5-2021

Accessing the Gray Area Between Phonetics and Phonology: the Development of Vowel Length as a Subphonemic Cue

Abby Fergus

Recommended Citation

Fergus, Abby, "Accessing the Gray Area Between Phonetics and Phonology: the Development of Vowel Length as a Subphonemic Cue" (2021). *Undergraduate Honors Theses.* William & Mary. Paper 1719.
https://scholarworks.wm.edu/honorstheses/1719

Accessing the Gray Area Between Phonetics and Phonology: the Development of Vowel Length as a Subphonemic Cue

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Arts in Linguistics from
William & Mary

by

Abby Fergus

Accepted for Honors
(Honors)

Caitlyn R
Dr. Kate Harrigan, Director

Anya Hogoboom
Dr. Anya Hogoboom

Dan Parker
Dr. Dan Parker

Dr. Brian Rabinovitz

Williamsburg, VA
May 14, 2021

Table of Contents

Abstract

Previous research has shown that speakers of English use vowel length as a subphonemic cue to the voicing of a following obstruent. Countless studies have demonstrated adults' ability to make a voicing judgement based upon vowel length but studies with children have provided mixed and sometimes conflicting results. In the present study, we sought to first determine whether adults would exhibit varying sensitivity to vowel length based upon whether it is found in a position where it is predictive of the phonemic status of another sound (i.e. serving as a subphonemic cue). Second, we removed top-down information in order to isolate the acoustic system and better understand children's sensitivity to subphonemic vowel length from 4 to 6 years of age. Both adults and children completed a sound discrimination task where they listened to pairs of nonce words that differed only in the length of their vowel and were asked to indicate their level of similarity. In the adult study (N=63), participants revealed higher levels of sensitivity to vowel length preceding an obstruent showing that a subphonemic position boosts perceptibility for vowel length. Results from the child studies (N=73, MEAN AGE=5;5.6) demonstrate that children from 4 to 6 years of age treat subphonemic vowel length quite differently from adults. First, children fail to show sensitivity at the same level as adults. Specifically, 5- and 6-year-olds require vowel length differences that are twice as large to show sensitivity and 4-year-olds do not show sensitivity even at the larger lengths. Second, children do not reveal varying sensitivity based upon whether the vowel is in a location where it could be used subphonemically. Together, this study reveals that children from 4 to 6 months of age are unable to demonstrate subphonemic sensitivity at least when tested in an explicit way similar to adults. This suggests that children have not fully developed their native phonology by the time they are 6-years-old.

Acknowledgements

First and foremost, I would like to thank my incredible advisors Dr. Kate Harrigan and Dr. Anya Hogoboom for not only their unending support throughout my thesis but also for helping me grow into the linguist, scientist, and person that I am today. Thank you for challenging me to ask deep questions about the world, inspiring my love for this topic, always believing in me, and teaching me how to be a kinder and more empathetic member of society. I would also like to thank the other members of my committee Dr. Dan Parker and Dr. Brian Rabinovitz for their support, guidance, and feedback throughout this process. Thank you to the William & Mary Linguistics Research Group for listening to the many iterations of this project and offering advice and support every step of the way. Thank you to the Child Language Lab at William & Mary for inspiring my love for acquisition and shaping me into the acquisitionist that I am today. To my fellow linguistics honors students and good friends, Abram Clear, Elizabeth Maneval, and Maddy Wade thank you for always being my biggest fans and inspiring me with your brilliance and incredibly impressive projects. A special thank you to my friend Valerie Bambha for her help in recruiting children. Most importantly, this project would not be possible without the participants, parents, and children that volunteered their time to participate in this study. Seeing the adorable faces of preschoolers and playing “the robot game” was always the best part of my day. I am forever grateful for the joy that they brought me during this uncertain time.

1. BACKGROUND

1.1 Phonemic vs. Phonetic

Consider the words ‘cab’ [kæb] and ‘cap’ [kæp] in English. Although they are quite different semantically, acoustically they are similar in many ways. They share the first two phonemes, and the final phoneme varies in only one dimension: voicing. Voicing is characterized by the presence (i.e. voiced) or lack (i.e. voiceless) of vocal fold vibration during articulation. Because a change in voicing would result in a different word, voicing is considered to be phonemic. Acoustic properties that are phonemic are critical for meaning and thus important dimensions for the listener to encode. However, not all acoustic properties are critical for meaning. For instance, the speed at which a word is spoken would not result in a different word. These acoustic properties that are not critical for meaning are considered to be phonetic. Historically, phonological theory has assumed that sounds belong to one of these two distinct categories based upon their importance for meaning.

1.2 Acoustic Cues

How does the perceptual system know whether there was vocal fold vibration to determine voicing? As it turns out, this is not an easy question to answer. There are a variety of acoustic cues that serve to indicate the phonemic status of a sound (Repp, 1982). These cues are not themselves phonemic because they would not constitute a different word, but they are critical for speech recognition because they provide information about the phonemic status of a sound. We will refer to these cues as subphonemic because they are meaningful but below the level of the phoneme. What makes this more challenging, however, is that a phonemic contrast is rarely, if ever, indicated by a single acoustic cue and oftentimes many, sometimes conflicting, cues serve to indicate a single phonemic contrast (McMurray et al., 2008). For instance, voicing is cued by voice onset time (Liberman et al., 1961), vowel length (Miller & Dexter, 1988), F1 onset frequency (Stevens & Klatt, 1974), and pitch (Haggard et al., 1970) all arriving at different time courses during speech (McMurray et al., 2008). Toscano & McMurray (2012) demonstrated that listeners weigh cues based on their reliability in determining phonological properties. Thus, more reliable cues will be weighed more strongly than less reliable cues.

Phonemic recognition not only relies on these ‘bottom-up’ acoustic cues, it also relies on ‘top-down’ contextual cues. For instance, Rubin et al. (1976) found that initial consonants were recognized more quickly when in words compared to nonwords. These results suggest that the context of being in a word can influence the use of acoustic cues. Likewise, well-known work by Ganong (1980) has suggested that an ambiguous speech sound is categorized based on other sounds in the word and crucially whether it would constitute a lexical item. For example, if an ambiguous sound between [d] and [t] was presented at the beginning of ‘ash’ [æʃ] listeners would be more likely to report having heard a [d] because ‘dash’ [dæʃ] is a word in English and ‘tash’ [tæʃ] is not. More recently, Brock & Nation (2014) investigated the influence of contextual top-down cues using eye-tracking and found that participants were quicker to look to the target word when semantic context relevant to that word was provided. When context was not provided, participants took longer to look to target suggesting that participants more strongly considered phonetically viable competitors. Finally, work on phonemic restoration, a phenomenon when listeners report having heard phonemes that were absent in the speech stream, indicates that restoring a lost phoneme relies on the combination of top-down and bottom-up cues (Samuel, 1981; Warren, 1970). Reaching a phonemic decision is a difficult feat as it involves the integration of many sources of information that are both acoustic and contextual in nature.

1.3 Voicing, Voice Onset Time, and Vowel Length

Recall, voicing is determined by whether there is vocal fold vibration during articulation. Voice onset time (VOT), the amount of time between the release of articulatory closure and the onset of voicing, has often been considered a model cue category system for voicing (Andruski et al., 1994; McMurray et al., 2002; Pisoni & Tash, 1974). In other words, VOT appears to be one of the closest examples in speech perception of a single cue indicating a phonemic contrast. However, vowel length (Denes, 1955; Hogan & Rozsypal, 1980; Miller & Dexter, 1988; Port & Dalby, 1982; Toscano & McMurray, 2012; Warren & Marslen-Wilson, 1988), F1 onset frequency (Stevens & Klatt, 1974), and pitch (Haggard et al., 1970) also cue voicing. This paper will focus on vowel length (VL) as a cue to voicing. VL participates in a trading relation with VOT such that when the phonemic information provided by one is ambiguous, the other can make up for the difference (Miller & Volaitis, 1989; Summerfield, 1981). In other words, if a listener is presented with an ambiguous VOT, VL can serve as a stronger cue to voicing. Some have considered the strength of VL in determining voicing to be strictly tied to its relationship with VOT. However, a study by Toscano and McMurray (2012) indicated that VL is not only useful in conjunction with VOT but serves as a separate cue itself indicating voicing. Similar to how phonemic contrasts are indicated by several acoustic cues, acoustic cues can serve to inform multiple phonemic contrasts. As such, VL also serves to cue vowel quality (Hillenbrand et al., 2000), manner (Miller & Liberman, 1979; Miller & Wayland, 1993; Shinn et al., 1985), and syllable structure (Salverda et al., 2003). Not only is VL an acoustic cue for a variety of phonemic contrasts, in some languages it is itself phonemic. In Czech for example, a change in the length of the vowel would result in a different word (Chládková et al., 2013) thus we would say that they have phonemic vowel length. Regardless of whether a language exhibits phonemic vowel length, VL serves as an important cue to many phonemic contrasts including voicing.

1.4 The Vowel Length Effect

It is well documented in English that vowels preceding voiced sounds are longer in duration than voiceless sounds (Allen & Miller, 1999; Chen, 1970; House, 1961; House & Fairbanks, 1953; Kessinger & Blumstein, 1998; Klatt, 1973; Lisker, 1957; Luce & Charles-Luce, 1985; Peterson & Lehiste, 1960; Sharf, 1962). Namely, the ratio of vowel length before voiced and voiceless consonants in English is 3:2 (Peterson & Lehiste, 1960; Sharf, 1962). Following Ko et al. (2009), we will call this pattern between VL and voicing where longer vowels are found preceding voiced sounds and shorter vowels are found preceding voiceless sounds the vowel length effect (VLE). Although it is not universal, the VLE has been documented in many of the world's languages including Hindi (Maddieson & Gandour, 1975), French, Russian, Korean (Chen, 1970), Danish (Fischer-Jorgensen, 1964), Dutch (Slis & Cohen, 1969), and Norwegian (Fintoft, 1961) to name a few. However, some languages do not exhibit the VLE such as Arabic (Flege & Port, 1981), Polish, and Czech (Keating, 1985).

Because the VLE is so common, many have proposed production-based explanations for its occurrence (several are reviewed in Kluender et al., 1988). For instance, Belasco (1953) proposed that more energy or 'physiological force' is required to articulate a voiceless consonant than a voiced consonant and as a result of the extra strain of articulating a voiceless consonant, the vowel is shorter. Others argue that vowels are lengthened in front of voiced consonants to allow time for laryngeal closure (Raphael, 1975). Kluender et al. (1988) instead offers a perceptual explanation. They propose that speakers produce vowels at different lengths to exploit the durational contrast of the following consonant. However, regardless of why the VLE exists, speakers of English are aware of its presence and utilize it for voicing determinations.

Countless perceptual studies have shown that English-speakers use vowel length as a cue to determine obstruent voicing (Denes, 1955; Hogan & Rozsypal, 1980; Port & Dalby, 1982; Toscano & McMurray, 2012; Warren & Marslen-Wilson, 1988). Perhaps the first of which was a study by Denes (1955) where participants were presented with synthesized 1-syllable words ending in voiceless fricatives and were asked to indicate whether the fricative was voiced or voiceless in a forced-choice task. Crucially, the length of the vowel preceding the fricative was varied. The results indicated that participants were more likely to perceive a voiced fricative when the vowel was longer even though there was no internal voicing. This finding was monumental at the time because it revealed that not only internal, but also external, acoustic cues can indicate a phonemic contrast. The VLE seems to have primary importance determining the voicing of an obstruent in coda position (Klatt, 1973, 1976; Sharf, 1962), but English-speakers have been shown to use VL as a cue to obstruent voicing word-finally (Denes, 1955; Hogan & Rozsypal, 1980; Warren & Marslen-Wilson, 1988), word-medially (Port & Dalby, 1982) and word-initially (Toscano & McMurray, 2012). Vowel length serves as a critical cue to the voicing of a following consonant in English as a result of the VLE.

1.5 Phonological Development

The current understanding of phonological development is that children are born "universal listeners" ready to acquire the phonology of any of the world's languages (Werker, 1995; Kuhl et al., 2005; Nakisa & Plunkett, 1998; Kuhl, 2004). This understanding claims that infants have the perceptual capabilities to discriminate any phonemic contrast that could exist in a language. Countless studies have demonstrated that young infants are able to discriminate phonemic contrasts for languages of which they are not exposed to supporting the claim that they are universal listeners. For instance, it has been shown that English-learning infants are able to discriminate phonemic contrasts in Zulu (Best et al., 1988), Hindi (Werker et al., 1981), Nthlakampx (Werker & Tees, 1984a), and Czech (Trehub, 1976) to name a few. Over time, however, infants' acoustic perception becomes tuned to the language(s) of which they are learning and as a result they lose the ability to distinguish nonnative contrasts (Bosch & Sebastián-Gallés, 2003; Burns et al., 2007; Kuhl et al., 2006; Polka & Werker, 1994; Trehub, 1976; Werker & Lalonde, 1988; Werker & Tees, 1984a). It is estimated that infants develop the phonemic categories for vowels between 4 and 6 months of age and for consonants between 10 and 12 months of age (Chládková & Paillereau, 2020).

