



PAPER

Accommodating variability in voice and foreign accent: flexibility of early word representations

Rachel Schmale¹ and Amanda Seidl²

1. Department of Psychological Sciences, Purdue University, USA

2. Department of Speech, Language and Hearing Sciences, Purdue University, USA

Abstract

In six experiments with English-learning infants, we examined the effects of variability in voice and foreign accent on word recognition. We found that 9-month-old infants successfully recognized words when two native English talkers with dissimilar voices produced test and familiarization items (Experiment 1). When the domain of variability was shifted to include variability in voice as well as in accent, 13-, but not 9-month-olds, recognized a word produced across talkers when only one had a Spanish accent (Experiments 2 and 3). Nine-month-olds accommodated some variability in accent by recognizing words when the same Spanish-accented talker produced familiarization and test items (Experiment 4). However, 13-, but not 9-month-olds, could do so when test and familiarization items were produced by two distinct Spanish-accented talkers (Experiments 5 and 6). These findings suggest that, although monolingual 9-month-olds have abstract phonological representations, these representations may not be flexible enough to accommodate the modifications found in foreign-accented speech.

Introduction

Before becoming sophisticated language learners, infants must master the process of word recognition. In word recognition, infants must first find a word in fluent speech and recognize it as distinct from other similar-sounding words. This preliminary step to acquiring a language may itself seem arduous; however, this task is further complicated by the inherent variability of spoken language. In any language, words vary in surface form depending on a talker's voice (e.g. age and gender), speaking rate, emotional state or affect (e.g. happy, angry), accent, social class, dialect and idiosyncratic pronunciations (Abercrombie, 1967). Swingley and Aslin (2000) suggest that this variation is 'a non-trivial problem because distinct instances of the same word can have very different acoustic manifestations' (p. 148). Thus, a central aspect of word recognition is contending with variable acoustic forms, in that infants must learn to cope with different types and degrees of surface-form variability and to recognize when variability is relevant to the task and when it should be ignored.

Singh (2008) suggests that it may be difficult for infants to learn what variation is linguistically relevant (e.g. the difference between *cut* and *cup*) and what variation is not linguistically relevant (e.g. the difference between the production of *cup* by a male and by a female talker). Unfortunately, categorizing sounds in a linguistically relevant way is complicated by the fact that infants encode

and store fine acoustic-phonetic details. Thus, infants' representations include talker-specific information as well as information about linguistic distinctions such as phonemes (e.g. Houston & Jusczyk, 2000; Singh, 2008). When processing speech, this attention to fine phonetic detail often works to infants' detriment by causing them to attend to linguistically irrelevant information (e.g. voice, affect). A further obstacle is that infants learn about the features of a language by hearing varying types of speech across different types of talkers and are thus exposed to a considerable amount of both irrelevant and relevant information. In a series of six experiments we examined how variability in voice and foreign accent impact the ability of English-learning infants to represent, recognize, and generalize words.

Traditionally, a non-native accent is thought to hinder language processing. A large body of research has demonstrated that acoustic variability can confuse native listeners (Lane, 1963; Munro & Derwig, 1995) and that it adversely affects the recognition and learning of words (Goh, 2005; Mullennix, Pisoni & Martin, 1989; Nygaard, Sommers & Pisoni, 1994) and the recognition of vowels (Verbrugge, Strange, Shankweiler & Edman, 1976), and reduces accuracy and slows processing speed in word recollection tasks (Craik & Kirsner, 1974; Goldinger, Pisoni & Logan, 1991; Martin, Mullenix, Pisoni & Summers, 1989). With regard to foreign-accented speech, Clarke and Garret (2004) suggest that non-native phonological rules or alternations apply to both segments and suprasegments,

and this on top of subphonemic variation may provide perceptual difficulty for listeners unfamiliar with the accent.¹ However, the view that variability in foreign accent is an impediment to language processing is derived from research on adults and does not necessarily address the potential effects of such speech on inexperienced language learners (infants). This is an especially relevant difference because we know that infants and adults do not process acoustic cues in the input in the same way (e.g. Werker & Tees, 1984). It therefore remains unclear to what extent variability in foreign accent affects infants' abilities to segment, recognize, generalize and subsequently learn words.

Some evidence suggests that variability in the input may *not* be an impediment to language processing for infants. For example, it has been shown that infants exhibit some ability to deal with talker changes (Jusczyk, Pisoni & Mullenix, 1992b; Kuhl, 1979, 1983). In her investigations, Kuhl (1979, 1983) demonstrated that, despite substantial talker variability across both age and gender, 6-month-olds discriminate a phonetic contrast between different vowels. This finding suggests that infants are capable of ignoring talker variability at a very young age. Jusczyk *et al.* (1992b) extended the findings of Kuhl by demonstrating that 2-month-olds are capable of coping with talker variability by detecting a syllable change when multiple talkers produce the same stimulus. However, infants could only do so when the test phase immediately followed the familiarization phase of this experiment. Infants failed to detect such changes when a 2-minute delay was introduced. In fact, even when only one talker produced different tokens of a syllable, infants did not detect a change after a delay. Thus, infants seem to be equipped with the basic ability to cope with variability in the input; however, when the variability is high, both within a talker and between multiple talkers, infants may fail to retain information about the phonetic structure of speech sounds. This seems to indicate a certain vulnerability and lack of abstraction in their early phonemic representations. Although these studies provide some insight into infants' early representations of speech sounds in discrimination tasks, less is known about infants' early representations as they begin to recognize words in continuous speech.

Before infants can link sound patterns to a particular meaning, they must first be able to discover those sound patterns in fluent speech. In a series of seminal studies using the Headturn Preference Procedure (HPP), Jusczyk and Aslin (1995) demonstrated that infants of 7.5 months of age successfully recognize words in continuous speech, whereas infants of 6 months of age do not. (Note that this is not the case when target words are flanked by familiar words – then word recognition is shown at 6

months; Bortfeld, Morgan, Golinkoff & Rathbun, 2005.) Jusczyk and Aslin (1995) familiarized infants with two words in isolation (e.g. *dog, cup*) and then tested the infants' listening preferences for passages containing familiar (e.g. *dog, cup*) or unfamiliar (e.g. *feet, bike*) words. The 7.5-month-olds demonstrated a listening preference for passages containing familiar over unfamiliar words, and were not fooled when presented with words that were phonologically similar to the familiar words but varied in a place feature (e.g. *bawg, tup*). These findings suggest that infants at this age are able to find and recognize words heard in familiarization, but do not accommodate phonemic variation. Furthermore, later studies of older infants' word-learning ability revealed a similar pattern, whereby infants are attentive to phonemic distinctions (Swingley & Aslin, 2000, 2002; although see Stager & Werker, 1997 and Fennell & Werker, 2003 for evidence for phoneme confusion in 14-month-olds). This sort of sensitivity may work to their detriment in perceiving words spoken in different accents or even dialects, provided that phonemic variation receives the same attention as subphonemic variation in young infants. Thus, despite their clear sensitivity to phonemic changes, it is still unknown whether infants accommodate the sort of subphonemic and suprasegmental variation found in different voices as well as the greater subphonemic and suprasegmental variation found in different accents.

Houston and Jusczyk (2000) examined the effects of voice variability on infant word recognition. They found that attention to phonetic cues sometimes works to infants' detriment, blocking their ability to extract abstract phonological structure in words spoken by multiple talkers. Although in Houston and Jusczyk (2000) 7.5-month-olds were found to attend longer to passages with familiar over unfamiliar words across two female or two male talkers, they were unable to generalize a word across one male and one female talker. Thus, the indexical differences of words produced across different genders (very different voices with large pitch differences owing in part to different vocal tract sizes) are too dissimilar for infants at this age to generalize. It is not until 10.5 months of age that infants can accommodate such talker variability by generalizing word tokens across male and female talkers.

Adult listeners can most often understand speakers with different voices and non-native accents by accommodating extraneous differences in both subphonemic and suprasegmental structure. One way that adults adapt to the variability of foreign accents is through top-down processing, which promotes lexical feedback and generalization (Norris, McQueen & Cutler, 2003). However, such adaptation may be difficult for young infants, given that they process speech differently, exhibit difficulty with talker variability (Houston & Jusczyk, 2000; Jusczyk *et al.*, 1992b), and are *overly* attentive to phonemic detail (Jusczyk & Aslin, 1995; Swingley & Aslin, 2000, 2002). Thus, because a non-native accent modifies many surface-level features of words for both consonants and vowels

¹ Subphonemic variation occurs at a level lower than the segment. For example, although both English and Spanish vowel inventories contain the phoneme [i], the F1 of Spanish [i] is higher than that of English.

(e.g. the length of voice onset time, VOT), such variability might be an impediment to word recognition. Furthermore, given that children weight acoustic cues differently from adults (e.g. Mayo & Turk, 2004), they may have more trouble dealing with foreign-accented speech. For example, because infants process speech more globally than adults (Seidl & Cristià, 2008), they might be more likely to be confused by slight subphonemic or suprasegmental variation (Singh, 2008; Singh, Morgan & White, 2004), whereas adults, who may be able to process acoustic cues more independently, may not be distracted by such variation. Thus, in order to approximate the processing abilities of adults, infants must be able to normalize across instances of words that vary suprasegmentally and subphonemically.

The question then becomes: which cues are modified in accented speech? Here, we focused on Spanish-accented speech. We chose to examine the effects of a Spanish accent, as opposed to some other accent, because the vowels and consonants present in Spanish are also present in English (even though some of these sounds may be contrastive/phonemic in one and non-contrastive/allophonic in the other). Given these similarities we expected that accommodating this accent might provide fewer challenges than accommodating an accent produced by a speaker whose language has a more distinct phonological inventory. In addition, because we eventually hope to test bilingual infants with these same stimuli, Spanish was an appropriate choice as Spanish-accented English is frequently produced in the immediate region.

