Using Speech Sounds to Guide Word Learning: The Case of Bilingual Infants

Christopher T. Fennell  
University of Ottawa

Krista Byers-Heinlein and Janet F. Werker  
University of British Columbia

Despite the prevalence of bilingualism, language acquisition research has focused on monolingual infants. Monolinguals cannot learn minimally different words (e.g., “bih” and “dih”) in a laboratory task until 17 months of age (J. F. Werker, C. T. Fennell, K. M. Corcoran, & C. L. Stager, 2002). This study was extended to 14- to 20-month-old bilingual infants: a heterogeneous sample (English and another language; \( N = 48 \)) and two homogeneous samples (28 English–Chinese and 25 English–French infants). In all samples, bilinguals did not learn similar-sounding words until 20 months, indicating that they use relevant language sounds (i.e., consonants) to direct word learning developmentally later than monolinguals, possibly due to the increased cognitive load of learning two languages. However, this developmental pattern may be adaptive for bilingual word learning.

The vast majority of experimental work in language development, particularly in the area of early lexical acquisition, has focused on monolingual infants. Yet, due to immigration, official language policies, and cultural preferences and norms, many infants are raised in a bilingual environment. The simultaneous exposure to two languages, each with different sound inventories and lexicons, leads to a very different set of acquisition challenges for bilingual infants. To date, it is unknown whether the burgeoning literature linking speech perception to early word learning in monolinguals will generalize to all infants. The set of experiments presented in this article will explore the theoretical and empirical claims about the relationship between speech perception and word learning by extending this line of research to infants raised in a bilingual environment.

The task of learning two languages from birth is potentially not an easy one. Bilingual infants must accomplish dual language learning in an environment that is not always overtly structured with respect to language exposure. Many infants hear their two languages from a multitude of sources (e.g., parents, other family members, close friends, surrounding community). Even though they may not have systematic language division, bilingual infants acquire words in both of their languages from the beginning of lexical acquisition (Pearson, Fernández, & Oller, 1995; Quay, 1995). When compared to monolinguals, they pass language milestones at similar ages (de Houwer, 1995; Oller, Eilers, Urbano, & Cobol Lewis, 1997; Pearson & Fernandez, 1994; Petitto et al., 2001) and have similar-sized vocabularies, when taking words from both languages into account (Pearson & Fernandez; Pearson, Fernández, & Oller, 1993; Petitto et al.). Although suggestive of important similarities between monolinguals and bilinguals, these outcome measures reveal little about the process by which bilingual infants build their lexicons, specifically how they use their speech perception skills to guide word learning. Do the steps bilingual-learning infants take in linking speech perception to word learning mirror those taken by monolinguals, or do bilingual-learning infants take a different path to achieve similar outcomes?

**Speech Perception**

Before beginning the process of word learning in earnest, infants must first refine their perceptual sensitivities to the sound contrasts that distinguish one word from another in their native language, and bilingual-learning infants must accomplish this goal for each of their native languages. For example, an English-learning infant needs to attend to the measurable and perceptible information that distinguishes [b] from [d], as these sounds can differentiate English words (e.g., “bad” from “dad”). The term *phonetic* denotes these physical properties of speech sounds;
Linking Speech Perception to Word Learning

The use of native-language phonetic categories to guide word learning is a crucial step linking perceptual development to language acquisition and can serve as a potentially sensitive juncture to reveal when and if there are differences in monolingual and bilingual development. Monolingual research demonstrates that 7.5-month-old infants are able to segment and recognize familiar word forms, discriminating these words from similar-sounding distracters (e.g., Jusczyk & Aslin, 1995). However, starting at around 11 months of age there are many tasks in which infants treat mispronunciations of known words as if they were in fact real words (Hallé & de Boysson-Bardies, 1996; Mills et al., 2004; Vihman, DePaolis, Nakai, & Hallé, 2004). Although these word segmentation and recognition studies are informative, they do not directly answer the question of how infants apply speech perception skills to the act of word learning.

To study the link between speech perception and word learning it is essential to use a methodology that requires the infant to demonstrate some evidence of associating a meaning with the word. Moreover, the task must be appropriate for use with infants who have established their native language perceptual...
categories, but are still at the beginning of the word learning period. Hence, we chose a task that requires the infant to learn the associative link between a word and an object. Associative word learning does not necessarily require the referential understanding that a word can stand for an object in the absence of the object (Hirsh-Pasek & Golinkoff, 1996), but it is a critical step in word learning that infants this age have mastered (Werker, Cohen, Lloyd, Casasola, & Stager, 1998).

We used the Switch, word-object associative task (Werker et al., 1998) as it has been used successfully with infants from 14 to 20 months of age (e.g., Werker, Fennell, Corcoran, & Stager, 2002). In this task, infants are habituated to two word–object combinations (Word A–Object A; Word B–Object B) and then tested on both a Switch (Word A–Object B) and a Same (Word B–Object B) trial. If infants have learned the words and the objects individually, but have not learned their associative link, the Switch and Same trials would be equally familiar, and should elicit equivalent looking times. Thus, if infants are treating the procedure as a straight speech perception task and attending only to the words, they will not succeed in the associative task. If, however, infants have learned the associative link, the pairing violation in the Switch trial will generate surprise and hence greater looking time than the Same trial.

In previous work using the Switch task (Werker et al., 1998), it has been shown that monolingual infants as young as 14 months successfully detect the “switch” in word–object pairings when the two words are not similar sounding (e.g., “lif” and “neem”). However, monolingual infants of 14 months fail if the words differ in only one sound (e.g., “bih” and “dih”) (Pater, Stager, & Werker, 2004; Stager & Werker, 1997). Indeed, it is not until 17 months of age that they successfully learn similar-sounding words such as “bih” and “dih” in the Switch task (Werker, et al., 2002). This failure to associate similar-sounding words with new objects is seen even though the infants easily discriminate the same word forms when they are not required to link the words to objects (Stager & Werker, 1997).