However, not all phonemic contrasts are created equally. For instance, Swedish infants have difficulty discriminating between [ɑ] and [a] but not [ɑ] and [ɔ] even though they are all phonemic contrasts of Swedish (Lacerda, 1992). Additionally, Kuhl et al. (2006) demonstrated that English-learning infants have difficulty distinguishing [ɹ] and [l] until about 10 to 12 months although these phonemes are native phonemic categories of English. Narayan (2020) proposed an acoustic perspective of infants' speech recognition arguing that infants' difficulty or delayed development of certain phonemic contrasts is a reflection of the acoustic similarity (see also Narayan et al., 2010). In other words, infants struggle to acquire phonemic contrasts that are very similar acoustically. Over the first year of life, infants' phonology becomes tuned to the language(s) for which they are learning. However, acoustic contrasts that are more perceptually similar might be more difficult to acquire.

1.5.1 Vowel Length Development

Vowel length might be an example of a contrast that is particularly difficult to acquire (Narayan, 2020). Support for this claim comes from studies where infants have difficulty discriminating vowel length even though it is a phonemic contrast native to their language. For instance, a study by Mugitani et al. (2009) tested vowel length discrimination for Japanese-learning infants where vowel length is phonemic in Japanese. Results show that 10-month-old Japanese infants were able to discriminate vowel length but 18-month-olds could not. The authors suggest that the Japanese 10-month-olds are discriminating based on low-level acoustic differences. At 18-months, Japanese infants' developing phonology interferes with their length discrimination. Another study found that 4- and 7.5-month-old Japanese-learning infants were not sensitive to vowel length differences while 9.5-month-olds were (Sato et al., 2010) suggesting that infants are not born with the ability to discriminate vowel length differences but rather that it develops over time. These findings are problematic for the idea that infants are universal listeners because they show evidence for certain phonemic contrasts (i.e. vowel length) that infants might not be born ready to acquire. Perhaps the challenging nature of acquiring phonemic vowel length can explain the difficulty of subphonemic vowel length.

The majority of the studies testing infants' knowledge of phonemes rely on implicit discrimination tasks due to the obvious limitations of testing this population. However, studying implicit discrimination is quite limiting and research in many areas of development indicates that children act differently in implicit and explicit tasks. For instance in theory of mind literature, 15-month-old infants show implicit sensitivity to the belief states of others (theory of mind) when measured with looking time (Onishi & Baillargeon, 2005) but consistently fail explicit false belief tasks until the age of four (Wellman et al., 2001).

1.5.2 Subphonemic Development

A few recent studies have tested children's subphonemic development of various acoustic cues. For instance, one such study found that 24- and 29-month-olds were sensitive to vowel-to-vowel coarticulation mismatches (Paquette-Smith et al., 2016). Of course, this paper focuses on children's development of vowel length as a subphonemic cue to voicing. Researchers have investigated subphonemic vowel length in children in three main ways: infant discrimination, child production, and child perception.

Infant discrimination. Research has generally indicated that young infants are able to show implicit discrimination of vowel length. However, their discrimination is less robust than adults and they are still beginning to develop awareness of the VLE. Eilers et al. (1984) conducted one of the earliest studies testing subphonemic vowel length in English-learning infants using a Visually Reinforced Infant Speech Discrimination (VRISD) paradigm which reinforces head turns towards a change in the stimulus. In the study, 5- to 11-month-old infants were tested on their ability to discriminate varying vowel lengths preceding word-final [d] in 1-, 2-, and 3-syllable words. The stimuli were created such that the vowels were either 300ms, 400ms, 500ms, or 600ms in length. Infants were able to marginally discriminate duration differences and their discrimination improved as the duration difference became larger leaving the authors to conclude that a 67% increase in vowel length was required for perception.

Although they were able to find marginal discrimination, it was much lower than the discrimination rate for adults. Additionally, the 300-600ms vowel lengths are extremely exaggerated and thus unrealistic for natural speech. The researchers also tested infants' discrimination of [mæd] and [mæt] without VL as a cue to voicing leaving internal voicing cues as the only available cues. Somewhat surprisingly, infants were unable to discriminate based-upon internal voicing cues which the authors speculate to suggest that vowel length is a primary cue to voicing determinations at this age.

Another study by Mugitani et al. (2009) used a habituation-switch discrimination design where infants were habituated to a nonce word with a long vowel (e.g. [ta:ku]) or short vowel (e.g. [taku]). Looking time was measured to determine if infants would display different rates of looking when presented with a stimulus of the same or different vowel length as the habituated word where different rates of looking would indicate that they were sensitive to the vowel length difference. Results indicated that 18-month-old English-learning infants were able to discriminate the nonce words, thus indicating that they were sensitive to vowel length in a subphonemic position when tested through an implicit measure.

Finally, a study by Ko et al. (2009) tested whether infants expected vowel length and voicing to pattern in accordance with the VLE. To test their awareness of the VLE, Ko et al. (2009) tested whether infants would detect a mismatch in VL and obstruent voicing by measuring their looking time. A 'mismatch' was characterized by a short vowel followed by a voiced stop (e.g. [bæg]) or a long vowel followed by a voiceless stop (e.g. [bæ:k]). A 'match' was characterized as patterns in accordance with the VLE (e.g. [bæ:g], [bæk]). Half of the infants were presented with long vowels followed by voiced and voiceless stops (e.g. [bæ:g], [bæ:k], [kʌ:b], [kʌ:p], [pɪ:g], [pɪ:k]) and half were presented with short vowels followed by voiced and voiceless stops (e.g. [bæg], [bæk], [kʌb], [kʌp], [pɪg], [pɪk]). Interestingly, the results indicated that 8-month-olds did not show different rates of looking time across matches and mismatches and 14-month-olds were sensitive to the mismatch only when the vowel was long but not when it was short. This work suggests that awareness to the VLE develops between 8 and 14 months. Research on infant discrimination has indicated that 5- to 11-month-olds are able to marginally discriminate vowel length in a subphonemic position (Eilers et al., 1984), 18-month-olds are able to discriminate vowel length in a subphonemic position (Mugitani et al., 2009), and 14- but not 8-month-olds are aware of the VLE (Ko et al., 2009). Together these results indicate that young infants are able to discriminate vowel length differences, at least when tested implicitly, and are beginning to develop sensitivity to the VLE.

Child production. Another line of work exploring children's development of subphonemic vowel length comes from their production. In general, production studies show that young children are relatively adept at using the VLE. For instance, studies have revealed that children produce patterning consistent with the VLE at 21 months (Naeser, 1970), 2 years (Ko, 2007; Buder & Stoel-Gammon, 2002), 3 years, and 4 years (Raphael et al., 1980). However, an earlier study by Stoel-Gammon & Buder (1999) reported that only 50% of the 2-year-olds produced patterns with the VLE suggesting that production of the VLE at younger ages might not be as robust.

Child perception. Studies investigating older children's perception of subphonemic vowel length have demonstrated mixed results. An early study by Greenlee (1980) tested children's voicing determinations of a word-final obstruent. In the study, children listened to familiar words and indicated which word they heard of a minimal pair that differed only in the voicing of the final obstruent (e.g. bag, back) by pressing a button corresponding to an image. The stimuli were manipulated to strengthen the influence of vowel length on voicing. For words with voiced final obstruents (e.g. bag, buzz) voicing internal to the consonant was removed and the preceding vowel was shortened. For words with voiceless final obstruents (e.g. back, bus) the preceding vowel was lengthened. Thus, the researchers were able to determine if vowel length could cause participants to make voicing determinations that were contrary to the original voicing of the obstruent. Results indicated that adults and 6-year-olds were able to make a voicing decision that was contrary to the original voicing but in line with the vowel length. However, 3-year-olds always responded based upon original voicing. These results suggest that the subphonemic use of vowel length develops between 3 and 6 years of age.

However, a subsequent study by Wardrip-Fruin & Peach (1984) produced conflicting results. In their study, they also tested 3- and 6-year-olds and adults using English words ending in stops. They manipulated the stimuli by removing various cues for the voicing of the final obstruent to determine the relative weight of formant transitions, internal voicing, and preceding vowel length. Participants were again asked to press a button corresponding to the image of the word that they heard. Contrary to Greenlee (1980), they found that the 3-year-olds weighed vowel length as a stronger cue to voicing than 6-year-olds and adults.

Adding further complication, Krause (1982) found that 3- and 6-year-olds were able to make a voicing decision based upon vowel length when it was the only available cue to voicing in 1-syllable words but demonstrated a different threshold than adults. Again, the participants were tested in an explicit manner where they responded by pressing a button corresponding to the picture of the word for which they heard. In order for children to switch from a voiceless to voiced decision, children required longer vowel lengths than adults demonstrating a different threshold. However, it is difficult to reconcile these results because it is impossible to determine the extent to which top-down cues influence perception because four of the test words were expected to be familiar to children (e.g. back, bag, bib, pot), one was expected to be unfamiliar (e.g. pod), and one was a nonce word (e.g. bip). It is difficult to reconcile the conflicting findings found with child perception. Perhaps the conflicting findings suggest that children's subphonemic vowel length is still developing and somewhat fragile at these ages. Taken together, work on subphonemic vowel length in children broadly suggests that infants are able to show implicit sensitivity, children produce patterns consistent with the VLE from a young age, and results are mixed on whether older children can make a voicing judgement based upon vowel length.

1.5.3 Word Learning and Word Recognition

More recently, researchers have considered how differing vowel length might interact with word learning or word recognition. In one such study, Dietrich et al. (2007) tested Dutch- and English-learning 1.5-year-olds' ability to learn words that differed only in the length of the vowel. They were taught the novel words [tam] and [ta:m] in a switch procedure. Dutch infants but not English infants were able to learn that [tam] and [ta:m] were distinct words suggesting that they treat vowel length as lexically contrastive while English infants did not. This is consistent with their native phonology as vowel length is phonemic in Dutch but not English. However, based on this study we cannot determine whether English-infants have subphonemic vowel length because the vowel was not before an obstruent.

A subsequent study by Swingley & van deer Feest (2019) tested infants' ability to recognize words with a subphonemic mismatch. 21-month-old English-learning infants were presented with familiar CVC words that, crucially, ended in an obstruent (e.g. bed, keys, dog, cup). The researchers used a mispronunciation design with language-guided looking which relies on the fact that children fixate to named images less when the pronunciation is noncanonical. The length of the vowels was manipulated such that some stimuli consisted of mismatches between vowel length and voicing and others consisted of matches between vowel length and voicing. Interestingly, infants did not fixate at different rates as a function of the mispronunciation showing that word recognition was not hindered by a subphonemic mismatch. We would not expect infants to reach a different lexical decision due to a subphonemic mismatch but it is surprising that they do not show different patterns of looking. This is interesting because studies of phonetic specificity show that children's recognition of familiar words is hindered by phonemic mispronunciations (Swingley & Aslin, 2000). Together, these results suggest that children's word recognition of familiar words is influenced by acoustic properties that are phonemic but not subphonemic. Note however, that the lack of hindrance due to subphonemic mismatch does not mean that children are unaware of subphonemic patterns but rather that they do not consider the subphonemic level when recognizing familiar words.

1.5.4 Phonetic Specificity and Upper-Level Cues

Although infants have developed the phonemic categories of their language between 10 and 12 months of age (Chládková & Paillereau, 2020) recent studies have shown that they fail to apply these skills when mapping novel words to meaning. For instance, Stager & Werker (1997) conducted a study where 14-month-olds were taught labels for two objects where the two words were either phonetically similar (e.g. 'bih' and 'dih') or phonetically dissimilar (e.g. 'lif' and 'neem'). Results showed that children were able to learn the word-object pairings for the phonetically dissimilar words but not the phonetically similar words. In a follow-up study both 8-month-old and 14-month-old infants were tested in a similar word-learning task where they were taught one label instead of two. Interestingly, they found that 8-month-old infants were able to notice the difference between the learned word 'bih' and a phonetically similar novel word 'dih' but, 14-month-old infants were not. These results suggest that when the older infants map sound to meaning they are less sensitive to phonemic differences and younger infants are showing more phonemic sensitivity because they aren't linking to meaning and are instead simply discriminating the sounds.

A subsequent study tested whether older children had a bias towards consonant or vowel identities in a word-learning task (Havy et al., 2011). In the task, French-learning 3- , 4- and 5-year-olds were taught two labels for two words that different in a consonant and a vowel (e.g. [byf] and [duf]). At test, children were presented with a word that matched in either the consonant or vowel of the previously learned words (e.g. [buf] or [dyf]) and crucially had to choose whether to map based on the matching consonant or vowel. The 3-year-olds exhibited a consonant-bias where they chose to map to the word with the matching consonant rather than matching vowel suggesting that they consider consonants to be more lexically contrastive than vowels. However, 4- and 5-year-olds did not exhibit this bias suggesting that by the time children are 4 years they assign equal lexical weight to consonants and vowels. Perhaps this initially lower lexical weight given to phonemic vowel contrasts could translate into subphonemic vowel length. Overall, these studies indicate that when children are mapping words to meaning they are often less specific with their phonemic representations even though they are aware of the phonemic contrasts that exist in their language.

