TRACKING ETHNICITY-LANGUAGE CO-OCCURRENCES AMONG
TEN-MONTH-OLD BILINGUAL INFANTS

by

Shun-Fu Hu

B.A. and B.Sc., National Taiwan University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in
The Faculty of Graduate and Postdoctoral Studies
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2017

©Shun-Fu Hu, 2017
Abstract

It is known that bilingual infants can separate languages based on linguistic cues and silent faces. It is also argued by Kandhadai, Danielson and Werker (2014) that bilinguals track the co-occurrences between language and other contextual cues. Given these findings, the current study asks if ethnicity of faces, as the contextual cue, selectively activates the discrimination of sound in the language most frequently experienced with that ethnicity. Specifically, we ask if 10-month-old Chinese-English learning bilinguals activate Chinese sound discrimination to a greater extent when primed with East Asian faces than when primed with Caucasian faces, and conversely, if Caucasian priming activates English sound discrimination to a greater extent than East Asian faces in bilinguals. To test this hypothesis, we recruited 48 bilinguals and 67 English learning monolingual 10-month-olds. Both groups were first shown static images of either Caucasian or East Asian faces to prime them with the ethnicity associated with one of their languages. Next, in the habituation phase, they were repeatedly presented with either one Chinese (non-English) or English (non-Chinese) syllable accompanied with a visual stimulus. After the infants’ looking time dropped significantly, the 2-trial test phase began: the “same” trial comprised the same syllables as in the previous phase, while the “switch” trial comprised a minimally-contrastive syllable in the same language. Infants’ discriminatory ability was inferred from the increased looking time to the “switch” trial from the “same” trial. The results for Chinese provided support for our hypothesis: bilinguals discriminated the Chinese contrast better when primed with East Asian faces than in the Caucasian priming condition. The results for English are more difficult to interpret: Caucasian faces did not facilitate English discrimination by the bilinguals, but even English monolingual learning infants failed to discriminate the English contrast. Analyses in the priming and habituation phase collectively suggest that the absence of an effect for English could be due to the paucity of exposure to Caucasian faces, or to the difficulty of the English contrast, but not to the inability to track language-ethnicity co-occurrences. The meaning of these results is discussed.
Lay Summary

Bilinguals are sensitive to the contexts under which languages occurs. The current study tests the hypothesis that bilinguals use their knowledge from tracking external cues that co-occur with languages – in this case, ethnicity – to selectively activate language processing in the form of sound discrimination. We recruited a group of 10-month-olds from primarily English speaking households and another group from homes where both Chinese and English (hereafter, bilinguals) are spoken. Infants were first shown either Caucasian or East Asian faces, and then tested on their ability to discriminate two similar sounding Chinese syllables or two English syllables. The results revealed that bilinguals shown East Asian faces better discriminated Chinese sounds than those shown Caucasian faces, but those shown Caucasian faces did not better discriminate the English sounds. Still consistent with our hypothesis that external cues would only help bilinguals, monolinguals did not benefit from being shown faces of either ethnicity.
Preface

The entire thesis is based on the work conducted in the Infant Studies Centre at UBC by Dr. Janet F. Werker, Jennifer Campbell, D. Kyle Danielson, and me. Jennifer Campbell developed the experimental stimuli, Kyle Danielson collected the first 50 samples of the dataset, and along with Dr. Werker and Jennifer Campbell, designed the experiment. Under the supervision of Dr. Werker, I completed collecting the dataset, coded the behavioural data, ran statistical analyses, and wrote the entire thesis. Some of the participants and coding were conducted by volunteer undergraduate research assistants. The research project was approved by UBC Children's and Women's Research Ethics Board and the certificate number is H04-80441.
List of Tables

Table 1. Acoustic properties of stimuli

Table 2. Infants’ looking times and pairwise t values
List of Figures

Figure 1. A conceptual framework from Kandhadai, Danielson and Werker (2014)

Figure 2. Infants’ language and ethnicity exposure

Figure 3. Experimental procedure

Figure 4. Looking times data
Acknowledgements

I offer my enduring gratitude to the faculty, staff and my colleagues at UBC, Department of Psychology, Early Development Research Group and Infant Studies Centre, who have inspired me to continue my work. I owe particular thanks to Dr. Janet F. Werker, whose supervision and extensive feedback contributed largely to the current Thesis, to Savannah Nijeboer for her continuing support in all regards, to Jennifer Campbell and D. Kyle Danielson, for the stimuli and technical advice, to undergraduate research assistants, including Gloria Ng, Lal Koyuncu, Cassie Tam and Laura Lu, and to participant family in the BC area.