It appears that the difficulty in distinguishing similar-sounding words stems from the increased computational difficulties that arise when infants begin listening for meaning and not just form (Nazzi & Bertoncini, 2003; Naigles, 2002) and is most strikingly evident when young word learners first attempt to use native phonetic categories to guide word learning (Stager & Werker, 1997). This explanation based on increased computational load has been labeled the resource limitation hypothesis (Stager & Werker, 1997; Werker & Fennell, 2004).

Many factors may make word–object associative tasks computationally demanding for novice word learners of 14 months, including the requirements of simultaneously listening to new words, seeing new objects, and establishing the link between the two (Werker & Fennell, 2004). The difficulty is compounded by the fact that infants arguably do not yet know that word meanings are contrasted by differences in speech segments (e.g., [b] vs. [d]), rather than by differences in speaker affect (Singh, Morgan, & White, 2004) or other indexical variables such as gender, age, and so forth (see Werker & Curtin, 2005). As is predicted by the resource limitation hypothesis, if the demands of the word learning situation are lightened, infants can successfully encode and retrieve the contrastive phonetic detail needed to distinguish the similar-sounding words. For example, infants of 14 months succeed in the following situations: when tested on words they already know, such as “ball” and “doll” (Fennell & Werker, 2003; Swingley & Aslin, 2002), when learning word–object pairings that consist of familiar objects (Fennell & Werker, 2004), and when tested in a task that reduces memory demands (Ballem & Plunkett, 2005).

Growing up bilingual might make performance in word–object associative tasks particularly challenging, thus increasing cognitive load. Bilinguals need to activate and use only the phonetic detail that is appropriate to the language in which the new words are being presented. Whether the establishment of native speech sound categories is temporarily delayed (the vowels in Bosch & Sebatsián-Gallés, 2003) or even if two categories are established in a tight phonetic space (consonant VOTs in Burns et al., 2007), it could be predicted that the computational demands of using these representations to guide word learning are greater for bilinguals. The structure of the bilingual lexical system(s) may make learning words in more than one language more computationally demanding in other ways as well. Bilingual infants have to acquire more object labels and language sounds than do monolingual infants either (1) by acquiring two separate lexicons and phonetic systems simultaneously, thus putting extra demand on their overall acquisition system (e.g., Paradis & Genesee, 1996); or (2) by having one undifferentiated system with additional cognitive demands due to the increased number of phonetic categories and object labels that need to be acquired (e.g., Volterra & Taeuschner, 1978). Although the current study is not designed to test these two hypotheses, it is important to note that either of these possibilities would increase the cognitive resources needed for word bilingual learning. Thus, we predict that bilingual infants will
show a later use of relevant phonetic detail in novel words as compared to their monolingual peers.

An alternative possibility is that, because bilingual infants have vocabulary sizes similar to those of monolinguals and achieve language milestones at comparable ages (e.g., Pearson & Fernandez, 1994; Petitto et al., 2001), bilinguals’ and monolinguals’ use of phonetic detail in word learning could follow a similar developmental trajectory and emerge at the same age. We tested the resource limitation hypothesis against this alternative prediction by examining the development of bilingual infants’ ability to associatively learn two words differing in only one consonant. We used the same Switch task, the same similar-sounding stimuli, and the same age groups as the Werker et al. (2002) study, which demonstrated that monolingual English-learning infants begin to use phonetic detail to guide novel word learning after 14 but before 17 months. Because there are no changes to the test design, any difference in the results from this study and the results found in the Werker et al. study will be attributable to growing up with two languages rather than one.

**Experiment 1**

**Method**

**Infants**

Forty-eight bilingual infants successfully completed the study, 16 infants (8 boys) in each of three age groups: 14, 17, and 20 months of age. The infants’ mean age (and range) in each group was 14 months, 12 days (13 m, 22 d to 15 m, 5 d), 17 months, 14 days (16 m, 24 d to 18 m, 10 d), and 20 months, 16 days (19 m, 24 d to 20 m, 31 d). An additional 54 infants were tested but not included in the analyses due to fussiness (12 at 14 months, 15 at 17 months, and 14 at 20 months), parental interference (2 at 14 months, 2 at 17 months, and 3 at 20 months), experimenter/technical error (2 at 14 months and 1 at 17 months), being off-camera (1 at 14 months and 1 at 17 months), or sleepiness (1 at 14 months). As race/ethnicity was not central to our research question, it was not included in the demographic questions asked to the parent(s) of the infants. However, we can safely state that the sample was very diverse with respect to race and ethnicity for two major reasons. First, the fact that this is a bilingual sample increased the ethnic diversity of the sample: 52% of the infants were learning a non-European language at home as one of their languages (see the Appendix), and even those learning a European language at home (e.g., French or Spanish) were not uniformly Caucasian (e.g., Haitian–French or Hispanic from South America). Second, the infants were all from the Vancouver area, one of the most racially and ethnically diverse cities in Canada. The last Canadian census (Statistics Canada n.d.) reported that 46% of the residents of Vancouver were born outside Canada and 49% of Vancouver residents belong to a visible minority group. This diversity was reflected in our sample.

All subjects were without apparent health or hearing problems and were at least 37 weeks gestation. The infants were selected from the infant database at the Infant Studies Centre at the University of British Columbia. Participating infants were given an “Infant Scientist” T-shirt and a diploma.

To be considered bilingual, an infant needed to have a maximum of 70% exposure to one language and a minimum of 30% exposure to the other. These limits are based on, but slightly more conservative than, the inclusion criteria for child bilingualism studies recommended by Pearson, Fernández, Lewedeg, and Oller (1997). Another requirement for inclusion in the study was exposure to the two languages from birth. In other words, all infants were native bilinguals.

