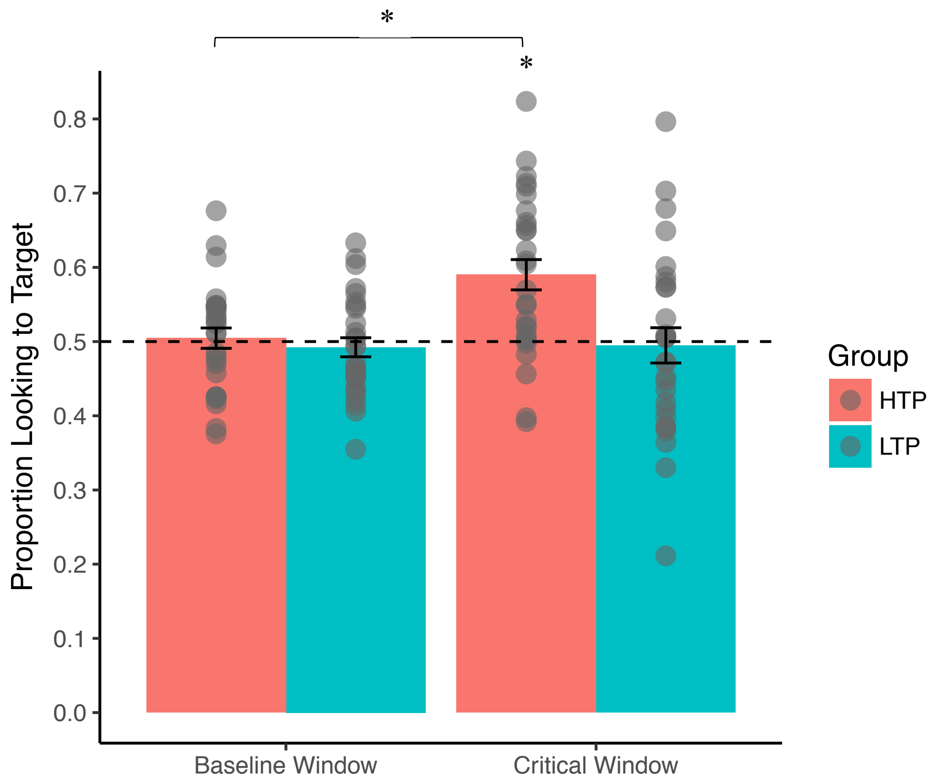

Figure 11. Mean proportion of time looking to the target object on HTP and LTP test trials averaged across the baseline window (-2000 to 0 ms from familiar word onset) and the critical window (300-2000 ms) in Experiment 1. Data points represent the proportion looking at the target for each infant averaged across trials. Error bars represent ± 1 *SE*. The dashed horizontal line at 0.5 marks equal looking to the target and distractor objects.
* $p < .001$.

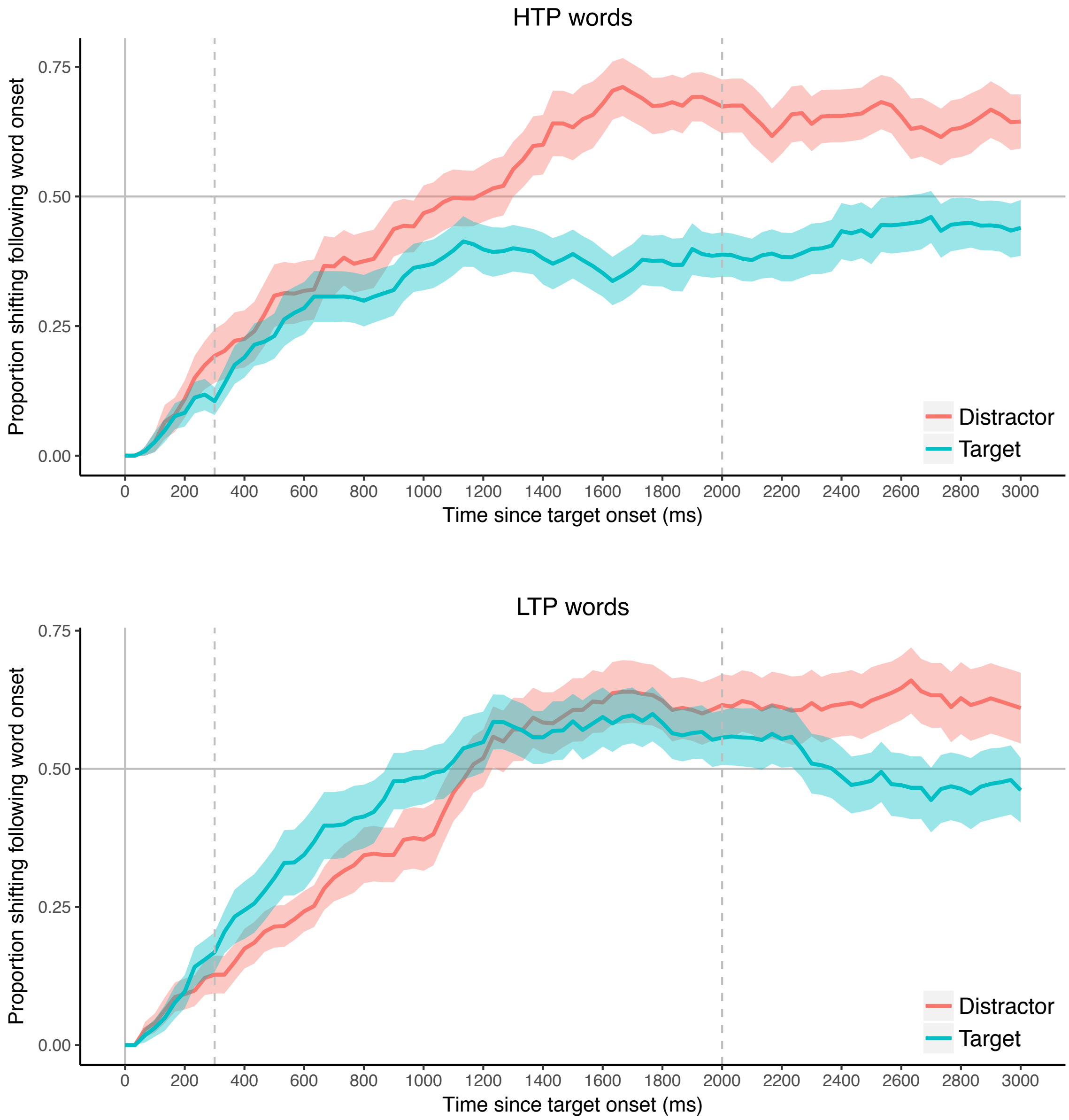

Figure 12. Onset-contingent plots for HTP (top) and LTP trials (bottom): Proportion shifting following word onset on target initial trials and distractor initial trials in Experiment 1. The ribbon around the line indicates ± 1 SE. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