Characteristics of Spanish-accented English

In a series of studies that compared Spanish-accented and native English vowels, Jongman and Wade (2007) found that non-native speakers exhibit a higher degree of variability in their use of vowel space compared with native speakers. Bradlow (1995) illustrated the differences in vowel space and inventory between Spanish and English vowel systems, which are probably responsible for the high degree of overlapping of adjacent vowel categories between Spanish and English vowel systems that is reported in the productions of speakers of Spanish-accented English (Jongman & Wade, 2007). Furthermore, multiple systematic temporal acoustic differences between Spanish-accented English and native American English have been found to influence native American English adults' perception of accentedness of non-native speech (Shah, 2004). For example, Shah (2004) reported that some of the subphonemic characteristics of the productions of speakers of Spanish-accented English include shorter VOT and longer flap and closure durations. In addition, suprasegmental distinctions were found: despite the fact that Spanish has lexical stress, there are no significant temporal distinctions between stressed and unstressed vowels, although Spanish-accented English has longer syllable duration. This may be a result of the facts that Spanish is most often called a syllable-timed

language as opposed to a stress-timed language like English, and that Spanish does not have vowel reduction in unstressed syllables (possibly resulting in longer unstressed vowels). However, it is important to note that other suprasegmental cues probably exist in force, especially given the absence of vowel reduction in Spanish. For example, as a result of its distinct metrical typology, lexical stress may have a different realization in Spanish than in English (e.g. more reliance on pitch and amplitude than on syllable duration). Most importantly, in Shah (2004) the acoustic differences found at the subphonemic level positively correlated with adults' perception of accentedness, and are also likely to influence infants' perceptions of subphonemic variability in tests of word recognition with Spanish-accented English.

The present research therefore examines how speaker differences in voice and accent impact word recognition in English-learning 9- and 13-month-olds. Specifically, how do English-learning infants cope with voice variability, accent variability, and voice *and* foreign accent variability? Furthermore, does experience with the native language minimize the impact of voice and accent variability?

Experiment 1

Houston and Jusczyk (2000) demonstrated that 7.5-month-olds cannot generalize a word across one male and one female talker owing to the significant differences of words produced across different genders (very different voices). Although these infants were able to recognize words across two female talkers, it is possible that the female talkers used in their experiment had highly similar voices, which facilitated token generalization. Thus, in this experiment, we used the HPP (Jusczyk & Aslin, 1995; Kemler Nelson *et al.*, 1995) to investigate whether English-learning 9-month-olds can recognize words when there is a close, but inexact, match between tokens in test and familiarization owing to large voice differences between talkers. We tested infants on their abilities to generalize words produced by two different talkers in Midwestern American English who were evaluated through talker similarity ratings as having highly dissimilar voices. We predicted that, if there was a large overlap in the productions of the tokens by the speakers, the infants, when familiarized with words produced by one English talker and tested on passages containing familiar or unfamiliar words by a different, highly dissimilar English talker, would listen significantly longer to passages containing familiar words than to passages containing unfamiliar words.

Method

Participants

Twenty-four monolingual English-learning 9-month-olds were tested (15 males). The infants had a mean age of

36 weeks, 0 days (range = 34 weeks, 3 days to 38 weeks, 0 days). Eleven additional infants were tested but were excluded for the following reasons: fussing or crying (7), sickness (1), at least 30% exposure to a language other than English (2), and parental interference (1). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Talker similarity ratings

It is unclear which acoustic dimensions are the most important in talker recognition (Gelfer, 1993; Houston, 2000). Thus, acoustic measurements may not be the best way to assess the perceptual similarity of two different talkers. Therefore, in order to select the female talkers to produce the stimuli for this experiment (as well as for the five subsequent experiments), we obtained talker similarity ratings among 10 talkers by obtaining adult perceptual similarity ratings. Eighteen undergraduate volunteers rated the perceptual similarity of 10 female talkers (five speakers of Midwestern English from the Northern Indiana or Southern Illinois region and five native speakers of Spanish from the Dominican Republic, Peru, and Colombia who spoke fluent Spanish-accented English).

The raters listened to two isolated words spoken by each of the 10 talkers. These words were presented both in natural speech and in sinewave speech. Voice and accent are not orthogonal domains and cannot be judged independently in natural speech (Remez, Fellowes & Rubin, 1997; Remez, Van Dyk, Fellowes & Rubin, 1998). Thus, in order to obtain ratings of voice similarity of talker pairs without the possibility that accent characteristics might interfere with the raters' judgments, the raters listened to all talker pairs in sinewave speech, as well as in natural speech. Sinewave speech is advantageous because it has been shown to eliminate natural voice quality while preserving idiosyncratic phonetic variation (Remez *et al.*, 1998). To produce sinewave speech, the natural speech sample is replicated by three sinusoids that 'imitate the course-grain spectrotemporal properties of a speech signal' (Remez *et al.*, 1997, p. 654). Although sinewave speech does not maintain the original prosodic characteristics of the speech, it 'preserves much of the complex temporal and prosodic information found in speech within a similar range of frequencies without retaining the distinctly biological quality of true speech' (Krentz & Corina, 2008, p. 2). Thus, sinewave speech maintains accentual characteristics, but *not* voice ones. Natural speech has both voice and accent characteristics, so in order to ascertain whether judgments were based on voice, we simply subtracted the sinewave ratings from the natural speech ratings. Therefore, two talkers were determined to have *similar* voices if there was little-to-no change between the mean squared distances in sinewave and in natural speech (e.g. .53 in natural speech, .54 in sinewave speech). By contrast, two talkers were determined

to have *dissimilar* voices if there was a large change between the mean squared distances in sinewave and in natural speech (e.g. 11.03 in natural speech, 6.08 in sinewave speech).

Ratings were obtained using an experimental program in PRAAT 4.4.30 (Boersma & Weenik, 2001). The raters judged the perceptual similarity (on a scale of 1 to 5) of each pair of talkers based on two tokens of two different target words (*kingdom* and *raptor*). Each talker was paired with each of the other nine talkers (as well as with their selves) once for each target word, and each word pair was separated by 500 milliseconds of silence. Raters were randomly assigned to one of two rating conditions. Each condition consisted of trials that were blocked by target word, accent group type (e.g. talker pairs in Midwestern English, talker pairs in Spanish-accented English, talker pairs across Midwestern English and Spanish-accented English), and speech form (e.g. natural speech or sinewave speech; see Table 1 for rater groups). The blocks were randomly ordered and talker pairs were counterbalanced.

Once we had obtained the raters' talker similarity ratings on all possible talker pairs in both natural and sinewave speech (both within and across accent groups), we conducted a multidimensional scaling analysis (ALSCAL, using SPSS) on the data. This analysis provided matrices of the perceptual dissimilarity (e.g. mean squared distance or Euclidean distance) between each pair of talkers. Multidimensional scaling (MDS) assembles a geometric representation of the data along a three-dimensional space by using matrices of relational data (e.g. proximities, distances, correlations and similarities; Houston, 2000). Thus, MDS detects significant underlying dimensions about a set of variables (e.g. speakers) by analysing a matrix of observed dissimilarities between the variables and configuring them to best represent the observed

Table 1 Order of trials in conditions from talker similarity ratings

	Word	Accent group	Speech form	
Condition 1	Kingdom	English	Natural	
	Kingdom	English/Spanish-accent	Sinewave	
	Kingdom	Spanish-accent	Sinewave	
	Kingdom	English/Spanish-accent	Natural	
	Kingdom	Spanish-accent	Natural	
	Kingdom	English	Sinewave	
	Raptor	English	Sinewave	
	Raptor	Spanish-accent	Natural	
	Raptor	English/Spanish-accent	Natural	
	Raptor	Spanish-accent	Sinewave	
	Raptor	English/Spanish-accent	Natural	
	Raptor	English	Natural	
	Condition 2	Raptor	English	Natural
		Raptor	English/Spanish-accent	Sinewave
Raptor		Spanish-accent	Sinewave	
Raptor		English/Spanish-accent	Natural	
Raptor		Spanish-accent	Natural	
Raptor		English	Sinewave	
Kingdom		English	Sinewave	
Kingdom		Spanish-accent	Natural	
Kingdom		English/Spanish-accent	Natural	
Kingdom		Spanish-accent	Sinewave	
Kingdom		English/Spanish-accent	Sinewave	
Kingdom		English	Natural	

Table 2 Passages with target words. Adapted from Jusczyk et al. (1999)**Candle Passage**

The candle in the kitchen was almost melted. So Annie bought another candle at the stationary store. She came home and put away the old candle. Fran gave that candle to you later. Then she made a place for the new big candle. Your candle is very pretty and smells nice too.

Hamlet Passage

Your hamlet lies just over the hill. Far away from here near the sea is an old hamlet. People from the hamlet like to fish. Another hamlet is in the country. People from that hamlet really like to farm. They grow so much that theirs is a very big hamlet.

Kingdom Passage

Your kingdom is in a faraway place. The prince used to sail to that kingdom when he came home from school. One day he saw a ghost in this old kingdom. The kingdom started to worry him. So he went to another kingdom. Now in the big kingdom he is happy.

Raptor Passage

The raptor saw you the other day. He's much younger than the old raptor. I think your raptor is very nice. He showed another raptor your pretty picture. That raptor thought you grew a lot. Maybe someday you'll be a big raptor.

distances. These observed distances are reported through the mean squared distance (i.e. average dissimilarity) of each talker pair, as determined by the adult talker similarity ratings. MDS analyses have been used in several experiments to determine judgments of talker similarity for the purposes of identifying how talker variability affects speech perception (Goldinger, 1996; Houston, 2000; Houston & Jusczyk, 2000; Wannemacher, 1995).

Although adult ratings consistently map onto acoustic measurements of speech (Remez et al. 1997, 1998, 2004), it is unknown which specific acoustic properties the adults attended to when conducting talker similarity ratings. It is also unclear whether infants focus on the same acoustic dimensions as adults in perceiving acoustic differences between talkers. Furthermore, because the raters were all native speakers of English, they may have paid attention to different acoustic characteristics when rating the similarity between the English speakers and the Spanish-accented speakers. Nonetheless, adult ratings may be the best approximations of perceptual similarity between speakers, even if adult ratings do not equal infant ratings. In order to ensure that perceptual similarity bore some relationship to acoustic similarity, we also conducted acoustic measurements. The next section discusses these results and attempts to relate them to the perceptual measures.

Stimuli

The stimuli used in this experiment were adapted from the stimuli used in Experiment 1 of Jusczyk, Houston and Newsome (1999), in which the researchers used four trochaic bisyllabic target words: *hamlet*, *doctor*, *kingdom* and *candle*. For the purposes of the present research, *doctor* was replaced with *raptor* in order to avoid the possibility that infants were familiar with the word *doctor*. The target word lists consisted of 14 different repetitions of each word, which were separated by 500 ms of silence. In testing, there was one passage for each of the target words, in which the target word appeared six times (once per sentence). To control for the role of the location of the words within the sentences (Seidl & Johnson, 2006), the target words appeared twice in sentence-initial position, twice in sentence-medial position, and twice in

sentence-final position. In addition, each target word was preceded by six different words within each passage: *your*, *that*, *old*, *the*, *another* and *big* (see Table 2 for passages).