The infants’ language exposure was measured by the Language Exposure Questionnaire, which has been used to classify bilinguals in previous published research (Bosch & Sebastián-Galleés, 1997). The questionnaire requires parents to provide precise estimates of the infant’s exposure to both languages. An estimate is given for each major caregiver in the infant’s life (e.g., parents, grandparents, childcare workers), which is critical for quantifying bilingual exposure (de Houwer, 1995). A full breakdown of participants’ languages and percent exposure to each language is found in the Appendix.

**Audio Stimuli**

A female native speaker of English recorded three nonsense words in a soundproof booth: “bih” [bɪ], “dih” [dɪ], and “pok” [pɒk]. The first two, “bih” and “dih”, were used as habituation and test stimuli. The phonetic detail by which these labels differ relates to the place of articulation of the initial consonant, the place in the mouth where contact is made to produce the consonant. The [b] is produced by the lips contacting one another and the [d] is produced by the tongue touching the alveolar ridge. These consonants are identical to those used in both Stager and Werker (1997) and Werker et al. (2002). The third label, “pok,” was used in the pre- and posttest trials. “Pok” is different from the habituation labels in syllable form.
(a final consonant), in the VOT of the initial consonant, and in the vowel, so it is unlikely to be confused with the test words. Seven exemplars of each label were recorded, all in an infant-directed manner of speech. Each exemplar was around 0.7 s in duration, with a silent 1.5 s interval between tokens. The first three were presented again at the end of the sequence, producing an audio file of 20.0 s in duration, as in Werker et al. (2002).

To make the study comparable to the previous monolingual study, it was important to ensure that bilingual infants could discriminate the English \([b]–[d]\) contrast in simple speech perception tasks, especially considering the mixed results from previous research with bilingual infants (i.e., the difficulties seen in Bosch & Sebastián-Galleés, 2003, and the success seen in Burns et al., 2007). In a preliminary study, Fennell (2005) confirmed that bilingual infants learning English as one of their native languages could discriminate this contrast at 14 months.

**Video Stimuli**

The visual stimuli used in the habituation and test phases of the study consisted of two attractive objects, “crown” and “molecule.” The crown object was made from red, blue, and yellow colored modeling clay (see Figure 1a), and the molecule object was made using the green and turquoise components from a chemistry set (see Figure 1b). A store-bought, multicolored toy water wheel was used for both the pre- and posttests (see Figure 1c). All three objects were videotaped against a black background and then transferred to laser disc format for the experiment with 14-month-olds and digitized for the experiment with 17- and 20-month-olds. The crown and molecule objects were filmed moving back and forth across the screen at a slow and constant speed. The water wheel was filmed with its arms moving around in a rotating motion. These visual stimuli were identical to the stimuli used in Werker et al. (2002).

**Apparatus**

For the groups of 17- and 20-month-olds, testing took place in a 2.8 m × 2.3 m quiet room, dimly lit by a shaded 60 W lamp situated 80 cm to the left of the infant at a 45° forward angle. Infants sat on parents’ laps facing a 27-in. video monitor that was approximately 1.2 m from the infants. The audio stimuli played at 65 dB, ±5 dB, over two speakers located below the monitor, one to the left and one to the right. The monitor was surrounded by black cloth, which stretched the width and height of the room. Infants were recorded using a digital video camera. The lens of the digital video camera peeked out of a hole in the cloth located 21 cm below the monitor. As a masking control during testing, parents wore headphones over which female vocal music was played from a CD player. A Macintosh Power PC G4 ran Habit 2000, a computer program developed by the Infant Cognition Laboratory at the University of Texas at Austin, to order stimuli presentation and collect looking time data. Both the visual and audio stimuli played from digitized files on the computer and were sent to the monitor and speaker in the testing room.

The experimenter, who was blind to the audio stimuli being presented and to whether a trial was a habituation or test trial, monitored infants’ looking times via a closed circuit television system from an...
adjacent testing room. A designated key was pressed on the computer keyboard during infant looks, which the Habit 2000 program recorded. The video record was used for subsequent reliability coding.

For the group of 14-month-olds, whose trials were completed prior to the other two groups, we used a slightly different setup, and only the differences will be highlighted. The experiment took place in a 2.1 m × 2.8 m room. The audio stimuli were presented from one speaker, located directly above the monitor. A Power Mac 8500/120, which was linked with a laser-disc player, ran the Habit program. The visual stimuli from the laser disc and the audio stimuli from a digitized audio file were synchronized and sent to the monitor and speaker in the testing room.

**Procedure**

Infants sat on parents’ laps facing the television screen, with the parents wearing headphones. After ensuring parents understood the instructions, the experimenter left the room. The experiment began when the experimenter could see infants’ eyes focusing on the television screen via a closed circuit television system. A flashing red light or animated oval preceded the first trial and all subsequent trials to draw attention to the video display. When infants looked at the screen, the researcher initiated a trial by pressing a computer key.

Infants were tested using the Switch procedure, originally designed by Werker et al. (1998). The specific Switch experimental design used in this study matched Werker et al. (2002). During the habituation phase, the infants were presented with two word–object pairings via the television and sound system. One pairing consisted of the crown object paired with the label “bih” (Pair A) and the other pairing was the molecule object paired with the label “dih” (Pair B). Each block of four trials consisted of two trials of each pairing type, resulting in six possible orders of the pairings within a block (e.g., ABAB, BAAB, etc.). The presentation of these six orders was randomized during the habituation phase. Presentation of trial blocks continued until one of two criteria were met: the infant’s total looking time in a block decreased to 50% of the total looking time during the first block or the 24 habituation trials were completed. All infants reached the looking time habituation criterion, with the exception of five 14-month-old participants who reached the second criterion of 24 completed habituation trials. Once one of the habituation criteria had been met, two test trials were presented. The Same test trial consisted of one of the pairings from the habituation phase (e.g. crown–“bih”). In the Switch trial, one object was labeled with the other object’s name (e.g. molecule–“bih”). If the infants have learned about the words and the objects but have not learned the associative link, the Same and Switch trials will be equally familiar and should attract equal looking times. However, if the infants have learned the link between the specific words and objects, the Switch trial, as a violation of that link, should attract greater looking time than the Same trial.