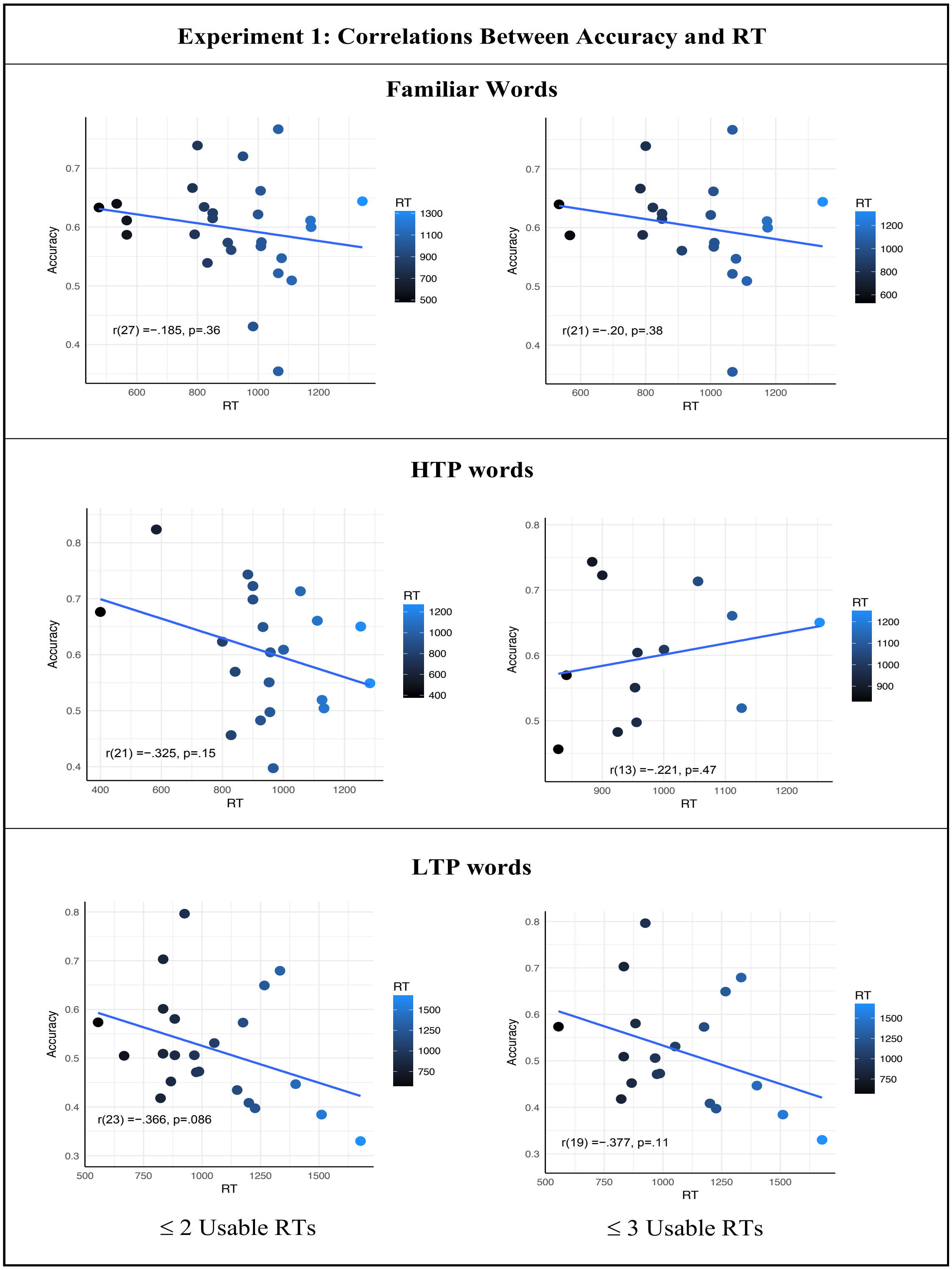

Figure 13. Correlations between accuracy and RT for infants who have at least 2 usable RT testing trials (left) and at least 3 usable RT testing trials (right) for familiar words (top), HTP (middle) and LTP words (bottom). The Blue lines represent the regression line.