I thank Social Sciences and Humanities Research Council for financial support that enabled this project, and Taiwanese Ministry of Education, Taipei Economic and Cultural Centre in Vancouver for overseas fellowship that supported me financially. Special thanks are owed to my parents, whose have supported me throughout my years of education, both morally and emotionally.
Dedication

To my parents and everyone who is interested in bilingualism
1. Introduction

Bilinguals are not two monolinguals in one person, who randomly switch between the two languages. They are, instead, fundamentally social beings sensitive to the demands of the situation (Grosjean, 1989; 2001). In public places, bilingual children and adults from immigrant families respond in the official language of the locality, while being at home makes their heritage language more readily accessible; they switch between the two languages in a marketplace within a predominantly bilingual neighbourhood (Grosjean, 2001). Anecdotes of bilingual French-English children suggest that they often link a person with a language and can be distressed if that expectation is violated (Grosjean, 2010). Following a Portuguese-English learning child from 1 year to 1.5 years of age, Nicoladis (1998) documented that the bilingual learner pragmatically differentiated the languages by selectively producing one of the two in accordance with the interlocutor. Even two-year-old bilinguals switch and intermix languages in a context sensitive way (Lanza, 1992; Genesee, Nicoladis & Paradis, 1995). In adults, the sociolinguistic sensitivity between the two languages is so prominent that even a brief priming with Chinese cultural artefacts suffices to set participants in the Chinese language mode, where they more readily name objects with direct translations from Mandarin than with standard English words, e.g. “happy nuts” for pistachios (Zhang, Morris, Cheng & Yap, 2013). The phenomenon that these non-linguistic cues can selectively enhance perception and production of one of the two languages may suggest, in line with Kandhadai, Danielson and Werker (2014), that bilinguals are keeping track of the contextual cues that co-occur with each of the languages (for a conceptual visualization, see Figure 1 from Kandhadai, Danielson & Werker, 2014) to activate
relevant “language mode(s)” (Grosjean, 2001) or “functional category(-ies)” (Byers-Heinlein, 2014) for language differentiation. The goal of this thesis is to study whether bilingual infants (defined as those who grow up being exposed to two or more languages, consistent with Byers-Heinlein, 2015) use one of these co-occurrences, ethnicity of faces, to set themselves in the mode of the relevant language most frequently experienced with that ethnicity. For the convenience of discussion, I will consistently use the term bilinguals to refer to infants growing up exposed to two or more languages simultaneously.

Figure 1. A conceptual framework from Kandhadai, Danielson and Werker (2014)

Bilinguals are hypothesized to track the co-occurrences between each of the languages and the corresponding contextual cues, such as faces, music, and costumes. This figure is included with permission from Kandhadai, Danielson and Werker (2014).
The acquisition of the requisite ability for sociolinguistic sensitivity is no easy feat, but it is safe to assume that language separation is one of the key prerequisites. Thus, a significant part of the Introduction of this thesis is dedicated to the types of cues infants can use to differentiate languages, from paralinguistic, phonetic and visual, to non-linguistic cues. Wherever pertinent, I will explain how language differentiation can be attributed to the interplay among biological endowment, language-specific input, and general learning mechanisms. After establishing language differentiation as a given in ten-month-old bilinguals, I will elaborate on the purpose of the current study to test the ability of bilinguals to use ethnicity of faces as a cue to selectively activate the relevant language processing mode.

1.1. Language differentiation based on paralinguistic cues

1.1.1. Both monolinguals & bilinguals use rhythmicity

As a prerequisite for language differentiation, monolingual neonates differentiate language from other auditory inputs by showing an attentional preference for language. Vouloumanos and Werker (2004; 2007) adopted the high amplitude sucking (HAS) procedure and presented neonates with both speech and an equally complex auditory clip of filtered speech that was matched with real speech on the first three formants, and measured the total duration of high-amplitude sucking as an index of interest or attention. Remarkably, despite the similarity of the filtered speech to the infants’ experience in
utero, neonates showed a preference – by producing significantly more high-amplitude sucks – for speech, suggesting, at least, an early template to attend to language.

Soon after birth, monolingual infants discriminate familiar from unfamiliar language based on rhythmicity. Through an infant-controlled procedure where audio recordings of either Spanish or English can be activated by infant sucking, Moon, Cooper and Fifer (1993) demonstrated that two-day-olds with monolingual Spanish- or English-speaking mothers activated the clips that featured their native language significantly more, showing language differentiation. The same phenomenon was also documented in 2-month-old French and English monolinguals, who oriented faster to their native language (Dehaene-Lambertz & Houston, 1998). As part of the motivation for Moon et. al.’s study, Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison (1988) had demonstrated that sensitivity to rhythm enabled language differentiation.

