Using the Headturn Preference Procedure to Assess Word Segmentation Skills in Dutch Infants

Master's Thesis – Linguistics (General Programme)

Radboud University

Lyan Porto

Supervisor: Prof. Dr. Paula Fikkert

Second reader: Dr. C. Junge

Table of Contents

Acknowledgments

Writing a thesis has been undoubtedly the single biggest challenge I have ever faced. The constant need to operate at one's maximum capacity for concentration and rational thought is something I did not expect to manage. And I'd be remiss to say that I've succeeded when, in fact, I may never have finished it without the help of many generous people, often purely out of the kindness of their hearts. First and foremost, Imme, for taking time in her busy schedule to show me the ropes and accompany me through most of the experiments, as well as for her invaluable advice and feedback every step of the way; I couldn't have asked for a better guide. On that topic, I must also give my sincerest thanks to Angela and the rest of the Babylab staff for all the help running the experiment and dealing with our participants' parents and siblings. Likewise, I am deeply grateful to Maaike, for still more help with parents and for serving as my second coder; I wish the best of luck to you moving forward. And one cannot overstate the impact of Julia's support and assistance, without which this thesis would be devoid of statistics.

But before testing could even begin, when defining our design, I could not have done without the guidance and feedback of Caroline and Titia, who, even from afar, were always quick, patient, and to the point. In the same vein, thanks to Paula and the FLA group for the support, the BRC meeting attendees for their questions and feedback, and Christina for the massive assistance that made some of the more daunting tasks manageable.

Also many thanks to Bárbara for her illustration work, and to the many wonderful people I met during the past year in this amazing country, who made the much-needed breaks something to look forward to. Finally, my sincere thanks to all the parents, caretakers and infants who participated. And last, but not least, the deepest gratitude for my family, without whose support this entire endeavor would not have even started.

Every one of these people made this work possible. Thank you.

Abstract
Among the many skills infants must develop in their first year of life in order to acquire language is the ability to segment words from fluent speech. Literature in the past decades has come a long way in uncovering how and when this ability emerges, much of it using the Headturn Preference Procedure (HPP). Though Dutch infants have been shown in EEG experiments to be able to perform word segmentation (Junge, Cutler, and Hagoort, 2012), there are few successful experiments with Dutch infants using the HPP. These studies suggest a delay in the emergence of segmentation skills in Dutch infants (Kuijpers, Coolen, Houston, and Cutler, 1998). It has been hypothesized that the order of presentation used in these studies (familiarization with words in isolation, test with passages) makes the task more difficult (Nazzi, Iakimova, Bertoncini, Frédonie, and Alcantara, 2006) than familiarization with words in passages. The present study aimed to test Dutch infants on a word-segmentation task using the HPP with familiarization with word in passages and testing with words in isolation. As hypothesized, Dutch infants were able to correctly segment words from fluent speech under these circumstances. Our results support the interpretation that there is a benefit to the passages-first order of presentation, though word segmentation tasks may still emerge later for Dutch than American infants.

Keywords: headturn preference procedure, word segmentation, Dutch infants, order of presentation

Introduction

During the first year of life, infants will develop an important set of skills to aid them in the process of language acquisition. Research in the last few decades has shed light on the development of some these skills (see Kuhl, 2004, for a review). Among the most important of these skills is word segmentation from fluent speech. Normal speech, even when directed at infants, does not usually or reliably present acoustic discontinuities, or silences, between words, making the identification of word boundaries a challenging task. Unlike adults acquiring a second language, who might make use of, among others, cognates, knowledge of syntactic structure and formal instruction to aid them in this task, infants must rely on other mechanisms. Several such mechanisms have been proposed, as we will see, each seeming to present a part of the solution. Moreover, it appears that many of the cues proposed occur at the same time, with their relative weights shifting during development (Nazzi, Mersad, Sundara, Iakimova, and Polka, 2014).

In order to test such proposals, many studies have been carried out, and one experimental method stands out: the Headturn Preference Procedure (HPP) (Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss and Kennedy, 1987; Jusczyk and Aslin, 1995). This method works broadly as follows: a child is placed in the center of a three-sided booth outfitted with red lights on the side walls and a blue light on the center wall. When the infant orients to the center light for a set minimum amount of time, the light is extinguished and one of the side lights begins flashing. When the infant orients to the flashing side light, the stimulus begins playing, and stops either when the child looks away for a set amount of time, or when it plays to completion. A typical segmentation experiment consists of two phases. In the first, the familiarization phase, the infant is exposed to one or more words. Then, in the test phase, they are then presented with these familiarized words and control words, and longer looking times to one kind of stimulus over the other provides evidence that they were able to recognize and thus segment the familiarized words.

While much research has been done using this method since Jusczyk and Aslin (1995), not much is known about the abilities of infants acquiring languages other than American English. Less than 1.5% of over two hundred HPP experiments included in a database of published word segmentation studies used in a meta-analysis by Bergmann and Cristia (2015) were conducted on Dutch infants.

American English-learning infants have been shown to be able to segment bisyllabic words at 7.5-months of age (Jusczyk, Houston, and Newsome, 1999, experiment 1). Dutch infants, on the other hand, were only able to perform this task at 9 months of age and not earlier (Kuijpers et al., 1998, experiment 3), suggesting that their word segmentation skills emerge later than for American English-learning infants.

Interestingly, in a subsequent experiment (experiment 2), Jusczyk et al. (1999b) found that if the familiarization phase was done with passages and the test phase with words in isolation (the opposite of both experiments mentioned above) infants showed an enhanced preference for familiar words relative to the results of experiment 1. This and similar findings (Nazzi et al., 2014) suggest that there is a benefit associated with the passages-first order of presentation that makes it easier for infants to perform segmentation.

In the present study, we run a word segmentation experiment using the HPP with 10-month-old Dutch-learning infants using the passages-first order. However, since easier test conditions have been associated with a preference for novel stimuli (see Background), it may be that the increased age and easier order of presentation relative to the experiments reported in Kuijpers et al. (1998) will lead to a switch in direction of preference. Therefore, our experiment attempts to answer the following questions:

1. Are 10-month-old Dutch-learning infants able to segment bisyllabic words from fluent speech if familiarization is done on passages?

2. If so, would the apparent added benefits of the passages-first order of presentation and increased age cause a different pattern of looking behavior, namely, a switch to a novelty preference?

The answers to these questions will further expand our knowledge of the emergence of segmentation abilities in Dutch infants, as well as our general understanding of the effects of changes in experimental conditions.

The structure of the remainder of this thesis is as follows: first, we present a review of the relevant literature using the Headturn Preference Procedure for word segmentation tasks, including its main findings, and a discussion of the differences between them and possible effects of these differences. This is followed by a description of the present study: our hypotheses, the method utilized for our investigation, and the results of our analyses. Finally, we discuss the implications of our findings, along with the limitations of this work and suggestions for further research.

Background

While the Headturn Preference Procedure (HPP) may be used for other tasks, since Jusczyk and Aslin's (1995) seminal work on speech segmentation, dozens of other studies have been conducted using the HPP to test different hypotheses of how and when speech segmentation emerges in infants. However, as we will see, there is little agreement over the specifics of its employment. Studies vary greatly in order of presentation (words first: Nazzi, Iakimova, Bertoncini, Frédonie, and Alcantara, 2006; passages first: Bosch, Figueras, Teixidó, Ramon-Casas, 2013), type of stimuli (natural language: Jusczyk and Aslin, 1995; pseudo-words: Mattys, Jusczyk, Luce, and Morgan, 1999; synthetic speech: Aslin, Saffran, and Newport, 1998), length of familiarization criteria (30 seconds: Barker and Newman, 2004; 45 seconds: Jusczyk et al., 1999b; 2 minutes: Saffran, Aslin, and Newport, 1996), among many others aspects. The impact of these differences on results is not yet fully understood. Nevertheless, the impact of studies using this method on the infant language acquisition literature is undeniable. What follows is an overview of some of the most important findings in the speech segmentation literature, the differences between studies and their possible effects on results.

Cues to segmentation

Early studies such as Jusczyk and Aslin (1995) established that successful segmentation is performed on monosyllabic words, and subsequent studies aimed to discover which cues infants used, or how detailed are the representations that they use. One of the first proposals for how infants begin segmenting words from fluent speech was related to rhythmic cues (Jusczyk, Cutler, and Redanz, 1993; Cutler, 1994). According to this proposal, infants would attend to strong syllables as an indication of a word onset, allowing them to segment both monosyllabic and bisyllabic words. Jusczyk et al. (1999b) found that American English-learning infants familiarized with words having a trochaic (i.e., strong-weak) stress pattern (*kingdom*, *hamlet*, *doctor*, *candle*) looked significantly longer at passages containing the familiar words than passages containing other words at 7.5 months of age. Subsequent experiments revealed that 7.5 month-olds showed no preference between familiarized and novel words if the familiarized words had an iambic (i.e., weak-strong) stress pattern, and that only at 10.5 months of age were infants successful in segmenting iambic words. However, evidence suggests that the preference for a given stress pattern is not innate, but rather language-dependent and must be acquired sometime in the first six months of life (Nazzi et al., 2006; Höhle, Bijelac-Babic, Herold, Weissenborn, and Nazzi, 2009).

Another important finding of HPP studies is that infants can use statistical cues, in particular transitional probabilities, to aid them in the task of segmenting words. Saffran, Aslin, and Newport (1996) used synthetic speech stimuli to test if infants were sensitive to differences in the probabilities of certain syllables following others. The authors argued that syllables within words are more likely to follow one another than syllables across word boundaries, that is, in a pair such as *pretty baby*, *ty* is more likely to follow *pre* than *ba* is to follow *ty*. This difference could cue infants in on the location of word boundaries. After being familiarized on speech streams of concatenated 'nonsense' trisyllabic words controlled for their transitional probabilities and devoid of any prosodic cues, infants were tested on lists of words whose syllables had high (familiar) or part-words with low (novel) transitional probabilities. Indeed, infants oriented reliably longer to novel (part-)words, indicating that successful segmentation had taken place using only this cue. Additionally, statistical cues have been shown to be both limited by uniformity of word length (Johnson and Tyler, 2010) and aided by adjacent known words (Mersad and Nazzi, 2012).