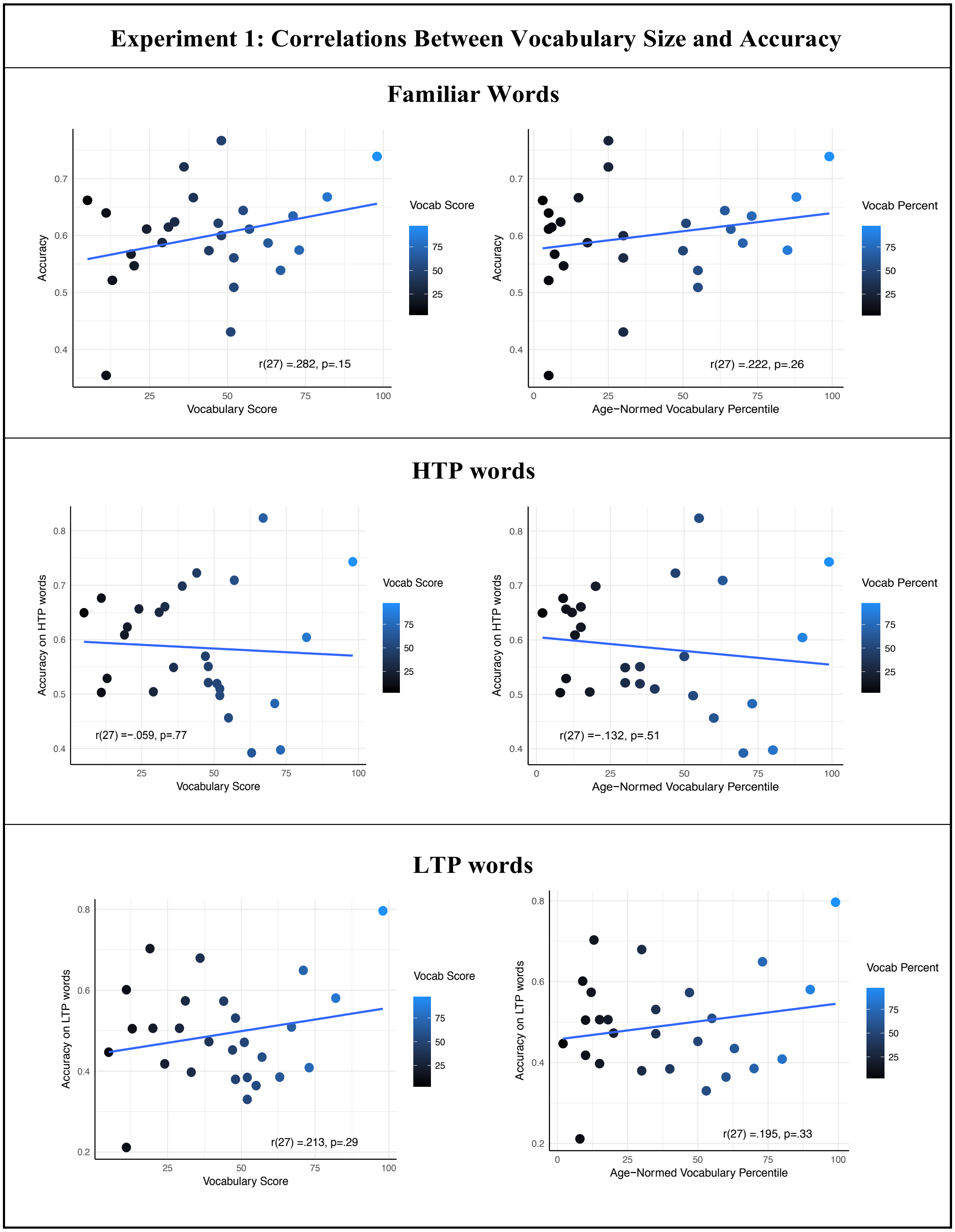

Figure 14. Correlations between vocabulary score and accuracy (left) and age-normed vocabulary percentiles and accuracy (right) for familiar (top), HTP (middle) and LTP words (bottom). Blue lines represent the regression line.

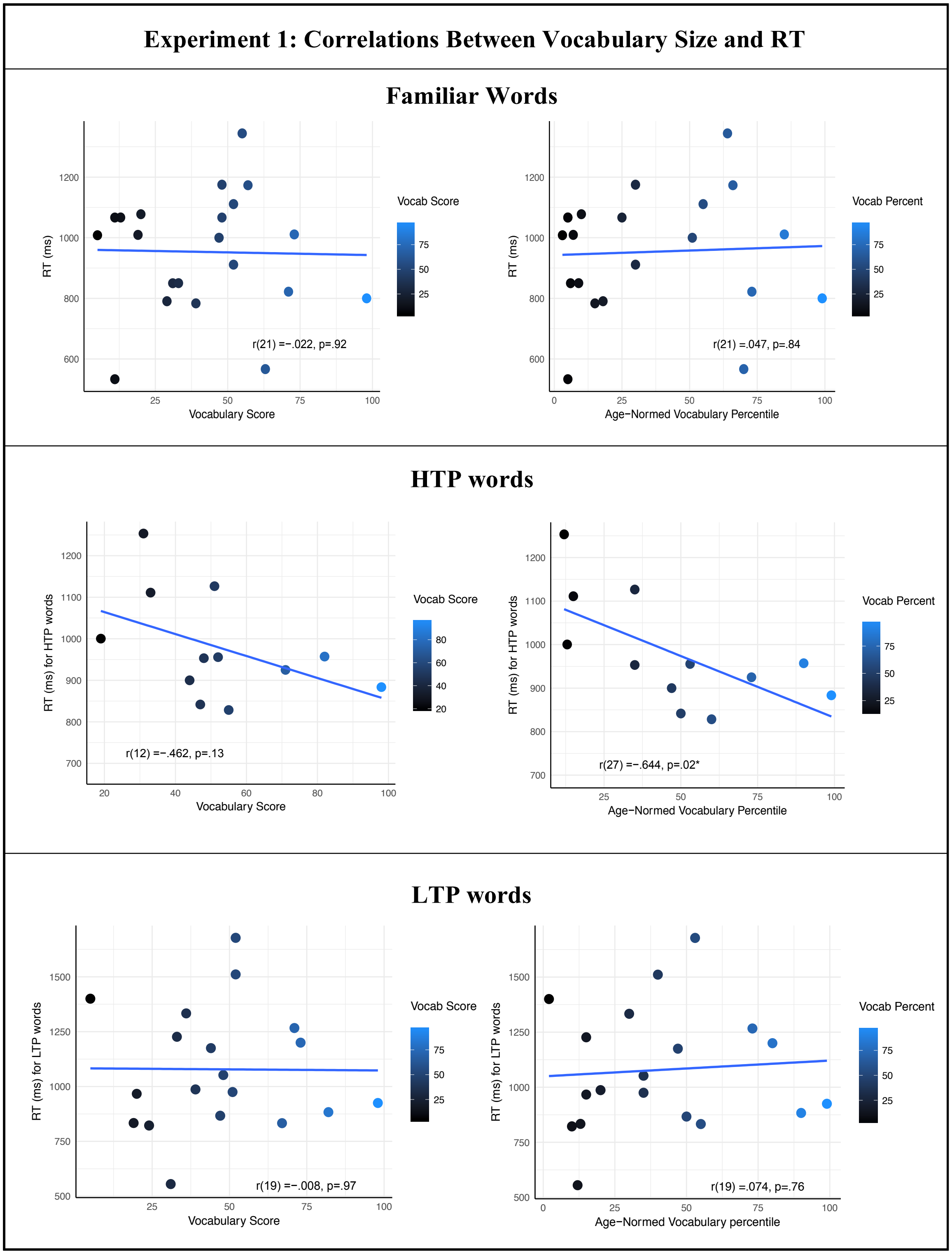

Figure 15. Correlations between vocabulary score and RT (left) and age-normed vocabulary percentiles and RT (right) for familiar (top), HTP (middle) and LTP words (bottom). Blue lines represent the regression line. RT includes only infants who have at least 3 usable test trials. *p<.05.

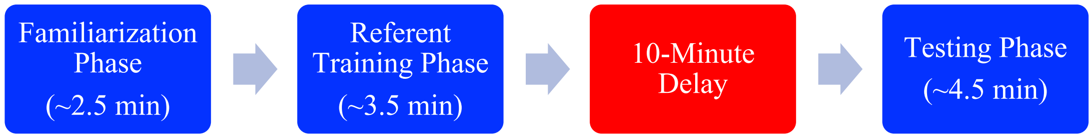

Figure 16. Overview of the experimental design at Experiment 2.

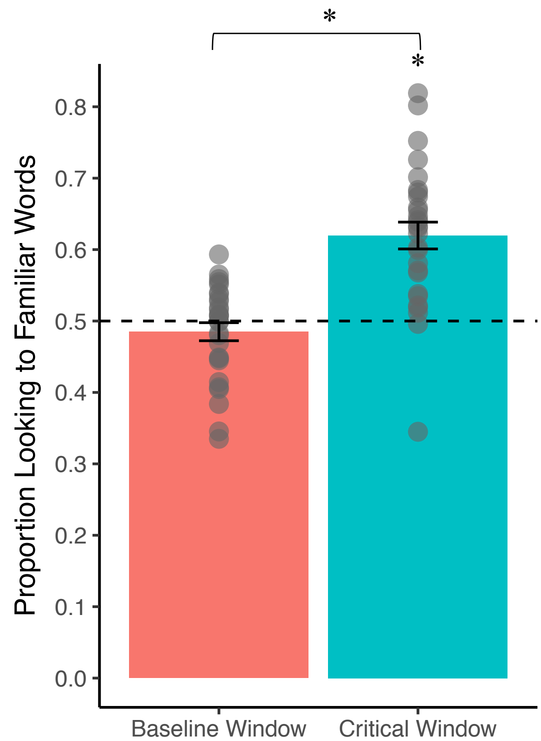

Figure 17. Mean proportion of time looking to the target object on familiar word test trials averaged across the baseline window (-2000 to 0 ms from familiar word onset) and the critical window (300-2000 ms) in Experiment 2. Data points represent the proportion looking to the target object for each infant averaged across trials. Error bars represent ± 1 *SE*. The dashed horizontal line at 0.5 marks equal looking to the target distractor objects. * p <.001.