The researcher assigned one male and one female infant to each of eight possible presentation orders in the test phase of the experiment. These orders counterbalanced which trial type occurred first (Same or Switch) and which sound–object combinations were presented for each of these trials. The water wheel–“pok” combination, a completely novel word–object pairing, was presented after the test trials to ensure that the infants recovered to a large change and were not fatigued or generally disinterested in the task. This combination had also preceded the habituation phase as a pretest.

**Coding**

For online coding, the experimenter pressed a computer key when the infant was looking at the visual display and released it when the infant looked away from the display. A second coder then performed reliability tests off-line by scoring 25% of the infants’ videotaped responses. A Pearson correlation of online (first coder) and off-line (second coder) test trial scores was calculated and found to be significant, $r(22) = .997, p < .001$.

**Results**

All looking time measures in this experiment and Experiment 2 are reported in seconds.

A paired sample $t$ test comparing total looking time in the first block of four habituation trials ($M = 15.09$, $SD = 3.53$) to that in the last block of four habituation trials ($M = 7.41$, $SD = 2.87$) was conducted for the group of 14-month-olds, as 5 infants in this group did not reach the habituation criterion of a 50% decrease in looking time, but instead reached the criterion of completing 24 habituation trials. There was a significant reduction in looking time from the first to last block, $t(15) = 6.60, p < .001$, demonstrating that the 14-month-old infants had habituated to the stimuli as a group. All 17- and 20-month-old infants reached the looking time habituation criterion. We then conducted a 2 (recovery: last habituation block vs. posttest) × 3 (age: 14, 17, 20 months) mixed ANOVA to test recovery to the posttest from the last block. The main effect of
recovery was the only significant effect, $F(1, 45) = 850.19, p < .001$: infants looked significantly longer to the posttest ($M = 18.65, SD = 1.99$) than to the last block ($M = 6.52, SD = 2.48$), thus indicating that they were not fatigued or generally disinterested in the task.

To directly test the question of whether bilingual infants can learn similar-sounding words and whether this ability changes across development, a mixed 2 (test trial type: same vs. switch) × 2 (gender: male vs. female) × 3 (age: 14, 17, and 20 months) ANOVA was conducted. Gender was included as a factor because one previous Switch study revealed a gender difference in performance (Werker et al., 1998). Bilinguals did not look longer to Switch ($M = 7.69, SD = 5.35$) than to Same ($M = 6.73, SD = 3.90$) as a group, $F(1, 45) = 1.27, p = .27$. There were no main effects of gender or age. However, the Age × Test Trial interaction approached significance, $F(2, 45) = 2.96, p = .07$. Because our primary motivation was to ascertain the age at which bilingual infants succeed in this task, this warranted a further examination of the bilinguals’ ability to notice the Switch at each age level.

A series of $t$ tests were run to determine which age group, or groups, looked significantly longer to the Switch trials. At 14 months, there was no significant difference between Same ($M = 7.08, SD = 3.55$) and Switch ($M = 7.68, SD = 5.21$) trials, $t(15) = -.50, p = .62$. At 17 months, Same ($M = 6.72, SD = 4.92$) and Switch ($M = 5.36, SD = 3.99$) also did not significantly differ, $t(15) = .78, p = .44$. However, bilingual infants of 20 months looked significantly longer at Switch ($M = 10.06, SD = 5.85$) over Same ($M = 6.50, SD = 3.49$), $t(15) = -2.74, p = .015$ (see Figure 2).

An analysis of the relationship between level of bilingualism and performance on the task was conducted. Level of bilingualism was determined by subtracting the percentage exposure to the nondominant language from that of the dominant language (e.g., 62% exposure to one language and 38% exposure to the other results in a score of 24). Therefore, scores closer to zero ($50 - 50 = 0$) reflect more balanced exposure to the infant’s two languages. Task performance was computed by subtracting the looking time to the Same trial from that of the Switch trial. Higher positive scores on this measure indicate better performance on the task (longer looking times to Switch). There were no significant overall correlations between level of bilingualism and task performance when all three age groups were combined, $r(46) = .015, p = .92$, nor were the correlations between these variables significant in any age group alone (all $ps > .6$).

A final analysis was done to probe for a relationship between amount of exposure to English and task performance. There was no overall correlation when all ages were combined. Examining this relationship at each age revealed only one significant correlation, that at 14 months, $r(14) = .531, p = .03$. Although it is possible that English exposure only aids younger bilinguals, the lack of significant correlations at older ages and of an overall correlation make any strong conclusions premature.

**Discussion**

The results from Experiment 1 indicate that bilingual infants begin using phonetic detail to guide learning of novel words at a later age than their monolingual peers. The bilingual infants did not successfully use phonetic detail to guide associative word learning until 20 months of age, whereas the monolingual infants in Werker et al. (2002) did so at 17 months of age. Thus, these results are consistent
with the predictions of the resource limitation hypothesis (e.g., Fennell & Werker, 2003; Werker & Fennell, 2004): The added demands of bilingual word learning (i.e., more object labels and phonetic contrasts than monolinguals, multiple category boundaries in confined perceptual space) lead to a protracted period of difficulty encoding and retrieving relevant phonetic detail in novel words.