1.6 Other Uses of Vowel Length

1.6.1 Glides

Recent work suggests that speakers use vowel length subphonemically to identify a glide in vowel hiatus (Hogoboom, 2020). Vowel hiatus is a configuration that exists in some languages where two vowels are found adjacent to one another across a syllable boundary (V.V). Interestingly, when a high front vowel (e.g. [i]) precedes a lower, backer vowel (e.g. [ə], [ɑ]) the transition between the vowels exhibits the acoustic signature of a glide (i.e. [j]) resulting in what we will call an 'acoustic glide.' Figure 1 below demonstrates the acoustic similarity between an acoustic glide that arose from transition between [i] and [ə] in vowel hiatus (left) and an actual glide (right).

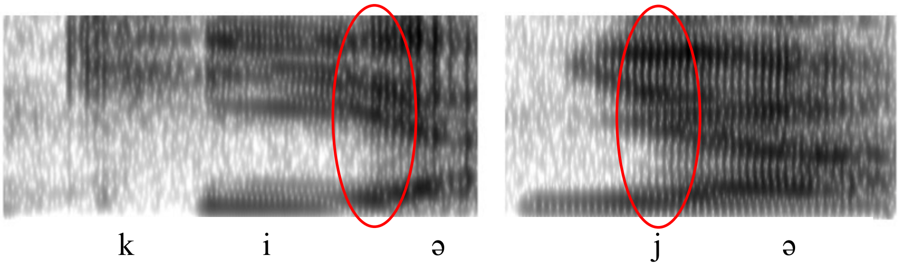

Figure 1. Spectrogram of [kiə] and [jə]
The spectrogram above illustrates an instance of vowel hiatus (left) and a glide (right) where the acoustic patterning is almost identical (circled in red).

Because of the similar acoustic patterns of the transition between [i] and [ə] and a [j], [iə] and [ijə] are confusable to speakers. Work with English-speakers has demonstrated the they are more likely to perceive a glide in hiatus when the preceding vowel is long rather than short (Hogoboom, 2020). This pattern is stronger between words (i.e. where English contrasts [i#jə] and [i#ə]) than within a word (i.e. where English does not contrast [ijə] and [iə]). Some languages, such as Mandarin and Korean, likely exhibit this contrast within a word (Hogoboom, 2020). For instance, Korean has a minimal pair [kiʌ] 'crawl' and [kijʌ] 'contribution' demonstrating that they must contrast these forms. Analyzing production from Korean and Mandarin speakers supports the claim that like English-speakers, they use vowel length subphonemically to identify a glide (Hogoboom, 2020). Hogoboom (2020) found that the speakers did not produce durational differences between the transition between [i] and [ə] and a glide. However, she found that speakers of both languages produced longer vowels preceding forms that were said to contain a glide than preceding forms without glides.

1.6.2 Final Lengthening

Final lengthening is a cross-linguistic phenomenon that occurs in many of the world's languages wherein sounds at the end of a prosodic domain are lengthened (Crystal & House, 1988; Johnson & Martin, 2001; Lunden, 2017; Nakai et al., 2009). Although final lengthening might have its origins at higher prosodic domains, such as phrase finally, it has been noted to occur at the word level. Many have speculated about the potential usefulness of final lengthening for speakers. Klatt (1976) suggests that final lengthening serves to indicate the end of a phrase or domain. There has also been evidence that final lengthening cues can facilitate word segmentation (Elordieta et al., 2005). In English, White and Turk (2010) note that final lengthening might be particularly strong as a result of its stress system. Although final lengthening is not subphonemic, as it does not indicate the phonemic status of a sound, it is a tendency that occurs in English that speakers are likely aware of.

1.7 The Current Study

The current study seeks first to determine whether adults will be differentially sensitive to vowel length based upon whether it is found in a subphonemic position. Unlike previous studies, we asked not whether adults could make a voicing determination based upon vowel length but rather if the length would be more perceptible in a position where the length is predictive. We test adults' sensitivity in four environments two of which are not subphonemic (word-final and baseline), one which might be subphonemic (before a glide), and one which has been shown to be subphonemic in English (preceding an obstruent).

We also compare the adult sensitivity in baseline and the obstruent condition to children, improving upon the previous child literature in several ways. We chose crucially to test sensitivity and not phonemic decision making in order to eliminate the top-down cues that would be involved in testing children with known words and the mapping to meaning required to map to a picture. This allowed us to isolate the acoustic system and gain a better understanding of children's subphonemic vowel length from 4 to 6 years. The study itself consisted of an explicit sound discrimination task that could be compared across children and adults using nonce words. We suggest that the lower level of phonetic detail children provide when using upper-level word learning skills might explain children's lack of sensitivity to subphonetic (as in Dietrich et al., 2007) and subphonemic (as in Swingley & van deer Feest, 2019) vowel length. Additionally, perhaps the top-down cues as a result of using familiar words and mapping to meaning (i.e. by requiring children to press a button corresponding to a picture) can explain the conflicting results we find in Greenlee (1980), Wardrip-Fruin & Peach (1984), and Krause (1982).

2. ADULT STUDY

2.1 Purpose and Hypotheses – Adult Study

The adult study was created to determine whether vowel length would be differentially perceptible based upon its predictiveness as a subphonemic cue. Specifically, whether vowel length differences may be more perceptible before an obstruent where adults have been shown to use it as a subphonemic cue to voicing. The study was designed to provide a comparison for children to determine their use of subphonemic vowel length from 4 to 6 years. Crucially, by using nonce words we eliminated top-down cues in order to isolate the acoustic perceptual system. We made two main hypotheses regarding the adult study. First, we hypothesize that larger length differences will be more perceptible than smaller length differences overall. The second hypothesis is that adults would show varying perceptibility of vowel length based on its position within the word. Specifically, we expect vowel lengths that are found before an obstruent to be more perceptible than other places in the word. The hypotheses for the adult study are illustrated below.

> *Hypothesis 1*: Larger vowel length differences will be more perceptible than shorter vowel length differences.
> *Hypothesis 2*: Adults will show varying perceptibility of vowel length based on whether it is found in a subphonemic position.

Hypothesis 1 predicts that participants will provide less *yes perfect match* responses at higher length difference levels. Hypothesis 2 predicts that participants will respond *yes perfect match* significantly less in the OBSTRUENT environment compared to other environments. See method section below for more details.

2.2 Method – Adult Study

Participants.

Sixty-six native English-speaking adults between the ages of 18 and 57 (MEAN: 21.83 years) were recruited from the William & Mary SONA participant pool and social media. Of those recruited 43 identified as female, 22 as male, and 1 as nonbinary. All participants were native speakers of English and two were also native speakers of Spanish, one of Urdu, three of Mandarin, one of Vietnamese, one of Tagalog, and one of Korean. Those recruited through SONA consisted of William & Mary students enrolled in introductory Psychology and Linguistics courses who received class credit for their participation. All participants were required to provide written consent to participate in the study. Of the 66 adults recruited, 3 were excluded for completing the study more than once leaving 63 participants included in the study (N=63).

Design.

The study consisted of a sound discrimination task where participants listened to pairs of phonemically identical nonce words and were asked to indicate the level of similarity across the pair. The design was within-subjects where all participants were exposed to all levels of the independent variables. Two independent variables were manipulated in this study: **ENVIRONMENT** and **LENGTH**. ***Response*** was recorded as the dependent measure.

ENVIRONMENT (OBSTRUENT, GLIDE, WORD-FINAL, BASELINE). The environment of the target-vowel (i.e. the vowel that was changing in length) was manipulated as a within-subjects factor. This factor tests the hypothesis that speakers will treat vowel length differently dependent on the position and surrounding phonemes within the word. Thus, environment was determined by the location of the target-vowel in comparison to other phonemes. There were four environments: OBSTRUENT, GLIDE, WORD-FINAL, and BASELINE. The first segment (a consonant) was randomly assigned in every nonce word to provide variety (see stimuli section for details on initial consonants). The remaining segments were the same within each environment to provide consistency. These final segments constitute the environment. In the OBSTRUENT environment, the target-vowel was found in a location where English-speakers typically use it subphonemically: before an obstruent. The final segments for the OBSTRUENT position were [ikə] with target-vowel [i]. In the GLIDE environment, the target vowel was found in a location where its length might warrant an acoustic glide: in vowel hiatus before a [ə]. The final segments for the GLIDE environment were [iə] with target-vowel [i]. The final segments for the WORD-FINAL position were [ɑzə] with target-vowel [ə]. The vowel [ɑ] and fricative [z] were chosen here to provide variety. Finally, the baseline environment was a neutral environment with no reason for subphonemic vowel length: before a nasal. The final segments for the BASELINE environment were [inə] with target-vowel [i].

LENGTH (0, 1, 2, 3, 4, 5). The difference in length of the target-vowels across the nonce word pair was also manipulated as a within-subjects factor. For the OBSTRUENT, GLIDE, and BASELINE environments, the length of the target-vowel ranged from 80ms to 180ms using 20ms increments. This resulted in six lengths: 80, 100, 120, 140, 160, and 180ms. For the WORD-FINAL environment, the lengths were longer as segments are longer at the ends of words as a result of final lengthening (Crystal & House, 1988; Johnson & Martin, 2001; Lunden, 2017; Nakai et al., 2009). The length of the WORD-FINAL target-vowels ranged from 120ms to 220ms using 20ms increments leaving six lengths: 120, 140, 160, 180, 200, and 220ms. The length difference was calculated as the absolute value of the target-vowel's length in the second word of the pair minus the target-vowel's length in the first word of the pair. There were six levels of length difference: 0 (0ms), 1 (20ms), 2 (40ms), 3 (60ms), 4 (80ms), and 5 (100ms). Pairs were created such that there were four pairs at each length difference level within each environment. Both orders of every pair were included.

Response (*yes perfect match*, *no not quite*). The response of the participant was recorded as the dependent measure. Participants were given a forced choice response based on the similarity of the words within the pair of *yes perfect match* and *no not quite*. These response options were chosen to reflect the phonemic flexibility with which we wanted participants to categorize the stimuli since the pairs were phonemically identical.

Stimuli.

Test Stimuli. The test stimuli for the study consisted of 96 pairs of nonce words created using MBROLA (Dutoit & Pagel, 1995) and Praat (Boersma & Weenink, 2018) software. The individual nonce words were first synthesized through MBROLA using the US1 voice. After the nonce words were synthesized, they were adjusted to correct any discrepancies and concatenated into pairs through Praat. Nonce words took the form CVCV for the OBSTRUENT, WORD-FINAL, and BASELINE environments and the form CVV for the GLIDE environment. The first consonant in the word was always 80ms and the second consonant, when applicable, was always 60ms. These lengths were chosen because they are typical in speech. For the OBSTRUENT, GLIDE, and BASELINE environments, the first vowel was the target-vowel and the second vowel was a [ə] of length 170ms, the mid-range length for the WORD-FINAL vowels. For the WORD-FINAL environment, the second vowel was the target-vowel and the first vowel was [ɑ] of length 130ms, the mid-range length for the non-word-final vowels. Recall, the final three segments within each condition were identical to provide a consistent environment for the target-vowel but the first consonant was varied to provide variety. The following initial consonants were randomly assigned to pairs [t, p, f, v, z, s, ʃ, t͡ʃ, n, m, r, d͡ʒ].These initial consonants were chosen in order to provide a wide range of manners, voicing, and places. Once nonce words at every length had been created, they were paired and concatenated through Praat. The two words across the pair were phonemically identical and the pitch of the second word was 20hz higher than the first in order to simulate two different "voices." We chose to use varying pitches across the word pair instead of different voices in MBROLA to avoid any potential idiosyncrasies within the voices that could confound our data. The amount of time between the first and second word in the pair, the interstimulus interval (ISI), was 200ms. We chose 200ms because speakers show better discrimination using lower ISIs (Tyler & Fenwick, 2012).

Practice stimuli. Practice stimuli were created using the same process as the test stimuli but were instead single CVC syllables. We chose to use CVC words instead of the CVCV used at test to cue participants into the length and level of specificity required of the task without the specific context of what we were testing. There were eight different practice pairs: four of the smallest length difference (0) and four of the largest length difference (5). Four pairs contained the target-vowel [ɑ] and four contained the target-vowel [i] to provide variety. Segments for the initial consonant were [p, f, v, z, ʃ, t͡ʃ, r, d͡ʒ]. The final consonant of the word was a nasal to provide a neutral environment. The place of the nasal was varied where four pairs contained [m] and four contained [n].

Procedure.

The study was run remotely through Ibex farm (Drummond, 2016), a web-based platform for deploying psycholinguistics experiments. Participants received a link to the experiment to complete the study. Participants were instructed to find a quiet place free of distractions and the use of headphones was recommended. Before the study could begin, the participants were required to read and sign a consent form. Once the consent form had been signed, participants were directed to a sound check page where they could play a syllable and were instructed to adjust their volume. There were two phases of the study: practice and test followed by a brief demographic form and debrief.