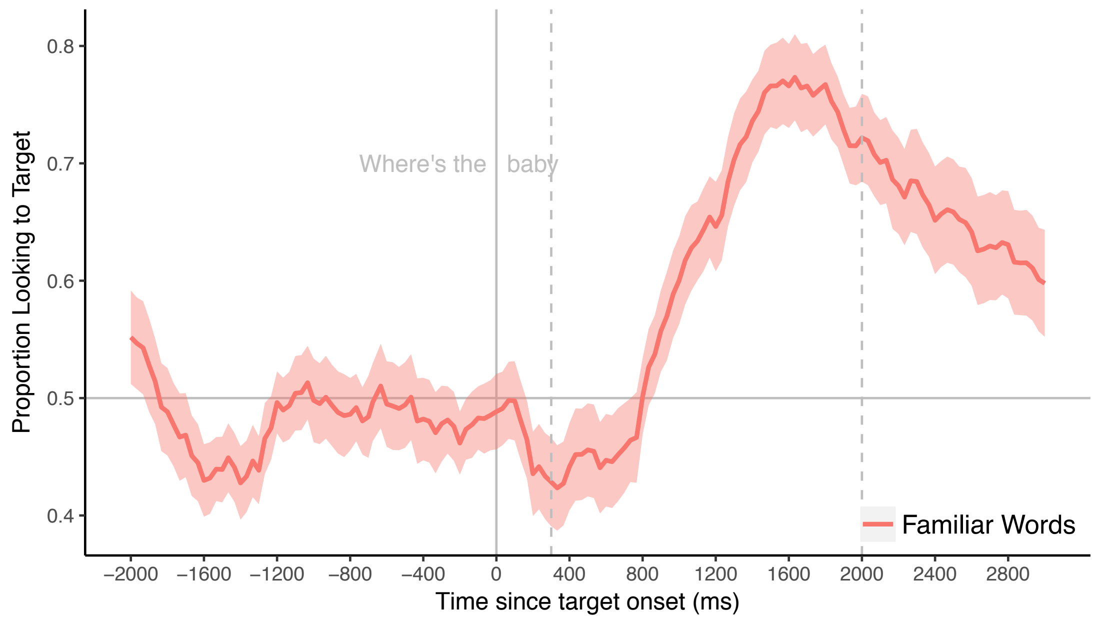

Figure 18. Mean proportion of time looking to the target object on familiar word test trials at each frame (33 ms interval) as a function of time (since label onset) in Experiment 2. The red line represents the proportion of fixations to the target familiar object in 33 ms increments averaged across infants. The ribbon around the line indicates ± 1 *SE*. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

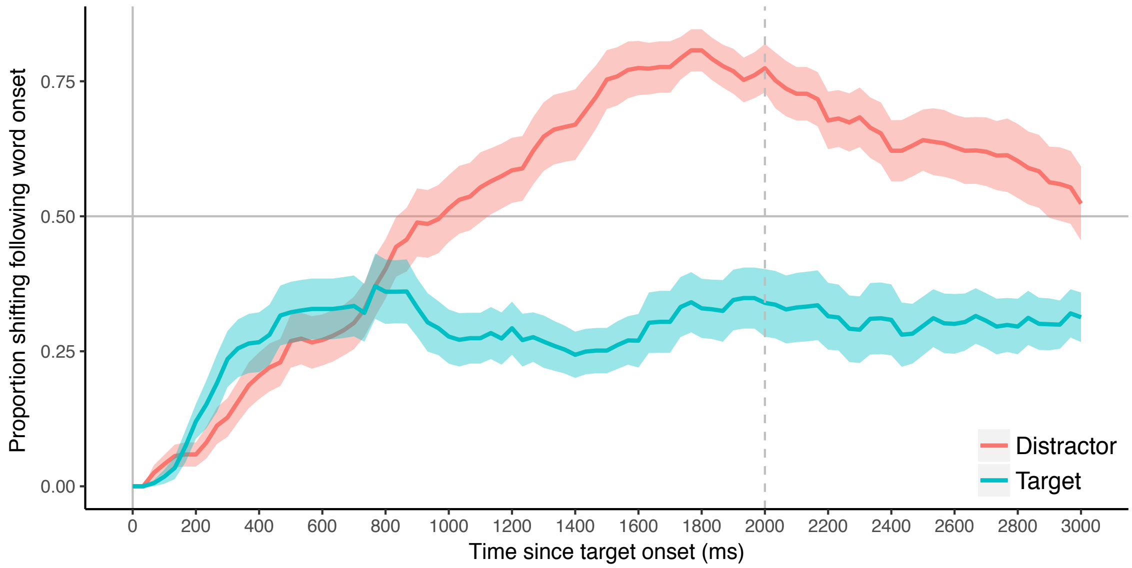

Figure 19. Onset-contingent (OC) plot: Proportion of shifting following familiar word onset on target initial trials and on distractor initial trials in Experiment 2. The ribbon around the lines indicate ± 1 *SE*. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the familiar label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

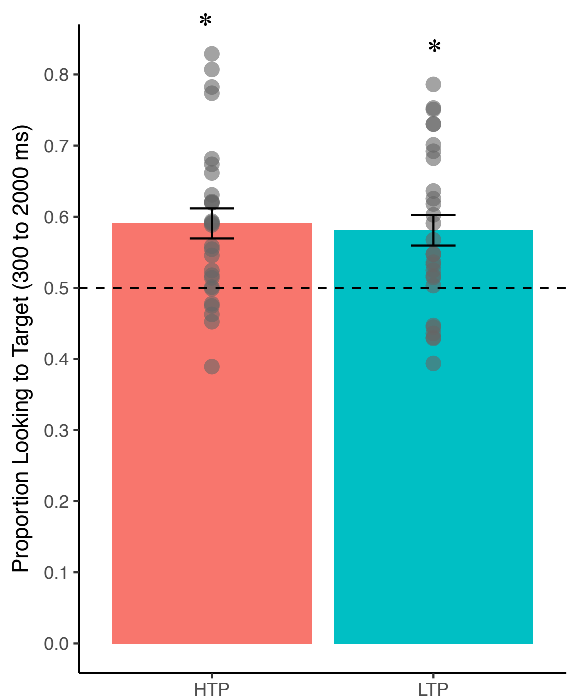

Figure 20. Mean proportion of time looking to the to the target object on HTP and the LTP test trials averaged across the critical window (300-2000 ms) in Experiment 2. Data points represent the proportion for each infant averaged across trials. Error bars represent ± 1 *SE*. The dashed horizontal line at 0.5 marks equal looking to the target and distractor objects. * $p < .001$.