To confirm the difference between monolingual and bilingual development, we compared the data from Experiment 1 with the data from Werker et al. (2002). Specifically, infants’ looking times were compared in a 2 (test trial: Same vs. Switch) × 2 (language exposure: monolingual vs. bilingual) mixed ANOVA at each age: 14, 17, and 20 months. The results of these ANOVAs are summarized in Table 1 and the looking times are presented in Figure 2. As expected, there were no significant main effects at 14 months and no interaction between test trials and language exposure, confirming that both groups failed to use the relevant phonetic detail in novel words at this age. At 20 months, there was a significant main effect for test trial, but no group interaction. Both monolinguals and bilinguals looked significantly longer to the Switch ($M = 12.5$ and $10.1$, respectively) than to the Same trial ($M = 9.5$ and $6.5$, respectively), showing that 20-month-old infants, irrespective of language background, successfully encoded and retrieved phonetic detail present in novel words. The key group was the 17-month-old infants, as this is the age at which it appears that monolingual and bilinguals differ in their development. The analysis confirmed this difference. There was no main effect for test trial, but the interaction between groups very closely approached significance. Paired $t$ tests showed that monolinguals looked significantly longer, $t(15) = 2.14, p = .049$, to the Switch ($M = 11.7$) than to the Same trial ($M = 8.5$) whereas bilinguals had similar looking times to the Same ($M = 6.6$) and Switch ($M = 5.3$) test trials, $t(15) = .76, p = .44$. There was also a main effect of language exposure at 17 months, with monolinguals’ total looking time to the test trials ($M = 20.22$) significantly higher than that of bilinguals ($M = 11.93$). Similarly, the main effect for language exposure approached significance in the 20-month-old group, with the monolinguals again looking longer overall to both test trials. Without further research we can only speculate as to whether overall looking time differences are informative as to underlying cognitive processing. What is important to note is that the pattern of looking to the Switch and Same trials was identical across the monolingual and bilingual groups at 14 and 20 months, and only differed at 17 months.

The possibility remains that the linguistic variability in our sample (i.e., the various languages across bilinguals) influenced the outcome. Studies with older bilingual children have revealed language group differences in the processing of phonetic contrasts (e.g., Spanish – English vs. Cantonese – English bilinguals in Bialystok, Majumder, & Martin, 2003). The precise phonetic realization of any specific consonant or vowel varies across languages, and it is quite likely that this variability was expressed in different ways in the various non-English languages to which the infants were exposed. It would be informative to restrict the participant samples to specific bilingual groups whose languages differ in the cross-linguistic overlap of the specific sounds that distinguish the similar-sounding words tested. This would allow us to explore if the results from Experiment 1 are replicated in homogeneous groups of bilinguals.

### Table 1

<table>
<thead>
<tr>
<th>Source</th>
<th>14 months</th>
<th></th>
<th>17 months</th>
<th></th>
<th>20 months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>$p$</td>
<td>$F$</td>
<td>$p$</td>
<td>$F$</td>
<td>$p$</td>
</tr>
<tr>
<td>Test trial (same vs. switch)</td>
<td>0.04</td>
<td>.84</td>
<td>0.70</td>
<td>.41</td>
<td>12.59*</td>
<td>.01*</td>
</tr>
<tr>
<td>Language exposure (monolingual vs. bilingual)</td>
<td>1.17</td>
<td>.29</td>
<td>12.90*</td>
<td>.01*</td>
<td>3.21+</td>
<td>.08+</td>
</tr>
<tr>
<td>Test Trial × Language Exposure</td>
<td>0.10</td>
<td>.75</td>
<td>3.95*</td>
<td>.06*</td>
<td>0.11</td>
<td>.75</td>
</tr>
</tbody>
</table>

Note. The $df$ for all statistics presented in the table are 1, 30.  
*p < .10; *p < .05.

### Experiment 2

To obtain homogenous groups of bilinguals, we augmented the sample sizes of the two largest subgroups of bilingual infants found in Experiment 1,
Chinese–English (including both Mandarin and Cantonese, which share the language sounds being tested) and French–English bilinguals. We tested both groups of bilinguals at 17 and 20 months of age using the same task and stimuli as in Experiment 1. Because no group of infants had succeeded in the task with novel words at 14 months, we did not test infants at this younger age.

Infants in Experiment 2 were tested on the same English [b]–[d] contrast as in Experiment 1. As argued above, the extent to which [b] and [d] overlap across languages may lead to performance differences on this task. The place of articulation of the [b] sound is the same (released from the lips) across the three languages represented among the participants: English, French, and Cantonese (and Mandarin). However, the place of articulation of the [d] sound differs across the languages. When compared to English, Cantonese and Mandarin have a similar place of articulation: The tongue touches the alveolar ridge just behind the front teeth (Chan & Li, 2000; Dow, 1972; but see Bauer & Benedict, 1997, for an argument that the Cantonese [d] is produced with the tongue tip closer to the teeth than the English [d]). Unlike English, the French [d] is fully dental: the tongue tip actually touches the teeth (Picard, 1987). The VOT for the sounds corresponding to [b] and [d] is also more similar across English and Cantonese than across English and French (Caramazza et al., 1973; Lisker & Abramson, 1964; MacLeod & Stoel-Gammon, 2005). It should be noted that [b] and [d] sounds in Cantonese and Mandarin are realized as unaspirated (i.e., lacking a strong burst of air when released) [p] and [t] sounds, respectively.

Although here we have characterized some differences and similarities of the consonants of interest in English, French, and Chinese, it is a considerable challenge to envisage how these may interact in language development. There has been considerable research comparing the impact of shared versus different properties of language sounds on cross-linguistic speech perception (e.g., Escudero & Boersma, 2004; Peperkamp & Dupoux, 2002), but there is no bilingual research examining how such similarities and differences impact the development of successful word differentiation in infancy. Thus, it is difficult to even speculate as to whether the greater differences between the French and English consonants will help or hinder the bilinguals’ use of the English contrasts in a word learning study. For these reasons, we do not have a directional hypothesis with respect to the language groups. Nonetheless, the differences between the groups with respect to target consonants allow us to explore whether language-specific information can influence use of consonants in early word learning.

Method

Infants

A total of 53 infants successfully completed the study. The infants’ details are given below, grouped by language background: Chinese–English and French–English. All subjects in Experiment 2 were without apparent health or hearing problems and were at least 37 weeks gestation at birth. Infants were recruited in the same manner as in Experiment 1. As stated in the previous participants section, race/ethnicity was not included as a demographic question to the parents because it was not central to our research question. However, we can safely state that an overwhelming majority of the Chinese–English sample had at least one parent who would identify as Asian–Canadian. Although the French–English sample was mostly from European–Canadian backgrounds, it was not homogeneous, as some parents, in the course of our conversations with them after the experiment, mentioned that they came from other ethnic backgrounds (e.g., Haitian–French).