Like monolinguals, bilinguals likely start with selective attention to language in general (e.g. Curtin, Byers-Heinlein & Werker, 2011). Moreover, they show equal preferences for their two familiar languages at birth and can also differentiate the two based on the rhythm (Byers-Heinlein, Burns & Werker, 2010). With an infant-controlled HAS procedure, Byers-Heinlein and colleagues (2010) presented 0- to 5-day-old bilingual English-Tagalog learning infants a speech stream alternating between Tagalog and English sentences and recorded the number of high-amplitude sucks elicited during the respective phases of English and Tagalog. As predicted, Tagalog-English bilinguals showed a much more dispersed preferential pattern, indicating that they preferred both,
than monolingual English learning neonates, who uniformly preferred English. To rule out the possibility that the bilinguals were simply confused, Byers-Heinlein and colleagues recruited Chinese-English bilinguals, for whom Tagalog is neither a familiar nor a completely foreign language due to its similarity to Chinese in rhythmic class. Confirming their hypothesis, the Chinese-English bilinguals demonstrated a pattern in between, preferring Tagalog less than Tagalog-English bilinguals but more than English monolingual newborns. To investigate if infants could differentiate the languages, they habituated a new group of monolingual and a new group of bilingual infants with low-pass filtered speech in one or the other language. After the habituation phase, infants were presented new sentences in either the same or the other language. Extending Mehler et. al.’s (1988) finding that monolinguals can distinguish languages based on rhythmicity, Byers-Heinlein et al. (2010) added to the literature that both monolinguals and bilinguals succeeded in this type of language differentiation. They also excluded chance recovery, whereby the increase in post-habituation looking time is due to chance as an alternative explanation. Similar findings have been reported using other methods. If the rhythms of the two languages are similar, such as Catalan and Spanish, bilingual infants are slightly later, though still early, in being able to tell the two languages apart, at around 4 months of age (Bosch & Sebastián-Gallés, 1997). Hence, it appears that bilinguals share the early proclivity for language, and are quick to selectively attend to their two familiar languages, comparable to their monolingual counterparts.
1.1.2. Only bilinguals can flexibly use prosody to distinguish word order

Rhythm may not be the only cue by which infants, especially bilinguals, differentiate languages. While both monolingual- and bilingual-learning infants can use prosody to discriminate languages (Christophe et al., 2003) and to segment sentences (Bernard & Gervain, 2012), only bilingual infants can capitalize on prosody (e.g. intonation and duration) together with frequency to guide the segmentation of noun phrases differentially in each of their two languages (Gervain & Werker, 2013). In their work, Gervain and Werker recruited 7-month-old bilinguals learning both a VO (verb-object word order) language, where the verb usually precedes the object, and an OV language, where the reverse order is more common. VO languages, such as English, tend to have high-frequency word classes, such as prepositions, occurring before infrequent word classes, such as nouns, and OV languages are inclined to have high-frequency classes after the nouns. Thus, Gervain & Werker (2013) reasoned that a natural noun phrase in a VO language would consist of a frequent word followed by an infrequent word, while the reverse is true for OV languages, with evidence: monolingual infants can use word frequency to parse phrases. Infants exposed only to VO languages, such as Italian, parse the artificial language into frequent-first phrases while infants exposed only to OV languages, such as Japanese, do the opposite (Gervain, Nespor, Mazuka, Horie & Mehler, 2008). Given that an ambiguous stream alone could be parsed either way by bilinguals growing up with both a VO and an OV language, Gervain & Werker (2013) hypothesized that additional cues, such as prosody (pitch and duration, which also specifies word order differently in the two types of languages) would be necessary for bilinguals to decide between frequent-initial or frequent-final modes of parsing.
Familiarizing infants with either prosody on streams of alternating frequent and infrequent tokens, Gervain & Werker (2013) demonstrated that only bilingual infants could flexibly parse the streams in accordance with the two sets of cues. The use of prosody thus provides additional evidence that bilingual infants are sensitive to many cues that support language differentiation.

1.2. Infants could potentially use phonetic cues for language differentiation

In addition to the paralinguistic cues discussed above, infants could also hypothetically use phonetic cues for language differentiation. The following paragraphs review the establishment of native phonetic categories in infancy through the symphony of biological endowment, language-specific input, and general learning mechanisms. These sound categories, used to contrast meaning by adults, are phonemes. As examples of phonemes, Hindi speakers perceive voiced dental stop [d̪], where the articulatory closure is at the front teeth before the release of air, and its retroflex version [ɖ], which requires the back of the tongue to touch the alveolar ridge prior to the release, as two distinct categories, whereas English users consider them a single sound category, /d/. More importantly, /d̪al/ (दाल) and /ɖal/ (डाल), contrasting only the [d̪] and [ɖ], correspond to different meaning in Hindi: lentil and branch, respectively. To process only the acoustic dimensions used contrastively, and to ignore those that are not, is paramount for learning one’s native language. Since there is little evidence that 10-month-old infants’ representations of these sound categories are used to contrast meaning, the term native phonetic categories is preferred over phonemes.
Perceptual attunement is the most dominant pattern

In the establishment of the native phonetic categories in infancy, substantial evidence shows that both monolingual and bilingual infants undergo a process by which perceptual sensitivities for distinctions that the infant was not exposed to during a window of high malleability become reduced or pruned out, while discrimination of those distinctions that do occur in the input is maintained (e.g. Werker & Tees, 1984) or even sharpened (Kuhl et. al., 2006). This process, termed perceptual narrowing or perceptual attunement (used hereafter, to include both the process by which lack of experience leads to a decline while the presence of experience enhances the perception of native phonetic categories), is at least experience-expectant, if not maturational. The role of experience is merely in triggering or “wedging” (Petitto, Berens, Kovelman, Dubins, Jasinska & Shalinsky, 2012; Peña, Werker & Dehaene-Lambertz, 2012), but not inducing the sensitivities (Maurer & Werker, 2014), as discrimination according to phonetic category boundaries is evident without prior listening experience (e.g. Eimas, et al, 1971; Werker & Lalande, 1988). Strong evidence that the timing of perceptual attunement in establishing sound categories of the native language is likely maturational came from Peña, Werker & Dehaene-Lambertz (2012), who found that preterm infants, allowed more language experience, performed in similar manner to full-term infants matched with gestational, rather than postnatal, age. As will be discussed later, perceptual attunement is also evident across several domains, such as face perception and visual language discrimination (e.g. Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés & Werker, 2007), and can even be multisensory (Lewkowicz & Ghazanfar, 2009; audiovisual matching, Pons, Lewkowicz, Soto-Faraco & Sebastián-Gallés, 2009).
Moreover, adopting a within-subject design, Lalonde & Werker (1995) found strong correlations between infants’ developmental progress in speech perception and the ability to categorize visual stimuli based on feature correlations. All these suggest that perceptual attunement can be conceptualized as a manifestation of a general neurological process (or more generally, “perceptual cognitive,” in Lalonde & Werker’s words, 1995) that prepares infants to optimally categorize many kinds of ‘biologically relevant’ input, including language (Maurer & Werker, 2014; Lewkowicz & Ghazanfar, 2009, for a review of multisensory perceptual narrowing).