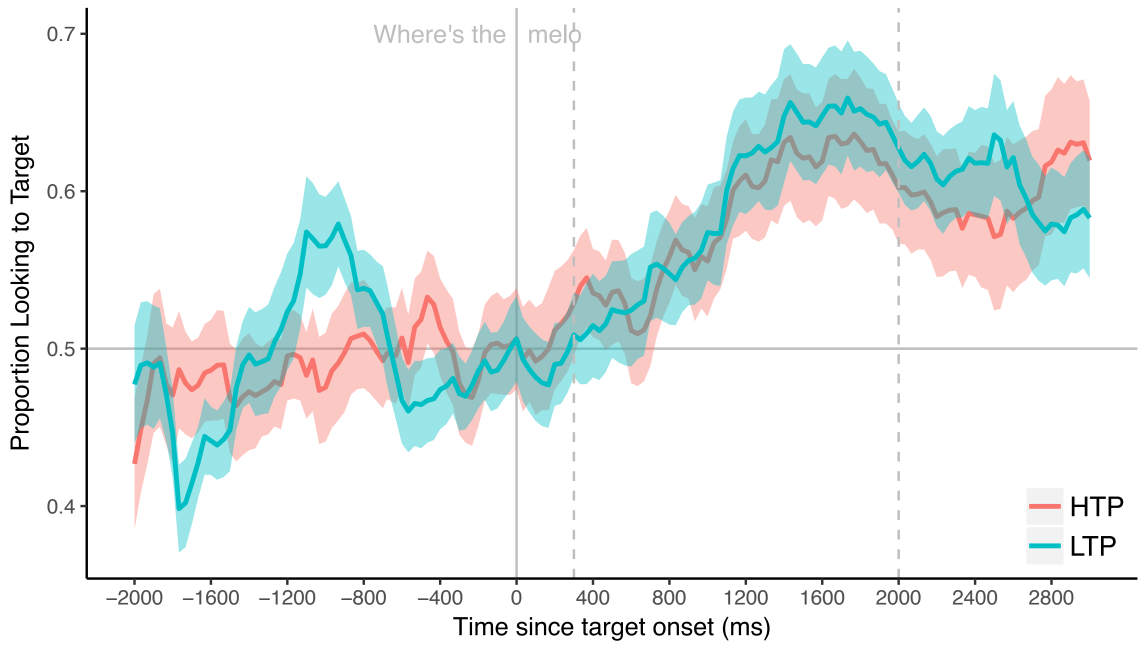

Figure 21. Mean proportion of time looking to the target object on HTP and LTP test trials at each frame (33 ms interval) as a function of time (since label onset) in Experiment 2. The red line represents the proportion of fixations to the target HTP object and green line represents the proportion of fixations to the target LTP object in 33 ms increments averaged across infants. The ribbon around the lines indicate ± 1 *SE.* The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the label. The dashed vertical lines represents the onset (300 ms) and offset (2000 ms) of the critical window.

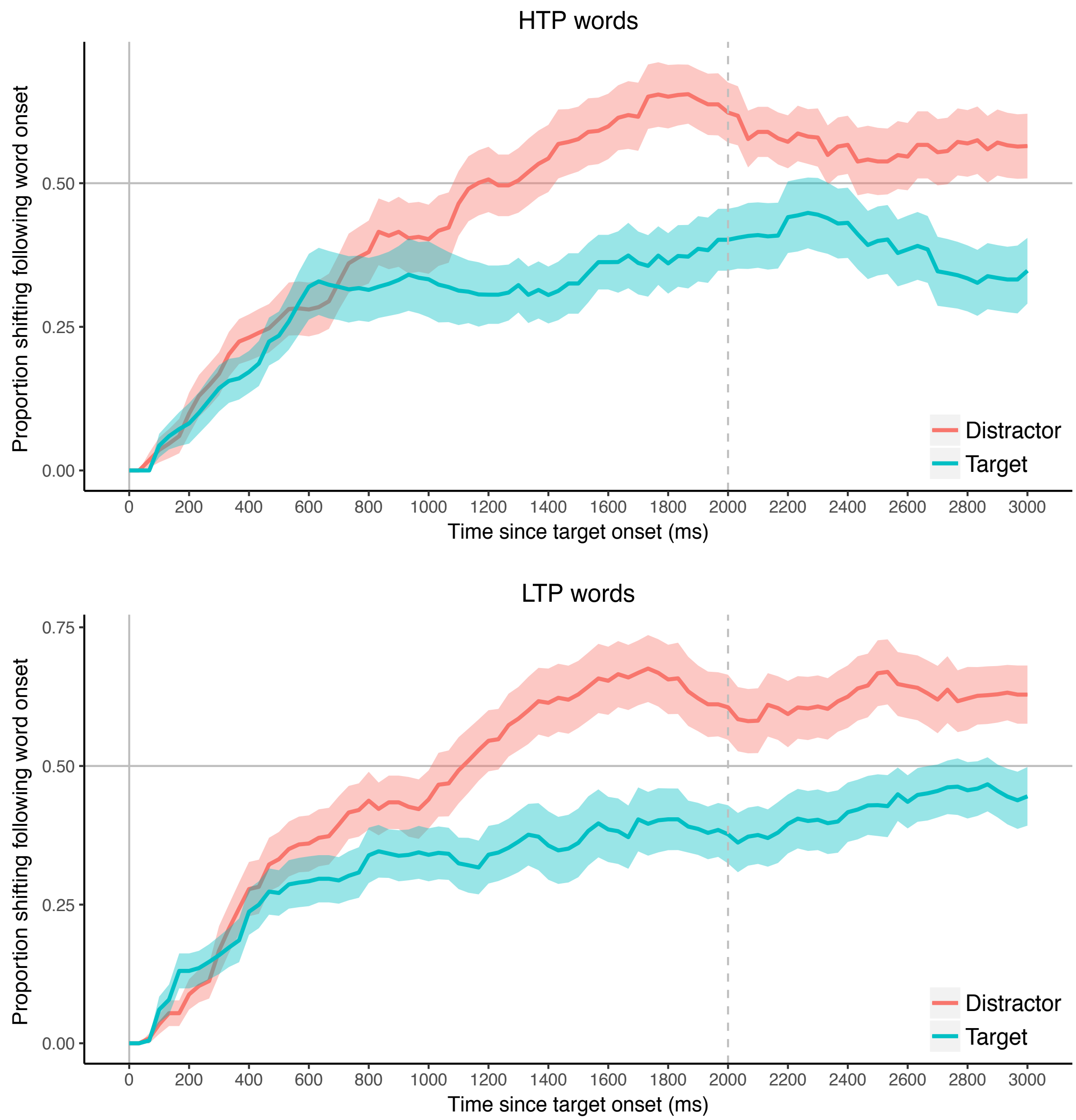

Figure 22. Onset-contingent plots for HTP (top) and LTP words (bottom): Proportion shifting following word onset on target initial trials and distractor initial trials in Experiment 2. The ribbon around the lines indicate ± 1 *SE*. The solid horizontal line represents the 0.5 chance level. The solid vertical line represents the onset (0 ms) of the label. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