Other cues have also been proposed and supported by empirical data (allophonic cues: Jusczyk, Hohne, and Bauman, 1999; Mattys and Jusczyk, 2001; phonotactic cues: Mattys et al., 1999), all of which serve to underscore the importance of this method to adding pieces to the puzzle of how infants segment words.

Order of presentation and cross-linguistic differences

The HPP usually consists of two consecutive phases, the familiarization phase, when the stimuli are presented, and the test phase, when the familiar stimuli are presented alongside novel stimuli. In the familiarization phase words can be presented either in isolation, or in passages. This allows for at least two possible presentation orders: words-first, wherein the familiarization phase is done with repetitions of the target words in isolation, and the test phase is done with passages that either include the target words or not; or passages-first, wherein the familiarization phase is done with passages and the test phase is done with words in isolation. In a meta-analysis by Bergmann and Cristia (2015), based on a survey of word segmentation studies with natural language, it seems that the words-first order is more common, being used in 159 individual experiments, whereas the passages-first order was used in 39. Nonetheless, studies that have tested both directions found largely converging results (Jusczyk and Aslin, 1995); Jusczyk et al. (1999b) found a stronger preference for familiar words using the passages-first order relative to the words-first order. Likewise, Van Heugten and Johnson (2012) argue that a shift from words-first to passages-first may have helped their participants to successfully segment words when familiarization stimuli were recorded by a female talker and test stimuli by a male one, where Houston and Jusczyk's (2000) age-matched participants had failed under otherwise similar conditions.

More direct evidence for the enhanced effect of familiarization with words in passages comes from the findings that French-learning infants were only able to segment words at 16 months of age in a task using the words-first order, while evidence for segmentation could be seen starting at 8 months if familiarization was done with passages (Nazzi et al., 2014). This remarkable 8-month difference led the authors to argue that the words-first order is more cognitively demanding, as it requires two tasks during the test phase (segmenting the passages and comparing them to familiarized words) as opposed to a single task in the passages-first order (comparing the isolated tokens to previously segmented and memorized words in the test phase).

The fact that such a strong advantage was not found for English-learning infants (Jusczyk et al., 1995) appears to be related to the rhythmic class of the language tested, which has been shown to modulate the results of word segmentation studies using the HPP. The syllables of syllable-timed languages like Spanish and French all have roughly the same duration, regardless of stress. The syllables in stress-timed languages like English and Dutch, on the other hand, have varying duration depending on their stress. The results of a series of experiments by Nazzi et al. (2006) corroborated the authors' hypothesis that the relevant unit for segmentation in syllable-timed languages is the syllable, whereas the relevant unit for that task in stress-timed languages is the (trochaic) foot. In this study, 12-month-old French-learning infants oriented longer to bisyllabic words after having been familiarized to the individual syllables of these words, suggesting that for them segmentation happens for each syllable, rather than for the rhythmic foot as a whole. In support of this, Bosch et al. (2013) found that the onset of segmentation of monosyllabic words takes place at 6 months of age for infants acquiring Spanish and Catalan, while infants acquiring English can only perform this task at 7.5 months (Jusczyk and Aslin, 1995). Thus, since reliance on transitional probabilities is greater for infants acquiring syllable-timed languages and transitional probabilities are easier to exploit in the passages-first word order, these infants should benefit more from this order of presentation than infants acquiring stress-timed languages (Nazzi et al., 2014).

Further support for this proposed cross-linguistic difference is found in a study by Höhle et al. (2009). Although German infants did not show the capability to segment words from fluent speech at 6 months (Höhle and Weissenborn, 2003), they showed a bias for trochaic units, unlike in their French-learning peers, even though the latter were capable of discriminating between the trochaic and iambic stress patterns. The fact that no bias is not found in French-leaning infants and that the trochaic bias is present before the onset of segmentation by German-learning infants (which is thought to be between 6 and 9 months; Höhle and Weissenborn, 2003) suggests three things. First, that the trochaic bias is not, as suggested by Allen and Hawkins (1978), innate (otherwise the French-learning infants would show it); second, that it is not the result of correct identification of the predominant stress pattern in that language (otherwise the French-learning infants would show an iambic pattern); and third, that the abovementioned identification is not the result of analyses made possible by successful segmentation (otherwise it would only emerge after and not before the onset of segmentation). What is still not yet understood is why the bias appears to be always trochaic and never iambic, in correspondence with the predominant stress pattern of the stress-timed languages studied. More research on different languages is needed to disentangle the presence of a bias from its stress pattern.

Characteristics of the stimuli and experimental criteria

Stimuli used in HPP studies can present a wide range of characteristics. The most common type of stimuli found in the literature is real words taken from the language being acquired (Jusczyk and Aslin, 1995; Jusczyk et al., 1999b; Nazzi et al., 2005). Authors are seldom clear on the criterion for selecting words (but see Kuijpers et al., 1998), thus one cannot rule out the possibility that at least some of these words (e.g., *feet* in Jusczyk and Aslin, 1995; *visit* in Nazzi et al., 2005) have been heard before. If that is the case, the preference shown in segmentation tasks could be influenced by previous familiarity.

Pseudo-words (also called *novel words* or *nonwords*) and artificial languages remove these suspicions. Pseudo-words are here defined as individual words either presented in isolation only, or within a sentential context in the child's native language. Though these are common in word-learning and object-naming tasks, they have not been used extensively in the speech segmentation literature. Artificial languages, on the other hand, have been used in several studies (e.g., Saffran et al, 1996; Thiessen, Hill, and Saffran, 2005), many of which also make use of artificial speech. Studies with artificial speech often find novelty preferences, which has led authors to suggest that this is due to the fact that artificial speech possesses far less variability than natural speech, allowing infants a quicker familiarization (Bosch et al., 2013; Pelucchi, Hay, and Saffran, 2009), a topic to which we will return later.

Studies also differ in terms of whether stimuli are monosyllabic or bisyllabic. As argued in the previous section, infants acquiring syllable-timed languages appear to be at an advantage when segmenting monosyllabic words (Bosch et al., 2013; Jusczyk and Aslin, 1995), though little is known about their abilities to segment bisyllabic words. In an artificial language paradigm with synthesized speech, bisyllabic words have been shown to be segmented by 5.5-month-old Dutch-learning infants (Johnson and Tyler, 2010), but interestingly, no segmentation was found for an artificial language having both bisyllabic and trisyllabic words under the same conditions, even at 8-months. These results suggest that transitional probabilities, although a powerful mechanism in early speech segmentation, has its limitations. In stark contrast to these findings, Thiessen et al. (2005) found that 7-month-old infants were able to segment artificial languages with varying word length when the stimuli were not synthesized and spoken in a child-directed manner, suggesting that the limitations of word length on the use of statistical cues for word segmentation can be modulated by other factors, including other cues found in natural speech.

Other experimental criteria that vary across studies and have been shown to impact results are length of familiarization and minimum looking angle until a headturn is considered. Like the choice of words, the choice for a particular length of familiarization is seldom explicitly motivated, though it is possible to note that studies tend to increase the amount of time to reach criterion when the familiarization phase is done on passages rather than words. This tendency goes back to Jusczyk and Aslin (1995), where experiments 1 through 3 used lists of isolated words in the familiarization phase and only 30s of accumulated looking time to each word. Experiment 4, in contrast, familiarized infants with passages, and the criterion was raised to 45s. This follows from the fact that passages contain much fewer instances of the target words than lists, so an extended criterion ensures that in both conditions infants are exposed to a comparable number of target word tokens.

This criterion for looking time appears to have been taken up by studies ever since (Jusczyk et al., 1999b, exp. 1, words-first: 30s; exp. 2, passages-first: 45s; Barker and Newman, 2004, words-first: 30s, Bosch et al., 2013, passages-first: 45s). A shorter criterion of 20s is found in Nazzi et al. (2006), which also tested words-first, while Nazzi et al. (2014) uses 30s in both orders of presentation. Even though the infants in this latter study were therefore exposed to fewer tokens in the passages-first condition than their peers in the words-first condition, there was still an apparent benefit to former word order, as we discussed in the previous section.

Studies using artificial languages tend to use much longer familiarization phases. This design choice could be explained by the fact that, in these studies, familiarization is usually not contingent on child behavior: stimuli are played simultaneously from both sides for a given amount of time, regardless of the direction of the child's gaze. Thus, added length is needed to account for the fact that infants are familiarized with two or more words simultaneously and are not always attentive. These values range from 1 minute to 3 minutes (Thiessen et al., 2005; Saffran et al., 1996; Aslin et al., 1998). It is worth noting that these studies are strongly associated with a novelty preference (but see Pelucchi et al., 2009), consistent with findings from studies examining the effect of longer habituation (Thiessen et al., 2005; Roder, Bushnell, and Sasseville, 2000) (see Direction of Preference below for more).

When it comes to the minimum turning angle for a sufficient headturn, most studies agree on 30°, a value that has been in use at least since Fernald (1985). It is generally assumed that this angle is counted from the starting position, which is the infant looking straight into the center light directly in front of him or her. Most studies that report this amount do so by generally saying that a headturn of "at least 30° in the direction of the loudspeaker" was required (e.g., Jusczyk and Aslin, 1995, p. 8). However, Seidl and Johnson (2006) use an alternate description by saying that a headturn was considered sufficient until the infant "looked more than 30 degrees away from the light" (Seidl and Johnson, 2006, p. 569). It is unclear whether these two descriptions are equivalent; that would require the infant to be placed in such a position within the booth that, from his or her position, the center lamp and the side lamps were at a 60° angle from one another, which is unlikely to be the case in every experimental set-up (Kemler Nelson, Jusczyk, Mandel, Myers, Turk, and Gerken, 1995). An alternative possibility is that a 30° angle from starting position is required to signal the start of a sufficient headturn, but a headturn of 30° away from the target side light signals the end of the headturn, creating an intermediate area in which the infant's gaze is considered to be at neither center nor side light. Hollich (2006) requires a stricter 70° headturn, revealing even more variation, if not in set-ups, then in criteria.