Two native speakers of English from the Midwest produced the target word lists and passages. The first English talker (hereby referred to as E1) was from the Indianapolis, Indiana region and had a deep, raspy voice. The second English talker (hereby referred to as E2) was from the Chicago, Illinois region and had a soft, breathy, childlike voice. These talkers were selected for this experiment based on the dissimilarity ratings provided by the MDS matrices of E1 and E2. These two talkers had a mean squared distance of 11.03 for natural speech and of 6.08 for sinewave speech, indicating that their voices were *dissimilar* because there was a significant change in the mean squared distance between natural and sinewave speech (see Table 3 for proximity matrices of mean

Table 3 Mean squared distances (squared Euclidean distances) of all talker pairs in sinewave and natural speech
Proximity Matrix: Natural Speech

	E1	E2	E3	E4	E5	S1	S2	S3	S4	S5
E1	0.00									
E2	11.03	0.00								
E3	1.89	4.05	0.00							
E4	1.36	6.74	0.46	0.00						
E5	0.47	6.97	0.50	0.58	0.00					
S1	5.88	14.66	5.31	2.84	5.62	0.00				
S2	3.17	5.12	0.68	0.38	1.65	2.65	0.00			
S3	0.53	6.75	0.44	0.54	0.002	5.58	1.56	0.00		
S4	14.00	7.58	7.15	6.64	10.60	5.31	3.91	10.35	0.00	
S5	7.09	8.93	3.81	2.47	5.42	1.11	1.29	5.29	1.70	0.00

Proximity Matrix: Sinewave Speech

	E1	E2	E3	E4	E5	S1	S2	S3	S4	S5
E1	0.00									
E2	6.08	0.00								
E3	1.05	2.09	0.00							
E4	3.31	0.47	0.64	0.00						
E5	3.40	1.73	1.34	1.39	0.00					
S1	9.66	2.79	5.17	2.83	8.01	0.00				
S2	1.56	7.90	2.57	4.57	7.51	7.33	0.00			
S3	0.54	9.61	2.78	5.89	6.63	12.03	1.02	0.00		
S4	9.74	6.71	6.45	5.40	12.18	1.33	5.21	10.48	0.00	
S5	2.80	2.67	1.16	1.13	4.36	2.18	1.73	3.97	2.35	0.00

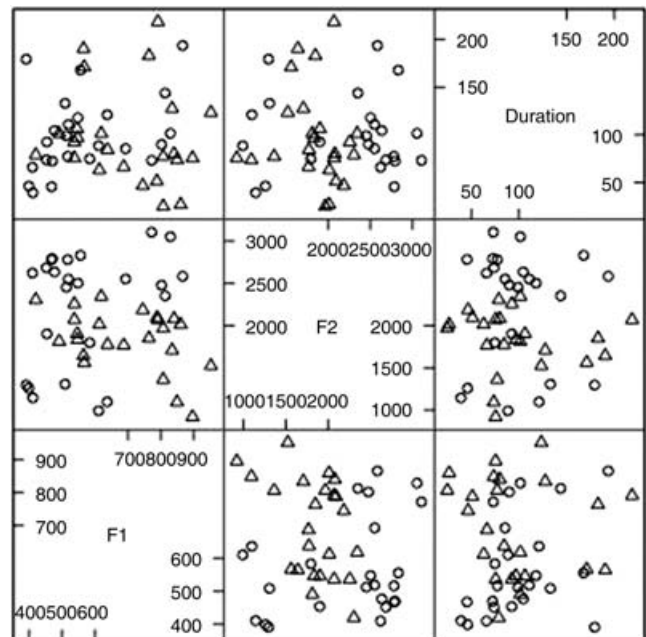
Table 4 Average acoustic measurements and standard deviations (in italics) of stressed and unstressed vowels of target words in passages produced by E1 and E2

Syllable	E1		E2	
	Stressed	Unstressed	Stressed	Unstressed
F0 (Hz)	220.58 <i>59.39</i>	207.83 <i>45.78</i>	273.3 <i>62.90</i>	231.55 <i>87.96</i>
F1 (Hz)	506.41 <i>72.61</i>	638.45 <i>138.31</i>	635.33 <i>179.34</i>	753.25 <i>141.02</i>
F2 (Hz)	2591.25 <i>250.17</i>	1905.67 <i>218.54</i>	1872.25 <i>795.55</i>	1774.67 <i>459.54</i>
Duration (ms)	92.74 <i>31.35</i>	87.07 <i>51.67</i>	107.00 <i>48.77</i>	107.61 <i>48.47</i>

squared distances of all talkers). Although familiar accents and languages always receive higher similarity ratings in sinewave speech than do unfamiliar accents and languages (Sheffert, Pisoni, Fellowes & Remez, 2002), 4.95 (11.03 –6.08) was the largest difference found by our raters among the Midwestern American English talkers (see Table 3 for matrices of mean squared distances of all talkers).

The recordings of E1 and E2 were conducted in the same double-walled soundproofed booth with an Audio-Technica 100HE Hypercardioid dynamic microphone. The final passages and lists of target words were digitized at 44,100 Hz and normalized to an amplitude of approximately 70 dB. The target word lists produced by E1 had an average duration of 17.94 seconds across all lists (the range was 16.53 s to 19.41 s) and the passages had an average duration of 19.81 seconds (the range was 18.31 s to 20.59 s). The lists of target words produced by E2 had an average duration of 19.35 seconds (the range was 17.87 s to 20.31 s) and the passages had an average duration of 19.81 seconds (the range was 18.23 s to 20.98 s).

As mentioned, in order to explore whether perceptual measures were similar to acoustic measures, we conducted acoustic measurements of F0, F1, F2, and of the duration of the stressed and unstressed vowels of the target words in passages of all speakers (see Table 4 for average measurements and standard deviations produced by E1 and E2). Although the relative intensity of the first and second syllables is a possible cue to English trochaic words, we chose not to measure this cue, given that our speakers did not reliably produce any intensity differences between first and second syllables. The measurements of F0, F1, F2 and duration support the adult ratings in revealing large voice differences between the speakers. For example, average F2 measurements for the stressed and unstressed syllables of the target words produced by E2 were substantially smaller and more variable than those of E1. A scatter plot presented in Figure 1 further illustrates the differences in F1, F2, and in the duration of stressed and unstressed vowels of the target words produced by E1 and E2.

**Figure 1** Average measurements of F1, F2 and duration of stressed and unstressed vowels of target words in passages produced by E1 and E2. Circles represent stressed vowels and triangles represent unstressed vowels. Labels of scatter-plot axes are indicated within the squares.

Design

All infants were randomly assigned to one of two experimental conditions. Out of 24 infants tested, 12 infants were assigned to Condition 1 (familiarized with repeated tokens of *kingdom* and *hamlet*) and the other 12 infants were assigned to Condition 2 (familiarized with repeated tokens of *candle* and *raptor*). All infants were tested on the same six-sentence passages containing *kingdom*, *hamlet*, *candle* and *raptor*. The infants were also randomly assigned to two Familiarization orders that varied by which talker produced the stimuli in familiarization and test (e.g. E1 in familiarization and E2 in test and vice versa). Out of 24 infants tested, six infants were assigned to Condition 1 (familiarized with repeated tokens of *kingdom* and *hamlet*) and heard the E1 Familiarization order (E1 produced the target words and E2 produced the passages), six infants were assigned to Condition 2 (familiarized with repeated tokens of *candle* and *raptor*) and heard the E1 Familiarization order, six infants were assigned to Condition 1 and heard the E2 Familiarization order (E2 produced the target words and E1 produced the passages), and six infants were assigned to Condition 2 and heard the E2 Familiarization order.

Once each infant reached a 30-s criterion to each target word in familiarization, the test phase began. In the test phase, the infants heard four passage testing blocks consisting of a total of 16 total test trials. Each passage occurred once in each test block, and the computer randomized the order of passages within the blocks. The infants' orientation times to the passages containing

familiar words were compared with those to the passages containing unfamiliar words.

Apparatus

Infants were tested in a three-sided testing booth made out of three white pegboard panels, which each stood approximately 140 cm high. White curtains hung from the ceiling to the top of the pegboard panels. Except for three openings in the front panel, the pegboard panels were completely backed with cardboard in order to prevent the infant from being distracted by any movements behind the panels. One opening was for a video camera, which was used to record each experimental session. The other two openings were used for viewing purposes: one for the experimenter to look through while coding the infants' orientation patterns and the other for another observer (e.g. parent, research assistant) to observe the session. A green light was mounted to the centre of the front panel, and a red light and speaker were mounted to the centre of each of the side panels. The lights were located at the approximate eye level of an infant seated on the lap of their caregiver in a chair in the centre of the booth. A speaker was hidden behind each red light and attached to the rear of each side panel. The computer, amplifier, video camera and response box were located behind the front panel, hidden from the infant.

Procedure

All infants were tested in the HPP (Jusczyk & Aslin, 1995; Kemler Nelson *et al.*, 1995). In this version of the procedure, the infant sits on his/her caregiver's lap on a chair in the middle of the testing booth facing the front panel. An experimenter conducts the experimental session from a position hidden behind the front panel of the booth and records the infant's orientation patterns via a response box by looking through a small opening in the pegboard. A computer program randomly specifies the side-light activation and auditory stimulus selection. In order to prevent the caregiver and experimenter from interfering with the infant's performance during the experiment, both the caregiver and the experimenter wear sealed aviator headphones that play loud, continuous music with low-level white noise. This music and noise serve to mask the auditory stimuli played during the experiment.

In this experiment, all familiarization and test trials began with a green light flashing in the centre of the front panel. As soon as the infant oriented towards the green light, the green light would extinguish and one of the red lights on the side panels would begin to flash (randomly chosen by the computer program). When the infant oriented at least 90 degrees in the direction of the flashing red light, a speech stimulus played from the speaker directly behind that light. The speech continued to play until the infant looked at least 30 degrees away from the flashing light for more than two consecutive seconds, or the auditory stimulus played in its entirety.

Once that trial ended, a new trial began with the flashing green light. For each trial, the computer recorded the length of time the infant oriented towards the light while the stimulus was playing.

The experiment consisted of two phases: familiarization and test. During familiarization, two lists of 14 isolated target words played alternately until the infant reached an orientation time criterion of 30 s to each word list. If the infant reached criterion on one word list, but not on the other, the word lists continued to alternate randomly until the infant reached criterion on both. Once the infant reached the listening time criterion on both word lists, the test phase began. During the test phase, the infant heard four blocks of randomly ordered test trials. So that the infants heard each passage only once within each block, there were four trials in each block, for a total of 16 test trials. The order of sentences within each passage was fixed. In each block, the infant heard two passages containing the target words presented in the familiarization word lists and two passages containing unfamiliar words not presented in the familiarization word lists. For example, if an infant heard *kingdom* and *hamlet* in the familiarization word lists, the passages containing the unfamiliar words were the passages containing *candle* and *raptor*.

