Infants’ Perception of Non-Contrastive Variation

Dena Krieger

Master of Arts

Department of Psychology
University of Toronto

2012

Abstract

Newborns were once viewed as having the ability to distinguish between all possible speech contrasts present in the world’s languages. However, recent work suggests that this view is an oversimplification of young infants’ perceptual abilities. This study used the Stimulus Alternation Preference Procedure (SAPP) to examine 6- and 10- month-old infants’ ability to discriminate between voiced and voiceless stops (lexically contrastive in English) as well as voiced and pre-voiced stops (not lexically contrastive in English). Six-month-olds distinguished between voiced and voiceless stops, but not between voiced and pre-voiced stops. Ten-month-olds failed to discriminate either contrast. We tentatively conclude that 1) pre-voicing may be a subtle contrast requiring experience to perceive, and 2) the SAPP might not be an ideal method to examine discrimination abilities in 10-month-olds.
Acknowledgments

I would like to thank my supervisor, Dr. Elizabeth Johnson, for her infinite patience, enthusiasm, and good humor, and for all of her guidance in designing and running these experiments. Thank you to Dr. Marieke van Heugten, Melissa Paquette-Smith, and Huiwen Goy, for all of their help and advice in this project, for helping me in the lab, and for their wonderful company. Thank you to all of the members of the Infant Language and Speech Lab for their help scheduling and running participants. Thank you to all of the families that took the time out of their busy schedules to come to our lab, and Naomi Krajden, for recording my stimuli. Finally, thank you to my friends and family for their support and encouragement throughout this project.
# Table of Contents

Acknowledgments........................................................................................................ iii

Table of Contents ........................................................................................................ iv

1 Introduction ............................................................................................................... 1
   1.1 Reexamining the Universal Listener Hypothesis .............................................. 1
   1.2 Infants’ Discrimination of Non-Contrastive Variation .................................... 3
   1.3 Voicing .............................................................................................................. 5

2 Experiment 1 ............................................................................................................. 7
   2.1 Method ............................................................................................................. 8
      2.1.1 Participants .............................................................................................. 8
      2.1.2 Stimuli .................................................................................................... 8
      2.1.3 Procedure .............................................................................................. 10
      2.1.4 Design ................................................................................................... 12
   2.2 Experiment 1a: Results ................................................................................... 13
   2.3 Experiment 1b: Results ................................................................................... 13
   2.4 Discussion ....................................................................................................... 14

3 Experiment 2 ........................................................................................................... 14
   3.1 Method ........................................................................................................... 15
      3.1.1 Participants ............................................................................................ 15
      3.1.2 Stimuli .................................................................................................. 15
      3.1.3 Procedure ............................................................................................ 16
      3.1.4 Design .................................................................................................. 16
   3.2 Experiment 2a: Results ................................................................................. 16
   3.3 Experiment 2b: Results ................................................................................. 16
   3.4 Discussion ....................................................................................................... 16

4 General Discussion ................................................................................................ 17

References..................................................................................................................... 23

Figures.......................................................................................................................... 28

Appendix...................................................................................................................... 34
List of Figures

Figure 1. Sample stimuli used in Experiment 1a and 2a. As indicated by the red box, the ‘ba’ syllable on the left begins with a pre-voiced stop (negative VOT). The ‘ba’ syllable on the right lacks pre-voicing (zero VOT). The latter syllable was created by splicing the pre-voicing off of the syllable on the left. In other words, the voiced and pre-voiced stimuli were acoustically identical in all respects besides the presence or absence of pre-voicing.

Figure 2. Sample stimuli used in Experiment 1b and 2b. As indicated by the red box, the ‘pa’ syllable on the left contains an aspirated portion that makes this token have positive VOT. The ‘ba’ syllable on the right lacks this aspiration, which makes it a zero VOT token. The latter syllable was created by splicing the aspirated portion off of the syllable on the left. In other words, the voiced and pre-voiced stimuli were acoustically identical in all respects besides the presence or absence of aspiration.

Figure 3. The Results of Experiment 1a. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.

Figure 4. The Results of Experiment 1b. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.

Figure 5. The results of Experiment 2a. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.

Figure 6. The results of Experiment 2b. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.
Appendices

Appendix- Adult Rating Results
1 Introduction

A longstanding view in the field of infant speech perception has been that young infants enter the world as universal listeners, who can distinguish between all speech sounds that exist in human language. Over time, infants gradually learn to perceive speech in a language-specific manner, attending only to their language’s phonemes, or the speech sounds that are lexically meaningful in their native language. By 6 months, infants show evidence of language-specific perception of vowels (Kuhl, 1979), and by 8 to 10 months, they demonstrate language-specific perception of consonants (Werker & Tees, 1984). In the current study, I examine English-learning infants’ perception of pre-voiced, voiced, and voiceless stops. My findings contribute to a growing body of work questioning traditional views of early speech perception. That is, my findings suggest that although infants perceive many speech contrasts at birth, acquiring the phoneme inventory of one’s native language likely requires much more than simply learning to ignore those contrasts that are not lexically distinctive in the native language.