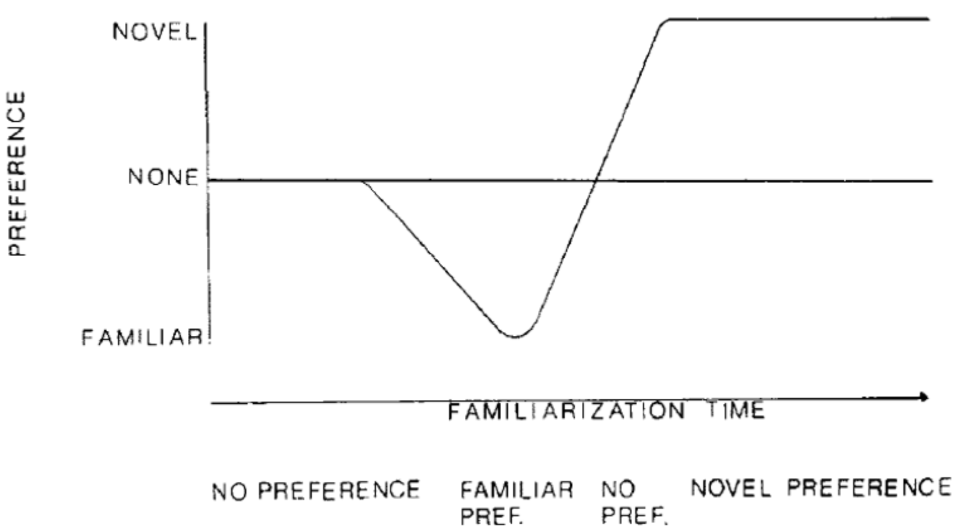
Few, if any, studies with real participants control for differences in minimum looking angle criteria; however, a computational model approach has found that changes in this criterion, which in their model corresponded to the strictness of the experimenter, may cause looking preferences to reverse (Bergmann, ten Bosch, Fikkert, and Boves, 2013).

Direction of Preference

In most studies, the direction of preference, that is, whether infants look longer at the familiarized words (familiarity preference or effect) or at the novel words (novelty preference or dishabituation effect) is not of primary concern. Either preference is taken by authors as evidence of successful segmentation, and the absence of either preference, that is, a non-significant difference in looking time to the different kinds of stimuli, is interpreted as a sign that infants were unable to identify the familiarized words during the test phase. Nevertheless, authors have observed that different experimental settings have led to opposite preferences, and have attempted to explain this effect.

Hunter and Ames' (1988) model argues that an infant who is presented with a stimulus will attempt to match it to representations of previous experience stored in memory. If no match is found, or if it is partial or vague, attention is drawn to it until an accurate neuronal representation of it can be established. Once that process is complete, infants then orient to a novel stimulus. In other words, a familiarity preference is thought to be the result of an early stage of processing, and a novelty preference occurs once processing of the stimulus is nearing completion. In this view, a familiarity preference necessarily precedes a novelty preference. This model has found direct support in studies in the visual and audio-visual domains (Roder et al., 2000; Houston-Price, 2002).

Figure 1 – Hunter and Ames' (1988) model.



The differences in direction of preference observed in the studies presented so far can be explained in terms of difficulty of the task, which would cause participants to go through an early familiarity preference towards a novelty preference at varying speeds.

Recall that studies with characteristics considered "easier" tend to find novelty preferences; that is the case of most artificial language studies (e.g., Saffran et al., 1996; Aslin et al., 1998; Johnson and Tyler, 2010), but not of studies using natural languages (e.g., Jusczyk and Aslin, 1995; Jusczyk et al., 1999b, Nazzi et al., 2014). Moreover, studies with large changes in difficulty across experiments occasionally found a switch in preferences. Such is the case of Bosch et al. (2013), when the same procedure performed by 6- and 8-month-olds yielded a familiarity and novelty preference, respectively. Likewise, Thiessen et al.

(2005) found that a one-month increase in age of participants along with a longer familiarization period (up from 1 minute to 2) across experiments, caused a change in preference from familiar to novel.

Though these findings are all consistent with Hunter and Ames' (1988) model, one element predicted in this model was not supported by Roder et al.'s (2000) investigation: the presence of an intermediate point between familiarity and novelty preferences, where no preference is exhibited. Therefore, more research is needed to establish whether there is, in fact, such a point, or as Roder et al. (2000) have found, the switch from one preference to the other is near instantaneous. The presence of this moment of transition might help explain studies which have found no preference between different types of stimuli. In any case, individual differences in processing speed may lead to an apparent absence of preference if the data from faster-processing (i.e., already showing a novelty preference) and slower-processing infants (i.e., still showing a familiarity preference) are collapsed together (Houston-Price and Nakai, 2004). These considerations underscore the importance of attending to individual differences and changes in preference over the course of the experiment.

The Headturn Preference Procedure and Dutch infants

Despite the prevalence of the HPP in word segmentation studies worldwide, the overwhelming majority of studies have been conducted on American English-learning infants. According to Bergmann and Cristia's (2015) database (consisting of natural language studies only), experiments with American English-learning infants make up 152 of 202 experiments, roughly 75%.

Experiments of this kind with Dutch-learning infants, on the other hand, total three, all of which were reported in Kuijpers et al. (1998), and one of which was reported again in Houston, Jusczyk, Kuijpers, Coolen, and Cutler (2000). Despite electrophysiological evidence that Dutch infants perform word segmentation (Junge et al., 2012; Kooijman, Hagoort, and Cutler, 2005; Kooijman, Hagoort, and Cutler, 2009), the main finding of these few HPP experiments was that there seems to be a developmental delay in Dutch infants relative to their age-matched American English-learning peers. While 7.5-month-old American infants were able to successfully segment bisyllabic trochaic words when presented with words in isolation first and test on words in passages (Jusczyk et al., 1999b, experiment 1), Dutch infants failed at the same age, and succeeded only at 9 months. The authors suggest that this delay may be result of acoustic differences between target words in the English and Dutch materials. Specifically, the ratios of mean F0 and standard deviation of mean F0 (pitch variation) between the strong and weak syllables of target words were higher in English than in Dutch. Since Dutch infants also failed to segment words using the English materials (Kuijpers et al., 1998) but the English-learning infants successfully performed the task using the Dutch materials (Houston et al., 2000), it is likely that Dutch infants are not as sensitive to these values as American infants. Therefore, natural characteristics of infant-direct English and Dutch may be responsible for the delay reported.

Another interesting finding was that much higher dropout rates were reported with Dutch infants compared to American infants, the majority of which due to behavioral problems. While dropout rates are approximately 44, 35 and 52% for the three Dutch experiments, they average 17.6% across the 152 studies with American infants surveyed (consistent with Kemler Nelson et al., 1995). Unfortunately, there is too little data from studies with Dutch infants to speculate on the cause for these increased dropout rates.

Finally, there is one study with Dutch infants using an artificial language paradigm, akin to Saffran et al. (1996). As reported in the previous section, Johnson and Tyler (2012) found that Dutch 5.5-month-olds are just as able as 8-month-olds (and 6-month-old American infants) to segment bisyllabic words when no cues other than transitional probabilities are present. This lends further credence to the possibility that the delay reported by Kuijpers et al. (1998) is a result of characteristics of natural, infant-direct Dutch.

The Present Study

The experiment we designed aimed to answer the following questions:

1. Are 10-month-old Dutch-learning infants able to segment bisyllabic words from fluent speech if familiarization is done on passages?

2. If so, would the apparent added benefits of the passages-first order of presentation and increased age cause a different pattern of looking behavior, namely, a switch to a novelty preference?

To answer the first question, we will compare mean looking times to test trials consisting of lists of words with which participants had been familiarized in the first phase (target words) to words they had not (control words). Thus, our null hypothesis is that the true difference between these values is equal to zero; that is, they would not be able to segment words from passages in the familiarization phase, thereby showing no preference during the test phase. The alternative hypothesis is that the true difference between these values is not equal to zero, but either less than or greater than zero. This listening preference to one type of word during the test phase is taken to indicate that they were able to segment the words from passages during the familiarization phase and then recognize them in the test phase. For that purpose, a two-tailed paired t test will be performed on these two lists, i.e., the mean looking times to trials containing familiar words and the mean looking times to trials containing novel words for each participant during the test phase. We hypothesize that this test will reveal a significant ($p < 0.05$) difference between the two values. This is based on the fact that Dutch 9-month-olds were able to segment words when tested in a words-first order (Kuijpers et al., 1998). Since studies have shown an improvement when using the passages-first order relative to the words-first order (Nazzi et al., 2014; Jusczyk et al., 1999b), the Dutch 10-month-olds in the present experiment should succeed.

While any difference would indicate successful segmentation, and many factors have been shown to influence word segmentation by infants (see Background), we further hypothesize that infants will orient reliably longer to familiar words than novel words. We expect the benefits of the passages-first order of presentation and the increased age of our participants to have a facilitative effect. However, since our passages are longer than those of comparable studies, our average number of tokens heard in 45s is lower (see Table 1), possibly making the task more difficult; thus, we expect this facilitation to be somewhat offset by the lower number of tokens.

Table 1 – Number of tokens and findings across different experimental settings.