Bilingual exposure was assessed in the same manner as in Experiment 1; however, we changed the bilingual criteria slightly in order to increase our sample size (due to the increased difficulty of obtaining homogeneous language groups). Infants in Experiment 2 were considered bilingual if they had a maximum of 75% exposure to one language and a minimum exposure of 25% to their other language. These criteria are still within the language exposure limits recommended by Pearson et al. (1997). See the Appendix for detailed information about participants’ language exposure.

Twenty-eight Chinese–English bilingual infants, 21 Cantonese and 7 Mandarin, successfully completed the study: 16 infants at 17 months (10 boys) and 12 at 20 months (6 boys). Data for 11 of the infants came from Experiment 1 (7 at 17 months, 4 at 20 months). An additional 18 Chinese–English infants were tested for Experiment 2 but were not included because they were fussy (5 at 17 months, 6 at 20 months), off-camera (2 at 20 months), their parent interfered in some way during a test trial (1 at 17 months), or there was a technical or experimenter error (1 at 17 months, 2 at 20 months). One additional 17-month-old participant did not habituate within 24 trials and was excluded from the analyses. This was done to match the sample to that of Experiment 1, where all 17-month-old participants reached the looking time habituation criterion within 24 trials. The infants’ mean age (and range) in each group was 17 months, 17 days (17 m, 4 d to 18 m, 10 d) and 20 months, 19 days (20 m, 6 d to 21 m, 5 d).
Twenty-five French–English bilingual infants successfully completed the study, 14 infants at 17 months (5 boys) and 11 at 20 months (6 boys). Data for 9 of the infants came from Experiment 1 (6 at 17 months, 3 at 20 months). An additional 6 French–English infants were tested for Experiment 2 but not included because they were fussy (2 at 17 months, 2 at 20 months), their parent interfered in some way during a test trial (1 at 17 months), or there was a technical or experimenter error (1 at 20 months). The infants’ mean age in each group was 17 months, 13 days (16 m, 24 d to 18 m, 7 d) and 20 months, 16 days (19 m, 24 d to 21 m, 0 d).

Audio and Video Stimuli

The stimuli were the same as in Experiment 1.

Apparatus and Procedure

The apparatus and procedure were identical to those used when testing the older groups in Experiment 1.

Coding

Because the video camera used in Experiment 1 was replaced with a digital video camera for Experiment 2, we were able to use frame-by-frame coding of the data for the infants who were tested specifically for Experiment 2. The video recordings were transformed into QuickTime movies using Final Cut Pro. A highly trained coder examined the pretest, posttest, and test trials (i.e., Same and Switch trials) frame by frame (1 frame = 33 ms) to determine whether the infant was looking toward or away from the screen.

Results

All infants reached the looking time habituation criterion, so a habituation analysis was unnecessary. To ensure that the infants in these two age groups had recovered to the posttest (M = 18.36, SD = 3.19) from the last block (M = 5.96, SD = 2.04), we conducted a 2 (recovery: last habituation block vs. posttest) × 2 (age: 17 vs. 20 months) × 2 (language: Chinese vs. French) mixed ANOVA. There was a main effect of recovery, F(1, 49) = 518.80, p < .001, but no other main effects or interactions. The infants’ posttest recovery indicates that they were not fatigued or disinterested in the task.

To directly test the question of whether the two homogeneous groups of bilingual infants used the English phonetic detail in the same manner across ages, a mixed 2 (test trial type: same vs. switch) × 2 (language: Chinese vs. French) × 2 (age: 17 vs. 20 months) × 2 (gender: male vs. female) ANOVA was conducted. The results of this ANOVA are presented in Table 2. There was no main effect of trial type: As a group, the infants in this study were not able to use phonetic detail to guide word learning. There was also no main effect of language or interaction between language and test trial. Thus, it appears that the non-English language of the bilingual did not have a language-specific impact on the use of English consonants in word learning. A significant Test Trial × Gender interaction was found, as well as a significant three-way interaction between test trial, gender, and age. This indicates that the interaction between test trial and gender was not the same for each age group.

Table 2

<table>
<thead>
<tr>
<th>Source</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (A): 17 vs. 20 months</td>
<td>0.01</td>
<td>.94</td>
</tr>
<tr>
<td>Gender (G): male vs. female</td>
<td>0.05</td>
<td>.82</td>
</tr>
<tr>
<td>Language (L): Chinese vs. French</td>
<td>1.36</td>
<td>.25</td>
</tr>
<tr>
<td>A × G</td>
<td>0.42</td>
<td>.52</td>
</tr>
<tr>
<td>A × L</td>
<td>2.17</td>
<td>.15</td>
</tr>
<tr>
<td>G × L</td>
<td>0.11</td>
<td>.74</td>
</tr>
<tr>
<td>A × G × L</td>
<td>0.002</td>
<td>.96</td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test trial (T): same vs. switch</td>
<td>2.36</td>
<td>.13</td>
</tr>
<tr>
<td>T × A</td>
<td>0.10</td>
<td>.75</td>
</tr>
<tr>
<td>T × G</td>
<td>4.83*</td>
<td>.03*</td>
</tr>
<tr>
<td>T × L</td>
<td>0.26</td>
<td>.61</td>
</tr>
<tr>
<td>T × A × G</td>
<td>6.86*</td>
<td>.01*</td>
</tr>
<tr>
<td>T × A × L</td>
<td>0.01</td>
<td>.90</td>
</tr>
<tr>
<td>T × G × L</td>
<td>1.21</td>
<td>.28</td>
</tr>
<tr>
<td>T × A × G × L</td>
<td>0.39</td>
<td>.53</td>
</tr>
</tbody>
</table>

Note. The df for all statistics presented in the table are 1, 45.
*p < .05.
Switch (M = 5.61, SD = 2.84) trials for boys at 20 months, t(11) = 2.04, p = .07. However, 20-month-old female infants did look significantly longer at the Switch (M = 9.12, SD = 5.52) than to the Same (M = 5.10, SD = 2.89) test trial, t(10) = −2.58, p = .03. Therefore, in the homogeneous samples, it appears that only female infants of 20 months successfully encoded and retrieved relevant phonetic detail in novel words (Figure 3).