1.1 Reexamining the Universal Listener Hypothesis

The universal listener hypothesis is based on well-established findings that young infants can discriminate between many non-native speech sounds that adults cannot. For example, 6- to 8-month-old infants acquiring Canadian English discriminate between dental and retroflex voiceless stop consonants, which are contrastive in Hindi but not in English (Werker & Tees, 1984). Older infants and English speaking adults, however, do not distinguish this contrast. Likewise, young infants distinguish between many other non-native contrasts that older infants do not discriminate (e.g., Tsao, Liu, & Kuhl, 2006; Mattock, Molnar, Polka, & Burnham, 2008).
These findings showcase infants’ remarkable perceptual abilities, and contribute to our understanding of how infants learn native language categories.

However, the universal listener hypothesis paints an incomplete picture of infants’ perceptual abilities, as research suggests that infants continue to differentiate their native language’s speech sounds as their language becomes more familiar. For example, American English-learning infants’ ability to distinguish between [ra] and [la] improves over time, with English-learning 10- to 12- month-olds being better able to distinguish the contrast than their 6- to 8-month-old counterparts (Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006). This suggests that although infants are able to perceive many contrasts early in life, infants’ perceptions of these contrasts may require experience to reach adult-like levels of competency in the discrimination of native language speech sounds.

Also, a growing body of research suggests that there are some speech contrasts that young infants cannot perceive until after they have acquired some experience with their native language. For example, [na] and [ŋa] are contrastive sounds in Filipino, but not in English. Six- to 8-month-old infants, regardless of whether they were exposed to English or Filipino, were not able to distinguish this contrast, despite showing evidence of distinguishing between [ma] and [na] (Narayan, Beddor & Werker, 2010). English-learning 10- to 12-month-olds were also not able to distinguish between [na] and [ŋa], but Filipino infants of the same age were able to distinguish this contrast (Narayan, Beddor & Werker, 2010). This finding suggests that over the course of the first year, Filipino infants become sensitive to this contrast. This pattern has been shown with some English contrasts as well. Twelve- to 14-month-old infants discriminate English speech sounds such as [as] and [az], as well as [at] and [ad], but their 1-3-month-old and 6-8-month-old counterparts do not (Eilers, Wilson, & Moore, 1971). Some consonant contrasts
in English, such as the [fi]- [0i] contrast, have been shown to be difficult to discriminate, even for infants as old as 12 months (Eilers et al., 1977). Thus, these findings provide evidence that some subtle contrasts may have to be learned, and are not merely maintained on the basis of a pre-existing sensitivity.

Taken together, the evidence reported in this section demonstrates that newborn infants may be universal listeners, but in a different sense than researchers originally supposed. It appears that young infants are able to discriminate the same contrasts, regardless of language input, but are not sensitive to all contrasts that exist in human language. First, research has shown that infants’ ability to discriminate native speech sounds improves over time. Second, it appears that infants require some experience with their native language to be able to detect some native language contrasts. Therefore, learning of native language sounds does not only occur through a process of loss and maintenance, as infants require some familiarity with their language input in order to be able to perceive native speech sound contrasts as adults do.

1.2 Infants’ Discrimination of Non-Contrastive Variation

This revised view of infants’ perceptual abilities has implications for how infants cope with variability in the speech signal. Although each language has a set of distinct speech sounds, the way each sound is pronounced can differ significantly depending on various factors, including the placement of a sound in a word or sentence, speech rate, and a speaker’s emphasis and tone. These kinds of variations in pronunciation are not lexically contrastive, and are known as allophones. For example, English speaker’s pronunciation of [l] differs dependent on whether it is at the beginning of a word (known as a ‘light L’) or at the end of a word (known as a ‘dark L’). Despite this variation, adults do not have trouble interpreting the speech signal. Our extensive experience with our native language has affected how we perceive speech sounds,
making this variation relatively easy to ignore (Whalen, Irwin & Best, 1997). However, given that infants’ perceptual abilities are qualitatively different from adults, it is unclear how this variation affects how they perceive the speech signal.

There is evidence, for example, that very salient non-contrastive variation can hinder infants’ ability to perceive speech sounds in a language-specific manner. For example, vowel length is not lexically contrastive in English, but is in Dutch. Thus, words that differ in vowel length could be different words in Dutch, but not in English. However, in a word classification task, English-learning 7.5- to 12-month-old infants considered long and short vowels to be different speech sounds (Deitrich, 2006). Canadian English infants’ performance on this task suggests that infants not only perceive vowel length, but they use this information when classifying different vowel sounds. Although perceiving vowel length may be a potentially useful cue for English-learning infants, as it can be indicative of lexical stress and therefore word boundaries in English (Redford, Davis, & Milkkulainen, 2004), infants actually treat this speech cue as a different speech sound category. Canadian-English infants have been shown to classify longer vowels as distinct from shorter vowels until sometime between 12 and 18 months of age (Deitrich, Swingley, & Werker, 2007; Deitrich, 2006).

However, although this work shows that infants may misclassify non-contrastive variation into their second year, it is unclear how infants perceive and classify other kinds of non-contrastive variation present in the speech signal. Vowel length is a particularly salient speech sound, and is one that English-speaking adults have no trouble perceiving (Deitrich et al., 2007). Other speech sound contrasts that are allophonic in English but phonemic in other languages are subtler than vowel length distinctions. Thus, it is important to understand if infants perceive more subtle differences in pronunciation. If English-learning infants are able to perceive
them, then they may misclassify them, as they do with vowel length distinctions. Examination of
how other non-contrastive sounds are perceived will clarify the effect that these variants have on
infants’ ability to determine native language categories. This issue can be addressed by testing
discrimination of acoustically distinct speech sounds that are present in the signal, but map onto
the same speech sound category in English. One such contrast is the distinction between negative
VOT and zero VOT stop consonants, or pre-voiced and voiced stops.