Results and discussion

The infants' mean listening times to each of the four passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. The infants listened for an average of 7.98 seconds ($SD = 3.17$ s) to passages with familiar words and for an average of 7.08 seconds ($SD = 2.71$ s) to passages with unfamiliar words. A repeated measures analysis of variance (ANOVA) on Word Familiarity (familiar, unfamiliar) as repeated measure and Condition (1, 2) and Familiarization Order (E2 first, E1 first) as factors revealed no significant main effect of Condition ($F(1, 20) = 1.27, p < .273$) or Order ($F(1, 20) = .308$), and no significant two-way interactions ($F_s(1, 20) < 3.22, p_s > .087$) or three-way interactions ($F(1, 20) = .007$), but a main effect of Word Familiarity ($F(1, 20) = 5.99, p < .024$). The estimated effect size was a small-to-medium effect of Word Familiarity ($d = .33$) as determined by Cohen (1988). This main effect of familiarity resulted because 18 out of 24 infants listened longer to passages containing familiar words than to those containing unfamiliar words, regardless of which words they heard or which voice they heard first. These results are shown in Figure 2.

The infants listened significantly longer to passages containing familiar words than to passages containing unfamiliar words, indicating that they generalized words produced by an English familiarization talker to the same words produced by another dissimilar English test talker, thus replicating and extending to a new age group the findings of Houston and Jusczyk (2000). Furthermore, this was found despite the fact that the talkers were rated to have

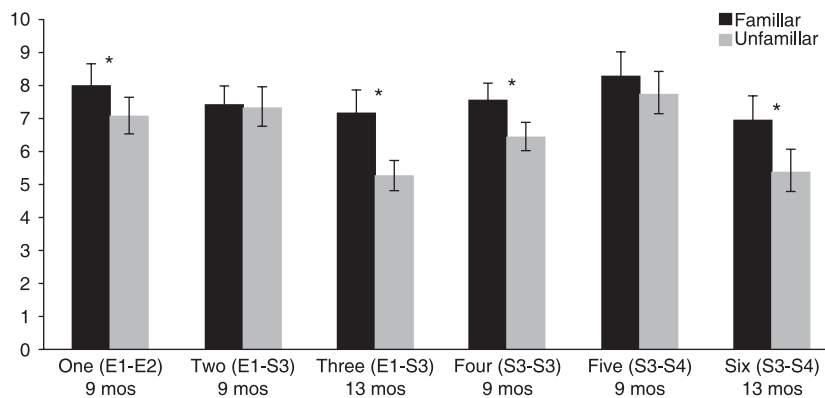


Figure 2 Mean listening times to passages containing familiar and unfamiliar items across all six experiments with error bars showing standard error. * = $p < .05$.

very distinctive voices. Although the voices of these talkers were dissimilar, it is plausible that the infants' experience or familiarity with the subphonemic and suprasegmental features of Midwestern American English facilitated talker generalization in this experiment, as it is possible that the more familiarity infants have with the words of their native language, the more generalizable they are. Thus, if familiarity with an accent promotes talker generalization, how would familiarity with Midwestern American English impact generalization to words in Spanish-accented speech? We explored this question in Experiment 2, in which we tested infants' abilities to generalize a word spoken by a familiarization talker in Midwestern American English to the same word spoken by a test talker in Spanish-accented English (and vice versa).

Experiment 2

If familiarity with Midwestern American English helped infants to generalize across talkers with dissimilar voices in Experiment 1, we predicted that familiarity with Midwestern English would hinder word recognition and generalization in Spanish-accented English. Although 9-month-old infants could accommodate variability in voice across two talkers in their native dialect, they may be unable to extract the relevant features and ignore irrelevant variation between words in Spanish-accented and Midwestern English in order to generalize the same abstract phonological structure of the target words. We explored this question in this experiment, in which we familiarized 9-month-olds with words in either Spanish-accented English or Midwestern English and tested them with passages containing familiar or unfamiliar words in a different accent to that with which they were familiarized. For example, infants familiarized with words in Spanish-accented English were tested on passages in Midwestern English, and infants familiarized with words in Midwestern English were tested on passages in Spanish-accented English. We selected the two talkers from our cross-accent talker similarity ratings that were rated to have the most similar voices of all possible pairings of talkers, including pairs of

talkers within a given accent group. We predicted that if infants at this age could recognize words across talkers and accents, they might be approximating the abilities of adults when accommodating word forms that deviate subphonemically and suprasegmentally from native-like targets by referring to their native phonological system to generalize the same abstract structure for the target words. Alternatively, because a Spanish accent modifies both the subphonemic and suprasegmental structure of English words, it is also possible that infants might assume a different structure for the same word in Spanish-accented English than in Midwestern American English, thereby exhibiting difficulty in this task.

Method

Participants

Twenty-two monolingual English-learning 9-month-olds were tested (10 males). The infants had a mean age of 36 weeks, 2 days (range = 35 weeks, 0 days to 37 weeks, 5 days). Five additional infants were tested but excluded for the following reasons: fussing or crying (4), and sickness (1). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Stimuli

The stimuli used in this experiment were identical to those of Experiment 1, with one exception: one Midwestern American English speaker (E1 from Experiment 1) and one Spanish-accented English speaker (hereafter referred to as S3) produced the target word lists and passages. S3 was a university-educated female native speaker of Spanish from the Dominican Republic who acquired English after the age of 12, and who spoke English with an intelligible Spanish accent (as judged by adult raters). These talkers were selected for this experiment based upon the perceptual similarity ratings provided by the multidimensional scaling matrices of S3 and E1. These two talkers had a mean squared distance of .53 in natural

speech and of .54 in sinewave speech, indicating that their voices were highly *similar* as there was almost no change between the ratings in natural and sinewave speech (see Table 3 for matrices of mean squared distances of all talkers).

The recordings of S3 and E1 were conducted in the same double-walled soundproofed booth with an Audio-Technica 100HE Hypercardioid dynamic microphone. The final passages and lists of target words were digitized at 44,100 Hz and normalized to an amplitude of approximately 70 dB. The target word lists produced by S3 had an average duration of 17.00 seconds (the range was 16.01 s to 18.11 s) and the passages had an average duration of 24.75 seconds (the range was 23.41 s to 26.00 s). See Experiment 1 for average durations of target word lists and passages for E1.

Although the mean squared distances derived from the talker similarity ratings reveal that S3 and E1 had highly similar voices, acoustic measurements of F0, F1, F2 and the durations of the stressed and unstressed vowels of the target words produced by E1 and S3 show large differences in F1, F2 and duration (see Table 5 for acoustic measurements and standard deviations). For instance, S3 seems to have more variability in F2 for unstressed syllables (larger standard deviations), more variability in both F1 and F2 for stressed syllables, less variability in F1 for unstressed syllables, and longer vowel durations for stressed syllables and shorter vowel durations for unstressed syllables compared with E1. Furthermore, measurements of VOT ([k] in word-initial position for *candle* and *kingdom*) demonstrate a substantial subphonemic difference between the speakers, with the VOT duration of E1 equivalent to nearly twice the VOT duration of S3 (79.44 ms and 39.45 ms, respectively; see Table 6 for average acoustic measurements and standard

Table 5 Average acoustic measurements and standard deviations (in italics) of stressed and unstressed vowels of target words in passages produced by E1 and S3

Syllable	E1		S3	
	Stressed	Unstressed	Stressed	Unstressed
F0 (Hz)	220.58 <i>59.39</i>	207.83 <i>45.78</i>	237.75 <i>50.09</i>	235.64 <i>87.33</i>
F1 (Hz)	506.41 <i>72.61</i>	638.45 <i>138.31</i>	649.42 <i>334.03</i>	529 <i>89.77</i>
F2 (Hz)	2591.25 <i>250.17</i>	1905.67 <i>218.54</i>	2295.5 <i>427.64</i>	1679.58 <i>340.79</i>
Duration (ms)	92.74 <i>31.35</i>	87.07 <i>51.67</i>	142.42 <i>33.83</i>	90.75 <i>42.88</i>

Table 6 Average acoustic measurements and standard deviations (in italics) of voice onset time duration of all speakers

	E1	E2	S3	S4
Duration (ms)	79.44 <i>23.95</i>	99.02 <i>37.87</i>	39.45 <i>15.38</i>	50.22 <i>14.02</i>

deviations of VOT of all speakers).² Although the similarity ratings revealed by the talker similarity ratings reveal that these speakers are similar, according to these acoustic measurements the speakers selected for this experiment (E1 and S3) actually seem to be quite different (e.g. VOT differences, longer vowel duration of stressed syllables and shorter vowel duration of unstressed syllables). Thus, although the acoustic differences between E1 and S3 are quite distinct, the adult perceptual measures revealed the opposite result, suggesting that adults may not be sensitive to the differences revealed through the acoustic measurements. Furthermore, despite the fact that perceptual similarity ratings were similar for adult raters, these substantial acoustic differences may make it difficult for infants, who are sensitive to different acoustic cues than are adults, to generalize across talkers.

Design

This was identical to that of Experiment 1, except that accent was varied as well as voice.

Apparatus and procedure

These were identical to those of Experiment 1.

Results and discussion

Once again, the infants' mean listening times to the four different passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. The infants listened for an average of 7.42 seconds ($SD = 2.53$ s) to passages with familiar words and for an average of 7.35 seconds ($SD = 2.78$ s) to passages with unfamiliar words. A repeated measures ANOVA on Word Familiarity (familiar, unfamiliar) as repeated measure and Condition and Familiarization Order (Spanish accent first, English first) as factors revealed no main effect of Condition ($F(1, 18) = 3.24, p < .088$), Word Familiarity ($F(1, 18) = .009$), or Order ($F(1, 18) = 2.15, p < .159$), and no two-way interactions ($F_s(1, 18) < 2.18, p_s > .158$) or a three-way interaction ($F(1, 18) = 3.56, p < .071$). Furthermore, there was determined to be no effect of Word Familiarity ($d = .03$) in a calculation of effect sizes. This pattern of results suggests that the infants could not generalize the same abstract structure for words produced by a Spanish-accented familiarization talker to the same words produced by a Midwestern English test talker (and vice versa), even though the talkers were of the same sex and were rated to have the most similar voices of all talkers from the talker similarity ratings.