In the case of monolingual English learning infants, the broad-based sensitivity for native and non-native phonetic categories becomes narrowed to the contrasts used in English by 12 months of age, as demonstrated in Werker & Tees’s (1984) classic study of cross-language phonetic discrimination. Adopting a head-turn procedure with an animated toy as the reinforcer, they showed that the ability to distinguish non-native contrasts, from both Thompson and Hindi, was present in 6- to 8-month-old English learning infants, but was not maintained in 10- to 12-month-olds, suggesting that the perceptual reorganization (attunement) likely occurred during the first year of life. Similar patterns have been documented for lexical tone (e.g. Mattock & Burnham, 2006), stress (Jusczyk, Cutler & Redanz, 1993), vowel distinctions (e.g. the monolingual groups in Bosch & Sebastián-Gallés, 2003a), and many other consonants (for a review and a conceptual framework, see Werker & Curtin, 2005).

While the bilingual results are slightly mixed (for a review, see Sebastián-Gallés,
Díaz & Costa, 2013), some groups of bilinguals showed a similar pattern of perceptual attunement of phonetic categories to monolinguals (e.g. Burns, Yoshida, Hill & Werker, 2007). Burns and colleagues (2007) took advantage of the differences in voice onset time (VOT) boundaries, an often language-specific criterion, to define phonetic categories, which happen to differ in English and French for the /b/ vs /p/ distinction. By habituating infants of three age groups to an intermediate token, [p], and subsequently presenting them a new token [b] or [pʰ], they found, again, that while 6-month-old monolinguals were able to detect both differences, monolingual English and monolingual French 10-month-olds’ VOT boundaries started to resemble those of adult speakers of their respective language. Importantly, bilinguals at all three ages (6, 8, and 10 months) retained the flexibility to set their phonetic boundary either way, confirming the aforementioned pattern of perceptual attunement, where experience served to maintain language-specific sensitivities.

On the other hand, Bosch & Sebastián-Gallés’s (2003a) results with the Catalan (non-Spanish) /e/ - /ɛ/ distinction among Spanish-Catalan bilingual infants revealed an unpredicted nadir of this ability: discrimination was present in infants aged both 4 and 12 months, but not at 8 months of age. The authors attributed the lack of discrimination at 8 months to the rarity of /ɛ/ and the potential overlap between the Spanish and Catalan phonetic space across the two languages, and thus a reasonable consequence of the heightened malleability to language-specific input around 6 months, as part of perceptual attunement. The success among Spanish-Catalan bilinguals at 8 months in discriminating the two sounds when a more sensitive measure, Anticipatory Eye Movement, is used
(Albareda-Castellot et. al., 2011) further supports this conclusion. Last but not least, the finding that Spanish-English bilingual 8-month-olds could discriminate the same contrast points to the possibility that the U-shaped development is specific to either the language group or the task, not to be taken as evidence that bilinguals do not undergo perceptual attunement as monolinguals do (Sundara, & Scutellaro, 2011).