1.3 Voicing

Testing infants’ sensitivity to voicing contrasts can potentially provide information about
infant’s perceptual abilities, as voicing contrasts are highly variable in the speech signal and
realized differently across languages (Cho & Ladefoged, 1999; Lisker & Abramson, 1967;
Lisker & Abramson, 1964). Voicing refers to the timing of the movement of the vocal cords in
relation to the initial release of air in stop consonant sounds, and occurs on a continuum. English
speakers produce distinct sounds along this continuum, using zero voice onset time (VOT; where
the initial release of air beginning the consonant and the onset of voicing approximately co-
occur- as in the case of [ba] and [da]) and positive VOT (where the release of air occurs before
the vocal chords begin to vibrate- as in the case of [pa] and [ta]).

Pre-voiced stops, or stops with negative VOT, occur when the vocal chords begin to
vibrate before the initial stop release (Lisker & Abramson, 1967). Adult English speakers can
perceive the acoustic difference between negative VOT and zero VOT stops, but will classify
both as the speech sound [b] when they hear either one within a word (Burnham et al., 1991;
Zlatin, 1974). Also, English speakers produce consonants with negative VOT, and although there
are some regularities as to when it is produced, it is not consistent (Lisker & Abramson, 1967).
Negative VOT seems to be dependent on the speaker, with some speakers producing negative
VOT stops quite regularly, and some speakers only producing zero VOT stops (Lisker & Abramson, 1967). When speakers enunciate their stop consonant, negative VOT stops are produced more frequently, but they can also occur in fluent speech (Lisker & Abramson, 1967). Thus, negative VOT stops are present in the speech signal in English, but are not lexically contrastive or consistently produced.

Although English speakers do not classify negative VOT and zero VOT stops differently, speakers of other languages do. Like English, languages including Dutch, Spanish, and Hungarian have two speech sound categories for stop consonants, [b] and [p]. However, speakers of these languages pronounce these consonants with more negative VOT than English speakers for the same phonemes. In fact, some languages distinguish [b] from [p] based on the presence of negative VOT, as [b] is pre-voiced, and [p] is realized with voice onset times that are close to zero (van Alphen & Smits, 2004; Ladefoged, 2001; Lisker & Abramsen, 1964). Also, there are other languages that use three-way contrasts, such as Thai, which has stops with negative VOT, zero VOT, and positive VOT (Cho & Ladefoged, 1999; Lisker & Abramson, 1964). Therefore, as the distinction between negative and zero VOT stops are contrastive in other languages, this contrast must be reasonably salient, and infants may be able to perceive it when it is present in the speech signal.

Studies have in fact examined if infants discriminate the negative to zero VOT contrast, but results are inconclusive (Aslin, Pisoni, Hennessy, & Perey, 1981; Eilers, Gavin & Wilson, 1979; Lasky, Sydral-Lasky & Klien, 1975). Some evidence suggests that Spanish-learning 6- to 8-month-olds are able to distinguish between negative VOT and zero VOT stops (Eilers et al., 1979), while other research has reported that Spanish-learning 4- to 6.5-month-old infants are not able to distinguish between slightly negative and slightly positive VOT stops (Lasky et al.,
In a study that examined the perceptual abilities of 6- to 12-month-old English-learning infants, it was found that they are able to distinguish the negative to zero VOT contrast, but are better able to distinguish between the zero to positive VOT contrast (Aslin et al., 1981). Regardless of whether the distinction between negative VOT and zero VOT is lexically contrastive in the infant’s native language, results do not conclusively show that infants attend to this contrast.

Previous studies on infant discrimination of this contrast have used synthesized speech with fixed VOT as their test stimuli. As the focus of this study is to clarify how English-learning infants perceive negative VOT that is present in their input, actual realizations of negative VOT stops by an English speaker make up the stimuli in this study. Natural speech, and its inherent variation, may be a more representative method of examining how infants perceive this contrast as it occurs in an English-learning infant’s language input, and is more appropriate for this study.

2 Experiment 1

In Experiment la, 6-month-old infants are tested on their ability to distinguish between a bilabial stop consonant with negative VOT (a pre-voiced stop) and a bilabial stop consonant with zero VOT (a voiced stop consonant), which is a non-contrastive distinction in English. English speakers perceive both of these speech sounds as [b] (Zlatin, 1974). In Experiment 1b, a different group of infants will be tested on their ability to distinguish between a bilabial stop consonant with positive VOT and aspiration (a voiceless stop consonant) and a bilabial stop consonant with zero VOT (a voiced stop consonant), which is a contrastive distinction in English. The first stimulus in the contrastive set in Experiment 1b corresponds to the English [p], whereas the second corresponds to the English [b].
This study will help determine how infants who are 6 months of age, who have not fully mastered which consonant contrasts are meaningful in their language, perceive one-category and two-category contrasts. Infants will hear test trials that either present both members of the contrast, or only one (Best & Jones, 1998). If infants are able to distinguish the sound pair, they should show a consistent preference for one type of test trial over the course of the experiment. If infants are not able to distinguish these sounds, there should be no consistent preference for either type of trial during the testing session.