² As pointed out by one reviewer, this VOT difference between E1 and S3 means that some of the Spanish speakers' tokens with voiceless VOT targets may have been miscategorized by English-learning infants as voiced, because they fell within one English VOT category. This VOT misclassification would have rendered the task even more difficult.

One possible interpretation of the infants' difficulty with this task is that, although they might have abstract phonological representations of the target words (as demonstrated by their success in Experiment 1 in generalizing across two talkers with dissimilar voices), their representations may not be sufficiently flexible to accommodate the subphonemic variation of foreign-accented speech in generalization tasks. Consequently, when the language-specific subphonemic cues (e.g. VOT) with which infants are familiar are modified, they might have difficulty extending their representations to a new phonological system. For example, these infants may have learned the VOT of Midwestern English, so they may not be able to extend to a new VOT system (as in the case of Spanish-accented English). Furthermore, because the acoustic measurements revealed large differences between E1 and S3 at both the subphonemic and the suprasegmental level, the Spanish-accented words may have deviated too much from their native phonological system for infants to map the structure of the accented target words onto that of the native target words. It is unclear, however, whether these infants are attending to differences at the subphonemic level, the suprasegmental level, or both. Thus, because infants seem to progress from larger units to smaller units in their phonological processing of the speech stream (Hirsh-Pasek *et al.*, 1987; Jusczyk *et al.*, 1992a; Soderstrom, Seidl, Kemler Nelson & Jusczyk, 2003), older infants may be more successful in recognizing a word across speakers with different voices and accents, as generalization may not be disrupted by variability in voice as well as in accent. We explored this question in Experiment 3, in which we investigated 13-month-olds' abilities to generalize words across a speaker in Spanish-accented English and a speaker in Midwestern English.

Experiment 3

Although infants at 9 months of age do not appear to have sufficiently flexible phonological representations to generalize the structure of the same word in Spanish-accented English and in Midwestern American English, it is possible that older infants may have more abstract phonological representations, which may allow for greater ease of generalization of words in the face of accent and voice variability. It is well documented that infants' perceptual sensitivities become more language-specific and abstract throughout development, probably as a result of more experience and practice with the target language. In word recognition, 'younger infants require a greater degree of similarity than do older infants in dimensions such as voice, gender, coarticulation, and stress' (Werker & Curtin, 2005, p. 209). Thus, older infants who are just beginning the task of word learning (13-month-olds) may be better at processing variable subphonemic and suprasegmental information with substantial dissimilarity across representations compared with younger infants, who have a more tenuous hold on their word representations.

We explored this question in Experiment 3, in which we tested English-learning 13-month-olds using the same design, stimuli and procedure as in Experiment 2. We expected that these older infants would better accommodate subphonemic and suprasegmental variability by generalizing the same abstract structure for the target words in Spanish-accented English and in Midwestern American English.

Method

Participants

Twenty monolingual English-learning 13-month-olds were tested (9 males). The infants had a mean age of 52 weeks, 0 days (range = 48 weeks, 6 days to 52 weeks, 5 days). Eleven additional infants were tested but excluded for the following reasons: fussing or crying (6), sickness (1), falling asleep (1), hearing a language other than English more than 30% of the time (2), and parental discomfort (1). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Stimuli

These were identical to those of Experiment 2.

Design

This was identical to that of Experiment 2.

Apparatus and procedure

These were identical to those of Experiments 1 and 2.

Results and discussion

Once again, the infants' mean listening times to each of the four passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. The infants listened for an average of 7.16 seconds ($SD = 3.09$ s) to passages with familiar words and for an average of 5.27 seconds ($SD = 2.03$ s) to passages with unfamiliar words. A repeated measures ANOVA with Word Familiarity (familiar, unfamiliar) as a repeated measure and Condition (1, 2) and Familiarization Order (Midwestern American English familiarization, Spanish-accented familiarization) as factors revealed no main effects of Familiarization Order ($F(1, 16) = .224$) or Condition ($F(1, 16) = .682$), and no interactions with Familiarization Order or Condition (all $F_s(1, 16) < .81$), but a main effect of Word Familiarity ($F(1, 16) = 7.46, p < .015$). This was determined to be a large effect of Word Familiarity ($d = .76$) in an analysis of effect size. Thus, regardless of which order the infants heard a Spanish-accented production in, they still listened longer to the passages containing familiar over unfamiliar words (see Figure 2 for results).

These results suggest that, whereas 9-month-olds rely on the similarity of the sound pattern of words across speakers, 13-month-olds more readily abstract across variable instances of a word. As a result, these older infants seem to have a more robust system that approximates that of adults, given that they are able to accommodate the increased suprasegmental and subphonemic variability present in the productions of a Spanish-accented speaker. It may be that, for younger infants, such variability is interpreted as phonemic, thus contributing to their inability to succeed in this task. Furthermore, this pattern of results suggests that these older infants cannot be doing mere pattern-matching when recognizing words in the speech stream, but must instead be constructing more abstract phonological representations of words that allow for word recognition in the face of both voice and accent variability. This inevitably contributes to more successful generalization across dissimilar surface forms and to the development of a mature lexicon.

Experiment 4

Although infants at 9 months of age seem to be able to accommodate variability in voice when generalizing words across speakers (Experiment 1), they fail to do so when dealing with variability in both voice and accent across speakers (Experiment 2). Can English-learning 9-month-olds accommodate accent variability at all, or is dealing with such variability cumbersome at this age? We explored this question in Experiment 4, in which we familiarized infants with repetitions of two isolated words produced by a Spanish-accented speaker (S3 from Experiments 2 and 3) and tested them on passages containing familiar or unfamiliar words produced by the same speaker.

By 7.5 months of age, infants are sensitive to the rhythmic structure of their native language (Jusczyk, Cutler & Redanz, 1993) and use language-specific properties to facilitate word segmentation (Jusczyk *et al.*, 1999; Polka & Sundara, 2003; Polka, Sundara & Blue, 2002). English-learning 9-month-olds can successfully segment words from fluent Dutch speech (Houston, Jusczyk, Kuijpers, Coolen & Cutler, 2000), probably achieving this by relying on the predominant rhythmic structure of words in their native language because the Dutch phonological inventory differs from English in many ways and infants are thus unfamiliar with the precise subphonemic cues of Dutch. Consequently, when segmenting words from foreign-accented speech with varying subphonemic cues, 9-month-old infants may be able to approximate the abilities of the 9-month-olds in Houston *et al.* (2000).

Alternatively, because adults are sensitive to differences in foreign accent marked at the level of the segment (Lane, 1963; Rogers, 1997; Strange, 1995) and infants are highly sensitive to phonemic differences in the speech they segment (Jusczyk & Aslin, 1995), it is perhaps just as likely that the subphonemic changes present in foreign-accented speech will make word recognition difficult for infants

at 9 months of age. Although adults most often adapt quickly to such subphonemic differences (Nygaard & Pisoni, 1998; Nygaard *et al.*, 1994), infants and toddlers may not because they have highly detailed representations, which may be inflexible to such variation.

Method

Participants

Twenty-eight monolingual English-learning 9-month-olds were tested (14 males). The infants had a mean age of 37 weeks, 6 days (range = 34 weeks, 4 days to 40 weeks, 6 days). Ten additional infants were tested but excluded for the following reasons: parental interference (2), fussing or crying (5), at least 30% exposure to a language other than English (1), and experimenter error (2). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Stimuli

The stimuli used in this experiment were produced by S3 (Experiments 2 and 3). Identical to the case for Experiments 1, 2 and 3, the recordings were made in a double-walled sound-proofed booth with an Audio-Technica 100HE Hypercardioid dynamic microphone. The final passages and lists of target words were digitized at 44,100 Hz and normalized to an average amplitude of 70 dB. See Experiments 2 and 3 for average durations of target word lists and passages for S3.

Design

The design of this experiment was identical to that of Experiments 1, 2 and 3 with one exception: infants were randomly assigned to one of two experimental conditions instead of to one of four conditions. Out of 28 infants tested, 14 infants were assigned to Condition 1 (familiarized with repeated tokens of *kingdom* and *hamlet*) and the other 14 infants were assigned to Condition 2 (familiarized with repeated tokens of *candle* and *raptor*). All infants were tested on the same six-sentence passages containing *kingdom*, *hamlet*, *candle* and *raptor*.

Apparatus and procedure

These were identical to those of Experiments 1, 2 and 3.

Results and discussion

The infants' mean listening times to the four passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. Eighteen out of 28 infants listened longer to passages containing familiar words than to passages containing unfamiliar words. The infants listened for an average of 7.57 seconds ($SD = 2.58$ s)

to passages with familiar words, compared with an average of 6.44 seconds ($SD = 2.36$ s) to passages with unfamiliar words. A repeated measures ANOVA on Word Familiarity (familiar, unfamiliar) as repeated measure and Condition (1, 2) as factor revealed no significant main effect of Condition ($F(1, 26) = .56, p < .459$), no significant interaction between Word Familiarity and Condition ($F(1, 26) = 1.88, p < .182$), but a main effect of Word Familiarity ($F(1, 26) = 4.99, p < .034$), thereby indicating a significant difference in listening time between passages with familiar and unfamiliar words. An analysis of effect size revealed a medium effect of Word Familiarity ($d = .48$). Thus, infants who were familiarized with isolated versions of Spanish-accented target words were able to recognize those words in Spanish-accented passages, despite the fact that the subphonemic and suprasegmental cues of Spanish-accented English are distinct from those of the infants' native language (Midwestern American English). Results are shown in Figure 2.

These results demonstrate that, although Midwestern English subphonemic cues (e.g. VOT, exact location of the vowels), for example, were modified in the Spanish-accented English, English-learning infants were still able to recognize words in a Spanish-accented speech stream. We therefore conclude that in the face of variability in foreign accent, English-learning infants may not be confused by subphonemic and suprasegmental cues that are distinct from those of their input language. This pattern of results furthermore suggests that, as long as infants at this age do not have to generalize across different voices, an intelligible Spanish accent may not hinder word recognition.

An alternative explanation for infants' ability to recognize words in an unfamiliar accent is that these infants were not really solving the invariance problem, but rather were performing well as a result of pattern matching. Thus, although they may have recognized that the speech differed substantially from their input language, they still performed well in the test phase because the target words, although not identical to the familiarization words, were nonetheless a good acoustic match. Even though average acoustic measurements of the stressed and unstressed vowels of the familiarization tokens reveal differences in duration and F0, when compared with those of the stressed and unstressed syllables of the passage tokens, the infants successfully generalized a word across familiarization and test. For example, the average durations of both stressed and unstressed vowels in the familiarization tokens were longer than those of both stressed and unstressed vowels in the passage tokens (e.g. .16 s and .13 s; .12 s and .07 s, respectively). Furthermore, the average F0s of both stressed and unstressed vowels in the familiarization tokens were larger than those of both stressed and unstressed vowels in the passage tokens (e.g. 293.5 Hz and 261.5 Hz; 242.2 Hz and 212.5 Hz, respectively). Although these differences suggest that infants were not reliant upon *exact* acoustic matching, it is likely that infants attended to speaker-

specific patterns when recognizing words across familiarization and test.