1.2.2. Distributional learning is one way infants can learn from language-specific input

In their seminal work, Maye, Werker and Gerken (2002) provided evidence for how the attunement from broad-based to language-specific phonetic categories might be based on infants’ abilities to track the distributional properties of the input. As a general learning mechanism, distributional learning should be present in both bilinguals (Curtin, Byers-Heinlein & Werker, 2011; Werker, 2012) and monolinguals. Maye et al. (2002) hypothesized that if a language contrasts meaning on a particular dimension, such as the place of articulation (i.e. the place of the most intense closure in the oral cavity before the release of air during the production of a consonant), then the corresponding sound tokens would roughly fall in a bimodal distribution where the two modes are the prototypical pronunciation of the sound categories. Tokens that deviate from the prototypes are to be categorized as one or the other depending on their perceptual proximity from each prototype. On the other hand, if the language does not contrast meaning on that dimension, then the sound tokens would form a unimodal distribution, where deviations are treated as merely variants of the prototypical sound, located at a single mode. Indeed, this distributional pattern was observed in two studies that involved recording infant-
directed speech from Japanese and English-speaking mothers (Vallabha, McClelland, Pons, Werker & Amano, 2007; Werker, Pons, Dietrich, Kajikawa, Fais & Amano, 2007). The utterance of /e/ and /e:/ (long /e/) – contrastive in Japanese but not English – by Japanese mothers, for example, fall in a bimodal distribution of vowel duration, while the tokens fell on a unimodal distribution in terms of vowel spectra across /e/ - /ɛ/, which are not contrastive in Japanese. Thus, Maye et al. (2002) predicted that providing a bimodal distribution that contrasts (atypical, in this case) unaspirated voiceless and voiced alveolar stops followed by the vowel /a/, i.e. /ta/ and /da/, for infants that are within the critical period of heightened sensitivity to phonetics would enable them to distinguish between /t/ and /d/, whereas a unimodal distribution would not. Testing 6- and 8-month-old monolinguals, they found exactly what they predicted: in 6-month-olds (and 8-months-olds), the bimodal distribution allowed for the distinction whereas the unimodal one did not. A similar pattern of results was found in a follow-up study by Yoshida, Pons, Maye & Werker (2010a), who, adopting a similar paradigm, additionally found that it took 10-month-old monolinguals longer familiarization to learn the discrimination, presumably because they had already begun to attune to the phonetic distinctions of the native language.

1.2.3. Bilinguals may use phonetic detail to differentiate languages

Sundara, Polka and Molnar (2008) found that, through an altered visual habituation procedure, 10- to 12-month-old French-English bilinguals retained the ability to differentiate alveolar and dental /t/ sounds, which are not contrastive in either French or English. This sensitivity to the acoustic dimensions not used contrastively in either
language, which Byers-Heinlein and Fennell (2014) termed “latent contrasts”, is meaningful for at least three reasons. First, since a prototypical /t/ produced by a French monolingual speaker tends to be dental and that by an English monolingual is inclined to be alveolar, the dental-alveolar distinction is potentially a useful cue to tell the two languages apart. Thus, the sensitivity mentioned may reflect the bilingual phonological awareness of the two languages, with evidence from magnetoencephalography (MEG) that 11-month-old bilinguals show stronger late mismatch response for both languages (Ferjan Ramírez, Ramírez, Clarke, Taulu & Kuhl, 2017). Second, the fact that the prevalence of /t/ in both languages enabled the distinction was argued as evidence that bilinguals are tracking both distributional information (where the perceptual boundary of the phonetic categories lies) and the frequencies of occurrences of the categories; if they had only learned from the distributional information, the non-contrastive nature of the dental-alveolar distinction should not have allowed them to keep the phonetic categories distinct (Sundara, Polka & Molnar, 2008). Finally, and most importantly, it could work in conjunction with, or as a mechanism for, language differentiation, to profile the separate sets of phonological rules and grammars of the two languages in order to facilitate acquisition.

1.3. Language discrimination based on visual cues in bilinguals

Language differentiation can also occur in the visual domain in both monolinguals and bilinguals. Weikum and colleagues (2007) habituated monolinguals and English-French bilinguals with videos of three bilingual speakers reciting the story “The Little Prince”, in either English or French, and then removed the audio, thus
presenting the talking faces in silence. Following habituation, they presented infants the same faces, but with a switch in the language to see if the infants noticed the change in language. Not only did they find direct evidence for visual language discrimination, but also perceptual attunement, in that monolingual 4-month-old infants, but not 8-month-olds, succeeded in the task, whereas bilinguals, due to language-specific input, maintained the ability at 8 months. Sebastián-Gallés, Albareda-Castellot, Weikum and Werker (2012) adopted an identical procedure to Weikum et al. (2007), but instead tested bilingual Spanish-Catalan bilinguals, for whom neither English nor French was familiar. While Weikum et al. (2007) found that bilinguals maintained the visual sensitivities for their relevant languages, Sebastián-Gallés et al. (2012) found perceptual attentiveness, unexplained by mere maintenance, to unfamiliar languages among 8-month-old Spanish-Catalan bilinguals. A further investigation by Pons, Bosch and Lewkowicz (2015) discovered that one possible mechanism, or by-product of it, was for bilinguals to maintain the inclination to attend to the mouth area, as opposed to the eyes, for longer in development (until 12 months of age), a propensity that seems to generalize to non-speech stimuli (Ayneto & Sebastián-Gallés, 2017).

Therefore, language differentiation can occur within the domain of language, through rhythm, prosody, and phonetics – and even through faces silently telling a story. The challenging environments of a divided and noisy input can lead bilinguals to a high level of perceptual attentiveness to phonetic, prosodic, and visual-facial information. But a parallel process of tracking the co-occurrences between language and external cues,
such as faces, may also be involved. This is the topic of the next section.

### 1.4. Tracking the co-occurrences between language and contextual cues

Although linguistic cues alone (e.g. prosody, see Gervain & Werker, 2013) may be sufficient for bilinguals to differentiate the languages, I propose, in line with Kandadhai, Danielson, and Werker (2014), that for bilinguals, especially for those growing up in a bicultural household, tracking the co-occurrences between the broader context and the language may also help with language differentiation and selective activation of the bilinguals’ native languages. Three mechanisms have been proposed to underlie this process (Kandadhai, Danielson & Werker, 2014).