2.1 Method

2.1.1 Participants

Thirty-eight 6-month-olds (5 months 15 days to 6 months 28 days; 18 females) participated in this study. Sixteen infants participated in Experiment 1a, and 22 infants participated in Experiment 1b. Prior to scheduling the appointment, parents of all participants were asked a series of questions to gauge their infant’s exposure to other languages and other variants of English. All infants who participated had at least 90% English input, and at least one parent who learned English in Canada before the age of five. All infants that participated in the study had not had an ear infection in at least a month, and were more than 38 weeks gestation at birth. An additional nine infants were tested, but were excluded from analyses due to fussiness (n=5), experimenter error (n=3), and failure to meet language requirements (n=1).

2.1.2 Stimuli

For Experiment 1a, Six CV tokens of syllables containing a bilabial voiced stop with negative VOT followed by a low back vowel (“ba”) were collected from a native English speaker who sometimes produces stops with negative VOT. Tokens were either presented in their full form (negative VOT), or with the negative VOT portion spliced off using Praat.
Without the negative VOT portion, the voicing and the stop release coincide, making the token have zero VOT (as in van Alphen & McQueen, 2006). An example of one such token, both in its full and manipulated form, is shown in Figure 1. The negative VOT portion that occurred before the stop release had an average duration of 165.8 msec (range: 121-212 msec) across the six tokens. In order to maintain perceptual uniformity in list presentation rate, the vocalic portions of the syllables were all separated by 750 msec. Therefore, tokens with negative VOT had a shorter inter-stimulus interval than zero VOT tokens. Two tokens from the six were used in the test phase, and the remaining four were used in the habituation phase. For the Alternating test trials, the same token was repeated, and alternated between its full and manipulated form.

In Experiment 1b, six syllables consisting of bilabial voiceless stop consonants followed by a low back vowel (“pa”) were collected from the same native English speaker. For the positive VOT tokens, the tokens were kept in their full form. Zero VOT tokens were made by first determining the stop release, aspirated portion, and vowel onset of each token. The aspirated portion of each token was edited out using the speech software Praat, leaving the stop release and the vowel. An example of a token, both in its full and manipulated form, is shown in Figure 2. The aspirated portion of the syllable was 116.83 msec (range: 105-130 msec) across the six tokens. As the speaker was asked to enunciate, the aspiration portion is longer in duration than found in previous literature (Lisker & Abramson, 1964). As in experiment 1a, 750 msec of silence preceded the onset of the vowel, making the inter-stimulus interval shorter for positive VOT tokens than for zero VOT tokens. Two tokens from the six were used in the test phase, and the remaining four were used in the habituation phase. For the Alternating test trials, the same token was repeated, and alternated between its full and manipulated form.
In order to ensure that the contrast in Experiment 1a mapped onto the same English speech sound, and the contrast in Experiment 1b mapped onto different speech sounds, adults classified each speech sound used in this experiment. Ten English-speaking adults were presented with the 24 tokens used in these experiments. The details and results of this experiment are presented in the Appendix.

2.1.3 Procedure

Participants were tested on their ability to discriminate between two separate sets of stimuli: Alternating trials, that were a set of alternating minimal pairs (e.g., zero and negative VOT), and Non-Alternating trials, that were a set of the same type of token as the participant heard in habituation (e.g., either zero or negative VOT). Phoneme discrimination was determined in a modified version of the Stimulus Alternation Preference Procedure (SAPP; Best & Jones, 1988; Bertoncini, Nazzi, Cabrera, & Lorenzi, 2010; Houston, Horn, Qi, Ting, & Gao, 2007; Maye, Werker, & Gerken, 2002). This method operates under the assumption that infants exhibit differential looking times towards a series of alternating speech sounds versus a series of the same speech sound. This procedure has been found to provide a valid measure of speech sound discrimination when examining the average looking times of a group of infants, as well as showing an individual infant’s ability to discriminate a contrast when his or her performance in previous trials are held constant (Houston et al., 2007). In this study, the procedure began with a habituation phase involving one member in the contrast pair (as in Houston et al., 2007). This habituation procedure was included so that infants, if able to hear the contrast, would be more likely to prefer the Alternating trials, as these trials would be both a novel and a more variable set of speech sounds than Non-Alternating trials.
In Experiment 1a, participants are habituated to either negative VOT tokens or zero VOT tokens. In Experiment 1b, participants are habituated to either zero VOT or positive VOT tokens. In this habituation phase, infant participants are seated on their parent’s lap in a sound attenuated booth. This booth contains three lights, one situated in front of where the infant is sitting, and two on either side. Above the two side lights sit two speakers. The experiment starts when the centre light begins to blink. Once the infant orientates towards this light, the centre light stops blinking and one of the side lights begins to blink. When this side light starts to blink, the habituation stimuli are played from both speakers in the booth, and play for the duration of the sound file (59 seconds in both experiments). In order to keep the habituation phase constant across participants, the sound file plays after the initial orientation to the centre light, regardless of the infants’ looking behaviour. Once the infant orientates toward the flashing light and then looks away for 2 consecutive seconds, the side light stops flashing and the centre light begins to flash again, and the sequence will repeat for the duration of the habituation phase. The infant directs the light activity in habituation so that they learn they can direct the lights with their orientation behaviour.