The same pattern of results emerged when the bilingual criteria were restricted to those used in Experiment 1 (i.e., maximum 70%, minimum 30%). However, due to the reduction in sample size from 53 to 46 infants, and the resulting loss of power, some of the relevant results under this criterion were significant at the .06 level, rather than .05.

There were no significant correlations between level of bilingualism and task performance (all ps > .1) for any group (i.e., all infants, boys, girls, 17 months, 20 months, Chinese – English, French – English). Level of English exposure was also not correlated with performance for any group (all ps > .1).

Discussion

The use of homogeneous samples of bilinguals supported the finding of Experiment 1: Bilingual infants use phonetic detail to guide associative word learning at a later age than their monolingual peers. Moreover, at least for French and Chinese, the specific other language being learned by the bilingual infant had no apparent impact on the infants’ ability to use phonetic detail in this task. As with the older infants in Experiment 1, a correlation analysis revealed no significant relationship between success in the word learning task and percent exposure to English. Further research is required to determine whether later use of phonetic detail in word learning is a universal fact of acquiring two languages or whether it is specific to the languages and contrasts being acquired (see Mattock, Polka, & Rvachew, 2006).

The homogeneous groups of bilinguals showed an unexpected gender effect. Only the female infants of 20 months in this study showed the pattern of looking longer to the Switch trial, thus demonstrating use of phonetic detail in this task. Whereas 20-month-old boys in Experiment 1 looked significantly longer to Switch, same-aged boys in Experiment 2 actually showed a trend in the opposite direction. Similar female advantage effects have been found in other infant language acquisition studies (e.g., Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Reznick & Goldfield, 1992; Woodward, Markman, & Fitzsimmons, 1994), including one other Switch experiment that has shown a female advantage in performance (Werker et al., 1998). In fact, a female advantage is seen in many areas of language learning and processing (e.g., Maccoby & Jacklin, 1974). However, most studies of word learning have demonstrated that when a female advantage is seen in word learning, the advantage is typically small (Fenson et al., 1994) and unstable across replications (Werker et al., 1998; Woodward et al., 1994).

General Discussion

Bilingual infants do not show successful use of phonetic detail in the Switch word learning task as early as their monolingual peers, who are able to use this detail at 17 months of age when tested with identical stimuli under identical conditions. This finding was replicated.
in three samples of bilingual infants: a heterogeneous sample of bilinguals (English and any other language) and two homogeneous groups of bilinguals (French – English and Chinese – English). As a group, the heterogeneous bilinguals successfully used the consonant contrasts to distinguish novel words at 20 months of age, whereas in the homogeneous groups only girls of 20 months were successful. The above results support the outcome predicted by the resource limitation hypothesis (Fennell & Werker, 2003, 2004; Stager & Werker, 1997; Werker & Fennell, 2004; Werker et al., 2002): The added demands of bilingual word learning postpone the use of phonetic detail to guide novel word learning in the Switch task.

It is useful to reconsider just how much more challenging the language acquisition task is for bilingual than for monolingual infants. As discussed before, bilingual infants must establish the phonetic properties of the consonants and vowels that distinguish words in each of their languages, even though in some cases the tight phonetic space may lead to temporary perceptual confusion (Bosch & Sebastián-Gallés, 2003). Further, bilingual infants need to learn two words at the same basic level for each object, as well as for properties, actions, and so forth. This dual labeling could raise specific challenges, for example, in applying the mutual exclusivity constraint (Byers-Heinlein & Werker, 2006). Finally, to use phonetic detail to guide word learning, the bilingual child must determine which phonetic system is required in each particular word learning context.

The resource limitation hypothesis can account for the fact that bilingual infants learn similar sounding words in an associative word learning task at a later age than do monolingual infants, but what allows infants to eventually succeed? What is it that makes the processing demands less onerous? One explanation that has been proposed is the emergence of phonemes—summary representations of the speech sound categories that distinguish meaning in the native language(s). Specifically, in a framework linking speech perception to word learning called PRIMIR, Werker and Curtin (2005; see also Curtin & Werker, in press) propose that, from the very beginning of word-form learning, infants pick up and represent words in rich detail, including not only all the phonetic properties of consonants and vowels, but also stress, intonation, speaker identity, visual correlates, syllable form, and so on. For the young infant, all of this detail is equally activated every time the infant listens to a word (see Singh et al., 2004). Thus, although discriminable, the contrastive phonetic information does not initially stand out for special attention. With the creation of a lexicon, via learning to link words with concepts, the phonetic regularities pop out as the properties that remain constant in distinguishing one meaningful word from another across variations in gender, affect, and other indexical variables. This is hypothesized to lead to the emergence of phoneme-like units. Once available, phonemes in turn serve to highlight the information (specifically, the phonetic detail) that distinguishes one meaningful word from another, while simultaneously moving indexical features, such as affect and speaker identity, into the background. In this way, the emergence of phonemes reduces the cognitive load in word learning and may allow the infant to succeed at linking similar-sounding novel words to unfamiliar objects.