Kandadhai, Danielson, and Werker (2014) argued that while some aspects of the context, such as traditional costumes, are only arbitrarily linked with language, others, like the rhythmical structure of music and language, bear deeper relationships with language. As an empirical example, Patel & Daniele (2003) compared French and English on the normalized variability in duration and found parallelism between music and diction as a function of language. Kandhadai, Danielson, and Werker labeled this mechanism “structural isomorphy,” where the inherent relationship between the rhythm of language and say, the music of the culture, makes one or the other cue redundant (e.g. see Yoshida et al., 2010b), because the regularities can be accounted for by one of the cues alone. This redundancy could nonetheless help reinforce the linkage between the broader culture and language, facilitating the process of tracking and selective activation.
Meanwhile, in the case where the language-context co-occurrence is arbitrary, associative mechanisms are broadly available – mechanisms like “acquired distinctiveness” (Hall, 1991), whereby two readily-discernable cues facilitate a subtler pair of distinctions when the elements of the easier distinction individually predict the occurrence of the individual elements of the subtler distinction. Thus, ethnicity, which 3-month-olds can already distinguish (Sangrigoli & De Schonen, 2004), when paired with different languages, can be hypothesized to help bilinguals differentiate languages through acquired distinctiveness. However, given the importance of language as a signpost for social grouping (Kinzler, Shutts & Correll, 2010), language and ethnicity pairings may be more than mere associations. In fact, evidence suggests that the language-ethnicity link, as opposed to some other potential co-occurrences (e.g. language-age or language-attire), may enjoy a special status, such that relations are more readily formed (Hirschfeld & Gelman, 1997) – this is the third mechanism, “privileged links”, proposed by Kandadhai, Danielson, and Werker (2014).

This ability to track ethnicity-language co-occurrence has been shown in monolinguals. Simultaneously presenting 6-month-olds infants with faces and languages, Uttley et al. (2013) found that 6-month-old Caucasian English-learning monolingual infants paired other-race faces with a non-native language, as indicated by longer looking time to the matching conditions (East Asian faces with Chinese language). They ruled out the account that the increased looking time was simply due to novelty by using backward speech. Consistent but slightly different findings were noted in May’s studies (2015, Experiment 5), where monolingual 11-month-old infants in Vancouver, Canada, looked
longer at the East Asian face, presented side-by-side with a Caucasian face, when the accompanying speech stimuli were in Cantonese. May (2015, Experiment 6) further investigated whether the 11-month-olds’ response pattern reflected an unfamiliar-race with unfamiliar-language pairing strategy (as in Uttley et al. above), or a more specific East Asian race with Cantonese language link, by running a control experiment where the face stimuli were the same as the previous experiment, but the speech stimuli, instead of English and Cantonese, were changed to English and Spanish. Since one would seldom encounter in Vancouver any Spanish-speaking individuals of East Asian ethnicity, Caucasian monolingual infants would only look at East Asian faces more during the Spanish trials if they were simply pairing unfamiliar ethnicity with any unfamiliar language. The results showed otherwise: there was no difference in looking time when the Spanish was played. Thus, May demonstrated that at least for the participants in Vancouver, which is the same as the current study, 11-month-olds could learn the specific pair between language and ethnicity for Chinese languages and East-Asian faces (2015). Given that the monolingual 11-month-olds spent most of the time only with their family members who were not East Asian, the fact that the scarce input of the East Asian faces from the neighbourhood sufficed for infants to learn the specific link between East Asian and the Cantonese language together could hypothetically speak to Kandadhai, Danielson, and Werker (2014)’s conjecture that the link between language and ethnicity may be privileged.

1.5. The current study

The literature summarized thus far provides evidence that bilingual 10-month-olds are capable of language differentiation based on rhythm, prosody, and phonetics.
From the evidence showing monolinguals can track ethnicity and language, I propose in this thesis that bilinguals can also track ethnicity-language links. Based on these assumptions, I set out to test if the ability to track ethnicity-language co-occurrences translates into the sociolinguistic sensitivity for bilinguals discussed in the opening paragraph. Specifically, I asked if brief exposure to East Asian and Caucasian faces sufficed for bilinguals to activate the relevant language mode, manifested as enhanced phonetic discrimination in that language. In the context of a bilingual Chinese-English environment, the sounds /fa/ - /va/ constituted an English-specific contrast, and hence were adopted to operationalize the testing of the activation of the English mode. Similarly, the Chinese-specific contrasts /tsa/ - /tsʰa/ was used to operationalize the activation of the Chinese mode. Thus, selective activation of the relevant mode to the ethnicity primes could be assessed by comparing bilinguals’ performance on the same contrast with the other priming condition. The ability to discriminate both the Chinese-specific and English-specific contrasts, given the evidence in perceptual attunement, was predicted: the more frequently the ethnicity is experienced with the language, the higher the degree of activation of the language. Thus, Caucasian faces, in the context of Vancouver, Canada, where the infants were tested, were hypothesized to activate the English mode to a greater extent than East Asian faces. On the other hand, East Asian faces were hypothesized to activate the Chinese mode to a greater extent than Caucasian faces. The expression “to activate a language” refers to a temporary status of high readiness, with facilitative effects, to process stimuli relevant to that language.
Phonetic discrimination was operationalized as the positive difference between post-habituation looking times at the trial that contained a new sound (the *Switch* trial) versus the one that contained only the sound to which infants had been habituated (the *Same* trial). Thus, the current study essentially was the Switch procedure (as in Experiment 4, Stager & Werker, 1997) preceded by an additional priming phase. Monolingual English-learning infants, who have had limited exposure to the East Asian ethnicity in their households, were also included for comparison, to shed light on the broader developmental questions, such as the role of specific experience in seeing ethnicity-language links in perceptual attunement, but monolingual infants’ sound discrimination were neither predicted to, nor not to, benefit from the ethnicity primes. Since fricatives, to which /fa/ and /va/ belong, are known to yield mixed results (Bosch & Sebastián-Gallés, 2003b), the monolingual group provides a test of whether the contrast is discriminable among 10-month-olds.