After habituation, participants immediately begin the test phase, and are tested on their ability to discriminate tokens with two types of test trials: Non-Alternating trials and Alternating trials. In one type of test trial, participants are presented with different tokens that are the same stimulus as habituation (Non-Alternating trials). For example, participants that heard negative VOT stops in habituation would hear negative VOT stops in these trials. In another test trial, participants are presented with an alternation between both members of the contrast (Alternating trials). Previous studies have used this paradigm to assess discrimination of different contrasts and words, and have found that infants have a preference for the Alternating trials (Best & Jones, 1998; Houston et al., 2007; Maye, Werker & Gerken, 2002; Bentocini et al., 2010). However, a
consistent preference for one type of trial over the other depends on the infants’ ability to detect the differences in the test trials, and therefore any consistent differences in looking times towards the different trial types is indicative of phoneme discrimination.

Infants were presented with three test blocks. Each block consisted of one Alternating test trial and one Non-Alternating test trial. The order of these two trials was randomized within each block. The test phase begins once the infant orients his or her gaze towards the centre light. When this occurs, the centre light turns off and one of the side lights begins to blink. The Alternating or Non-Alternating set of stimuli then begins to play once the infant orients towards the flashing side light. The stimuli set plays and the light continues to blink until the infant orient away from the light for more than 2 seconds or until the infant has heard all sixteen tokens (in both experiments, approximately 20 seconds). Therefore, unlike the habituation phase, the sound stops when the infant turns away, and his or her looking time can be used as an indication of interest in the test stimuli. When the trial finishes, the side light and sound file will stop, and the centre light will begin to flash again and a new trial will begin. Parents listened to masking music over noise-cancelling headphones throughout both the habituation and test phase so that they could not unintentionally bias the direction of their child’s gaze. Parents were told that their infants were participating in a study that aimed to determine if infants’ perception of speech sounds were tied to their native language.

2.1.4 Design

Participants in Experiment 1a are placed in one of two conditions: the negative VOT condition or the zero VOT condition. That is, they are habituated to either one minute of tokens with negative VOT or the same tokens with zero VOT. During the test phase, participants hear test trials that have an alternation in voicing, or the same type of tokens as their habituation,
dependent on condition. In Experiment 1a, those in the negative VOT condition hear trials with negative VOT tokens and trials that alternated between zero and negative VOT tokens. Participants in the zero VOT condition hear trials that consisted of zero VOT tokens and trials that alternated between negative and zero VOT tokens. This pattern is the same for Experiment 1b, but with positive VOT and zero VOT tokens. In order to help infants identify the alternating test trials, the novel token (for example, the negative VOT token for those in the zero VOT condition, and vice versa) is presented first. Therefore, the alternating test trials have the same stimuli across conditions, but the order in which the stimuli are presented was reversed for the different conditions.

2.2 Experiment 1a: Results

In Experiment 1a, 10 out of 16 infants listened longer to the Alternating trials. As it was unclear how long the novelty effect would last in this procedure if obtained, each block was analyzed separately for differences in looking times towards different trial types. Although many studies find an overall effect, some studies only find effects in the first block of trials (Bertoncini, Nazzi, Cabrera, & Lorenzi, 2010). These results are shown in Figure 3. Results show that there were no significant differences in looking times between the Alternating and Non-Alternating trials in the first ($M_{alt} = 8.205$, $SD = 5.528$; $M_{nonalt} = 8.777$, $SD = 5.272$; $t(15) = 0.412$, $p = .686$), second ($M_{alt} = 7.271$, $SD = 4.011$; $M_{nonalt} = 6.477$, $SD = 3.643$; $t(15) = -0.841$, $p = .413$), or third block ($M_{alt} = 7.851$, $SD = 4.266$; $M_{nonalt} = 7.632$, $SD = 5.078$; $t(15) = -0.156$, $p = .878$).

2.3 Experiment 1b: Results

In Experiment 1b, 6-month-old infants were tested in the same manner as Experiment 1a. Thirteen out of 22 infants looked longer towards the Alternating trials. Results from Experiment
1b can be seen in Figure 4. Results show that in the first block of trials, six-month-olds showed a consistent preference for the Alternating trials in the first block, ($M_{alt} = 13.328, SD = 5.918$; $M_{nonalt} = 10.876, SD = 5.503$; $t(21) = -2.114, p = .047$) but this effect disappeared in subsequent test blocks (second block: $M_{alt} = 9.278, SD = 5.383$; $M_{nonalt} = 9.208, SD = 4.892$; $t(21) = -0.049, p = .962$; third block: $M_{alt} = 8.733, SD = 5.715$; $M_{nonalt} = 8.205, SD = 5.457$; $t(21) = -0.328, p = .746$). Thus, in the first block, 6-month-old infants showed a preference for listening to trials that alternated between positive VOT and zero VOT tokens over trials that consisted of only one of these token types.