Language	Age (months)	Order of presentation	Length of familiarization	Average number of tokens	Result
American English	7.5	words-first	30s	21	familiarity preference[1]
	7.5	passages-first	45s	13	enhanced familiarity preference[2]
French	8	words-first	30s	31*	no preference[3]
	8	passages-first	30s	13*	familiarity preference[4]
Dutch	9	words-first	30s	23	familiarity preference[5]
	10	passages-first	45s	8	--[6]

[1] Jusczyk et al. (1999b), Experiment 1
[2] Jusczyk et al. (1999b), Experiment 2
[3] Nazzi et al. (2006), Experiment 1
[4] Nazzi et al. (2006), Experiment 2
[5] Kuijpers et al. (1998), Experiment 3
[6] Present study

* These values are more precise since this study reports the actual mean listening time for their participants. The others were calculated using the mean recording length and minimum length of familiarization criteria; thus, the actual values may be higher. For a more precise value for the current study, see Discussion.

Method

Participants

8 ten-month old Dutch infants (5 female and 3 male) took part in this study (mean age 298 days, range 286 – 318). All were born full-term and were reported not to have any hearing or language impairment. All came from monolingual Dutch-speaking families living in the Nijmegen area and were recruited via the Baby Research Center Nijmegen. American infants have repeatedly been shown to correctly segment words by using the HPP at 7.5 months of age (e.g., Jusczyk et al., 1999b), and since the only known successful study with Dutch infants using this method for this task was done on 9 month-olds (Kuijpers et al., 1998), they should have no problem completing this task at around 10 months. An additional 15 infants were tested but were excluded from the analysis due to experimental

error (3), crying (5), and failure to maintain attention (7). We intended to test 48 infants, but due to time constraints, this report includes only preliminary data.

Ethical approval for the study was obtained from the Ethiek Commissie Faculteit der Sociale Wetenschappen (ECSW) at Radboud University in Nijmegen.

Materials

Ten words were used in this study, all of which are trochaic pseudo-nouns abiding by Dutch phonotactics. Eight are target words: *basel* [ba:səl], *pasel* [pa:səl], *dasel* [ba:səl], *tasel* [ta:səl], *banno* [bɑnoʊ], *panno* [pɑnoʊ], *danno* [dɑnoʊ], and *tanno* [tɑnoʊ]; and two are control words: *fonie* [foʊni:] and *gemer* [χeɪməʀ]. These words were taken and expanded from a successful word-learning experiment with Dutch infants (Tsuji, Fikkert, Yamane, and Makuza, 2014, in prep.). Pseudo-words were chosen because real words might elicit preferences that are due to familiarity with the word rather than successful segmentation. Trochaic words were selected because it is the predominant rhythmic pattern in Dutch (Kuijpers et al., 1998), and infants acquiring stress-timed languages such as Dutch should more easily segment trochaic words due to a trochaic bias (Höhle et al., 2009). Nouns were selected because it has been shown that the segmentation of nouns happens earlier than segmentation of other word classes, such as verbs (Nazzi et al., 2005). The target words form two sets (*basel*, *pasel*, *dasel* and *tasel* vs. *banno*, *panno*, *danno* and *tanno*); words from the same set differ only in initial segment and are never presented together. The segments that form the words in each set, as well as those of the control words, were carefully selected to contain as little segmental overlap as possible. In other words, words from either set and control words were constructed to be as different as possible, otherwise infants might mistake one for another.

The carrier sentences used in the familiarization phase are simple Dutch sentences, as shown in Table 2.

Stimuli

A single native Dutch female speaker recorded all the audio stimuli that were used in the experiment. An effort was made to ensure that the recordings featured an enthusiastic voice, to grab and maintain the infant's attention. Thus the stimuli were recorded in a particularly exaggerated form of infant-directed speech. These recordings were digitized and stored as computer files (WAV file format, 705kbps bit rate), where they could be manipulated using the sound editing and analyzing software PRAAT (Boersma and Weenink, 2011).

For each target word, two passages consisting of six carrier sentences were recorded (see Table 2). The position of the target word in the carrier sentence and the preceding word are balanced across the two passages. Two sentences in each passage present the target word in the last position, meaning the word boundary and the sentence boundary are aligned, which we expect to provide a facilitative effect (Seidl and Johnson, 2006). The mean duration for these files was 33.16s (range 29.64s – 37.21s). The mean duration for all passages numbered 1 was 32.72s (range 32.09s – 33.27s); for passages numbered 2 it was 33.61s (range 29.64s – 37.21s).

Table 2 – Example of the two six-sentence passages recorded for every word (here, *basel*).

Passage 1	**Passage 2**
Jouw *basel* ligt al boven op zolder.	Jouw *basel* is erg geweldig.
De boer gebruikte vaak die *basel* als hij op bezoek kwam.	Iemand anders zag laatst een mooie oude *basel*.
Eens zag hij een mier bij de oude *basel*.	Mensen zien de *basel* dan niet eens.
De *basel* kreeg al snel meer gaten.	Een andere *basel* is vaak bruiner.
Dus wil hij een andere *basel*.	Mensen met die *basel* gaan graag naar buiten.
Met de nieuwe *basel* is hij blijer.	Vaak worden ze dan gezien als een nieuwe *basel*.

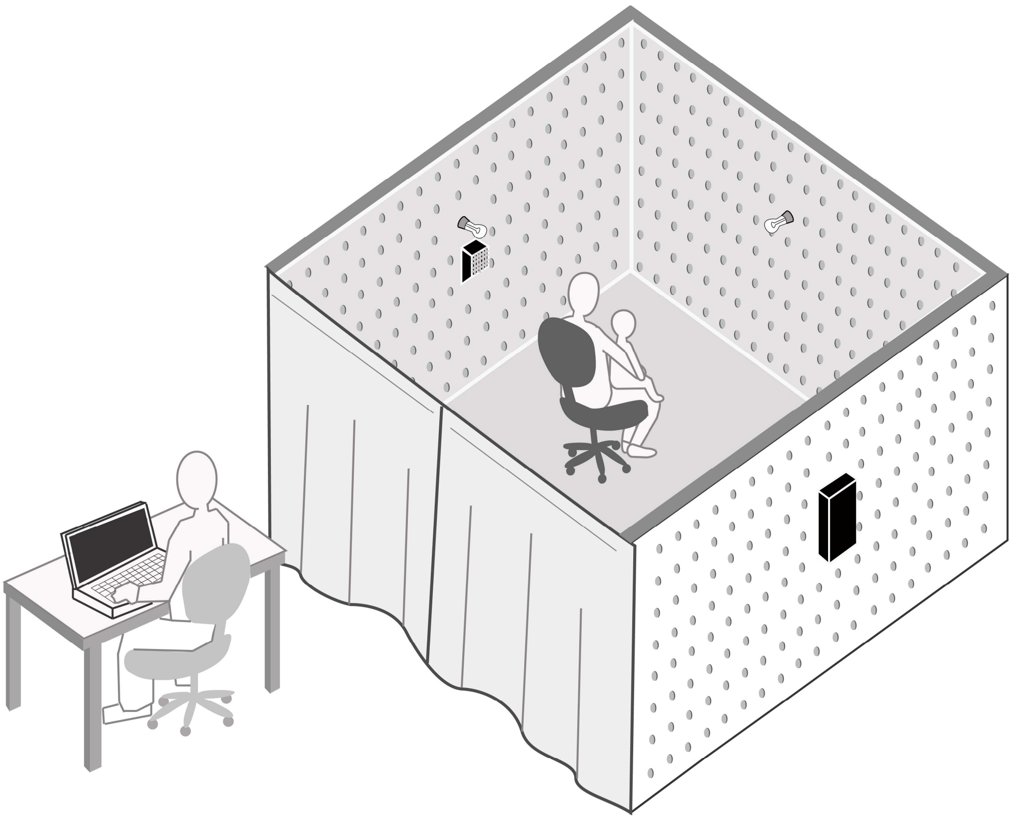
Additionally, for each target and control word, one recording was made containing a list of 16 repetitions of the same word in isolation for use in the test phase. Of these 16 tokens, 8 were unique and each was repeated twice in a random order such that a given token was never followed by its repetition. These tokens were separated by silences of around 1s, and each list began and ended with a 500ms silence. These lists had a mean duration of 35.53s (range 32.61s – 37.52s).

Apparatus

The test booth consists of a three-sided pegboard booth (see Figure 2). The wall directly in front of the child is outfitted with a blue light and the walls on either side are outfitted with red lights. Under the center light there is a 5cm diameter hole through which a camera (Mini CCTV camera MC 900-D12) monitors and records the child's headturns. Under the side lights there are two speakers (Canton LE-101) through which the stimuli are played. The sound files are stored on a computer (HP Compaq dc7700) and played through an amplifier (Sony TA-FE230) configured to output sound between 65 and 70 dBA, as measured by a decibel meter. The video recorded by the camera is captured on another computer (HP Compaq dc7700) after being converted (Canopus ADVC 300 video converter) using Virtual-Dub (freeware, retrieved from http://www.virtualdub.org/).

During the experiment, the fourth, back wall of the booth is closed by a curtain. The lights are turned off and the room is lit by five dim lights directly above the ceiling of the booth, which is made of a layer of extended cloth.

Figure 2 – The Headturn Preference Procedure booth.



Procedure

Upon arrival, the experiment was explained to parents, who provided informed consent. In addition, parents were asked to fill a questionnaire (N-CDI, Woorden en Gebaren, Zink and Lejaegere, 2002) regarding their child's comprehension and production of spoken language and gestures. The procedure used was an adaptation of the Headturn Preference Procedure (HPP) for word segmentation tasks (Hirsh-Pasek et al., 1987; Jusczyk and Aslin, 1995). In this procedure, the child sits on a parent or caretaker's lap on a chair in the center of the test booth. The parent or caretaker is instructed not to talk, point or look to the child or to the side lights, but to keep his or her gaze directed to the front for the entire duration of the experiment. Additionally, the parent or caretaker wears a pair of tight-fitting headphones (Sennheiser HMEC 300) playing masking music so that they are blind to the stimuli being played and cannot interfere with the experiment.