It was suggested in Jusczyk's word recognition and phonetic structure acquisition (WRAPSA) model (1993, 1997) that this is exactly how infants begin the task of word identification. More specifically, WRAPSA suggests that during the first few months after birth the infant is attuned to properties in the speech stream that are 'provided by the inherent organization of the human auditory system' (Jusczyk & Luce, 2002, p. 38) and are completely language-general, but then, as experience with the native language accumulates, the infant can re-weight cues to reflect meaningful differences within the native language. In this model, infants encode a large amount of surface-level detail about the utterances that they have heard (e.g. speaker-specific and perhaps also accent-specific information) and this allows them to then use this information in their recall of specific speakers' words (Houston & Jusczyk, 2000; see also Nygaard & Pisoni, 1998 for similar adult work).

Thus, although these findings provide some new information about how infants deal with foreign accent variability, they shed little light on the abstractness of their word representations at this age. That is, these results do not directly bear on the question of whether these infants can extract the relevant abstract phonological features of words, thereby ignoring any irrelevant variation between Spanish-accented and Midwestern-accented words (Experiment 2). Thus, the pattern of results of this experiment may be accounted for by the infants' attention to the specific productions of the Spanish-accented words, rather than by an ability to develop abstract word representations. For example, if the infants became familiar with the talker's particular accent and learned how she specifically produced each target word, they may have developed a very precise representation of each target word. As a result, it remains unclear whether infants at this age could generalize beyond specific representations of a single talker when accommodating extraneous differences in subphonemic and suprasegmental structure. To investigate this issue we conducted another experiment in which we tested infants' word recognition abilities when they were familiarized with words spoken by one Spanish-accented talker and tested on passages containing familiar and unfamiliar words spoken by a different Spanish-accented talker.

Experiment 5

The word recognition success exhibited by the 9-month-olds in Experiment 4 cannot be accounted for solely by exact acoustic matching, as items in familiarization are in isolation and items in test are in passages, hence providing different acoustic manifestations owing to sentential position and coarticulation. However, it is still likely that the 9-month-olds were doing some kind of pattern matching, given that they succeeded when there was

only within-talker accent variability (Experiment 4), but failed when there was across-talker accent variability *and* voice variability (Experiment 2). It is therefore unclear whether infants at this age would be better able to accommodate across-talker variability in accent and voice if the infants were more readily able to adopt a pattern-matching strategy with two talkers that were more similar in terms of accentual characteristics (e.g. two Spanish-accented talkers).

In their investigation, Houston and Jusczyk (2000) demonstrated that 7.5-month-old infants could generalize words produced by a familiarization talker to the same words produced by a test talker, so long as the talkers were of the same gender. Given these results and the results of Experiment 1, we expected that English-learning 9-month-olds would exhibit the same generalization success with two different female talkers with Spanish accents, as a pattern-matching strategy may be more effective when the patterns are more similar (e.g. similar voices, similar accents). Alternatively, we expected that if the two Spanish-accented talkers were too dissimilar in voice and accent and the infants at this age attended too much to surface-level detail, then the infants might exhibit difficulty with this more abstract task. In order to circumvent worries that the talkers might be too different, we obtained talker similarity ratings and selected the two Spanish-accented talkers who were rated to be the most perceptually similar in terms of voice. In this experiment, 9-month-old infants were familiarized with two target words spoken by one Spanish-accented talker and tested on passages containing familiar or unfamiliar words by the other Spanish-accented talker.

Method

Participants

Twenty-eight monolingual English-learning 9-month-olds were tested (14 males). The infants had a mean age of 36 weeks, 0 days (range = 34 weeks, 4 days to 38 weeks, 1 day). Seven additional infants were tested but excluded for the following reasons: parental interference (1), fussing or crying (4), and at least 30% exposure to a language other than English (2). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Stimuli

The stimuli used in this experiment were identical to those of Experiments 2 and 3 with one exception: two different university-educated native speakers of Spanish, who spoke English with an intelligible Spanish accent, produced the target word lists and passages. The first Spanish-accented talker (S3) produced the stimuli used in Experiments 2, 3 and 4. A second Spanish-accented talker, hereafter referred to as S4, was instructed to record the stimuli in the same manner as S3. S4 was from Peru and

learned English after the age of 11. This talker was selected for this experiment based upon the average dissimilarity ratings provided by the multidimensional scaling matrices of S3 and S4. These two talkers had a mean squared distance of 10.35 in natural speech and of 10.48 in sinewave speech, indicating that their voices were highly *similar* because there was almost no change in the average dissimilarity rating between natural and sinewave speech (see Table 3 for proximity matrices showing the mean squared distances among all talkers). These mean squared distances were probably higher values compared with previous experiments because the listeners who provided talker similarity ratings were all English monolingual speakers, and Sheffert *et al.* (2002) demonstrated that unfamiliar accents and languages always receive lower similarity ratings than familiar accents and languages. Because voice similarity has been shown to be important in word recognition tasks across different talkers (Houston & Jusczyk, 2000 found that young infants could not recognize words across genders (very different voices) until 10.5 months of age), these two talkers were chosen so as to provide the most optimal circumstances for word recognition (i.e., if the two talkers had been very dissimilar in voice, the experiment may have been too difficult for the infants).

Average acoustic measurements of F0, F1, F2 and durations of the stressed and unstressed vowels of the target words from the test passages of S3 and S4 validated the mean squared distances produced from the talker similarity ratings (see Table 6 for average acoustic measurements and standard deviations of VOT of S3 and S4, and Table 7 for average acoustic measurements and standard deviations of F0, F1, F2 and the durations of S3 and S4). Overall, the Spanish-accented speakers both seem to be more variable in their productions of stressed vowels and F2, which may have contributed to higher ratings of voice similarity, possibly owing to more overlap in their productions (see also Figure 3 for a scatter plot of F1, F2, and durations of stressed and unstressed vowels produced by S3 and S4). Furthermore, the VOT durations of S3 and S4 are similar (39.45 ms and

Table 7 Average acoustic measurements and standard deviations (*in italics*) of stressed and unstressed vowels of target words in passages produced by S3 and S4

Syllable	S3		S4	
	Stressed	Unstressed	Stressed	Unstressed
F0 (Hz)	237.75 <i>50.09</i>	235.64 <i>87.33</i>	276.67 <i>87.21</i>	198.9 <i>47.13</i>
F1 (Hz)	649.42 <i>334.03</i>	529 <i>89.77</i>	632 <i>194.73</i>	595.83 <i>50.00</i>
F2 (Hz)	2295.5 <i>427.64</i>	1679.58 <i>340.79</i>	1704.25 <i>769.51</i>	1690.75 <i>246.19</i>
Duration (ms)	142.42 <i>33.83</i>	90.75 <i>42.88</i>	111.01 <i>50.83</i>	102.50 <i>39.09</i>

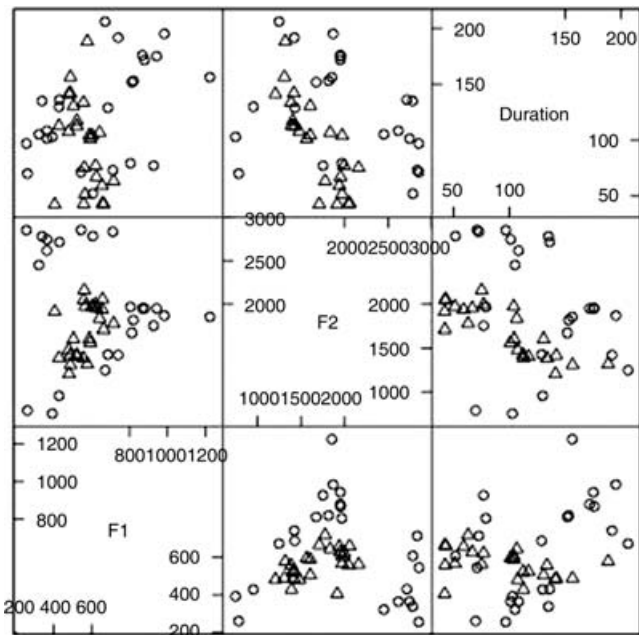


Figure 3 Average measurements of F1, F2 and duration of stressed and unstressed vowels of target words in passages produced by S3 and S4. Circles represent stressed vowels and triangles represent unstressed vowels. Labels of scatter-plot axes are indicated within the squares.

50.22 ms, respectively), compared with the VOT durations of E1 and E2 (79.44 ms and 99.02 ms, respectively).

The recordings of S4 were conducted in the same double-walled sound-proofed booth as used previously with an Audio-Technica 100HE Hypercardioid dynamic microphone. The final passages and lists of target words were digitized at 44,100 Hz and normalized to an approximate amplitude of 70 dB. The lists of target words produced by S4 had an average duration of 23.34 seconds (the range was 22.51 s to 25.71 s) and the passages had an average duration of 23.88 seconds (the range was 22.35 s to 26.24 s). See Experiment 2 for average durations of target word lists and passages for S3.

Design

The design of this experiment was identical to that of Experiments 1, 2 and 3, except that two Spanish talkers were used.

Apparatus and procedure

These were identical to those of Experiments 1, 2, 3 and 4.

Results and discussion

Once again, the infants' mean listening times to each of the four passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. The infants listened for an average of 8.28 seconds

($SD = 3.90$ s) to passages with familiar words and for an average of 7.76 seconds ($SD = 3.45$ s) to passages with unfamiliar words. A repeated measures ANOVA with Word Familiarity (familiar, unfamiliar) as repeated measure and Condition (1, 2) and Familiarization Order (S3 familiarization, S4 familiarization) as factors revealed no main effect of Condition ($F(1, 24) = 1.51, p = .23$), or Word Familiarity ($F(1, 24) = .55$), but a marginal effect of Familiarization Order ($F(1, 24) = 4.18, p = .052$), indicating that infants overall showed a preference for S4 over S3 (a distinction orthogonal to our question of interest). Furthermore, results revealed that there were no two-way interactions ($F_s(1, 24) < .89$). There was, however, a three-way interaction ($F(1, 24) = 4.37, p = .047$). Post-hocs within Condition and Order comparing listening time to familiar and unfamiliar items revealed that there were no significant differences between familiar and unfamiliar items within Condition and Order ($t_s(7) < 2.32, p_s > .06$). An analysis of effect size revealed that there was no overall effect of Word Familiarity ($d = .13$). This suggests that, despite the fact that infants showed different preferences in the different conditions and orders, they could not generalize words produced by a Spanish-accented familiarization talker to the same words produced by a different Spanish-accented test talker, even though the talkers were of the same sex and had been rated to have very similar voices.