2.4 Discussion

These data suggest that 6-month-olds do not discriminate bilabial stops with negative VOT from those with zero VOT (Experiment 1a), but do discriminate between bilabial stops with zero VOT from those with positive VOT (Experiment 1b). Results from Experiment 1b suggest that the results from Experiment 1a were not due to an inappropriate testing method for phoneme discrimination in 6-month-olds. These findings suggest that infants who are exposed to languages where negative VOT and zero VOT stops are both lexically contrastive may learn to perceive this contrast through language experience, like in other subtle contrasts (Narayan et al., 2011). It may be possible, however, that the procedure is not sufficiently sensitive to test discrimination of subtle contrasts.

3 Experiment 2

Ten-month-old infants, unlike 6-month-old infants, have developed native language speech sound categories and thus discriminate contrasts according to the sounds of their language (Kuhl et al., 2006; Werker & Tees, 1984). In English, stop consonants with negative VOT belong to the same speech sound category as zero VOT stops, but 10-month-old English-
learning infants would likely be exposed to some stop consonants with negative VOT from certain speakers or speech contexts. Do 10-month-olds, who have learned the speech sound categories in their language, nonetheless perceive negative VOT stops as distinct from zero VOT stops? Perhaps incidental and inconsistent exposure to stop consonants with negative VOT would prompt older infants to distinguish a negative VOT stop consonant from a zero VOT stop consonant. If the exposure they have received through the speech signal is enough to make them attend to the negative VOT, they will be able to discriminate the speech contrast. Thus, in Experiment 2a, 10-month-old infants were tested on their ability to discriminate between negative and zero VOT stop consonants, and in Experiment 2b, a different group of 10-month-old infants were tested on their ability to discriminate between aspirated bilabial stops with positive VOT and stops with zero VOT.

3.1 Method

3.1.1 Participants

Thirty-two 10-month-old infants (10 months, 1 day to 10 months 30 days; 12 females) participated in this study. Sixteen infants participated in Experiment 2a, and 16 participated in Experiment 2b. Participants were recruited in the same manner as Experiment 1. All participants were born after at least 37 weeks of gestation. Eight more participants were tested but were excluded from analysis due to fussiness (n=5), experimenter error (n=3), and being older than 11 months (n=1).

3.1.2 Stimuli

The stimuli in Experiment 2a were the same as Experiment 1a, and the stimuli in Experiment 2b were the same as Experiment 1b.
3.1.3 Procedure

The procedure in this study was the same as Experiment 1.

3.1.4 Design

The design of this study was the same as Experiment 1.

3.2 Experiment 2a: Results

Eleven out of 16 infants listened longer to the Non-Alternating trials. The results are shown in Figure 5. Results show that there were no significant differences in looking times between the Alternating and Non-Alternating trials in the first ($M_{alt} = 8.814, SD = 6.079; M_{nonalt} = 7.449, SD = 5.136; t(15) = -0.672, p = .512$), second ($M_{alt} = 7.496, SD = 6.631; M_{nonalt} = 8.863, SD = 5.779; t(15) = 0.987, p = .340$), or third block ($M_{alt} = 7.334, SD = 6.454; M_{nonalt} = 6.746, SD = 6.075; t(15) = -0.247, p = .808$).

3.3 Experiment 2b: Results

In Experiment 2b, 10-month-old infants were exposed to the same sets of stimuli in Experiment 1b. Eleven out of 16 infants looked longer to Non-Alternating trials. The results of Experiment 2b are shown in Figure 6. Results show that there were no significant differences in looking times between the Alternating and Non-Alternating trials in the first ($M_{alt} = 8.099, SD = 5.004; M_{nonalt} = 7.994, SD = 5.283; t(15) = -0.058, p = .955$), second ($M_{alt} = 7.520, SD = 4.758; M_{nonalt} = 6.126, SD = 4.413; t(15) = -0.674, p = .510$), or third block ($M_{alt} = 6.181, SD = 5.225; M_{nonalt} = 6.995, SD = 5.388; t(15) = 0.442, p = .665$).

3.4 Discussion

Given that 10-month-olds did not consistently prefer one trial over another in both Experiment 2a and 2b, the results from Experiment 2 are inconclusive. Ten-month-olds, with
some experience in their native language, should be able to discriminate between contrastive sounds in their native language, like the contrast in Experiment 2b, and have been shown to distinguish this contrast in from birth in past research (e.g., Best & Jones, 1998; Eimas, Siqueland, Jusczyk, Vigorito, 1971). As the method in Experiment 2b did not yield any consistent differences, it is unclear whether the results of Experiment 2a are due to the contrast or the method. Therefore, these results suggest that the testing paradigm may not be appropriate for infants who are 10 months of age, and the results from Experiment 2 are inconclusive.

4 General Discussion

In this study, I examined how infants perceive naturally occurring non-contrastive variation in the speech signal. Specifically, I tested how infants who are 6 and 10 months of age can perceive the distinction between pre-voiced stop consonants (which have negative VOT) and voiced stop consonants (with zero VOT). In English, negative VOT stops are not lexically contrastive from zero VOT stops, and English speakers classify these stops in the same way, as the English sound [b]. Six-month-old infants, who do not yet know the consonant categories in their native language, were not able to distinguish between pre-voiced and voiced stops, but were able to distinguish between voiced and voiceless stops (or positive VOT and zero VOT stops). These findings suggest that the contrast between pre-voiced and voiced stops may be too acoustically subtle for 6-month-olds to perceive. Ten-month-olds, however, failed to show any evidence of discriminating either contrast, suggesting that the SAPP was not an ideal method to test phoneme discrimination in infants of this age, as discussed in detail below.