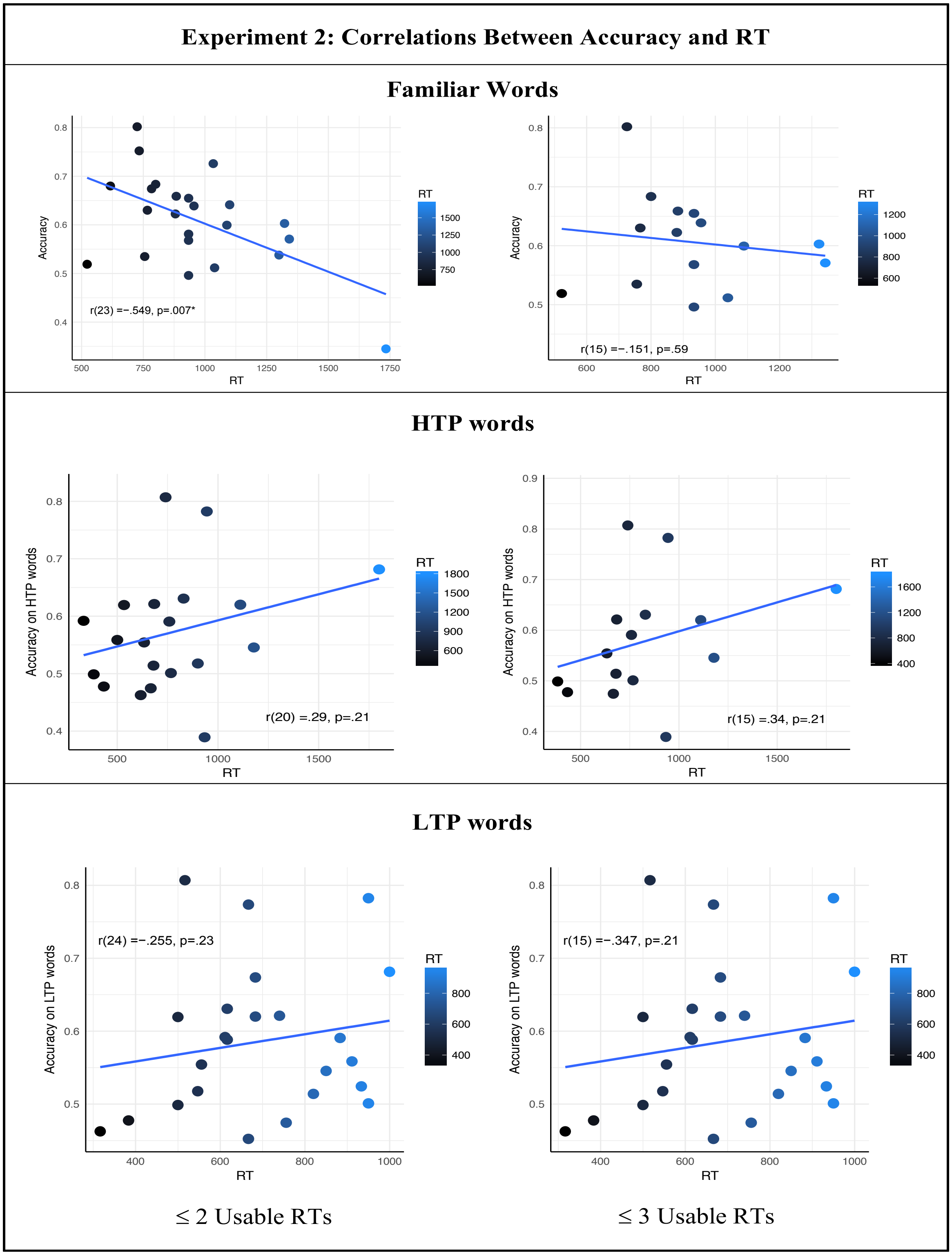

Figure 23. Correlations between accuracy and RT for infants who have at least 2 usable RT test trials (left) and at least 3 usable RT test trials (right) for familiar words (top), HTP (middle) and LTP words (bottom). Blue lines represent the regression line. *$p<.01$.