One possible interpretation of this result is that the talkers' Spanish accents were too dissimilar and the infants could not accommodate these differences in order to achieve token generalization. Houston and Jusczyk (2000) demonstrated that it was not until 10.5 months of age that infants could generalize tokens across a male and a female talker. Thus, it is possible that the modifications to the subphonemic properties of both the consonants and the vowels of the target words due to the Spanish accent led the infants to interpret the tokens produced by each talker as too distinctive and made it more difficult for them to process speech from one talker to the other because pattern-matching was not a possible strategy. Furthermore, it is entirely plausible that, despite the fact that our adult raters found the accents to be somewhat similar to each other, this was not true for infants, as they are more sensitive to subphonemic distinctions present in a non-native language, even at 9 months of age (Werker & Tees, 1984). These findings mirror those of Experiment 2, in that 9-month-olds seem to rely so heavily upon the similarity of sound patterns across speakers that they cannot abstract across tokens that deviate substantially in surface form. Thus, it is possible that, similar to the findings of Experiment 3, 13-month-olds may exhibit success with this more abstract task. We explored this question in Experiment 6.

Experiment 6

The developmental pattern revealed by Experiments 2 and 3 indicates that, although 9-month-olds have difficulty

recognizing words across dissimilar surface forms owing to a reliance on acoustic matching, 13-month-olds seem to be constructing more abstract phonological representations of words, thereby promoting generalization across talkers with variability in voice and accent. In Experiment 6, we investigated whether 13-month-olds could recognize a word when two different Spanish-accented talkers produced the stimuli in familiarization and test. We tested these infants on the same stimuli, design, apparatus and procedure as in Experiment 5. We predicted that if 13-month-olds have a more robust phonological system than 9-month-olds, which permits abstraction across variable instances of a word, they would successfully generalize the structure of a word across the two Spanish-accented talkers.

Method

Participants

Twenty monolingual English-learning 13-month-olds were tested (11 males). The infants had a mean age of 52 weeks, 2 days (range = 50 weeks, 1 day to 54 weeks, 1 day). Eleven additional infants were tested but excluded for the following reasons: fussing or crying (7), falling asleep (2), hearing a language other than English more than 30% of the time (1), and failure to locate the lights in the testing room (1). All infants' parents reported normal hearing and full-term status, and all infants were recruited from families in a Midwestern town.

Stimuli

These were identical to those of Experiment 5.

Design

This was identical to that of Experiment 5.

Apparatus and procedure

These were identical to those of Experiments 1, 2, 3, 4 and 5.

Results and discussion

The infants' mean listening times to the four different passages were calculated across the four test blocks for passages containing familiar and unfamiliar words. The infants listened for an average of 6.95 seconds ($SD = 3.69$ s) to passages with familiar words and for an average of 5.40 seconds ($SD = 3.36$ s) to passages with unfamiliar words. A repeated measures ANOVA on Word Familiarity (familiar, unfamiliar) as repeated measure and Condition (1, 2) and Familiarization Order (S3 first, S4 first) as factors revealed no significant main effect of Condition ($F(1, 16) = .125$) or Familiarization Order ($F(1, 16) = 1.451$, $p < .246$), and no interactions with Familiarization Order

or Condition ($F_s(1, 16) < 1.92$, $p_s > .185$), but a main effect of Word Familiarity ($F(1, 16) = 7.968$, $p < .012$). Results are presented in Figure 2. An analysis of effect size revealed that this was a medium effect of Word Familiarity ($d = .52$).

Thus, the infants listened significantly longer to passages containing familiar words than to passages containing unfamiliar words, indicating that they successfully generalized a word across two different Spanish-accented talkers. This finding provides additional support for the pattern of results found in Experiments 2 and 3, in that 13-month-olds seem to have a more robust phonological system, compared with 9-month-olds, that accommodates both suprasegmental and subphonemic variability by abstracting across variable productions of a word by two different Spanish-accented talkers.

General discussion

In a series of six experiments, we explored the impact of variability in voice, accent, and voice *and* accent on word recognition at 9 and 13 months of age. The present set of studies provides support that, in word recognition tasks, 9-month-old infants are at least somewhat dependent upon the subphonemic and suprasegmental similarity of words across familiarization and test (for similar results, see Houston & Jusczyk, 2000; Singh, 2008; Singh *et al.*, 2004). Although 9-month-olds were able to accommodate input variability in the form of a foreign accent when recognizing words (Experiment 4), this is probably because the infants were able to pattern-match the Spanish-accented words heard in familiarization to the Spanish-accented words heard in the passages because the speech was produced by a single talker and, consequently, showed little phonetic variability across familiarization and test. In contrast, in Experiments 2 and 5, 9-month-old infants failed to recognize words heard in isolation in the passages when there was a substantial amount of voice and accent variability across a Midwestern English talker and a Spanish-accented talker (Experiment 2) as well as across two different Spanish-accented talkers (Experiment 5). Although subphonemic and suprasegmental variability proved to be a clear impediment to word recognition in an unfamiliar accent, the 9-month-olds were still able to accommodate cross-talker voice variability in Experiment 1, in which the voices were rated to be highly dissimilar, but the accents were familiar (Midwestern American English). Thus, it seems that it may be possible to separate voice similarity from accent similarity, as infants at 9 months seem to be able to accommodate dissimilarity in the former, but not in the latter. Older infants (13 months of age), however, are able to deal with both kinds of variability, not only providing evidence that they have established both flexible and abstract representations, but also providing some preliminary evidence that suggests that variability in voice and accent may not significantly hinder word recognition ability.

Thus, 9-month-old infants seem to be unable to ignore subphonemic and suprasegmental variability (e.g. shorter VOT and increased syllable duration in Spanish-accented English), as illustrated by their failure to generalize words across one native Midwestern English talker and one Spanish-accented talker in Experiment 2 and across two Spanish-accented talkers in Experiment 5. In contrast, in order to achieve successful token generalization, 9-month-olds *are* able to ignore some subphonemic and suprasegmental variability (Experiment 4), while attending to the phonological properties of the words across familiarization and test. As suggested by WRAPSA (Jusczyk, 1999), Singh *et al.* (2004) state that infants' tendency to be overly attentive to indexical properties might be a result of the fact that 'early lexical representations preserve acoustic and phonetic detail with such precision that variation, whether it be phonemic or non-phonemic, impedes their recognition of novel exemplars as familiar types' (p. 186).

There are three possible explanations for 9-month-old infants' failure to recognize a word across talkers in Experiments 2 and 5: (1) suprasegmental variability at the sentence level, (2) suprasegmental variability at the word level, or (3) subphonemic variability, which impacts recognition at the word level. Although it may not be possible to firmly adjudicate among these possibilities with the current experiments, the acoustic and perceptual measures suggest that subphonemic factors may have been more to blame for their failure than suprasegmental factors. Although the subphonemic differences between the Spanish-accented and Midwestern English talkers are large (e.g. the VOT of the Midwestern English talkers is twice the duration of that of the Spanish-accented talkers), the suprasegmental differences at the word level were highly similar across Midwestern and Spanish-accented talkers (e.g. the pitch and duration of the stressed and unstressed syllables of the target words produced by the Spanish-accented speakers typically followed the patterns of English metrical stress; see Tables 4, 5 and 7). Thus, because the subphonemic variability present in the Spanish-accented speakers' productions was on a much larger scale compared with the suprasegmental differences, this might suggest that subphonemic cues are the cause of 9-month-olds' difficulty in recognizing Spanish-accented words. However, it is still possible that global or sentential prosody was to blame for the infants' failure. To examine this question we explored adults' judgments of global prosody of the same set of talkers in our first set of talker similarity ratings and found that the global prosodic differences were not dissimilar from the word-level prosodic differences.³ Nonetheless, even

³ Because the talker similarity ratings showed adults' ratings of the similarity of the isolated tokens between speakers, it was unknown whether global prosodic differences between the native English and Spanish-accented talkers contributed to the 9-month-olds' failure to generalize words across a Spanish-accented talker and a native English talker, as well as across two Spanish-accented talkers. Thus, in order to evaluate whether the different global prosodic properties of Spanish and English affected infants' performance, we conducted another

though the global prosodic differences between speakers were not found to be significant by adult raters, it is still unclear to what particular cues infants were attending. For example, even though adults' similarity judgments revealed no significant differences between the talkers in Experiments 2 and 5 (E1 and S3), 9-month-old infants responded in a different way by failing to generalize tokens across the talkers, probably because they represented them as too dissimilar.

Although at 9 months of age the subphonemic and suprasegmental variability provided by Spanish-accented English is sufficient to hinder word recognition and generalization, 13-month-olds do not exhibit the same effect (Experiments 3 and 6). Because infants at 13 months of age are able to generalize the structure of the same word in Spanish-accented English and Midwestern American English, as well as in two different Spanish accents, their word representations must be more sophisticated at this age so as not to be impeded by the subphonemic and suprasegmental variation of Spanish-accented speech. Thus, these experiments provide further support for the argument that young infants attend closely to acoustic-phonetic detail, and that this attention disrupts word recognition. However, at 13 months of age a loss of attention to subphonemic detail may have an impact on these infants' increased ability to ignore irrelevant information when recognizing words.

Another possible explanation for a developmental process to explain why 13-month-olds, but not 9-month-olds, can accommodate the variation of a non-native accent is experience. Younger infants do not process speech as quickly or as efficiently as older infants, possibly as a result of less experience with their native language. Thus, older infants might be better able to solve the invariance problem because they have acquired more knowledge about the various forms a word could take as a result of more language experience.