The results of Experiment 1a, that 6-month-old English-learning infants are not able to distinguish between stops with negative VOT and zero VOT, are consistent with past research that show young infants do not perceive this contrast (Lasky et al., 1975; Eilers et al., 1979).
These results suggest that English-learning infants do not perceive this contrast when it occurs in the speech signal, and thus young infants would not misclassify the members of this contrast as two different sounds. Although the results of this study only provide insight into how infants perceive one allophone in English, it may be that the pervasive variation that we can observe when analyzing the speech signal may not necessarily represent how infants perceive it. If they are insensitive to even some this variation, the speech signal may be easier to organize than it appears.

Young infants who are learning languages where the negative to zero VOT contrast is meaningful may only discriminate this contrast after some experience with their native language. For example, if 6- and 10-month-old Dutch infants were tested in these experiments, it may be that the older Dutch infants would discriminate the negative to zero VOT contrast, but the younger infants would not. The findings in this study, combined with others, show that 6-month-old infants are not sensitive to some speech sound contrasts (Narayan et al., 2010; Eilers et al., 1977). Therefore, this study provides further evidence that infants are not universal listeners in the classic sense. Rather, they may perceive speech sounds based on human audition, with some contrasts being more salient than others. The current study adds to the growing body of evidence suggesting that there are contrasts in many languages that are so subtle that humans do not initially perceive or attend to them. Future research will hopefully address how these speech sounds are learned.

In Experiment 2, 10-month-olds did not show evidence of being able to distinguish negative VOT stops from stops with zero VOT. However, 10-month-olds also failed to discriminate between aspirated stops with positive VOT, and non-aspirated stops with zero VOT, a phonemic contrast in English that infants of this age group have been shown to perceive in
previous research (e.g., Aslin et al., 1981). This pattern of results suggests that the variant of the SAPP used in this study may be inappropriate for 10-month-old infants.

The SAPP has been used successfully with 6- to 8-month-old infants (Houston, Horn, Qi, Ting, & Gao, 2007; Maye, Werker, & Gerkin, 2002), but little work has used this procedure with older infants (Archer & Curtin, 2011; Mattock et al., 2008). One study examined if 9-month-olds could distinguish between legal and illegal consonant clusters, which differed in both place and voicing, and thus was a more salient contrast (Acher & Curtin, 2011). Although 8- to 10-month-olds exhibited consistent preferences for the positive to zero VOT contrast using this paradigm, the habituation phase was very short, and was directed by the infants, rather than being consistent across infants (Best & Jones, 1998). Indeed, research on 11-month-olds using the SAPP did not yield any consistent preference for trial types in words that differ in different native language vowels (Phan & Houston, 2012). Therefore, the specific method used in this study may not have been ideal to test phoneme discrimination in infants older than 8 months.

An important issue that this study highlights is that we should not assume that when one method is appropriate for one age group, it is also for another. When testing infants of different age groups on the same task, it is important to have a control task like Experiment 1b and 2b, so we can determine that the results that we obtain are due to the contrast, rather than the method used. For example, in a study that tested 6- and 9-month-olds on lexical tone using SAPP, it was found that 6-month-olds discriminated sounds based on lexical tone, and 9-month-olds did not (Mattock et al., 2008). Without showing that 9-month-olds were able to respond consistently to another contrast, these results may not necessarily show that 9-month-olds are insensitive to lexical tone, as the method may have been inappropriate for this age group. Since the results of Experiment 2 and other studies show that infants older than 8 months may not show evidence of
discrimination in this procedure, older infants’ perceptual abilities should be confirmed by checking the method’s appropriateness.

As the results from Experiment 2 were inconclusive, future research will examine how 10-month-olds classify stops produced with negative VOT. A task where infants are able to employ classification categories may directly test whether non-contrastive variation affects infants. For example, even if 10-month-olds could perceive the negative to zero VOT contrast, it may be that they classify both as zero VOT stops. Therefore, it is of interest to see how these non-contrastive speech sounds are classified. Ten-month-old infants are currently being tested in a segmentation task where infants will hear a few sentences of fluent speech with a word presented many times. Then, in the test phase, this word will be presented in isolation. Previous work using this task has suggested that if infants are able to recognize this word from the speech stream, they will be more interested in this word than a new word that did not appear in the speech stream (Jusczyk & Aslin, 1995). The target words will either have zero VOT initial stops, as they did in the speech stream, or negative VOT initial stops. Infants are known to be sensitive to mispronunciations in segmentation, and may not recognize the word if the initial consonant has negative VOT rather than zero VOT (Jusczyk & Aslin, 1995). If this is the case, it may be that 10-month-olds misclassify negative VOT contrasts, as they no longer recognize the word from the speech stream. As discrimination patterns do not provide insight about how infants classify different speech sounds, classification tasks are crucial to understanding how infants cope with non-contrastive variation.