The unique pattern of both speech perception development and word learning in bilingual infants predicts a later emergence of phoneme-like units. As reviewed earlier, Spanish – Catalan bilingual infants temporarily confuse a Catalan vowel difference (Bosch & Sebastián-Gallés, 2003), and English – French infants have to set up two categories in a tight phonetic space where no known language has two boundaries (Burns et al., 2007). These challenges could cause difficulties in accurately attending to the regularities that give rise to phonemes. Additionally, it may take longer for phoneme-like units to emerge in bilingual infants than in monolinguals because, although bilingual infants know about as many words as their monolingual counterparts, they may only know half as many in each language. If a minimum number of meaningful words per language is required for phoneme-like units to emerge, it will necessarily take the bilingual infant longer than the monolingual to develop phoneme representations (for a similar argument regarding bilingual infants’ sensitivities to phonotactic constraints, see Sebastián-Gallés & Bosch, 2002). This is turn keeps the computational load heavy for the bilingual infant for a longer period of time when dealing with similar-sounding words.

In addition to increased challenges in phonemic and lexical development, the same learning situation could lead to very different task demands for monolingual and bilingual infants. In a task such as the Switch procedure, there are few cues available to signal what language is being spoken. This poses little problem for the monolingual infant, as only one language is being acquired. However, the bilingual infant has to learn that there are two languages in their world and has to first ascertain which language is being spoken in every language processing situation, thus increasing the computational load even more. A manipulation wherein the language being used is made less ambiguous, such as through the use of language-specific phonotactics or sentence frames, might aid the bilingual learner.
It would also be prudent to test bilinguals in various contrasts (e.g., vowel contrasts, consonant contrasts that differ in ways other than place of articulation) to determine if the developmental pattern seen in the current experiment holds for all contrasts. Preliminary results of a study by Mattoo et al. (2006) indicate that English–French bilingual infants may be able to use a [b]–[g] contrast in word learning at 17 months. However, it is an open question as to whether the bilingual infants succeeded due to the contrast used ([b] vs. [g] being more acoustically distinct than [b] vs. [d]) or because the words were taught, and tested, using both the French and English pronunciations. Presentation of the words with both pronunciations gave the bilinguals two ways to learn and retrieve the minimally different words and may also have alleviated the requirement of ascertaining the language being spoken. Further, monolingual infants of 17 months failed to use the phonetic detail in this study, even though previous research indicates that monolingual infants can succeed at this age (Werker et al., 2002). This lends support to the argument that it is the presentation of the words in both languages that affected the results. Because this mode of presentation would be unnatural to monolinguals, it might potentially impair their ability to use phonetic detail.

Could there be an advantage for word learners not fully utilizing phonetic detail in the initial stages of learning words? Novice word learners have to attend to, encode, and later retrieve detailed phonetic information while simultaneously establishing and maintaining the link between word and object. Werker and Fennell (2004; see also Fennell & Werker, 2003; Stager & Werker, 1997) appealed to the concept, common in developmental psychology, that “less is more” (Newport, 1990; see also Elman, 1993, on “the importance of starting small”) and postulated that not using all the detail in a novel word may be adaptive in that it allows the infant to divert cognitive resources toward learning a word–object link as rapidly as possible. Further, there would be little “cost” to this approach, as there appear to be very few similar-sounding words in the initial lexicons of infants (e.g., Caselli et al., 1995).

The “less is more” hypothesis may be especially important in understanding word learning in bilinguals. As discussed earlier, bilinguals and monolinguals have similar-sized vocabularies during infancy and reach language milestones at similar ages (e.g., Pearson & Fernandez, 1994; Oller et al., 1997; Petitto et al., 2001). Thus, bilinguals’ later use of phonetic detail does not appear to negatively impact their overall word-learning capabilities. Hence, rather than proposing a bilingual delay in word learning, we interpret our findings as indicating that bilinguals utilize this initial compensatory approach for a longer period of time than monolinguals, as they face more challenges with word learning. This approach would allow bilingual infants to remain on par with monolinguals by allowing for the quick establishment of word-referent links with little cost, as minimally different words are in all probability also rare in bilingual infants’ initial lexicons.

This set of experiments begins a line of research that extends the study of how basic speech perception affects word learning to a significant, but as yet poorly studied, portion of the population, namely those infants being raised in bilingual environments. The finding that bilingual infants follow a unique course of lexical acquisition shows that the conclusions that have been drawn from monolingual (often English) infants are not necessarily generalizable to all infants. Through studies with bilingual (and multilingual) infants from different populations and with languages that are similar and different in distinct ways, we can gain a deeper understanding of language development in all infants.

References


Bilingual Infants’ Use of Speech Sounds in Words 1523


### Appendix: Language Exposure Data for All Infants

Table A1

<table>
<thead>
<tr>
<th>Language Exposure Data for Each Infant in Experiment 1, With Group Averages in Bold at the Bottom of the Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 months</td>
</tr>
<tr>
<td>English (%)</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>54</td>
</tr>
<tr>
<td>55</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>58</td>
</tr>
<tr>
<td>51</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>62</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>

**Note.** The averages do not add to 100%, as some infants had minor exposure to a third language.
Table A2

Language Exposure Data for Each Infant in Experiment 2, With Group Averages in Bold at the Bottom of the Table

<table>
<thead>
<tr>
<th>English – Chinese Infants</th>
<th>English – French Infants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17 months</td>
</tr>
<tr>
<td></td>
<td>English (%)</td>
</tr>
<tr>
<td>25</td>
<td>75*</td>
</tr>
<tr>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td>57</td>
<td>40*</td>
</tr>
<tr>
<td>67</td>
<td>28</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>32</td>
<td>68</td>
</tr>
<tr>
<td>66</td>
<td>34</td>
</tr>
<tr>
<td>39</td>
<td>61</td>
</tr>
<tr>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>66</td>
<td>34*</td>
</tr>
<tr>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>37</td>
<td>63*</td>
</tr>
<tr>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td><strong>47.6</strong></td>
<td><strong>51.9</strong></td>
</tr>
</tbody>
</table>

Note. The averages do not add to 100%, as some infants had minor exposure to a third language. Infants also included in Experiment 1 are italicized. Infants hearing Mandarin, rather than Cantonese, are marked with an asterisk.