Practice phase. After the sound check, the participant entered into the practice phase of the task. The practice phase consisted of eight nonce CVC pairs where four were at the 0 length difference level and four were at the 5 length difference level. After each pair was played, the participant received feedback where *yes perfect match* was the correct response for the 0 length difference pairs and *no not quite* was the correct response for the 5 length difference pairs. Participants could listen to the practice words as many times as they liked before clicking to the next page for the answer. Instructions for practice rounds were “In the following study, you will be played pairs of nonce words (i.e. made up words) and asked whether the pronunciation of the second was the same as the first on a scale of *yes perfect match* and *no not quite.* You will now be given 8 practice pairings. Please listen to each pair, consider whether you would say they are a *perfect match* or *not quite*, then click to the following screen to see the correct answer.”

Test phase. Once the practice rounds were complete, the participant entered the test phase. In the test phase, the instructions were provided again and then the study began. For the test phase, the 96 pairs were randomly ordered for each participant and they were asked to respond whether the pronunciation of the second was the same as the first on a scale of *yes perfect* match and *no not quite*. Each test stimulus was only played once.

Demographic and debrief. Once the test phase was complete, the participants were directed to a brief demographic form where they were asked to indicate their age, gender, and list their native languages. After the demographic form, participants were directed to the debrief which contained information about the study and our contact information for questions. Once the participants clicked past the debrief page, their responses were sent to the Ibex farm server.

2.3 Results – Adult Study

We find that adults provide different rates of *yes perfect match* responses across the different lengths and environments. The proportion of *yes perfect match* responses across the different lengths and environments are listed in Table 1 below.

Table 1. Proportion *yes perfect match* responses in adult study

	LENGTH					
ENVIRONMENT	0	1	2	3	4	5
BASELINE	0.853	0.813	0.679	0.631	0.472	0.349
OBSTRUENT	0.845	0.833	0.687	0.544	0.329	0.202
GLIDE	0.817	0.794	0.778	0.575	0.476	0.369
WORD-FINAL	0.798	0.810	0.750	0.746	0.742	0.651

We used generalized linear mixed effects models to analyze the results. These models are well suited for analyzing categorical data (Baayen, 2007; Jaeger, 2008). The reported models have random intercepts. These models predict the probability of a specific response (*yes perfect match*) in the different environments (see Agresti, 2002; Jaeger, 2008). We ran a mixed-effect logit model with *yes perfect match* responses as the dependent measure, **LENGTH** (0, 1, 2, 3, 4, 5) and **ENVIRONMENT** (OBSTRUENT, GLIDE, WORD-FINAL, BASELINE) as fixed effects, and **SUBJECT** as a random effect. The model indicated a main effect of **LENGTH** [$X^2_{(5)} = 196.928$, $p < 0.0001$], showing that participants were responding *yes perfect match* at significantly different rates across the different **LENGTHS** and no main effect of **ENVIRONMENT** [$X^2_{(3)} = 3.631$, $p = 0.304$], showing that participants were not responding *yes perfect match* at significantly different rates across the different **ENVIRONMENTS**. We found an interaction between **LENGTH** and **ENVIRONMENT** [$X^2_{(15)} = 129.320$, $p < 0.0001$] showing significantly different patterns of responses based upon the combination of **LENGTH** and **ENVIRONMENT**.

In order to better understand what is driving the interaction, we ran pairwise comparisons, analyzing each level of **LENGTH** separately. Pairwise comparisons revealed no significant differences between any of the **ENVIRONMENTS** at lengths 0, 1, and 2. At length 3, we find that participants are significantly less likely to respond *yes perfect match* in the GLIDE ($p = 0.0002$), OBSTRUENT ($p < 0.0001$), and BASELINE ($p = 0.0217$) environments compared to the WORD-FINAL environment indicating that they have a lower sensitivity to vowel length differences in the WORD-FINAL environment comparatively. Crucially, at lengths 4 and 5 we find significant differences between all comparisons except BASELINE and GLIDE. Compared to BASELINE, participants are more likely to say *yes perfect match* in the WORD-FINAL environment ($p < 0.0001$), less likely in the OBSTRUENT environment ($p = 0.0040$), and equally likely in the GLIDE environment ($p = 0.9997$) at lengths 4 and 5. Figure 2 below illustrates the proportion of *yes perfect match* responses at each length and environment. Estimated marginal means and p-values for each environment at each length level are in Appendix A.

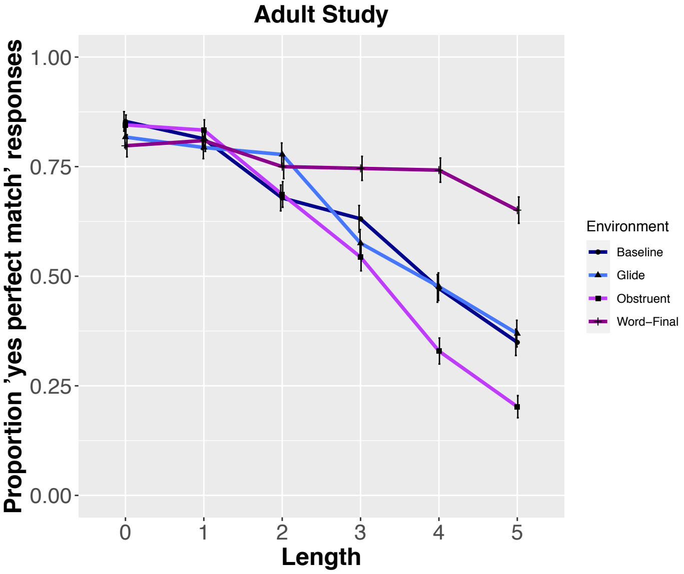

Figure 2. Proportion *yes perfect match* responses for all environments in the adult study.

2.4 Discussion – Adult Study

The main effect of **LENGTH** across all environments indicates that as vowel length differences became larger, participants were less likely to respond *yes perfect match* showing that they were more sensitive to the differences at these levels. This shows support for Hypothesis 1 that participants will show greater sensitivity to greater vowel length differences. The interaction between **LENGTH** and **ENVIRONMENT** indicates that participants are responding differently due to the combination of **LENGTH** and **ENVIRONMENT**. Pairwise comparisons allowed us to identify that the interaction was being driven by the larger lengths (3, 4, and 5). There were no significant differences in responses across environments at lower lengths (0, 1, and 2) indicating that this difference is too small for perception (or at least to warrant a *no not quite* response). Additionally, a subphonemic position (OBSTRUENT) did not boost sensitivity enough to show different responses at this level. It is possible that adults are capable of perceiving differences this small but perception is not strong enough to be different across the environments or to provide a *no not quite* response.

Responses were significantly different in WORD-FINAL compared to BASELINE, GLIDE, and OBSTRUENT at level 3 and between all environments except BASELINE and GLIDE at levels 4 and 5. These results indicate that vowel length differences were the least perceptible in WORD-FINAL. We suspect that the reason we see the least sensitivity in WORD-FINAL position is because perception is logarithmic (Kluender et al., 2003). Perceptual sensitivity is based on ratio and not the absolute difference between two stimuli; when the stimulus is larger, a greater magnitude of difference is required to perceive a difference. Because the vowel lengths were already comparatively larger in the WORD-FINAL environment to account for final-lengthening, a larger difference would be required to achieve the same ratio of change found in the other environments. If this is driving the differences we see in the WORD-FINAL environment, we would expect to see WORD-FINAL act similarly to BASELINE and GLIDE when the ratio of change is constant. Another possibility is that WORD-FINAL is a less perceptible location in the word.

Sounds at the ends of words often undergo phonetic patterns of devoicing (Hock, 1999), lengthening (Johnson & Martin, 2001; Lunden, 2017; Nakai et al., 2009), and tend to trail off. Because of word-final phenomena, speakers might be adept at ignoring differences in this position.

The differences at length levels 4 and 5 indicate that length differences are the most perceptible in OBSTRUENT, equally perceptible in GLIDE and BASELINE, and the least perceptible in WORD-FINAL. Participants are least likely to give a *yes perfect match* response in OBSTRUENT, indicating that their sensitivity to vowel length differences is influenced by a potential subphonemic use providing support for Hypothesis 2. In other words, the vowel length difference is most noticeable when it is found before an obstruent because that is a location where vowel length is predictive of a phonemic contrast. Because adults are treating vowel length differently based on its predictiveness we would say that they are using it subphonemically before an obstruent. Comparing GLIDE and BASELINE revealed that GLIDE was not a subphonemic position for English speakers. This is not surprising, however, because English does not contrast [ijə] and [iə] within a word (Hogoboom, 2020). Because there is no contrast word-internally, GLIDE is similar to BASELINE because there is no subphonemic use. If we were to test across word boundaries, where English does contrast [i#ə] and [i#jə], GLIDE might be more perceptible than BASELINE. In other languages, such as Korean and Mandarin (Hogoboom, 2020) where this contrast likely exists word-internally, we would expect GLIDE to act similarly to OBSTRUENT in English because before a glide is a subphonemic position for vowel length. Overall, these results indicate that a subphonemic position boosts perceptibility for vowel length.

3. CHILD STUDIES

Seventy-six children between the ages of 4;0 and 6;11 were recruited remotely from across the United States to participate in the child studies. Children were recruited through the William & Mary Child Language Lab Database, the Cornell Play & Learning Lab Virtual Child Database, social media, local pre-schools, and the Cognitive Development Society listserv. To participate in the study, all children were required to provide verbal assent and parents were required to give written consent. Of the 76 children, two were excluded due to excessive background noise and one was excluded due to parental interference leaving 73 children included in the studies (N=73, MEAN AGE=5;5.6).

Four different experiments were conducted to provide a comparison to the finding with adults showing subphonemic vowel length before obstruents. Experiment 1 served as a direct comparison to adults testing whether children were sensitive to vowel length differences before obstruents at the same threshold. Experiment 2 served as a control to see if children's responses in Experiment 1 were influenced by the context of the task. Experiment 3 tested whether children have a larger threshold than adults by using vowel length differences that were twice as large. Experiment 4 served as a comparison to Experiment 3 to see if children were responding differently based upon whether the position was subphonemic. A comparison between Experiment 3 and 4 can reveal whether children are using vowel length subphonemically.

Piloting indicated that the range of within subjects factors in the adult study would create an overwhelming setup for children, so we decided to simplify the adult design in several ways. These changes included: manipulating environment as a between subjects factor, pairing down the length difference level to two levels, making the pitch the same across the pairs, and using the same stimuli at practice and test. The reasoning behind simplifying the task for children was that regardless of the amount of scaffolding we provide, if children cannot perceive the length differences, they will fail regardless of the changes made from the adult experiment.

3.1 EXPERIMENT 1: OBSTRUENT

3.1.1 Purpose and Hypotheses – Experiment 1

The purpose of Experiment 1 was to serve as a direct comparison to the finding in adults of subphonemic vowel length before an obstruent. Because the literature suggests that children develop the phonological systems of their language by one year (Werker, 1995), we predict that children will act similarly to adults. Specifically, we expect children to be sensitive to vowel length differences in this experiment where the length is found in a predictive position (i.e. before an obstruent). Our hypothesis for this experiment is illustrated below.

> *Hypothesis 3*: Children will be sensitive to vowel length differences in a subphonemic position (before an obstruent).

This predicts that children will respond *same* significantly more at the SAME length difference than the DIFFERENT length difference.

3.1.2 Method – Experiment 1

Participants.

Twenty-three children between the ages of 4;0 and 6;11 (MEAN = 5;5.77) recruited remotely from across the United States participated in Experiment 1.

Design.

The study consisted of a sound discrimination task deployed through Zoom where participants listened to pairs of phonemically identical nonce words and were asked to indicate the level of similarity across the pair. The nonce words were spoken by two robot characters (Robot and Baby Robot) and children were asked to indicate whether Baby Robot said it the same or a little different from robot either verbally or by pointing to one of two buttons on the top of the screen. Baby Robot always spoke second as he was "practicing his robot sounds" by attempting to copy Robot. **LENGTH** was manipulated as a within-subjects factor and ***response*** was recorded as the dependent measure.

LENGTH (SAME, DIFFERENT). The difference in length of the target-vowels across the nonce word pair was manipulated as a within-subjects factor. Differently from adults, the length difference was paired down to two levels: SAME and DIFFERENT. SAME corresponds to the 0 length difference level (0ms) and DIFFERENT corresponds to the 5 length difference level (100ms) from the adult study. For the SAME level, target-vowels were either both 80ms or 180ms resulting in a 0ms difference across the pair. At the DIFFERENT level, one target-vowel was 80ms and the other was 180ms resulting in a 100ms difference across the pair. The order of the target-vowels was varied. Pairs were created such that there were 10 pairs at the SAME level and 10 at the DIFFERENT level. Within the SAME level, 5 pairs had target-vowels of length 80ms and 5 pairs had target-vowels of length 180ms. Within the DIFFERENT level, both orders of every pair were included.