This is certainly true for adults. It is well documented that experience and familiarity with variability in the target language (e.g. foreign-accented speech) can facilitate word recognition and generalization, talker identification, and speech intelligibility (Bradlow & Bent, 2003; McGarr, 1983; Weil, 2001). Bradlow and Bent (2003) demonstrated that adults could develop generalized representations of a novel foreign accent under training conditions of high talker

multidimensional scaling experiment on adults' ratings of the perceptual similarity of two representative passage sentences. Although the mean squared distances of the speakers across natural and sinewave speech were on average larger (probably because the sentences contain more information than the tokens, so the similarity ratings were more exaggerated), the distances within languages (e.g. Spanish-accented talkers) were approximately the same as those across languages (e.g. Spanish-accented and Midwestern English talkers) and were directly comparable to the mean squared distances calculated from the adult ratings on the tokens. Furthermore, as mentioned earlier, Houston *et al.* (2000) found that English-learning infants are able to segment words from Dutch. Although Dutch has similar word-level prosody to English, it has quite distinct sentential prosody and different weighting of acoustic cues to clausal units (Johnson & Seidl, 2008).

variability and repeated accent exposure. In addition, recent work suggests that exposure to variable input positively impacts listeners' speaker identification skills, in that adults exposed to a particular accent or dialect are better at talker identification and adapting to accented speech than are adults lacking exposure to the target accent or dialect (Clarke, 2000; Kerstholt, Jansen, van Amelsvoort & Broeders, 2006; Maye, Aslin & Tanenhaus, 2008; Remez, Wissig, Ferro, Liberman & Landau, 2004).

For infants learning multiple languages, the high degree of variability in their input may provide similar benefits. Infants growing up in multilingual environments not only hear two or more languages spoken with native phonology (e.g. native English, native Spanish), but also are likely to hear at least one of these languages spoken with a non-native accent (e.g. Spanish-accented English, English-accented Spanish; Fernald, 2006). Thus, familiarity with non-standard forms of an input language may influence the development of more abstract and flexible phonological representations. McLennan (2006) suggests that 'exposure to variable speech input contributes to the emergence of robust generalized representations' (p. 119) and may account for the possibility that bilingual infants might be better at extending their representations.

Furthermore, recent work by Singh (2008) suggests that exposure to words in variable forms facilitates word recognition by encouraging the development of broader lexical categories. Thus, mere exposure to the variable characteristics of another language might afford benefits in phonological acquisition. In this case, we would expect that 9-month-old Spanish-English bilingual infants or even English-learning infants exposed to a consistent form of Spanish-accented English would succeed in Experiments 2 and 5, as their phonological representations may be more accommodating to phonological variation as a result of the increased variability in the words heard in their input language(s). Such flexibility may also facilitate token generalization, perhaps accounting for the difficulty exhibited by the English-learning infants in the aforementioned experiments.

Our findings are the first to demonstrate how infants confront variability in the form of a foreign accent, a situation with which infants are often faced. These findings further provide important information about how English-learning infants come to approximate the generalization abilities of adults in accommodating accent and voice variability in the speech stream. Future research should explore the role of exposure and experience with input variability in impacting word recognition and generalization. To investigate this question, an experimental manipulation could be conducted in which English-learning 9-month-olds listen to Spanish-accented speech every day for one week and then are tested on Experiments 2 and 5 to examine whether exposure to the variability of foreign-accented speech facilitates the development of more robust and abstract lexical generalizations at a younger age. Such an experiment may provide insight into how exposure to variable forms of an input language affects

the development of representations in infants learning one or multiple languages.

Acknowledgements

This work was supported by a grant from the National Institute of Child Health and Human Development (RO3 HD046463) to A.S. and formed part of a doctoral dissertation submitted by R.S. to the Purdue University Graduate School. We thank Alejandrina Cristià, Ilana Feld, Ashley Tempel, and the undergraduate research assistants in the Purdue Infant Research Labs for assistance in testing infants.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, IL: Aldine.
- Boersma, P., & Weenik, D. (2001). *Praat 4.1.18: A system for doing phonetics by computer*. <http://www.praat.org>
- Bortfeld, H., Morgan, J.L., Golinkoff, R.M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, **16**, 298–304.
- Bradlow, A.R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, **97**, 1916–1924.
- Bradlow, A.R., & Bent, T. (2003). Listener adaptation to foreign-accented English. In M.J. Sole, D. Recasens and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2881–2884). Barcelona, Spain: Causal Productions.
- Clarke, C.M. (2000). Perceptual adjustment to foreign-accented English. *Journal of the Acoustical Society of America*, **107**, 2856.
- Clarke, C.M., & Garrett, M. (2004). Rapid adaptation to foreign-accented speech. *Journal of the Acoustical Society of America*, **116**, 3647–3658.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craik, F.I.M., & Kirsner, K. (1974). The effect of speakers' voice on word recognition. *Quarterly Journal of Experimental Psychology*, **26**, 274–284.
- Fennell, C.T., & Werker, J.F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, **46**, 245–264.
- Fernald, A. (2006). When infants hear two languages: interpreting research on early speech perception by bilingual children. In P. McCardle and E. Hoff (Eds.), *Childhood bilingualism: Research on infancy through school age* (pp. 19–29). Clevedon, UK: Multilingual Matters Ltd.
- Gelfer, M.P. (1993). A multidimensional scaling study of voice quality in females. *Phonetica*, **50**, 15–27.
- Goh, W.D. (2005). Talker variability and recognition memory: Instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 40–53.
- Goldinger, S.D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal*

- of *Experimental Psychology: Learning, Memory, and Cognition*, **22**, 1166–1183.
- Goldinger, S.D., Pisoni, D.B., & Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 152–162.
- Hirsh-Pasek, K., Nelson, D.G.K., Jusczyk, P.W., Cassidy, K.W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, **26**, 269–286.
- Houston, D.M. (2000). The role of talker variability in infant word representations (PhD thesis, Johns Hopkins University, 1999). *Dissertation Abstracts International*, **60**, 5802.
- Houston, D.M., & Jusczyk, P.W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1570–1582.
- Houston, D.M., Jusczyk, P.W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Both Dutch- and English-learning 9-month-olds segment Dutch words from fluent speech. *Psychonomic Bulletin and Review*, **7**, 504–509.
- Johnson, E., & Seidl, A. (2008). A cross-linguistic perspective on the detection of juncture in speech. *Infancy*, **18**, 440–455.
- Jongman, A., & Wade, T. (2007). Acoustic variability and perceptual learning. In O.S. Bohn and M.J. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 135–150). Amsterdam: John Benjamins Publishing Company.
- Jusczyk, P.W. (1993). From language-general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, **21**, 3–28.
- Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, **29**, 1–23.
- Jusczyk, P.W., & Luce, P.A. (2002). Speech perception and spoken word recognition: past and present. *Ear and Hearing*, **23**, 2–40.
- Jusczyk, P.W., Hirsh-Pasek, K., Kemler Nelson, D.G., Kennedy, L., Woodward, A., & Piwoz, J. (1992a). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, **24**, 252–293.
- Jusczyk, P.W., Pisoni, D.B., & Mullenix, J. (1992b). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, **43**, 253–291.
- Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687.
- Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39**, 159–207.
- Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., & Gerken, L.A. (1995). The Headturn Preference Procedure for testing auditory perception. *Infant Behavior and Development*, **18**, 111–116.
- Kerstholt, J.H., Jansen, N.J.M., van Amelsvoort, A.G., & Broeders, A.P.A. (2004). Earwitnesses: effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology*, **18**, 327–336.
- Kraljic, T., & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, **13**, 262–268.
- Krentz, U.C., & Corina, D.P. (2008). Preference for language in early infancy: the human language bias is not speech specific. *Developmental Science*, **11**, 1–9.
- Kuhl, P.K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, **66**, 1668–1679.
- Kuhl, P.K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, **6**, 263–285.
- Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America*, **35**, 451–453.
- McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, **26**, 451–458.
- McLennan, C.T. (2006). The time course of variability effects in the perception of spoken language: changes across the lifespan. *Language and Speech*, **49**, 113–125.
- Martin, C.S., Mullenix, J.W., Pisoni, D.B., & Summers, W.V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 676–684.
- Maye, J., Aslin, R.N., & Tanenhaus, M.K. (2008). The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science*, **32**, 543–562.
- Mayo, C., & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America*, **115**, 3184–3194.
- Mullenix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365–378.
- Munro, M.J., & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, **45**, 73–97.
- Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204–238.
- Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, **60**, 355–376.
- Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42–46.
- Polka, L., & Sundara, M. (2003). Word segmentation in monolingual and bilingual infant learners of English and French. In M.J. Solé, D. Recasens and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1021–1024). Barcelona, Spain: Causal Productions.
- Polka, L., Sundara, M., & Blue, S. (2002). The role of language experience in word segmentation: a comparison of English, French, and bilingual infants. Paper presented at the 143rd meeting of the *Acoustical Society of America*. Pittsburgh, Pennsylvania.
- Remez, R.E., Fellowes, J.M., & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 651–666.
- Remez, R.E., Van Dyk, J.L., Fellowes, J.M., & Rubin, P.E. (1998). On the perception of qualitative and phonetic similarities of voice. In P.K. Kuhl and L.A. Crum (Eds.), *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America* (pp. 2063–2064). New York: Acoustical Society of America.

- Remez, R.E., Wissig, S.C., Ferro, D.F., Liberman, K., & Landau, C. (2004). A search for listener differences in the perception of talker identity. *Journal of the Acoustical Society of America*, **116**, 2544.
- Rogers, C.L. (1997). *Intelligibility of Chinese-accented English* (PhD thesis, Indiana University, 1997). *Dissertation Abstracts International*, **58**, 9.
- Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, **11**, 596–606.
- Shah, A.P. (2004). Production and perceptual correlates of Spanish-accented English. *Proceedings of the MIT Conference: From sound to sense: 50+ years of discoveries in speech communication*, Cambridge, MA: MIT.
- Sheffert, S.M., Pisoni, D.B., Fellowes, J.M., & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 1447–1469.
- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, **106**, 833–870.
- Singh, L., Morgan, J.L., & White, K.S. (2004). Preference and processing: the role of speech affect in early spoken word recognition. *Journal of Memory and Language*, **51**, 173–189.
- Soderstrom, M., Seidl, A., Kemler Nelson, D.G., & Jusczyk, P.W. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *Journal of Memory and Language*, **49**, 249–267.
- Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, **388**, 381–382.
- Strange, W. (Ed.) (1995). *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press.
- Swingle, D., & Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, **76**, 147–166.
- Swingle, D., & Aslin, R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, **13**, 480–484.
- Verbrugge, R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, **60**, 198–212.
- Wannemacher, C.A. (1995). The relationship between processing of talker voice and phonetic information in speech perception (PhD thesis, University of New York at Buffalo, 1995). *Dissertation Abstracts International*, **56**, 2488.
- Weil, S.A. (2001). Foreign-accented speech: adaptation and generalization. Unpublished Master's thesis, The Ohio State University, Columbus, OH.
- Werker, J.F., & Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development*, **1**, 197–234.
- Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49–63.

Received: 27 July 2007

Accepted: 21 May 2008