The experimenter stands behind the curtain wearing a pair of headphones equal to that used by the parent or caretaker playing the same masking music, so that he or she is blind to the stimuli being played as well. The experimenter monitors a live feed of the child on a television set (Sony Trinitron) and presses buttons on the computer keyboard corresponding to the direction in which the infant is looking. This process, called 'on-line coding,' serves to control that stimuli are only played when the child's gaze is directed to the blinking light, which is taken to represent the child's attention. This monitoring also allows the experimenter to interrupt the experiment in some cases of equipment failure (e.g. faulty lights), or if the child is inattentive or crying. The button presses are recorded by the software LOOK (Meints and Woodford, 2008), which controls the execution of stimuli and appropriate illumination of the different lamps, and generates log files containing data on participants' headturns.

All trials begin with the center light flashing to grab the infant's attention. Once the infant orients to that light for one uninterrupted second, it extinguishes, and one of the two side lights begins to flash. When the infant turns his or her head at least 30° to the side of the flashing lamp, and the experimenter makes the corresponding button press, the stimulus starts playing from the speaker under the flashing light. If a child turns his or her head away from the flashing light for more than 2s, or if the stimulus is played to completion, the trial ends and the center light begins flashing again in preparation for the next trial. The sides on which the stimuli are played are pseudo-random and counterbalanced. If the child looks away from the flashing light for less than 2s before returning, the stimulus continues playing, but these non-looking times are not included in the analyses.

The experiment begins with a familiarization phase, during which a child is familiarized with two target words presented in carrier sentences (see Stimuli, below). The familiarization phase lasts for a maximum of 18 trials, or as soon as the child has accumulated 45s of looking time to each word, which required minimally 4 trials. If this criterion is reached for one word but not the other, only the trials with passages containing the non-familiarized word continue playing until this criterion is reached.

The test phase starts immediately after the end of the familiarization phase. In this phase, the child hears lists of repetitions of both target words heard in the familiarization phase, as well as lists of two new words. The test phase is organized in three blocks of four trials, two for target words and two for control words, for a total of 12 trials. The same list is never used twice consecutively, and the ordering of trials within each block is pseudo-random and counterbalanced across different versions. The test phase ends only when the infant reaches the end of the 12 trials. The total duration of the experiment depended on the child's behavior but usually fell between 5 and 7 minutes.

Design

The eight target words were bundled in four pairs, as follows: *basel* and *panno*, *pasel* and *banno*, *tasel* and *danno*, and *dasel* and *tanno*. Words belonging to the same pair start with segments sharing place of articulation but differing in voicing: *basel*-*panno* and *pasel*-*banno* form pairs starting with bilabial segments (*basel* and *banno* being voiced while *pasel* and *panno* are voiceless); and *tasel*-*danno* and *dasel*-*tanno* form pairs starting with alveolar segments (*dasel* and *danno* being voiced while *tasel* and *tanno* are voiceless).

For each of these pairs, four versions of the experiment were prepared, depending on the passages used during the familiarization phase. For the group *basel*-*panno*, for example, version 1 used passage number 1 for *basel* and 2 for *panno*, and started with *basel*. Version 2 used the same passages as version 1 (*basel* 1, *panno* 2), but started with *panno*. Version 3 used passage number 2 for *basel* and number 1 for *panno*, starting with *basel*. Version 4 used the same passages as version 3 (*basel* 2, *panno* 1), but started with *panno*. (See Table 3 for an illustration of this design.) This alternated design ensured that the two passages to which a child is exposed do not feature the same carrier sentences, which could bore the child, and it ensured counterbalancing of starting words.

Table 3 – First two of eighteen trials of all versions for the same word pair (here, *basel-panno*).

Version 1	Version 2	Version 3	Version 4
Basel 1	Panno 2	Basel 2	Panno 1
Panno 2	Basel 1	Panno 1	Basel 2

Each version also differs in the ordering of the lists of words in isolation played during the test phase. The lists themselves, however, are the same across all versions for a given word pair, and they all contain the same number of trials (12). This differentiation was done for counterbalancing purposes.

Data pre-processing

In order to ensure that the analysis includes only infants who were able to complete the experiment as intended, a number of exclusion criteria were used. The first and most obvious of these criteria was failure to maintain attention. If a child became increasingly restless during the experiment, moved around excessively or started crying, the experiment was interrupted before its completion and the data for that participant was discarded (6). The second criterion was failure to reach familiarization criteria. Because of the way the software that runs the experiment functions, it is possible for a child to reach the test phase without actually having been familiarized, simply by going through all 18 trials in the familiarization phase. In this case, the data were excluded after an analysis of looking times during this phase revealed that the criterion of 45s of accumulated looking time for each word had not been reached (1).

Off-line and reliability coding

After the experiments were conducted, all the videos (360x256 pixels, AVI file format, 25 frames per second) of the successful participants captured by the video camera in the booth were coded in ELAN (Wittenburg, Brugman, Russel, Klassmann, and Sloetjes, 2006) by the author of the present study. A randomly selected 25% was also coded by a second coder and the correlation between the values by both coders was calculated. Inter-coder agreement was over 99.9%. Coders were blind to the stimulus being played at any given point in the video, and marked, on a frame-by-frame basis, the direction in which the participant was looking, as well as the lengths of each gaze. This coding provided the precise data for looking time and direction used in the analyses (see Results).

Additionally, looking times are also recorded automatically by LOOK during the execution of the experiment. Though these values are less precise and therefore are not used in the analyses, they were calculated for each infant and correlated to the values generated by offline coding, in order to ensure that the latter process corresponded to the reality of the experiment. Agreement between the two sets of values for each participant ranged between 99.3% and 99.9%.

Results

Our sample consisted of eight participants, hereafter numbered 1 to 8. The numbers do not correspond to the order in which they were tested, but to the versions with which they were tested: participants 1-4 were tested with *basel-panno* versions 1-4 and 5-8 were tested with *pasel-banno* versions 1-4 (see Method for an explanation of different versions). No successful participants used any versions from other word pairs. For each one, total and mean listening times were calculated for trials containing each word in the familiarization (tables 4 and 5) and test phases (table 6 and 7). The mean looking time until familiarization was reached was 105.11s and the mean number of trials was 8. Figure 3 shows the mean looking times per trial for each infant.

Table 4 – Sum and mean of looking times in seconds to passages containing words in the familiarization phase for infants tested on versions using *basel* and *panno*.

Participant\Word	Basel		Panno		Total	
	Sum	Mean	Sum	Mean	Sum	Mean
1	64.43	21.48	47.79	15.93	112.22	18.7
2	47.48	23.74	48.05	9.61	95.52	13.65
3	60.7	30.35	46.73	9.35	107.43	15.35
4	46.88	5.86	50.76	6.35	97.64	6.1

Table 5 – Sum and mean of looking times in seconds to passages containing words in the familiarization phase for infants tested on versions using *pasel* and *banno*.

Participant\Word	Pasel		Banno		Total	
	Sum	Mean	Sum	Mean	Sum	Mean
5	52.44	13.11	51.49	17.16	103.93	14.85
6	48.54	12.14	51.38	25.69	99.92	16.65
7	69.81	23.27	57.02	11.4	126.83	15.85
8	48.77	16.26	48.66	12.16	97.43	13.92

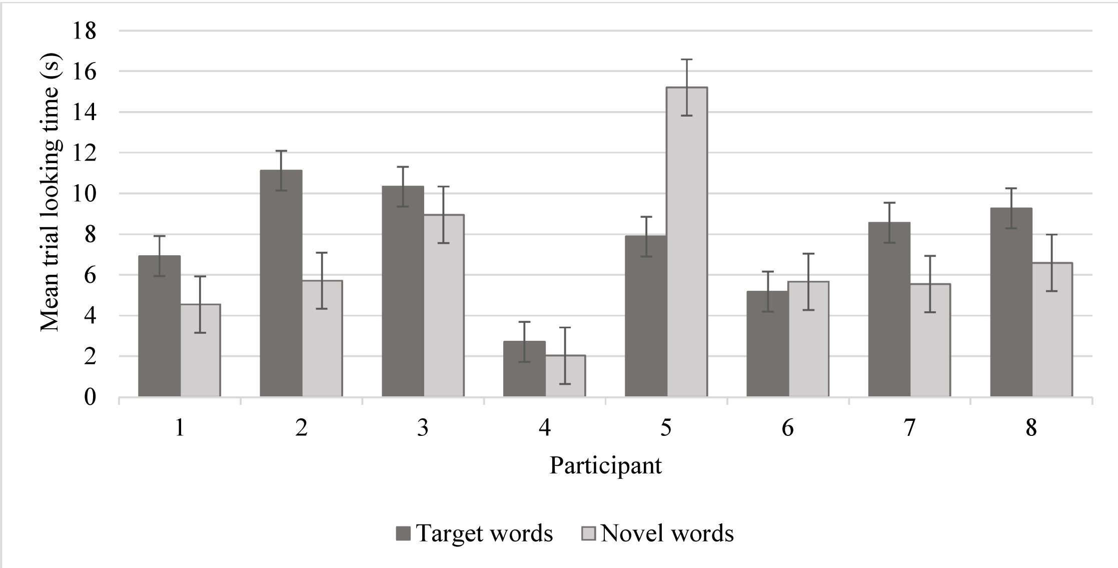
Table 6 – Sum and mean of looking times in seconds to lists of words in the test phase for infants tested on versions using *basel* and *panno*.