Response (*same, different*). The participant's response was recorded as the dependent measure. Participants indicated verbally or by pointing to a button on the top of the screen whether Baby Robot said the sound the *same* or *different* than Robot.

Stimuli.

Test Stimuli. The test stimuli for the study consisted of 20 pairs of nonce words created using MBROLA (Dutoit & Pagel, 1995) and Praat (Boersma & Weenink, 2018) software. The pairs were selected from adult tokens in the OBSTRUENT environment and adjusted to the appropriate lengths (when necessary) in Praat. Consonant lengths and non-target-vowel length was the same as adults. The final three segments were identical to the OBSTRUENT condition in adults to provide a consistent environment for the target-vowel. Again, the first consonant was varied to provide variety. The following initial consonants were randomly assigned to pairs [t, p, z, s, ʃ, t͡ʃ, n, m, r, d͡ʒ]. The ISI between the words was the same as adults.

Practice stimuli. Unlike adults, practice stimuli were chosen from the test stimuli. There were six different practice pairs: two SAME and four DIFFERENT. Segments for the initial consonant were [p, t, r, z, ʃ, m].

Materials and Procedure.

Children were tested remotely through Zoom. Parents were instructed to ensure a quiet location to complete the study and use either a high-quality speaker or headphones. The study was a sound discrimination task constructed as a game where the children were helping Baby Robot practice his robot sounds. The experimenter first introduced the children to the characters and then explained the layout of the game. Children were told that to help Baby Robot learn his sounds, Robot was going to say a sound and Baby Robot was going to try to copy it exactly as Robot said it. They were then introduced to two buttons on the top of the screen, one red and one green, corresponding to whether Baby Robot said it the *same* or *different* from Robot. The green button corresponded to a response of *same* and the red to a response of *different*. When each button was pressed, the button grew in size and played either a celebratory sound or a failure sound.

Practice phase. After the child had been introduced to the characters, buttons, and layout of the game, they entered a practice phase. The first two practice trials were modeled by the experimenter and the child was instructed to listen carefully and the experimenter will tell them whether Baby Robot said it the *same* or *different* by pressing the corresponding button[1]. After the two modeled pairs, the child was instructed to guess whether Baby Robot said it the *same* or *different* before the experimenter indicated the correct answer for the remaining four practice trials.

Test phase. Once the practice trials were complete, the child entered the test phase where they were "in charge" of which button is pressed. In the test trials, children listened to 20 pairs of

---

[1] This method of practice was implemented after the first 12 participants. Slightly different formats of practice were used previously but none were successful in helping children hear the vowel length differences.

nonce words and indicated whether they were the *same* or *different* verbally or by pointing to the buttons. After the child gave a response, the experimenter pressed the button that corresponded to the child's response. Responses were recorded by the experimenter in real time on a separate score sheet.

Pallet cleanse and resolution. Halfway through the test trials, there was a brief pallet cleanse to re-engage the child. They saw an animation where Baby Robot runs out of battery and Robot helps him recharge. Once the remaining test trials are completed, the child enters a resolution phase where Baby Robot thanks the child for their help and thinks about how his teacher is going to be so proud of everything he has learned.

3.1.3 Results – Experiment 1

We find that children are providing similar responses across the SAME and DIFFERENT trials. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 1 below.

Table 1. Proportion *same* responses in Experiment 1

SAME length	DIFFERENT length	*p* value
0.704	0.636	0.1706

To analyze the results, we ran a mixed-effect logit model with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect. The model did not reveal a main effect of **LENGTH** [$X^2_{(1)}$= 1.878, p = 0.171], showing that participants were responding *same* at similar rates across the SAME and DIFFERENT trials.

3.1.4 Discussion – Experiment 1

The lack of a main effect of **LENGTH** revealed that children were not sensitive to vowel length differences in this experiment. Thus, not supporting Hypothesis 3 that children would be sensitive to vowel length differences before obstruents as was seen with adults. We identified two possibilities for why children were not sensitive to vowel length in this experiment. The first possibility is that the context of the task is causing children to be less sensitive. Recall in the adult experiment, participants were presented with pairs of nonce words with no context. To make the child comparison appropriate for children, we turned the task into a game which involved the use of characters participating in a language exchange. It is possible that because the experiment closely resembled language, children were mapping to a lexical level in a way that adults were not. Previous research suggests that linguistic context can influence sensitivity to phonemic contrasts (Werker & Tees, 1984b). This possibility was addressed in Experiment 2 where the language context was removed. The second possibility is that children have a different threshold than adults. In other words, they might require larger length differences to show sensitivity. This second possibility is addressed in Experiment 3 where the length differences are twice as large as Experiment 1.

3.2 EXPERIMENT 2: NON-LANGUAGE OBSTRUENT

3.2.1 Purpose and Hypotheses – Experiment 2

The purpose of Experiment 2 is to serve as a control to see if the context of the task for children might be driving the different patterns seen in adults and children. Because the task involves two characters speaking in a manner that resembles a conversation, it is possible that children are considering the differences at the lexical level unlike adults. In other words, children could be less particular with Baby Robot's responses because a word with a longer vowel in English would not result in a different lexical item. Perhaps they are sensitive at a similar level to adults, but the context of the task is creating a language-bias. We predict that children will perform equally in Experiment 2 and Experiment 1 indicating that a language-bias was not influencing their sensitivity. See Hypothesis 4 below.

Hypothesis 4: Children's responses are not being influenced by a language-bias.

This hypothesis predicts that children will respond *same* at equal rates across the SAME and DIFFERENT length levels (as was seen in Experiment 1).

3.2.2 Method – Experiment 2

Participants.

Eleven children between the ages of 4;0 and 6;11 (MEAN = 5;5.7) recruited remotely from across the United States participated in Experiment 2.

Design.

The design was the same as in the previous three experiments except the pairs were played by a single computer instead of two robots. The stimuli were identical to Experiment 1. We chose to use the same stimuli as Experiment 1 to isolate context. By isolating context, we could clearly determine whether context was driving the different responses with children.

Materials and Procedure.

Children were tested remotely through Zoom. Parents were instructed to ensure a quiet location to complete the study and use either a high-quality speaker or headphones. The experiment was a sound discrimination task constructed as a game where the children were helping an owl named Ollie fix her computer. The children were told that Ollie is really smart and she created a computer that plays identical sounds. Lately, however, the computer hasn't been working very well and as a result Ollie is calibrating her computer using two buttons. When the computer plays two sounds that are exactly the SAME, Ollie presses the green button. When the computer plays two sounds that are a little DIFFERENT, Ollie presses the red button. The buttons were the same as in the previous experiments.

Practice phase. After the child had been introduced to Ollie, her computer, the buttons, and the layout of the game, they entered a practice phase. The first two practice trials were modeled by Ollie and the child was instructed to listen carefully and Ollie will tell them whether the computer played two sounds that were the SAME or DIFFERENT by pressing the corresponding button. After the two modeled pairs, the child was instructed to guess whether the sounds were the *same* or *different* before Ollie indicated the correct answer for the remaining four practice trials.

Test phase. Once the practice trials were complete, Ollie mentions that she has to go home to feed her baby birds and asks the child if they could finish calibrating the computer for her. At this point, the child entered the test phase where they were "in charge" of which button is pressed. In the test trials, children listened to 20 pairs of nonce words and indicated whether they were the *same* or *different* verbally or by pointing to the buttons. After the child gave a response, the experimenter pressed the button that corresponded to the child's response. Responses were recorded by the experimenter in real time on a separate score sheet.

Pallet cleanse and resolution. Halfway through the test trials, there was a brief pallet cleanse to re-engage the children. In the pallet cleanse, a warning message flashes on the computer indicating that the computer is out of battery and Ollie returns briefly to recharge it. Once the remaining test trials are completed, the child enters a resolution phase where Ollie checks to see if the computer is fixed. A green check mark appears on the screen of the computer to indicate that it is fixed and Ollie thanks the child for their help.

3.2.3 Results – Experiment 2

As in Experiment 1, we again find that children are providing similar responses across the SAME and DIFFERENT trials. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 2 below.

Table 2. Proportion *same* responses in Experiment 2

SAME length	DIFFERENT length	*p* value
0.540	0.491	0.4843

We ran a mixed-effect logit model with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect to analyze the data. The model found no main effect of **LENGTH** [$X^2_{(1)} = 0.489$, $p = 0.484$], showing that participants were responding *same* at similar rates across the SAME and DIFFERENT trials.

3.2.4 Discussion – Experiment 2

The lack of a main effect of **LENGTH**, indicated that children were not successful in perceiving the vowel length differences in this experiment showing support for Hypothesis 4. Thus, removing the context did not improve children's sensitivity for vowel length. The results from this study allow us to eliminate the possibility that the context of the child task was causing children to be less particular with vowel length differences than adults due to a lexical bias. By establishing that context is not driving child responses we are able to streamline the comparison between the adult and child studies.

3.3 EXPERIMENT 3: LONG OBSTRUENT

3.3.1 Purpose and Hypotheses – Experiment 3

The purpose of Experiment 3 was to test whether children have a different threshold than adults to show subphonemic use of vowel length. We predict that children will show sensitivity to vowel length with larger differences thus indicating a different threshold. See Hypothesis 5 below.

> *Hypothesis 5*: Children have a larger threshold than adults to show sensitivity to vowel length.

This predicts that children will respond *same* significantly more in the SAME than DIFFERENT length levels.

3.3.2 Method – Experiment 3

Participants.

Twenty children between the ages of 4;4 and 6;11 (MEAN = 5;5.65) recruited remotely across the United States participated in Experiment 3.

Design.

The design was the same as Experiment 1 except the length difference was twice as large. For the SAME level, target-vowels were either both 80ms or both 280ms thus resulting in a 0ms difference across the pair. At the DIFFERENT level, one target-vowel was 80ms and the other was 280ms resulting in a 200ms difference across the pair. To create the stimuli, the tokens from Experiment 1 were manipulated in Praat to acquire the appropriate target-vowel lengths.

Materials and Procedure.

The materials and procedure were identical to Experiment 1 except the pairs with the larger length differences were played instead of the pairs with the length differences in the adult study.

3.3.3 Results – Experiment 3

In contrast to experiments 1 and 2, we find that children are responding significantly differently across the SAME and DIFFERENT trials ($p < 0.0001$). Specifically, children are more likely to provide a same response when the vowel length is the SAME than when it is DIFFERENT. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 3 below.

Table 3. Proportion *same* responses in Experiment 3

SAME length	DIFFERENT length	*p* value
0.865	0.581	<0.0001

We ran a mixed-effect logit model with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect to analyze the data. The model revealed a main effect of **LENGTH** [$X^2_{(1)} = 34.708$, $p < 0.0001$], showing that participants are responding *same* at significantly different rates across the SAME and DIFFERENT trials. The probability of a *same* response in the SAME trials was 0.865 and the probability of a *same* response in the DIFFERENT trials was 0.581 indicating that children were more likely to provide a *same* response when the lengths were the SAME than when they were DIFFERENT.

To determine if there were age-related differences in children's sensitivity we ran three additional models by age (4, 5, and 6 years). All three models were mixed-effect logit models with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect. Interestingly, the models revealed a main effect of **LENGTH** for the 6-year-olds [$X^2_{(1)} = 12.041$, $p < 0.0001$] and 5-year-olds [$X^2_{(1)} = 27.185$, $p < 0.0001$] but not for the 4-year-olds [$X^2_{(1)} = 0.000$, $p = 1.000$] indicating that 5- and 6-year-olds are responding *same* at significantly different rates across the SAME and DIFFERENT trials but 4-year-olds are not.

Results for Experiments 1-3.

To determine why children failed to recognize vowel length differences in Experiment 1, we ran an additional mixed-effect logit model comparing all experiments in the OBSTRUENT environment (1, 2, and 3) with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) and **EXPERIMENT** (1, 2, 3) as fixed effects, and **SUBJECT** as a random effect. The model revealed a main effect of **LENGTH** [$X^2_{(1)} = 35.425$, $p < 0.0001$], showing that participants are responding *same* at significantly different rates across the SAME and DIFFERENT trials. The model did not find a main effect of **EXPERIMENT** [$X^2_{(1)} = 1.202$, $p = 0.548$], showing that participants are responding *same* at similar rates across the experiments. Finally, the model revealed an interaction of **LENGTH** and **EXPERIMENT** [$X^2_{(1)} = 16.779$, $p < 0.0001$] showing significantly different patterns of responses based on the combination of **EXPERIMENT** and **LENGTH**.

In order to better understand what was driving the interaction, we ran pairwise comparisons analyzing each level of **LENGTH** separately. At the SAME length, we find significant differences between Experiment 3 compared to Experiment 1 and 2 but no significant differences between Experiment 1 and Experiment 2. Compared to Experiment 1, participants are significantly more likely to respond *same* on the SAME trials in Experiment 3 ($p = 0.044$) but not in Experiment 2 ($p = 0.410$). There were no significant differences between the experiments at the DIFFERENT level showing that the interaction was driven by differences in the SAME level. The proportion of same responses across the obstruent experiments (1, 2, and 3) at each length is illustrated in Figure 3 below.