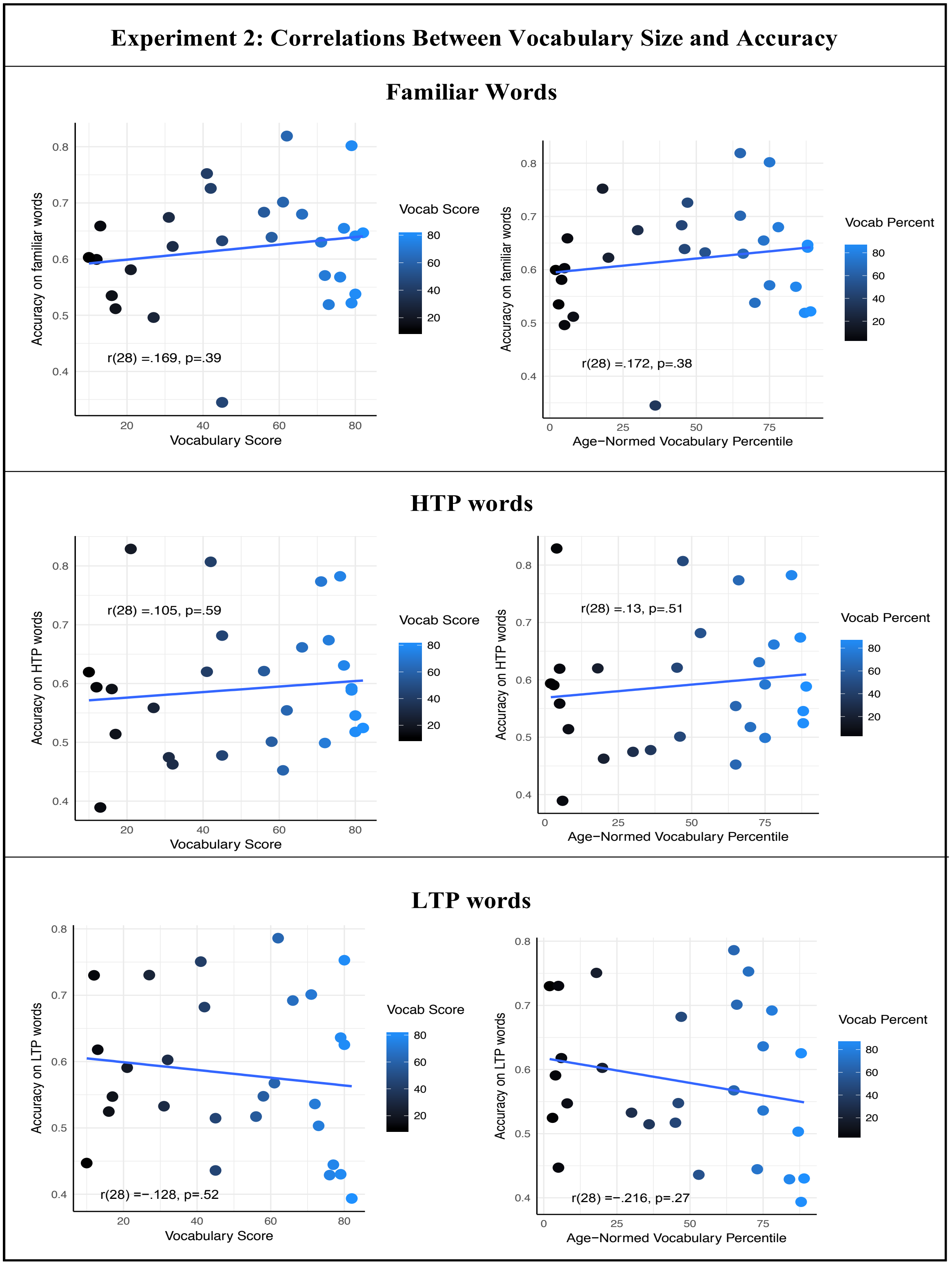

Figure 24. Correlations between vocabulary score and accuracy (left) and age-normed vocabulary percentiles and accuracy (right) for familiar (top), HTP (middle) and LTP words (bottom). Blue lines represent the regression line.

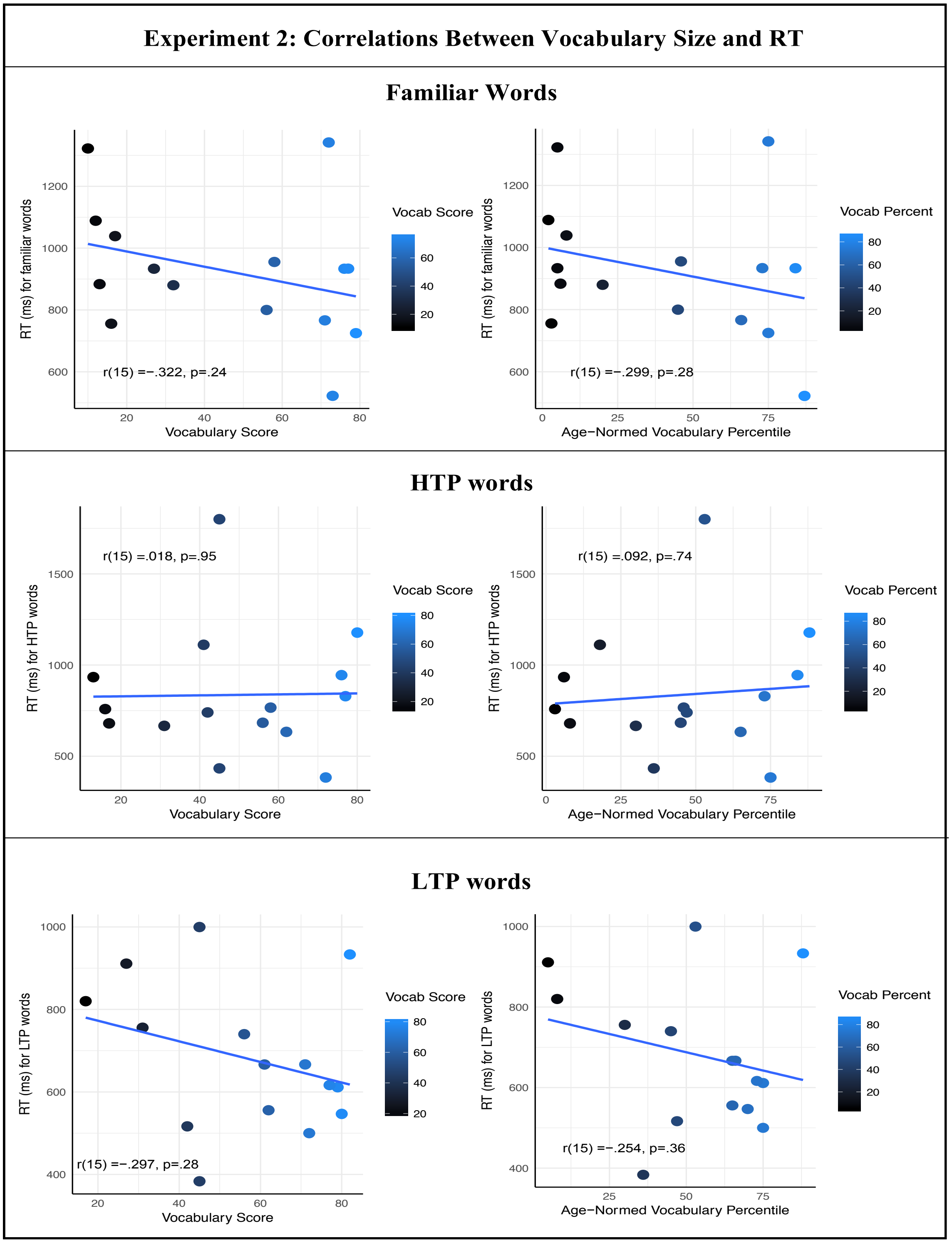

Figure 25. Correlations between vocabulary score and RT (left) and age-normed vocabulary percentiles and RT (right) for familiar (top), HTP (middle) and LTP words (bottom). Blue lines represent the regression line. RT includes only infants who have at least 3 usable test trials.