Taken together, the specific predictions were as follows: bilinguals primed with Caucasian faces would look longer to the switch trial than the same trial on the English contrast to a greater extent than those primed with East Asian faces. Conversely, bilinguals primed with East Asian faces were predicted to look more during the switch than the same trial to a greater extent than those in the Caucasian priming condition. Monolinguals were predicted to be able to discriminate the English specific contrast, /fa/-/va/, but not /tsa/-/tsʰa/, independent of the primes, but neither predicted to, nor not to,
benefit from Caucasian primes. Possibilities that some of the assumptions were violated were entertained.
2. Method

2.1. Participants

One hundred and fifteen 10-month-old infants (9;15 - 10;15, $M = 304$ days old, $SD = 8.31$) recruited at BC Women’s Hospital, BC, Canada, completed the study. Among them, 67 (35 females) were monolingual English learning infants (hereafter, monolinguals), and 48 (27 females) were bilingual English-Chinese learners (hereafter, bilinguals), who, further, comprised 33 (17 females) bilingual Cantonese-English learning infants and 15 (10 females) Mandarin-English learning infants. Sixty-six additional infants were tested but excluded due to equipment malfunction (9), experimenter error (3), parent’s failure to follow the instructions (11), fussing out (37), and failure to habituate (6).

Language and ethnicity exposure was measured via a structured interview of the infant’s primary caregiver(s) based on Bosch & Sebastián-Gallés’s (1997) language exposure questionnaire, adapted to include measures of ethnicity exposure, as in Appendix A. As shown in Figure 2, infants in the monolingual group were exposed to 90% or more English ($M = 98.84\%, SD = 2.35\%$) and no Chinese languages. On average, they additionally heard 1% non-Chinese languages other than English, such as French. In their time awake, they were exposed to Caucasian faces 91%, East Asian faces 1%, East Asian-Caucasian mixed <1%, and ethnic groups other than East Asian, Caucasian, or East Asian-Caucasian mixed 8% of the time.
Bilingual infants were exposed to at least 20% of both English (M = 49.22%, SD = 16.93%) and the Chinese languages combined (M = 50.47%, SD = 23.33%), though a small proportion of bilinguals additionally heard other Chinese languages, such as Hakka and Shanghainese. In their time awake, bilinguals were exposed to East Asian faces 88% (M = 6462.48 hours, SD = 2285.14 hr), Caucasian faces 10% (M = 724.95 hr, SD = 1170.55 hr), East Asian-Caucasian mixed-race faces 2% (M = 135.43 hr, SD = 602.69 hr) and other ethnic groups <1% (M = 30.62 hr, SD = 79.19 hr) of the time.

The average lifetime percentage of exposure to English and Chinese languages (Cantonese, Mandarin, Other Chinese languages) and Other Non-Chinese languages, among bilingual and monolingual infants. “Mixed” in the ethnicity exposure refers specifically to East-Asian and Caucasian mixed.
Importantly, the bilingual experience of the contingencies between ethnicity and language, assessed through point-serial correlations between the ethnicity of each source of the infant’s language input and the proportion of language spoken by that input (English, or any Chinese), were as follows: $r$ (Caucasian, English) = .46, $r$ (Caucasian, Chinese) = -.52, $r$ (East Asian, English) = -.47, and $r$ (East Asian, Chinese) = .54. Thus, for the current bilingual sample, Caucasian faces were positively predictive of the amount of English spoken to the bilingual infant, and negatively predictive of Chinese. Conversely, East Asian faces were positively predictive of the amount of Chinese spoken to the infant and negatively predictive of English. Also, given that the source of input (usually a person) is East Asian in ethnicity, there is an 88% chance that the person would be speaking to the bilingual infant in any amount of any Chinese languages and a 64% chance that the person would speak any amount of English. On the other hand, given the person is Caucasian in ethnicity, there is a 94% chance the person would be speaking to the bilingual infant in any amount of English, and 0% chance that the person would speak any Chinese in any amount to the infant.

The infants’ caregivers gave informed consent prior to the study, and received an infant scientist t-shirt, a mock degree diploma, and compensation for transportation upon completion.