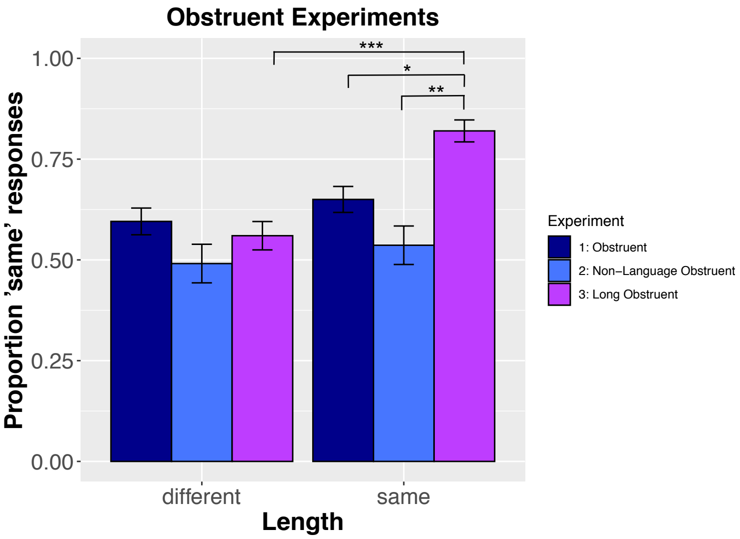

Figure 3. Proportion *same* responses for all experiments in the OBSTRUENT environment.

3.3.4 Discussion – Experiments 1-3

The main effect of **LENGTH** in Experiment 3 indicated that children require larger length differences to show sensitivity providing support for Hypothesis 5. The comparison between experiments 1, 2, and 3 indicated first that context was not driving the differences in responses and second that the larger length differences helped 5- and 6-year-old children succeed. However, even with the large length differences, 4-year-olds were not sensitive to vowel length. A comparison between Experiment 1 and 3 indicated that 100ms is too little for 4- to 6-year-olds and 200ms is enough for 5- to 6-year-olds to show sensitivity. It is unclear whether 4-year-olds might have an even larger threshold. However, 200ms is already longer than what would be found in natural speech so if 4-year-olds did show sensitivity at a length larger than 200ms, it would likely be due to a low-level acoustic strategy and not reflective of their developing phonology.

Interestingly, the different patterns of responses between Experiment 1 and 3 was driven by the SAME level not the DIFFERENT level. Thus suggesting that the increased threshold serves to make the SAME trials more salient. We suspect that the exaggerated vowel lengths at the DIFFERENT level serve to calibrate the children leaving the SAME trials even more salient. However, it's important to mention that this pattern is the opposite of what we found with adults. Recall, in the adult experiment the significant differences between environments were at the larger length levels (the equivalent to the DIFFERENT trials for children). This difference might suggest that children do not treat vowel length subphonemically. But, even if we do find subphonemic use with children at this threshold, the use is not adult-like.

In order to determine whether children are using vowel length subphonemically, they must show differing sensitivity based upon the position within the word. It is possible that their sensitivity is purely phonetic because of the larger differences and thus not influenced by the surrounding sounds. To determine whether their use was subphonemic, Experiment 4 was run as a comparison testing these larger length differences in a neutral environment.

3.4 EXPERIMENT 4: LONG BASELINE

3.4.1 Purpose and Hypotheses – Experiment 4

Experiment 4 was created to provide a comparison for Experiment 3 to determine whether children's sensitivity to vowel length in Experiment 3 is subphonemic or phonetic. This experiment used the same exaggerated vowel length differences from Experiment 3 but used the BASELINE environment instead of OBSTRUENT. If children perform equally well in this experiment as Experiment 3, they do not use vowel length subphonemically because they are not treating the length differently based upon whether it is predictive of another phoneme. On the other hand, if children perform worse in this experiment than in Experiment 3, children are treating vowel length subphonemically but they require a larger threshold to show subphonemic use. We predict that children will perform equally well in this experiment because the larger threshold indicates that their sensitivity is likely phonetic. The hypothesis and prediction for Experiment 4 is below.

Hypothesis 6: Children do not use vowel length subphonemically.

This hypothesis predicts that children will respond *same* at similar rates as Experiment 3 across the SAME and DIFFERENT length levels.

3.4.2 Method – Experiment 4

Participants.

Twenty children between the ages of 4;5 and 6;9 (MEAN = 5;5.2) were recruited remotely from across the United States to participate in Experiment 4.

Design.

The design was the same as Experiment 3 except the final segments were those from the BASELINE environment of the adult study instead of OBSTRUENT. To create the stimuli, the BASELINE tokens from the adult study were manipulated in Praat to acquire the appropriate target-vowel lengths. As with Experiment 3, the SAME length difference level was 0ms and the DIFFERENT length difference level was 200ms.

Materials and Procedure.

The materials and procedure were identical to Experiment 3 except the pairs with the larger length differences in the BASELINE environment were played instead of the pairs in the OBSTRUENT environment.

3.4.3 Results – Experiment 4

As in Experiment 3, we find that children are responding significantly differently across the SAME and DIFFERENT trials ($p < 0.0001$). Specifically, children are more likely to provide a same response when the vowel length is the SAME than when it is DIFFERENT. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 3 below.

Table 3. Proportion *same* responses in Experiment 4

SAME length	DIFFERENT length	*p* value
0.738	0.381	<0.0001

A mixed-effect logit model with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect was run to analyze the data. The model revealed a main effect of **LENGTH** [$X^2_{(1)} = 40.037$, $p < 0.0001$], showing that participants are responding *same* at significantly different rates across the SAME and DIFFERENT trials.

Results for Experiments 3 and 4.

To compare the results from Experiment 3 and 4, we ran another mixed-effect logit model with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) and **EXPERIMENT** (3, 4) as fixed effects, and **SUBJECT** as a random effect. The model revealed a main effect of **LENGTH** [$X^2_{(1)} = 40.768$, $p < 0.0001$], showing that participants are responding *same* at significantly different rates across the SAME and DIFFERENT trials, and a significant (although much smaller) main effect of **EXPERIMENT** [$X^2_{(1)} = 3.863$, $p = 0.049$] showing that participants are responding *same* at slightly significantly different rates across the SAME and DIFFERENT trials. Finally, the model revealed no interaction of **LENGTH** and **EXPERIMENT** [$X^2_{(1)}=0.007$, $p = 0.978$]

showing similar patterns of responses based upon the combination of **EXPERIMENT** and **LENGTH**. The proportion of *same* responses in Experiment 3 and 4 at each length is shown below in Figure 4.

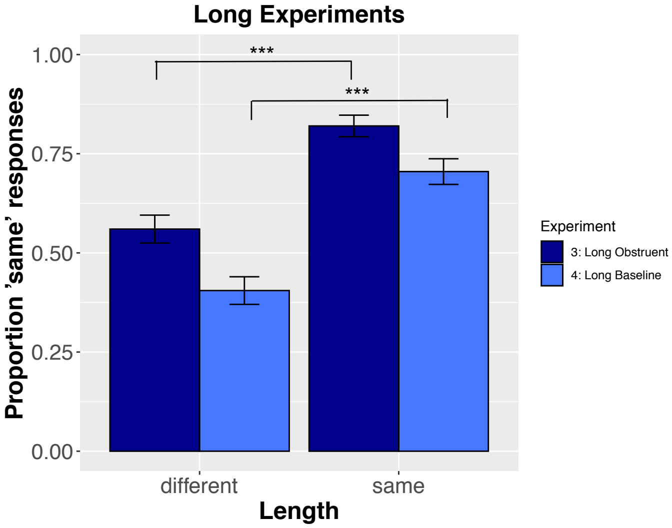

Figure 4. Proportion *same* responses for experiments with exaggerated vowel lengths

Finally, to determine whether there were age-related differences, we ran three additional models divided by age. All three models were mixed-effect logit models with *same* responses as the dependent measure, **LENGTH** (SAME, DIFFERENT) and **EXPERIMENT** (3, 4) as fixed effects, and **SUBJECT** as a random effect. The models revealed a main effect of **LENGTH** for 6-year-olds [$X^2_{(1)} = 29.924$, $p < 0.0001$] and 5-year-olds [$X^2_{(1)} = 16.753$, $p < 0.0001$] showing that 5- and 6-year-olds are responding *same* at significantly different rates across the SAME and DIFFERENT trials but not for the 4-year-olds [$X^2_{(1)} = 2.754$, $p = 0.097$] showing that 4-year-olds are responding *same* at similar different rates across the SAME and DIFFERENT trials. There was no main effect of **EXPERIMENT** and no interaction of **LENGTH** and **EXPERIMENT** for any of the ages indicating similar patterns of responses based upon the combination of **EXPERIMENT** and **LENGTH**.

3.3.4 Discussion – Experiments 3 and 4

The main effect of **LENGTH** in Experiment 4 indicated that children show sensitivity to exaggerated vowel length differences in a neutral position. The main effect of **EXPERIMENT** in the model with both long experiments (3 and 4) showed that there were different rates of *same* responses between experiments 3 and 4. However, the lack of an interaction between **LENGTH** and **EXPERIMENT** indicates that the different rates were not patterning based upon the **LENGTH** showing support for Hypothesis 6. In other words, children were not more or less likely to provide a *same* response based upon the **LENGTH** in one experiment or the other. Because children were not treating vowel length differences differently in these two experiments, they did not show subphonemic use. Instead, their sensitivity to vowel length is likely phonetic because it is not influenced by its predictiveness in determining the phonemic status of another sound.

Additionally, the fact that extreme vowel lengths were required to show any sensitivity indicates that the sensitivity was phonetic and not a reflection of their developing phonology.

Dividing the results by age revealed age-related differences in phonetic sensitivity but not subphonemic use. The lack of an interaction between **LENGTH** and **EXPERIMENT** in all of the age groups indicates that none of the age groups were using vowel length subphonemically as they would need to be treating the length differently based on its position. Interestingly, however, the 4-year-olds also did not show a main effect of **LENGTH** indicating that they were not sensitive to vowel length differences in either experiment. Thus, 5- and 6-year-olds have a phonetic sensitivity to the larger vowel length differences in experiments 3 and 4 but 4-year-olds do not.

4. GENERAL DISCUSSION

Results from the adult and child studies revealed that 4- to 6-year-old children treat vowel length quite differently than adults. Adults showed differential perceptibility of vowel length differences dependent on whether the vowel length was in a position where it could inform the phonemic status of another sound. Children's sensitivity to vowel length differed from adults in three ways: by indicating a different threshold, being phonetic in nature not subphonemic, and by showing differences in responses at the SAME length level. First, the results from Experiment 1 revealed that children were not sensitive to vowel length at the same threshold as adults. Further, children's sensitivity to vowel length differences in Experiment 3 highlighted the higher threshold required for perception for 5- and 6-year-olds. Namely, the 100ms difference that categorized the largest length level in the adult study and the DIFFERENT level in Experiment 1, was too little for child perception. 200ms, however, was large enough for 5- and 6-year-old children to show sensitivity. However, 4-year-olds never showed sensitivity to vowel length even at the 200ms level. Second, the results from experiments 3 and 4 highlighted the nature of children's use of vowel length. Adults' differential treatment of vowel length based upon its phonemic predictiveness revealed that they used vowel length subphonemically before an obstruent. On the contrary, due to children's similar patterns of responses across the two experiments, we can conclude that they were not using vowel length subphonemically. Finally, children showed significantly different responses at the SAME level while adults showed significantly different responses at the higher length differences. This pattern indicates that the vowel length served different purposes in the adult and child studies. For children, the larger length differences served to make the SAME trials more salient. For adults, the subphonemic position served to make the trials at higher length differences more salient.

5- and 6-year-old children's phonetic but not subphonemic sensitivity of vowel length in this study is consistent with work indicating that English-learning infants from 5 to 11 months (Eilers et al., 1984) and 18 months (Mugitani et al., 2009) are able to discriminate vowel length differences. Because the children required a larger threshold and were not treating vowel length differently as a function of its predictive potential, their ability to discriminate differences in experiments 3 and 4 was likely a result of low-level acoustic discrimination not their phonology. The explicit nature of this task builds off of previous findings by showing that 5- and 6-year-olds are able to show explicit discrimination of vowel length (as opposed to implicit in Eilers et al., 1984 and Mugitani et al., 2009). On the contrary, the study revealed that 4-year-olds are not able to demonstrate explicit discrimination of vowel length. However, their lack of explicit sensitivity does not necessarily mean that 4-year-olds lack the perceptual ability to discriminate vowel lengths at this level. We expect 4-year-olds to have the perceptual ability to notice these length differences because even young infants are able to discriminate vowel length (Eilers et al., 1984; Mugitani et al., 2009) but have difficulty applying those skills in an explicit task.