This study attempted to clarify how English-learning infants perceive voiced versus pre-voiced consonants, a contrast that is sometimes present in the speech input they receive. For this study, I recorded tokens from an English speaker who produces negative VOT stops
inconsistently. Therefore, these tokens may not necessarily be representative of how speakers realize negative VOT stops when it is a lexically contrastive cue. Although the duration of the negative VOT portions in Experiment 1a and 2a were the same duration as contrastive VOT tokens in other languages, there may be other elements to producing negative VOT stops when they are contrastive (van Alphen & Smits, 2004). English speakers realizations of negative VOT may have been too subtle for infants at 6 and 10 months of age to perceive. Perhaps if the same groups of infants were tested on distinguishing between negative VOT and zero VOT stops made by a Dutch speaker, for whom this contrast is lexically meaningful, infants from one or both age groups might be able to distinguish this contrast. Another possible direction is to see if infants with different language backgrounds perceive English pre-voicing and lexically contrastive pre-voicing differently.

Experiment 1b and 2b both yielded surprising results, which may have been due to how speech materials were produced. Six-month-old infants showed differentiation of voiced and voiceless consonants, but only in the first block, and 10-month-olds did not show evidence of perceiving this contrast in any of the three test blocks. This contrast has been well documented in the literature, with both newborns and 10-month-old infants displaying sensitivity to this contrast (Best & Jones, 1998; Eimas et al., 1971). Also, when adults were asked to rate these speech sounds, they rated the zero VOT tokens that were made for these experiments as “ba” in 70% of all trials. The findings obtained in this experiment may reflect that the zero to positive VOT contrast in this experiment was not as discriminable as when this contrast is produced naturally. Therefore, it may have been difficult to discriminate, resulting in a weak effect for the 6-month-olds, and no effect for the 10-month-olds.
Also, the habituation procedure, meant to drive infants towards preferring the Alternating test trials, may have been ineffective in motivating infants towards a consistent preference. It may be that if infants were just exposed to the test phase, they would be more interested in the stimuli, which would have prompted more consistent results. Another issue may have been that habituation was consistent over all infants, and perhaps if habituation was measured, rather than assumed, more consistent results would have been obtained. There are documented age differences in the time required for habituation, and the habituation procedure used in this study may have been somewhat appropriate for 6-month-olds, but not for 10-month-olds (Hunter & Ames, 1988).

The current study suggests that 6-month-olds discriminate positive VOT stops from zero VOT stops, two contrastive sounds that research has shown infants can easily discriminate (Eimas et al., 1971), but are not sensitive to the distinction between negative VOT stop consonants from zero VOT stop consonants. This study also suggests, based on the 10-month-old’s performance on the salient contrast that the testing method used may not have been appropriate for testing phoneme discrimination with this age group. This study also indicates that what may be an effective testing tool for one age group may not necessarily be appropriate for another, and studies examining different age groups should ensure methodological validity for all age groups whenever possible. Finally, if 6-month-olds are not able to distinguish between negative VOT and zero VOT stops, then they will not perceive the distinction in natural speech. Therefore, for the English-learning infant, their language input may be easier to organize than appears when considering analysis of the speech signal.
References


Figure 1. Sample stimuli used in Experiment 1a and 2a. As indicated by the red box, the ‘ba’ syllable on the left begins with a pre-voiced stop (negative VOT). The ‘ba’ syllable on the right lacks pre-voicing (zero VOT). The latter syllable was created by splicing the pre-voicing off of the syllable on the left. In other words, the voiced and pre-voiced stimuli were acoustically identical in all respects besides the presence or absence of pre-voicing.
Figure 2. Sample stimuli used in Experiment 1b and 2b. As indicated by the red box, the ‘pa’ syllable on the left contains an aspirated portion that makes this token have positive VOT. The ‘ba’ syllable on the right lacks this aspiration, which makes it a zero VOT token. The latter syllable was created by splicing the aspirated portion off of the syllable on the left. In other words, the voiced and pre-voiced stimuli were acoustically identical in all respects besides the presence or absence of aspiration.
Figure 3. The Results of Experiment 1a. Looking time differences (sec) for Alternating (striped) versus Non-Alternating trials (white) in the three test blocks.
Figure 4. The Results of Experiment 1b. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.
Figure 5. The results of Experiment 2a. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.
Figure 6. The results of Experiment 2b. Looking time differences (sec) for Alternating (striped) versus Non-Alternating (white) trials in the three test blocks.
Appendix

Adult Rating Results

Ten English-speaking adults (age range: 20-29) who all learned English before the age of five were asked to classify sounds as “ba”, “pa”, or neither syllable. Participants completed 24 trials (all tokens in Experiments 1 and 2) where one of each kind of token was presented randomly in six blocks. Negative VOT tokens from Experiment 1a and 2a were rated as sounding like “ba” in all but one trial across all participants, and Positive VOT tokens from Experiment 1b and 2b were rated as sounding like “pa” in all trials across all participants. The zero VOT tokens that were made from splicing off the negative VOT portion and used in Experiment 1a were rated as sounding like “ba” in all but one trial across all participants. Those with zero VOT that were made from splicing out the aspiration and used in Experiment 1b were rated as “ba” as 70% of all trials, as “pa” in 11.8% of trials, and as neither syllable in the remaining 18.2% of trials.