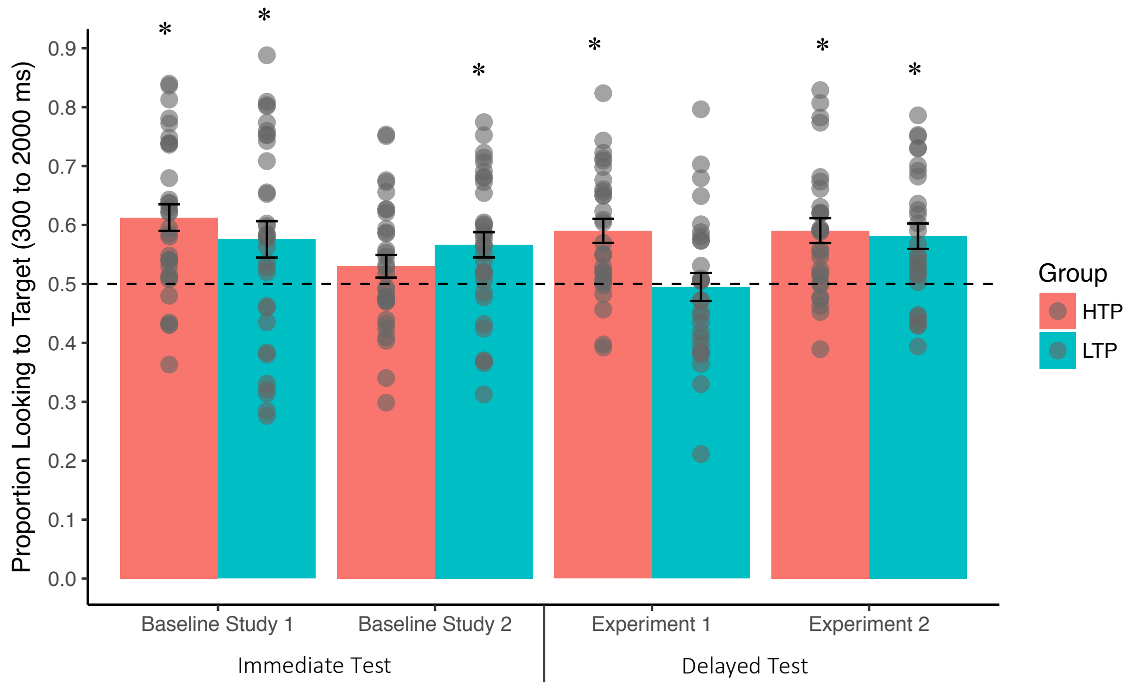

Figure 26. Mean proportion of time looking to the target objects on HTP and LTP test trials averaged across the critical window (300-2000 ms) in Baseline study 1 and 2 (Hay et al., 2017) and Experiment 1 and 2. Data points represent the proportion of looking to the target object for each infant averaged across trials. Error bars represent ± 1 *SE.* The dashed horizontal line at 0.5 marks equal looking to the target and distractor objects. * $p < .001$.

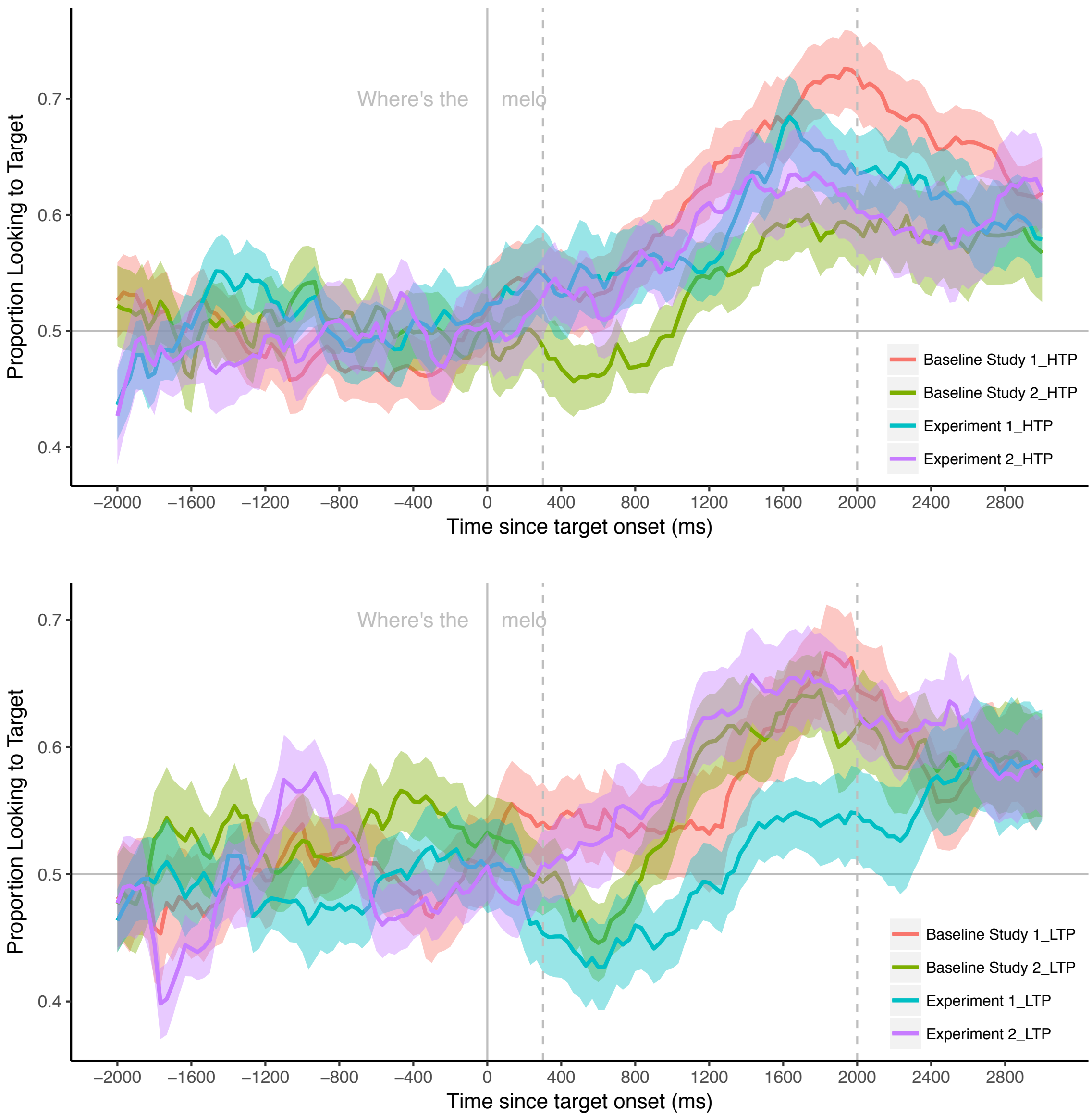

Figure 27. Mean proportion of time looking to the target objects on HTP and LTP test trials at each frame (33 ms interval) as a function of time (since the onset of the target familiar word). Red line represents the proportion of fixations to the target HTP object and green line represents the proportion of fixations to the target LTP object in 33 ms increments averaged across infants. Ribbon around the lines indicated ± 1 SE. The solid horizontal line represents the 50% chance level. The solid vertical line represents the onset (0 ms) of the target familiar word. The dashed vertical lines represent the onset (300 ms) and offset (2000 ms) of the critical window.

Vita

Ferhat was born in Salihli, Manisa, Turkey to Esma Karaman and Yasar Karaman. He attended Salihli High School. Following high school, he attended the Sakarya University in Turkey, graduating in 2009 with a bachelor's degree in Counseling Psychology. Then, he started the master's program in Counseling Psychology in Sakarya University. At the same time, he was working as a psychological counselor in Geyve Dogantepe Boarding School in Sakarya, Turkey. While in the Master program at the Sakarya University, he won a prestigious scholarship from the Turkish Ministry of National Education to study abroad as a graduate student. From 2010 to 2011, he attended an Extensive English Program at the University of California, Davis. In 2011, Ferhat entered graduate school at the Florida Atlantic University, where he earned a Master of Arts degree in Cognitive Psychology with a focus on language and cognition. Following completion of his master's degree, he entered the doctoral program in Experimental Psychology at the University of Tennessee, where he worked on early language acquisition in Dr. Jessica F. Hay's Infant Language and Perceptual Learning Lab. During the second year of his Ph.D., Ferhat received another master's degree in Developmental Psychology. Beginning in early 2019, Ferhat will begin serving as a faculty in Department of Psychology in Usak University in Turkey.