We identify two potential explanations for why 4-year-olds did not show vowel length discrimination in this task. First, being cognitive overload. It is possible that the cognitive demand of the task is too great for 4-year-olds to also discriminate vowel length differences. We suspect by testing 4-year-olds using an implicit measure and thus lowering the cognitive demand of the task will allow them to show sensitivity. The second explanation is that 4-year-old children might have a phonemic bias. Although Experiment 2 demonstrated that the context was not causing a language-bias, it's possible that hearing sounds that displayed the phonotactics of English caused 4-year-olds to consider differences at the phonemic-level. It is quite adaptive for children at this age to pay particular attention to the phonemic level because as they are rapidly expanding their vocabularies, the phonemic level is the proper level to encode. This explanation lies at the intersection of word recognition work showing that children's phonetic representations of familiar words are phonemic (Swingley & Aslin, 2000) but not subphonemic (Swingley & van deer Feest, 2019) and word learning work showing that children's phonetic representations of new words is underspecified (Stager & Werker, 1997).

The results found with children can also help to resolve some of the conflicting results we find when children are tested on their use of vowel length as a subphonemic cue to voicing (in Greenlee, 1980; Wardip-Fruin & Peach, 1984; Krause, 1982) and sensitivity to a subphonemic mismatch when recognizing familiar words (in Swingley & van deer Feest, 2019). By using nonce words that crucially were not mapped to meaning, we were able to isolate the acoustic system and gain a better understanding of how children treat vowel length in the absence of top-down information. Our results indicated that children did not show subphonemic sensitivity to vowel length. It is difficult to reconcile why 3-year-olds but not 6-year-olds in Wardrip-Fruin & Peach (1984) and 6-year-olds but not 3-year-olds in Greenlee (1980) were able to make a voicing distinction based upon vowel length as they are in direct conflict with one another. However, our results suggest that even if children are weighing vowel length as a potential cue to voicing in these previous studies, their use of vowel length is not adult-like. Thus, children's subphonemic use of vowel length is not fully developed by 6 years. Finally, recall Krause (1982) revealed that children required larger vowel length differences to make a voiced decision than adults. Our results are in some ways consistent with this finding but in some ways inconsistent. In Krause (1982), children needed larger vowel lengths to show subphonemic use but in our study children needed larger vowel lengths to show phonetic sensitivity. Together, both suggest that children need larger vowel length differences in general but the nature of their use differs across the studies. We suggest that the differing nature of children's use is a result of the top-down information available in Krause (1982) and lack thereof in our study. Finally, the lack of subphonemic use found in our study is consistent with children's lack of subphonemic sensitivity in a word recognition task (Swingley & van deer Feest, 2019).

5. FUTURE DIRECTIONS

This work lends itself naturally to several directions for future inquiry. One being expanding to other languages. Specifically, it would be of interest to compare speakers of languages that use the VLE as a cue to voicing and languages that do not to determine whether they show different patterns of sensitivity to VL. We would expect languages with the VLE to pattern similarly to English and languages without the VLE to not exhibit differences between

BASELINE and OBSTRUENT as OBSTRUENT is not a subphonemic position for them. Further, we wonder whether speakers will show differing levels of boosted perceptibility in the OBSTRUENT environment dependent on the extent to which their language exhibits and utilizes the VLE. Testing additional languages would also allow us to explore other environments where vowel length is used subphonemically. For instance, we might expect languages that use vowel length to contrast [iə] and [ijə] to show increased sensitivity to the GLIDE environment in a similar way as OBSTRUENT in this study.

Another natural direction for this work is testing children using an implicit measure. Our results showing that children were unable to demonstrate explicit subphonemic sensitivity does not indicate that they are not sensitive at all. By testing children using an implicit measure such as looking time, we would be able to better understand the nature for why children are failing to recognize subphonemic vowel length at the same threshold as adults in this study.

Future work could also seek to explore the influence of exaggerated VLE patterns in children's input on their sensitivity. For instance, it has been demonstrated that mothers exaggerate the VLE by as much as doubling the length of vowels preceding voiced consonants when speaking to children (Ratner & Luberoff, 1984). It is possible that children whose input consists of child-directed speech with exaggerated VLE might show increased sensitivity to subphonemic vowel length.

This work also leaves the question of why adults displayed the lowest sensitivity in the WORD-FINAL environment unanswered. Because VL differences were of the same absolute difference across all environments (ranging from 0 to 100ms) regardless of the base length of the vowel, the ratio of change is lower in the WORD-FINAL environment compared to other environments. In a follow-up study we would make the ratio of change consistent across all environments rather than the absolute difference. This would allow us to determine whether the low sensitivity in WORD-FINAL was due to the logarithmic nature of perception or the WORD-FINAL position itself.

6. CONCLUSIONS

Overall, this study uncovered information about the nature of subphonemic vowel length in English-speaking adults and children. The study with adults demonstrated that adults showed greater sensitivity to vowel length based on whether the vowel length is found in a position where it could be predictive of a phonemic contrast. The study with children demonstrated that subphonemic vowel length in 4- to 6-year-olds is not adult-like. Namely, 5- and 6-year-old children demonstrated a phonetic sensitivity to vowel length differences at a threshold that was twice as large as adults. However, 4-year-olds never demonstrated sensitivity. Together, these results serve to indicate that children have not fully developed their native phonology by 6 years.

References

Agresti, A. (2002). *Categorical data analysis. 2nd edn*. New York: Wiley.

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 2031–2039. https://doi.org/10.1121/1.427949

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187. https://doi.org/10.1016/0010-0277(94)90042-6

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Belasco, S. (1953). The Influence of Force of Articulation of Consonants on Vowel Duration. *The Journal of the Acoustical Society of America*, *25*(5), 1015–1016. https://doi.org/10.1121/1.1907201

Best, C., McRoberts, G., & Sithole, N. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology. Human Perception and Performance*, *14*, 345–360. https://doi.org/10.1037//0096-1523.14.3.345

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.1.09). http://www.praat.com/

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Brock, J., & Nation, K. (2014). The Hardest Butter to Button: Immediate Context Effects in Spoken Word Identification. *Quarterly Journal of Experimental Psychology*, *67*(1), 114–123. https://doi.org/10.1080/17470218.2013.791331

Buder, E. H., & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *The Journal of the Acoustical Society of America*, *111*(4), 1854–1864. https://doi.org/10.1121/1.1463448

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716407070257

Chen, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica*, *22*(3), 129–159. https://doi.org/10.1159/000259312

Chládková, K., Escudero, P., & Lipski, S. C. (2013). Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain and Language*, *126*(3), 243–252. https://doi.org/10.1016/j.bandl.2013.05.020

Chládková, K., & Paillereau, N. (2020). The What and When of Universal Perception: A Review of Early Speech Sound Acquisition. *Language Learning*, *70*(4), 1136–1182. https://doi.org/10.1111/lang.12422

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Syllabic stress. The Journal of the Acoustical Society of America 83(4), 1574-1585.

Denes, P. (1955). Effect of Duration on the Perception of Voicing. *The Journal of the Acoustical Society of America*, *27*(4), 761–764. https://doi.org/10.1121/1.1908020

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*(41), 16027–16031. https://doi.org/10.1073/pnas.0705270104

Drummond, A. (2016). *Ibexfarm* (Version 0.3.7) https://github.com/addrummond/ibexfarm

Dutoit, T., & Pagel, V. (1997). MBROLA Synthesizer (Version 3.02b). FPMs TCTS Lab – Multitel ASBL. https://github.com/numediart/MBROLA

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America*, *75*(4), 1213–1218. https://doi.org/10.1121/1.390773

Elordieta, G., Frota, S., & Vigário, M. (2005). Subjects, objects and intonational phrasing in Spanish and Portuguese. *Studia Linguistica*, *59*(2-3), 110-143.

Fintoft, K. (1961). The Duration of some Norwegian Speech Sounds. *Phonetica*, *7*(1), 19–39. https://doi.org/10.1159/000258096

Fischer-Jorgensen, E. (1964). Sound Duration and Place of Articulation. *STUF - Language Typology and Universals*, *17*(1–6), 175–208. https://doi.org/10.1524/stuf.1964.17.16.175

Flege, J. E., & Port, R. (1981). Cross-Language Phonetic Interference: Arabic to English. *Language and Speech*, *24*(2), 125–146. https://doi.org/10.1177/002383098102400202

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Greenlee, M. (1980). Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception*. *Journal of Child Language*, *7*(3), 459–468. https://doi.org/10.1017/S0305000900002786

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a Voicing Cue. *The Journal of the Acoustical Society of America*, *47*(2B), 613–617. https://doi.org/10.1121/1.1911936

Havy, M., Bertoncini, J., & Nazzi, T. (2011). Word learning and phonetic processing in preschool-age children. *Journal of Experimental Child Psychology*, *108*(1), 25–43. https://doi.org/10.1016/j.jecp.2010.08.002

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022. https://doi.org/10.1121/1.1323463

Hock, H. H. (1999). Finality, prosody, and change. *Proceedings of LP 98*, 15-30.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America*, *67*(5), 1764–1771. https://doi.org/10.1121/1.384304

Hogoboom, A. (2020, Feb 7-8). *Phonological vs. acoustic glides*. [Poster given]. Berkeley Linguistics Society Workshop on Phonological Representations, University of California, Berkeley, CA, United States.

House, A. S. (1961). On Vowel Duration in English. *The Journal of the Acoustical Society of America*, *33*(9), 1174–1178. https://doi.org/10.1121/1.1908941

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113. https://doi.org/10.1121/1.1906982

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4).434– 46. DOI: 10.1016/j.jml.2007.11.007.

Johnson, K., & Martin, J. (2001). Acoustic vowel reduction in Creek: Effects of distinctive length and position in the word. *Phonetica 58*(1-2), 81-102.

Keating, P. A. (1985). Universal phonetics and the organization of grammars, *Phonetic*

Linguistics, Academic Press, 115-132.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117–128. https://doi.org/10.1006/jpho.1997.0069

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, *54*(4), 1102–1104. https://doi.org/10.1121/1.1914322

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. https://doi.org/10.1121/1.380986

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, *41*(1), 59–69. https://doi.org/10.1016/S0167-6393(02)00093-6

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, *16*(2), 153–169. https://doi.org/10.1016/S0095-4470(19)30480-2

Ko, E. (2007). Acquisition of Vowel Duration in Children Speaking American English. *INTERSPEECH*.

Ko, E.-S., Soderstrom, M., & Morgan, J. (2009). Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *J. Acoust. Soc. Am.*, 7.

Krause, S. E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America*, *71*(4), 990–995. https://doi.org/10.1121/1.387580

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the “Critical Period.” *Language Learning and Development*, *1*(3–4), 237–264. https://doi.org/10.1080/15475441.2005.9671948

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Lacerda, F. (1992). Young Infants Prefer High/Low Vowel Contrasts. *PERILUS*, *15*, 85–90.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379–388. https://doi.org/10.1037/h0049038

Lisker, L. (1957). Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English. *Language*, *33*(1), 42–49. https://doi.org/10.2307/410949

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America*, *78*(6), 1949–1957. https://doi.org/10.1121/1.392651

Lunden, A. (2017). Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1).

Maddieson, I., & Gandour, J. (1975). Vowel length before stops of contrasting series. The Journal of the Acoustical Society of America, 58(S1), S61–S61. https://doi.org/10.1121/1.2002224

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. https://doi.org/10.3758/PBR.15.6.1064

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-42. https://doi.org/10.1016/s0010-0277(02)00157-9

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(3), 369–378. https://doi.org/10.1037//0096-1523.14.3.369

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*(6), 457–465. https://doi.org/10.3758/BF03213823

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505–512. https://doi.org/10.3758/BF03208147

Miller, J. L., & Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, *54*(2), 205–210. https://doi.org/10.3758/BF03211757

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental psychology*, *45*(1), 236–247. https://doi.org/10.1037/a0014043

Naeser, M. A. (1970). *The American Child's Acquisition of Differential Vowel Duration. Part 2 of Two Parts* [Microform]. Distributed by ERIC Clearinghouse.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of phonetics, 37*(1), 29-45.