Participant\Word	Target						Control					
	Basel		Panno		Total		Fonie		Gemer		Total	
	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean
1	12.56	4.19	28.95	9.65	41.51	6.92	14.18	4.73	13.05	4.35	27.23	4.54
2	22.68	7.56	44.05	14.68	66.73	11.12	16.77	5.59	17.46	5.82	34.23	5.71
3	27.08	9.03	34.87	11.62	61.95	10.33	23.17	7.72	30.51	10.17	53.67	8.95
4	8.73	2.91	7.56	2.52	16.28	2.71	5.24	1.75	6.95	2.32	12.19	2.03

Table 7 – Sum and mean of looking times in seconds to lists of words in the test phase for infants tested on versions using *pasel* and *banno*.

Participant\Word	Target						Control					
	Pasel		Banno		Total		Fonie		Gemer		Total	
	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean	Sum	Mean
5	22.38	7.46	24.9	8.3	47.28	7.88	43.23	14.41	48.04	16.01	91.27	15.21
6	22.07	7.36	9.01	3.01	31.08	5.18	14.02	4.67	19.92	6.64	33.93	5.66
7	35.83	11.94	15.55	5.18	51.38	8.56	11.85	3.95	21.47	7.16	33.31	5.55
8	18.92	6.31	36.71	12.24	55.63	9.27	22.93	7.64	16.63	5.54	39.56	6.59

Figure 3 – Mean looking times in seconds per trial for each participant by type of trial (target vs. control) during the test phase.



Six of the 8 infants showed a familiarity preference. Of the two that did not, one subject (number 5) showed a remarkable difference in looking behavior relative to the others. When we generated a density plot to verify that the *t* test would be applicable to our sample, this single infant caused a clear deviation from normality. Figures 4 and 5 show the distributions with and without this participant.

Figure 4 – Density plot with all participants.

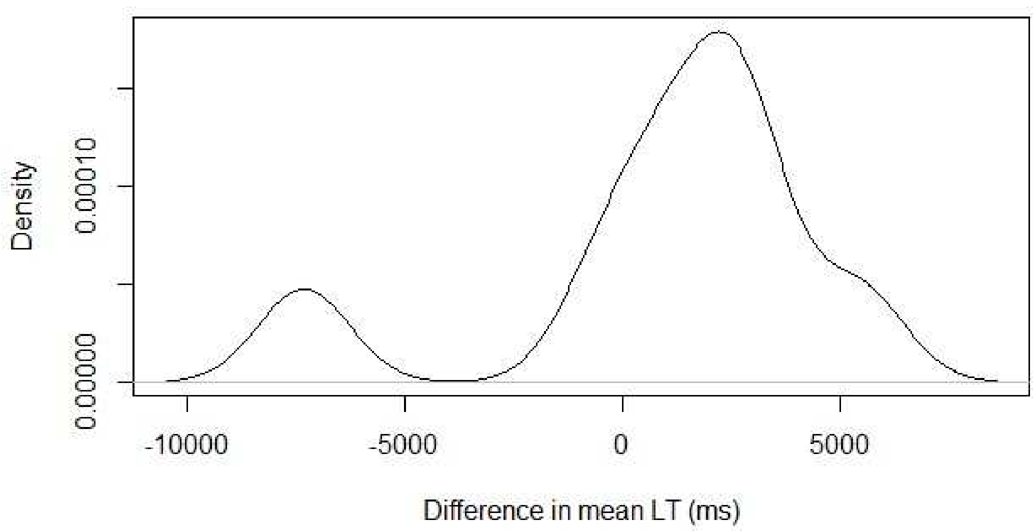
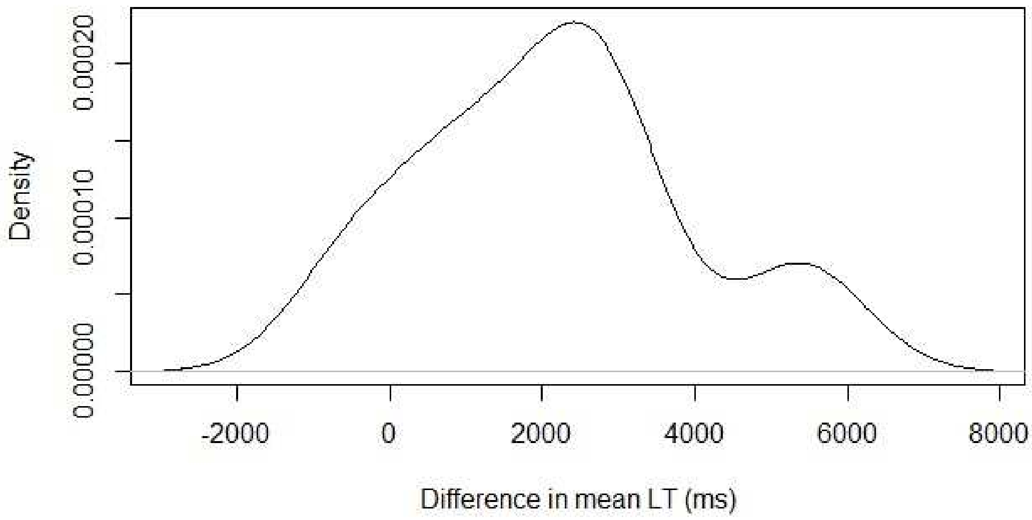


Figure 5 – Density plot with outlier removed.



Based on this finding, we opted to remove this participant on the basis that she constituted an outlier and was not representative of the population. Subsequent statistical analyses supported this decision. Across all subjects, the mean listening time to trials containing target words was 7.75s (SD = 2.59s) and control words was 6.78s (SD = 3.66s); with this subject removed, they were 7.73s (SD = 2.77s) and 5.57s (SD = 1.92s), respectively. Although the second curve is still not a normal distribution, the *t* test is not as sensitive to deviations from normality as others statistical tests (Bartlett, 1935; Wiedermann and von Eye, 2013), thus we considered it applicable to our sample. A paired *t* test including all infants' average looking times showed that the difference was not significant ($t(7) = 0.7$, $p = 0.5$). The same test without her revealed that the difference was significant ($t(6) = 3$, $p = 0.02$). This confirms our hypothesis that the infants in the experiment would be successful in segmenting words from the passages in which they appeared, and show a familiarity preference. The outlier was likely also able to segment the word, but she showed a novelty preference.

Discussion

Individual differences and direction of preference

Different studies have warned against the possibility of individual differences between infants severely affecting results (Houston-Price and Nakai, 2004; Bergmann et al., 2013), and this was precisely our case. Participant 5 showed a very strong novelty preference, contrary to that of most other participants. Recall that Hunter and Ames' (1988) model predicts that a novelty preference will be displayed after a familiarity effect has occurred and the infant has formed an accurate mental representation of the stimulus. It is possible, then, that participant 5 was simply faster in forming the representation than the average participant, and therefore was able to reach a novelty preference after the same amount of familiarization. In support of this interpretation is the fact that the total looking time she accumulated during familiarization was 103.93s over 7 trials, for a mean of 14.85s per trial. For comparison, the average total of the other participants was 105.28s over an average of 8.14 trials. Thus, her familiarization was comparable to that of the other participants and cannot account for the difference in behavior displayed in the test phase.

One possible explanation can be found in the careful examination of her looking behavior during the test phase. As explained in the Method section, a trial ends when it is played to completion or when the infant looks away for longer than 2 uninterrupted seconds. During the third trial of the first test block, a list containing repetitions of the word *pasel* played, but she looked away from the light almost immediately after looking at it just long enough to trigger the start of the trial. Time spent looking away, even if the sound is playing, is not included in our analyses, which caused her to register a mere 65ms of looking time during that trial. Occurrences like this are somewhat rare (we registered only 3 other looking times under 1 second), but can be easily explained by normal, temporary distraction on the part of the infant, without signifying that the infant was too distracted to perform the experiment. However, since the word in question was a target word, this had the effect of lowering her mean looking time to target words. Thus, new values were calculated for this infant excluding this trial to gauge the size of this impact. The total looking time to target words for this participant discarding this one trial was 47.22 (65ms less than the reported sum of 47.28s in Table 5), and the new mean is 9.44s (1.56s longer than the reported mean of 7.88 in Table 5). The original values revealed a 7.33s longer average look to novel trials than familiar trials, while the new values show a 5.77s difference. These news values decrease the size of her novelty preference, but still place over four standard deviations from the group's average of 2.16s longer looking times to familiar trials. Thus, we consider that this abnormal trial does not account for her difference in looking behavior. Likewise, her age cannot account for it, as she was 291 days old on the day of testing, compared to an average of 298.2 days for the other participants. Therefore, we maintain that her individual abilities allowed her to reach a novelty preference sooner than the other infants.

Two other participants show somewhat deviant patterns. Participants 4 and 6 showed only small preferences to familiar and novel words, respectively. Participant 4's means during familiarization are lower than average as he required many more trials than the others to accumulate 45s of looking time to each word due to having short looking times to individual trials. Similarly, his looking times during the test phase are also shorter. However, he reached familiarization criteria and successfully completed the experiment, eventually showing a slight familiarity preference. Participant 6 performed the task as well as the others, but showed only a slight preference for novel words overall. It is possible to argue that these two infants were unable to perform the task, and that their small differences are simply due to chance. While we cannot rule out this possibility, an alternative explanation was that they too familiarized faster than the others and were close to – but on opposite sides of – the intermediate point between familiarity and novelty preferences as proposed by Hunter and Ames (1988) (but see Roder et al., 2000; Background, above). In any case, their values did not cause such extreme deviances in our data, so we opted not to remove them. Had we done so, an even more significant difference would have emerged, but it would likely be unrepresentative of the population, as we would only be selecting those participants whose behavior conforms to our hypothesis.