Nakisa, R. C., & Plunkett, K. (1998). Evolution of a Rapidly Learned Representation for Speech. *Language and Cognitive Processes*, *13*(2–3), 105–127. https://doi.org/10.1080/016909698386492

Narayan, C. R. (2020). An acoustic perspective on 45 years of infant speech perception. II. Vowels and suprasegmentals. *Language and Linguistics Compass*, *14*(5), e12369. https://doi.org/10.1111/lnc3.12369

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science (New York, N.y.)*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Paquette-Smith, M., Fecher, N., & Johnson, E. K. (2016). Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, & Psychophysics*, *78*(8), 2329–2340. https://doi.org/10.3758/s13414-016-1186-4

Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703. https://doi.org/10.1121/1.1908183

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, *32*(2), 141–152. https://doi.org/10.3758/BF03204273

Raphael, L. J. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics*, *3*(1), 25–33. https://doi.org/10.1016/S0095-4470(19)31284-7

Raphael, L. J., Dorman, M. F., & Geffner, D. (1980). Voicing-conditioned durational differences in vowels and consonants in the speech of three-and four-year old children. *Journal of Phonetics*, *8*(3), 335–341. https://doi.org/10.1016/S0095-4470(19)31483-4

Ratner, N. B., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics*, *12*(3), 285–289. https://doi.org/10.1016/S0095-4470(19)30875-7

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110. https://doi.org/10.1037/0033-2909.92.1.81

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, *19*(5), 394–398. https://doi.org/10.3758/BF03199398

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1124–1131. https://doi.org/10.1037/0096-1523.7.5.1124

Sato, Y., Ito, Y., & Mazuka, R. (2010). Discrimination of Phonemic Vowel Length by Japanese Infants. *Developmental Psychology*, *46*, 106–119. https://doi.org/10.1037/a0016718

Sharf, D. J. (1962). Duration of Post-Stress Intervocalic Stops and Preceding Vowels. *Language and Speech*, *5*(1), 26–30. https://doi.org/10.1177/002383096200500103

Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, *38*(5), 397–407. https://doi.org/10.3758/BF03207170

Slis, I. H., & Cohen, A. (1969). On the Complex Regulating the Voiced-Voiceless Distinction II. *Language and Speech*, *12*(3), 137–155. https://doi.org/10.1177/002383096901200301

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382. https://doi.org/10.1038/41102

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*(3), 653–659. https://doi.org/10.1121/1.1914578

Stoel-Gammon, C., & Buder, E. H. (1999). Vowel length, post-vocalic voicing and VOT in the speech of two-year-olds. *Proceedings of the XIVth International Congress of Phonetic Sciences, 3*, 2485–2488.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(5), 1074–1095. https://doi.org/10.1037//0096-1523.7.5.1074

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Swingley, D., & van deer Feest, S. V. H. (2019). A Cross-linguistic Examination of Toddlers' Interpretation of Vowel Duration. *Infancy*, *24*(3), 300–317. https://doi.org/10.1111/infa.12280

Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301. https://doi.org/10.3758/s13414-012-0306-z

Trehub, S. E. (1976). The Discrimination of Foreign Speech Contrasts by Infants and Adults. *Child Development*, *47*(2), 466. https://doi.org/10.2307/1128803

Tyler, M. D., & Fenwick, S. E. (2012). Perceptual assimilation of Arabic voiceless fricatives by English monolinguals. *Interspeech 2012: Spoken Language Processing and Biomedicine: 13th Annual Conference of the International Speech Communication Association : September 9-13, 2012, Portland, Oregon*. https://researchdirect.westernsydney.edu.au/islandora/object/uws%3A14814/

Wardrip-Fruin, C., & Peach, S. (1984). Developmental Aspects of the Perception of Acoustic Cues in Determining the Voicing Feature of Final Stop Consonants. *Language and Speech*, *27*(4), 367–379.

Warren, P., & Marslen-Wilson, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics*, *43*(1), 21–30. https://doi.org/10.3758/BF03208969

Warren, R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, *167*(3917), 392–393.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866–1878. https://doi.org/10.1121/1.390988

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In *Language: An invitation to cognitive science, Vol. 1, 2nd ed* (pp. 87–106). The MIT Press.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development*, *52*(1), 349–355. https://doi.org/10.2307/1129249

Werker, J. F, & Lalonde, C. E. (1988). *Cross-Language Speech Perception: Initial Capabilities and Developmental Change*. 12.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, *38*(3), 459–471.

Appendix A

Estimated marginal means and *p*-values for adult study

LENGTH 0

Probability *yes perfect match* response		*p*-value
BASELINE 0.867	WORD-FINAL 0.812	0.3358
BASELINE 0.867	GLIDE 0.832	0.6864
BASELINE 0.867	OBSTRUENT 0.859	0.9942
WORD-FINAL 0.812	GLIDE 0.832	0.9385
WORD-FINAL 0.812	OBSTRUENT 0.859	0.4833
GLIDE 0.832	OBSTRUENT 0.859	0.8298

LENGTH 1

Probability *yes perfect match* response		*p*-value
BASELINE 0.828	WORD-FINAL 0.824	0.9994
BASELINE 0.828	GLIDE 0.808	0.9397
BASELINE 0.828	OBSTRUENT 0.847	0.9328
WORD-FINAL 0.824	GLIDE 0.808	0.9681
WORD-FINAL 0.824	OBSTRUENT 0.847	0.8917
GLIDE 0.808	OBSTRUENT 0.847	0.6462

LENGTH 2

Probability *yes perfect match* response		*p*-value
BASELINE 0.691	WORD-FINAL 0.765	0.2613
BASELINE 0.691	GLIDE 0.793	0.0503
BASELINE 0.691	OBSTRUENT 0.700	0.9973
WORD-FINAL 0.765	GLIDE 0.793	0.8750
WORD-FINAL 0.765	OBSTRUENT 0.700	0.3620
GLIDE 0.793	OBSTRUENT 0.700	0.0824

LENGTH 3

Probability *yes perfect match* response		*p*-value
BASELINE 0.642	WORD-FINAL 0.761	0.0217
BASELINE 0.642	GLIDE 0.583	0.5512
BASELINE 0.642	OBSTRUENT 0.549	0.1671
WORD-FINAL 0.761	GLIDE 0.583	0.0002
WORD-FINAL 0.761	OBSTRUENT 0.549	<.0001
GLIDE 0.583	OBSTRUENT 0.549	0.8794

LENGTH 4

Probability *yes perfect match* response		*p*-value
BASELINE 0.472	WORD-FINAL 0.757	<.0001
BASELINE 0.472	GLIDE 0.476	0.9997
BASELINE 0.472	OBSTRUENT 0.319	0.0040

WORD-FINAL 0.757	GLIDE 0.476	<.0001
WORD-FINAL 0.757	OBSTRUENT 0.319	<.0001
GLIDE 0.476	OBSTRUENT 0.319	0.0029

LENGTH 5

Probability *yes perfect match* response		*p*-value
BASELINE 0.340	WORD-FINAL 0.662	<.0001
BASELINE 0.340	GLIDE 0.361	0.9632
BASELINE 0.340	OBSTRUENT 0.188	0.0009
WORD-FINAL 0.662	GLIDE 0.361	<.0001
WORD-FINAL 0.662	OBSTRUENT 0.188	<.0001
GLIDE 0.361	OBSTRUENT 0.188	0.0001

Appendix B

Script for Experiments 1, 3, and 4

Introduction

Can you see both of the buttons on the top of the screen? What colors are they? So we have two robots we are going to playing with today! This is Robot [*click*] and this is Baby Robot [*click*]. Is that what baby robots look like? So the game we are going to be playing is Baby Robot is practicing his robot noises. And to help him out, Robot is going to say and word and Baby Robot is going to try to copy it exactly as Robot says it. And when Baby Robot says it the exact same we have this green button [*click green button*]. Did you hear that? But sometimes Baby Robot's going to say it a little differently and that why we have this red button [*click red button*] and that's how Baby Robot knows to do better next time. Does that make sense? Now Baby Robot is pretty good so even it he says it a little differently we're going to press the red button okay?

Practice Phase

Modeled practice

Okay [*child's name*] so on these first few I'm going to be in charge of the button and I'll tell you whether Baby Robot said it the same or a little different and it's your job to listen super close. Do you think you could do that? Are you ready for the first one? Go ahead Robot! [*play practice words]* Okay let's see how he did! *[play button]*

> Same: Oh wow Baby Robot said that one the exact same! Great job Baby Robot!
> Different: Not quite, that was a little different. Maybe Baby Robot can do better next time! Baby Robot is pretty good right? Even if it's just a little different the red button will go off

Ready for the next one?

Engaged practice

Okay so on these next few, I'm still going to be in charge of the button and I'll tell you whether Baby Robot said it the same or a little different but I wonder if you could guess before I tell you? Do you think you could do that? Great! Thank you so much! You're going to be so good at this. [*play practice words]* Were those the same or a little different? Okay let's see! *[play button]*

> Correct: Oh wow you're right that was [a little different/the same]! You are such a good listener!
> Incorrect: Almost, that one was [a little different/the same]. Good guessing though!

Ready for the next one?

Test Phase (1-10)

I heard that you're really smart [*child's name*]… is that true? Wow well since you're so smart and I know you're a good listener I wonder if maybe you could be in control of the button instead of me? Do you think you could do that? [*child responds*] Wow! Thank you so much! I have a feeling you're going to be so good at this. Can you practice pointing with your finger towards the green button on the screen? [*play green button as child points*]. Wow! Look at that! Okay now practice pointing to the red button [*play red button as child points*] Good job!

Okay so that's how you're going to press the buttons. Can you remind me, which one are you going to point to if Baby Robot says it the same? That's right! The green button. How about if he says it a little different? You're right again! The red button. You're so good at this! Now remember, Robot is going to say a word and Baby Robot is going to try to copy it exactly as Robot says it. It's your job to point to the green button when he says it the same and the red button if it's a little different. Does that make sense? Let's see how Baby Robot does! Are you ready? Okay go ahead Robot! [*play words*] Was that the same or a little different? [*press button corresponding to child's response*]

Pallet Cleanse

[*click screen to start animation*] Oh no what happened to Baby Robot! Hm… I wonder if he's getting sleepy? What should we do? I wonder if maybe Robot knows how to help him… [*click screen]* Wow it looks like Robot is going to help him get recharged so that he can keep learning! Thank you Robot! It looks like Baby Robot is feeling much better. What do you think? Does he look like he's feeling better?

Test Phase (11-20)

Okay now that baby robot is all recharged and ready to go he wants to keep learning! Do you think you could help him with a few more? Let's see how Baby Robot does! Are you ready? Robot and Baby Robot are you ready? Okay go ahead Robot! [*play words*] Was that the same or a little different?

Resolution

Wow! Baby Robot learned so much today his teacher is going to be so proud. Thank you so much for helping us out today you did an amazing job!

Appendix C

Script for Experiment 2

Introduction

Can you see both of the buttons on the top of the screen? What colors are they? This is my friend Ollie! Ollie the owl! Ollie is super smart. She made this really cool computer. And what the computer does is it copies sounds. It plays two sounds that are the exact same right after each other. But… for some reason it hasn't been working very well lately. And Ollie's been using these two buttons to fix it. So to fix it, the computer is going to play two sounds and if the sounds are the exact same, Ollie is going to press this green button [*click green button*]. Could you hear that? But when it's a little different, she's going to press the red button [*click red button*]. And that's how she's going to fix the computer! Does that make sense?

Practice Phase

Modeled practice

Okay [*child's name*] so on these first few Ollie is going to press the button to fix the computer based on whether it plays sounds that are the same or a little different. It's job to listen super close. Do you think you could do that? Are you ready for the first one? Let's hear the first one [*play practice sounds]* Okay let's see what Ollie says! *[click for Ollie to press button]*

> Same: Oh wow those sounds were the same!
> Different: Oh those sounds were a little different! The computer is pretty good right? So if it's just a little bit different Ollie will press the red button.

Ready for the next one?

Engaged practice

Okay so on these next few, Ollie is still going to press the buttons based on whether the computer plays two sounds that are the same or a little different but I wonder if you could guess before she tells us? Do you think you could do that? [*child responds*] Great! Thank you so much! You're going to be so good at this. [*play practice sounds]* Were those the same or a little different? [child responds] Okay let's see! *[play button]*

> Correct: Oh wow you're right that was [a little different/the same]! You are such a good listener!
> Incorrect: Almost, that one was [a little different/the same]. Good guessing though!

Ready for the next one?

Test Phase (1-10)

Ollie just told me she has to go home and feed her baby birds! I was wondering if maybe you could take over the buttons and finish fixing the computer? Do you think you could do that? [*child responds*] Wow! Thank you so much! I have a feeling you're going to be so good at this. Can you practice pointing with your finger towards the green button on the screen? [*play green* *button as child points*]. Wow! Look at that! Okay now practice pointing to the red button [*play red button as child points*] Good job!

Okay so that's how you're going to press the buttons. Can you remind me, which one are you going to point to if they are the same? That's right! The green button. How about if they're a little different? You're right again! The red button. You're so good at this! Now remember, the computer is going to play two sounds and it's your job to point to the green button when they are the same and the red button if they are a little different. Are you ready? Okay listen close! [*play sounds*] Was that the same or a little different? [*press button corresponding to child's response*]

Pallet Cleanse

[*click screen to start animation*] Oh no… Did you see that [child's name]? What happened to the computer! Hm… What should we do? Let's see if maybe Ollie knows [*click screen]* Wow I guess it was just running low on battery! Thank you Ollie! She's so good with computers.

Test Phase(11-20)

Okay now that the computer is all recharged, do you think you could help fixing it with a few more? Thank you so much for your help! Are you ready? Listen close! [*play sounds*] Was that the same or a little different?

Resolution

Wow! Thank you so much for your help! Let's call Ollie over and see if the computer is fixed. [*click screen*] Wow! Look at that! You fixed it. You did such a great job!