Differences in direction of preference have long been observed, but studies attempting to shed light on this particular question are few, especially in the auditory modality. As such, the enhanced familiarity effect Jusczyk et al. (1999b) found for the passages-first order, and the familiarity effect Nazzi et al. (2014) reported for the same order when none was found in the opposite order, cannot doubtlessly be attributed to facilitation. In fact, as Table 1 shows, the passages-first order generally familiarizes infants with fewer tokens than the words-first order. The reduced number of tokens from 31 in Experiment 1 to 13 in Experiment 2 in Nazzi et al. (2014) could, contrary to the authors' argument, make the task more difficult (Lew-Williams, Pelucchi, and Saffran, 2011), and the difference between experiments is possibly the participants moving "back" up the curve (see Figure 1), rather than moving forward on it towards a novelty preference.

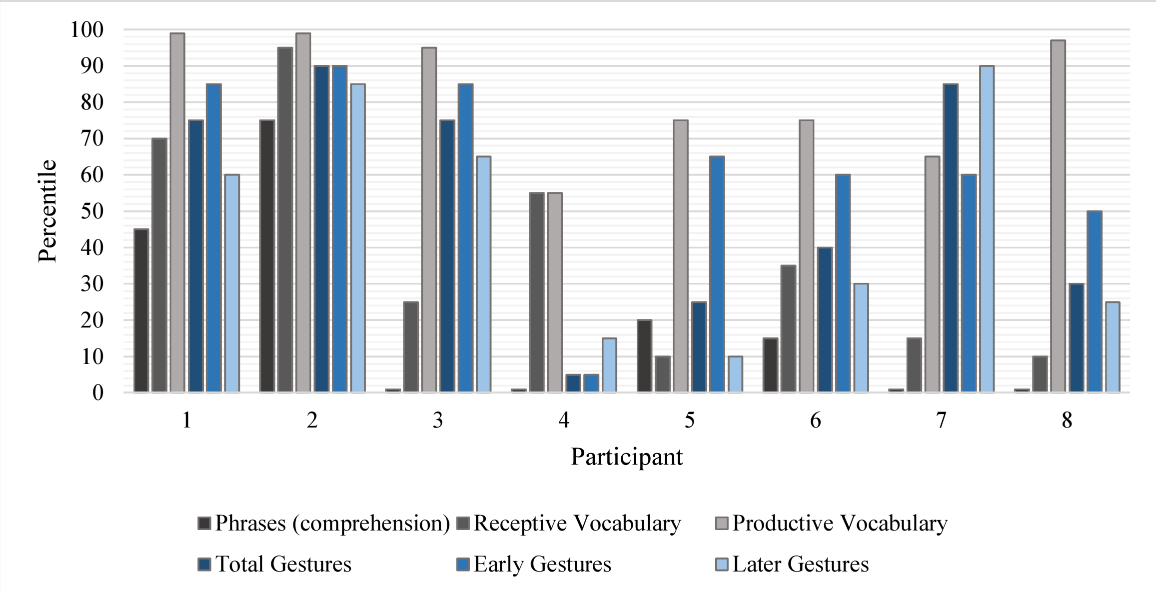
We find this explanation unsatisfactory for the following reasons. First, Nazzi's argument concerning the increased number of tasks during the test phase when using the words-first order (segmenting and comparing) when compared to the passages-first order (simply comparing, as segmentation happened earlier), is compelling. Furthermore, Bosch et al. (2013) (and to some extent, Thiessen et al., 2005) showed that increases in age make the task "easier", i.e., more likely to reveal a novelty preference. With this finding in hand, the fact that Nazzi et al. (2014) found that 8- and 12-month-olds equally showed no preference makes it unlikely that the 8-month-olds were at the intermediate point between the familiarity and novelty preferences together with the 12-month-olds, as the latter group should be ahead of the former if both can complete the task. Thus, in Experiment 2, 8-month-olds were likely further down the curve rather than earlier on it than the 8-month-olds in Experiment 1. As the second experiment does not include other age groups, it is not possible to make assertions beyond what these logical implications allow us. Likewise, the enhanced familiarity effect found in Jusczyk et al.'s (1999b) Experiment 2 relative to Experiment 1 is also likely due to the task being easier rather than more difficult.

Thus, in spite of the lower number of tokens – around 9 in our case –, the passages-first order provides enough of a facilitation effect that allows infants to overcome this reduced exposure. For this reason, we also expected our results to be stronger than that reported in Kuijpers et al. (1998, Experiment 3). Though direct comparisons are difficult due to differences in sample size and age, we indeed found a larger mean difference than the one they reported: 2.16s against 0.69s.

The findings of the present study also support the view that Dutch infants develop their word segmentation skills with a couple of months' delay when compared to infants acquiring other languages. As hypothesized, our participants on average did not reach a novelty preference even with the added benefits of presentation order and increased age, whereas Spanish- and Catalan-learning infants showed a novelty preference at 8 months of age with a words-first order (Bosch et al., 2013). It has been suggested that cultural differences in the characteristics of child-directed speech across languages may affect the onset of word segmentation abilities in acquirers of different French dialects (Nazzi et al., 2006), and the same could hold for Dutch.

N-CDI questionnaires

Figure 6 – Percentiles of scores for each participant given their age and gender.



The questionnaire scores do not appear to correlate with the performance on the task the infants performed. One problem is that the percentiles for each score are based on data from the Flemish population, which may be different from the Dutch population. More importantly, most parents reported having difficulty filling out the questionnaire since 10-month-olds generally have not yet begun producing words, which limits the capacity of parents and caretakers to gauge infants' understanding of language directed at them. Thus, many parents or caretakers filled out the questionnaire only partially or skipped certain sections altogether, which explains the many low scores in the "Phrases (comprehension)" category. On the other hand, some parents or caretakers may have exaggerated their child's abilities, indicating that they are able to speak a large number of words, when at this age this is unlikely. This explains why so many scores in the "Productive Vocabulary" category are at near ceiling: even low absolute scores are high for their age. Two interesting cases emerge from the data, however. One is of participant number 2, who obtained the highest scores in nearly every category and also showed the largest familiarity preference. The other is of participant number 4, who has low overall scores and also showed a small preference. While it is possible that, for these infants, the scores actually reflect their abilities, we cannot be certain that this is the case.

Version differences

We also investigated the possibility that the different versions used were somehow different in ways that affected our results. Recall that versions 1 and 2 used the same passages in reversed order, and the same was true for the versions 3 and 4 of a given word pair. The figures below show the total looking times in familiarization and test trials for the two word pairs. There appears to be no correlation between behavior during the familiarization phase and behavior during the test phase. Also, there appears to be no correlation between the behaviors of infants exposed to the same files during familiarization. The lack of correlations indicates that the files were comparable and no differences in behavior are likely to be due to characteristics of the stimuli.

Figure 7 – Total looking times to trials in familiarization and test phases per word grouped by infant/version – *basel-panno*.

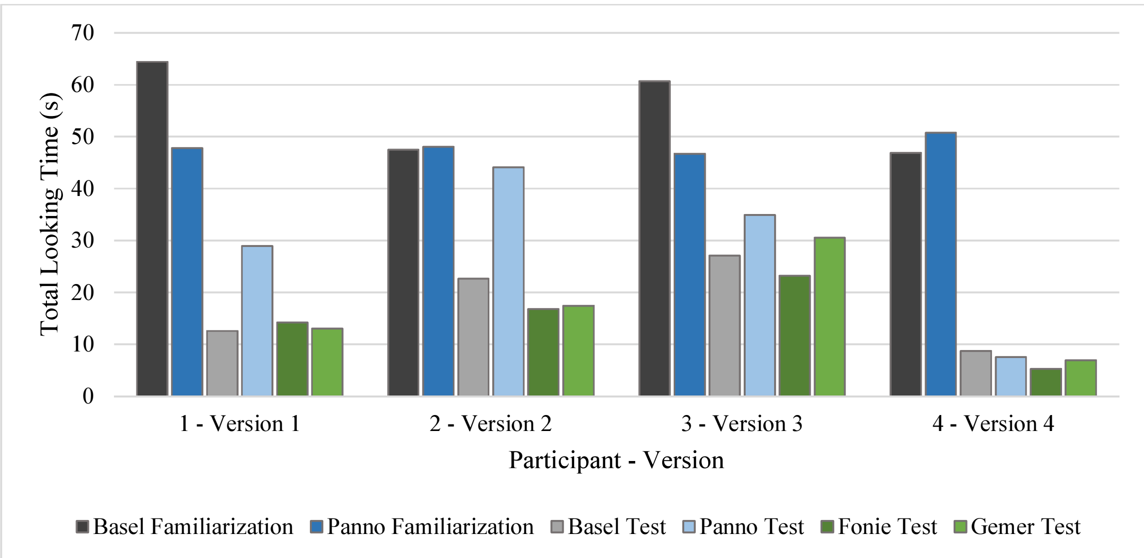
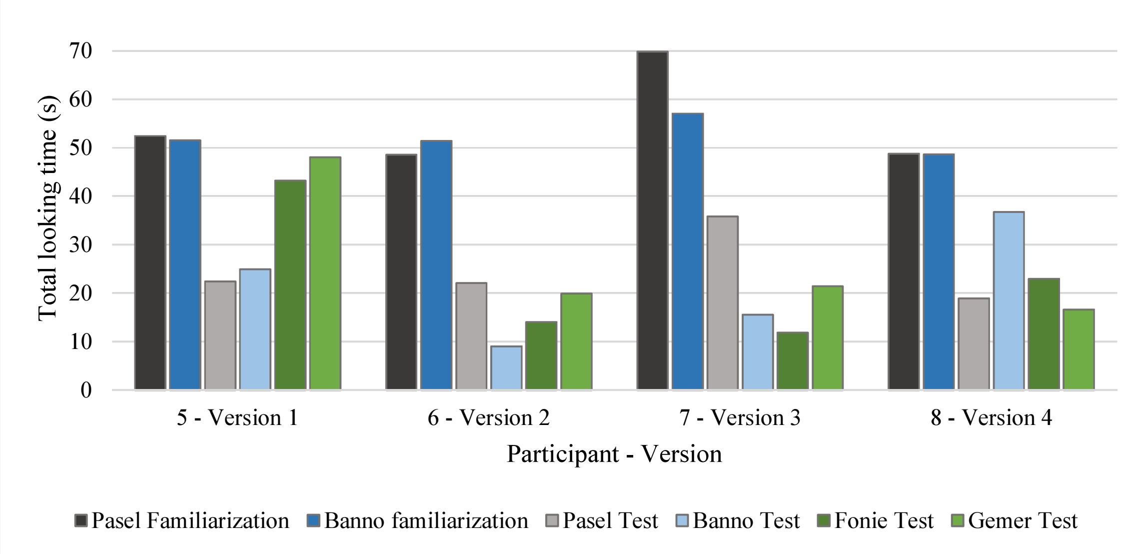


Figure 8 – Total looking times to trials in familiarization and test phases per word grouped by infant/version – *pasel-banno*.



The fact that no other versions were successful (those using *tasel-danno* and *dasel-tanno*) was partly due to coincidence and partly by design. At first all versions were being tested, but as the testing was nearing its end, we noticed a pattern had emerged and sought to complete the set of four versions for the pairs which had been most successful before moving on to other pairs. It is unlikely that participants were somehow more distracted with the other versions, as some participants were not excluded due to inattention, but experiment failure, coding error and crying before the experiment began. Moreover, pitch and duration measurements indicate that all versions were comparable.

Limitations

The most important limitation of the present study is the low number of participants. As reported in the Method section, this was due to a large number of dropouts, with 15 additional infants being tested but discarded. The reasons for such a high dropout rate are not clear, though other studies have reported similar rates with Dutch infants (Kuijpers et al., 1998; but see Johnson and Tyler, 2010 for normal dropout rates with Dutch participants). In the case of the present study, we hypothesized that having multiple people engage the infant before the experiment might have led him or her to become more distracted during the experiment. With our last participants, an attempt was made to limit the number of people involved and the duration of these pre-experiment interactions, and the dropout rate due to inattentiveness improved significantly. Still, unfortunate events such as one infant who fell and hit his head and online coding mistakes by the experimenter still caused the need for some participants to be discarded.

The small number of participants severely hinders our capacity to make wider and more meaningful inferences about our data, as fewer correlations are likely with such a small sample. Nevertheless, the finding of a significant difference in looking times suggests that the experiment was successful, and a larger sample is likely to support and extend the current findings.

Another limitation and difference between the current study and others of its kind is that the target words are never used as novel words. This keeps us from confidently saying that differences in behavior are not due to segmental characteristics of the novel words. Our novel words were designed to be maximally different from our target words, which is why its starting segments are different in manner of articulation (/f/ and /χ/ are fricatives while /p/, /b/, /t/ and /d/ are all stops). Studies have suggested that the nature of the starting segment may aid or hinder speech segmentation (Mattys and Jusczyk, 2001; Nazzi et al., 2005), but hardly any studies look directly into this question. Thus, a familiarity preference could be merely the result of a preference for stops, and not of successful familiarization. That said, since we have not observed an overwhelming preference to either kind of word across all infants, it is unlikely that segmental differences have uniformly impacted the segmentation abilities of the infants tested.

Conclusion

The present study aimed to answer the following two questions as established in the Introduction:

1. Are 10-month-old Dutch-learning infants able to segment bisyllabic words from fluent speech if familiarization is done on passages?

2. If so, would the apparent added benefits of the passages-first order of presentation and increased age cause a different pattern of looking behavior, namely, a switch to a novelty preference?

As we hypothesized based on previous literature, our participants were successful in segmenting words from passages in the familiarization phase and matching them to tokens of words in isolation in the test phase. Moreover, probably due to a small number of tokens and a delay in the emergence of segmentation abilities in Dutch infants, the benefits of the passages-first order of presentation and the increased age of our participants, though visible, were not enough to cause most of our participants to display a novelty preference.

Though statistical power was severely impaired by the small number of participants, the study still managed to yield results that both answers these questions and raises interesting questions. Importantly, it showed the importance of looking at the behavior of individual participants and how these relate to the expected patterns of looking preference. Because of our low number of participants, one displaying a very strong novelty preference greatly affected our results. Thus, it is possible that results that fail to look at individual differences across participants, or fail to check the density distribution of their differences, may be grouping together results from infants who perform the task very differently. Furthermore, our experience suggests that having more than one person engage the infant prior to the experiment, especially if play time in a separate room is involved, may lead to decreased attention during the experiment, and therefore a higher dropout rate than expected.

The current study touches upon the question of direction of preference and the factors that influence it, but direct investigation of this topic was not its primary objective. Research that directly tests the benefits of different orders of presentation with equal number of tokens is still lacking, as are studies that directly test the assumptions of Hunter and Ames'

(1988) model with auditory stimuli. These data are crucial to fine-tune the design choices and therefore accurately interpret the results of future experiments.

References

Allen, G. D., & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell, & J. B. Hooper (Eds.), *Syllables and segments.* Amsterdam, North-Holland.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probabilities Statistics by 8-month-old Infants. *Psychological Science, 9*(4), 321-324.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker famiilarity in infant streaming. *Cognition, 94*, B45-B53.

Bartlett, M. S. (1935). The effect of non-normality on the t-distribution. *Proceedings of the Cambridge Philosophical Society*, *31*, pp. 223-231.

Bergmann, C., & Cristia, A. (2015). Development of infants' segmentation of words from native speech: a meta-analytic approach. *Developmental Science*, 1-17. doi:10.1111/12341

Bergmann, C., ten Bosch, L., Fikkert, P., & Boves, L. (2013). A computation model to investigate assumptions in the headturn preference procedure. *Frontiers in Psychology, 4*.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. (Version 5.4.18). Retrieved September 7, 2015, from http://www.praat.org/

Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology, 4*.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92*, 81-104.

Fernald, A. (1985). Four-Month-Old Infants Prefer to Listen to Motherese. *Infant Behavior and Development, 8*, 181-195.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269-286.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science, 6*(2), 122-127.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior & Development, 32*, 262-274.

Hollich, G. (2006). Combining Techniques to Reveal Emergent Effects in Infants' Segmentation, Word Learning, and Grammar. *Language and Speech, 49*(1), 3-19.

Houston, D. M., & Jusczyk, P. W. (2000). The Role of Talker-Specific Information in Word Segmentation by Infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1570-1582.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review, 7*(3), 504-509.

Houston-Price, C. (2002). The acquisition of object names in explicit and ambiguous referential contexts. D.Phil. Thesis, University of Oxford, UK.

Houston-Price, C., & Nakai, S. (2004). Distinguishing Novelty and Familiarity Effects in Infant Preference Procedures. *Infant and Child Development, 13*, 341-348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research.*

Johnson, E. K., & Tyler, M. D. (2010). Testing the Limits of Statistical Learning for Word Segmentation. *Developmental Science, 13*(2), 339-345.

Junge, C., Cutler, A., & Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neurophysiologia, 50*, 3702-3712.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology, 29*, 1-23.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' Preference for the Predominant Stress Pattern of English Words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999a). Infants' sensitivity to allophonic cues for word segmentation. *Preception & Psychophysics, 61*, 1465-1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999b). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology, 39*, 159-207.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The Head-Turn Preference Procedure for Testing Auditory Perception. *Infant Behavior and Development, 18*, 111-116.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research, 24*, 109-116.

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy, 6*, 591-612.

Kuhl, P. K. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature Reviews: Neuroscience, 5*, 831-843.

Kuijpers, C., Coolen, R., Houston, D., & Cutler, A. (1998). Using the Head-Turning Technique to Explore Cross-Linguistic Performance Differences. *Advances in Infancy Research, 12*, 205-220.

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

Mattys, S. L., & Jusczyk, P. W. (2001). Do Infants Segment Words or Recurring Contiguous Patterns? *Journal of Experimental Psychology, 27*, 644-655.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology, 38*, 465-494.

Mersad, K., & Nazzi, T. (2012). When Mommy Comes to the Rescue of Statistics: Infants Combine Top-Down and Bottom-Up Cues to Segment Speech. *Language Learning and Development, 8*, 303-315.

Nazzi, T., Dilley, L. C., Jusczyk, A., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning Infants' Segmentation of Verbs from Fluent Speech. *Language and Speech, 48*, 279-298.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic evidence. *Journal of Memory and Language, 54*, 283-299.

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language, 41*(3), 600-633.

Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics, 58*(8), 1145-1156.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development, 80*(3).

Roder, B. J., Bushnell, E. W., & Sasseville, A. (2000). Infants' Preferences for Familiarity and Novelty During the Course of Visual Processing. *Infancy, 1*(4), 491-507.

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

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition, 69*, 105-134.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignments facilitates target extraction. *Developmental Science, 9*(6), 565-573.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7*(1), 53-71.

Tsuji, S., Fikkert, P., Yamane, N., & Mazuka, R. (2014). Specifying the underspecified: Young children's perceptual asymmetries are both language-specific and universal. Manuscript in preparation.

van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional abilities to natural language acquisition. *Journal of Memory and Language, 63*, 197-209.

van Heugten, M., & Johnson, E. K. (2012). Infants Exposed to Fluent Natural Speech Succeed at Cross-Gender Word Recognition. *Journal of Speech, Language, and Hearing Research, 55*, 554-560.

Wiedermann, W., & von Eye, A. (2013). Robustness and power of the parametric t test and the nonparametric Wilcoxon test under non-independence of observations. *Psychological Test and Assessment Modeling, 55*(1), 39-61.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.* Retrieved from http://tla.mpi.nl/tools/tla-tools/elan/

Zink, I., & Lejaegere, M. (2002). *Lijsten voor Communicatieve Ontwikkeling. Aanpassing en hernormering van de MacArthur CDIs van Fenson et al.* Leuven (Belgium)/Leusden (Netherlands): Acco.