2.2. Stimuli

The priming phase comprised three videos of static, full-colour female faces of Caucasian and East Asian adults (all the East Asian adults were either Cantonese-English or
Mandarin-English bilinguals). Each set featured faces of three different individuals of the same language-ethnic group, against a 35% grey background, with a resolution of 1280 by 720 pixels. The videos were all 16s long, and were accompanied with a 65 dB oceanic sound clip throughout the entire duration. Each participant was only assigned one of these two sets of faces.

As used in Stager & Werker (1997), the visual stimulus used in the habituation and test phase was a red-and-white checkerboard. The audio stimuli were sequences of syllables produced by each of two speakers: a female bilingual Cantonese-English speaking adult and a female bilingual Mandarin-English speaking adult. From each speaker, 3 instances were selected of each of the 4 consonant-vowel syllables, /fa/, /va/, /tsa/, and /tsʰa/, produced with the high-flat Chinese tone (55). The selected syllables were 466 to 811 msec in duration, followed by 976 to 1303 msec of silence, with the silence chosen to ensure onset to onset of sequential syllables of approximately 2 sec. Sequences comprised 9 repetitions (3 repetitions of each utterance) of a single syllable type, yielding sequence durations of 13.4 to 16.0 sec.

The stimulus for the pre-test and post-test phase was a slowly spinning waterwheel accompanied by continual ringing sounds of a bell. In between trials, the attention getter was a colour-changing and shape-changing ball against a white background (both as in Werker, Cohen, Lloyd, Casasola & Stager, 1998).
When displayed on the screen, all the stimuli were 52 cm (width) by 44 cm (height), occupying a visual angle of approximately 27.82 degrees, at a viewing distance of 105 cm. The empirically measured amplitude of the sound stimuli ranged between 63.10 to 67.20 dB, with an average dB of 65.3. As summarized in Table 1, the physical properties other than the diagnostic feature, the Voice Onset Times, were all similar across the phonetic categories.

### Table 1. Acoustic properties of stimuli

<table>
<thead>
<tr>
<th></th>
<th>/fa/</th>
<th>/va/</th>
<th>/tsa/</th>
<th>/tsha/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BiC</td>
<td>BiM</td>
<td>BiC</td>
<td>BiM</td>
</tr>
<tr>
<td>Total duration (ms)</td>
<td>695.27</td>
<td>627.29</td>
<td>762.57</td>
<td>598.27</td>
</tr>
<tr>
<td>Mean amplitude (dB)</td>
<td>64.97</td>
<td>65.11</td>
<td>64.98</td>
<td>64.98</td>
</tr>
<tr>
<td>Peak amplitude (dB)</td>
<td>71.21</td>
<td>72.18</td>
<td>71.77</td>
<td>69.25</td>
</tr>
<tr>
<td>Mean pitch (HZ)</td>
<td>253.22</td>
<td>255.87</td>
<td>252.43</td>
<td>245.29</td>
</tr>
<tr>
<td>Maximum pitch (HZ)</td>
<td>261.90</td>
<td>327.30</td>
<td>267.35</td>
<td>260.83</td>
</tr>
<tr>
<td>Vowel duration (ms)</td>
<td>324.77</td>
<td>356.58</td>
<td>321.22</td>
<td>393.45</td>
</tr>
<tr>
<td>Voice Onset Time (ms)</td>
<td>131.80</td>
<td>139.23</td>
<td>1.60</td>
<td>-44.25</td>
</tr>
<tr>
<td>Vowel F1 (HZ)</td>
<td>932.25</td>
<td>945.51</td>
<td>977.45</td>
<td>867.19</td>
</tr>
<tr>
<td>Vowel F2 (HZ)</td>
<td>1502.21</td>
<td>1558.34</td>
<td>1552.74</td>
<td>1555.96</td>
</tr>
</tbody>
</table>

The average physical properties across the 3 instances of each syllable type as produced by the Bilingual Cantonese-English speaker (BiC) and by Bilingual Mandarin-English speaker (BiM).

### 2.3. Apparatus

The study took place in a dimly-lit, sound-attenuated room, with a monitor screen, 69 cm in width and 41.6 cm in height, set 105 cm in front of a chair for caregiver and the infant. A camera (NTSC – KG-632EX) located below the screen streamed and recorded infants’ gaze behaviour, according to which, an experimenter, in a separate room, controlled the study using Habit X (Cohen, Atkinson & Chaput, 2004). While allowed to see the visual stimuli, the experimenter was not allowed to listen to the sound stimuli to
remain partially blind to the progress of the experiment. The recorded infant videos were then coded off-line using the Psycode software (http://psy.ck.sissa.it/Psycode/Psycode.html) by trained research assistants who were naïve to the experimental conditions.

2.4. Design & Procedure

As shown in Figure 2, the current study employed the Switch procedure to test speech sound discrimination (as in Experiment 4, Stager & Werker, 1997), but preceded by a priming phase.

Figure 3. Experimental procedure

The procedure of the experiment, exemplars of priming stimuli the visual, auditory stimuli

The experiment started with one pre-test trial, followed by either three different static Caucasian faces as primes for half of the infants in each group, or three East Asian faces for the other half. Following and counterbalanced to the priming phase, half of the infants were tested on the English speech sound contrast and half on the Chinese contrast. Those assigned to the Chinese contrast condition were habituated to either /tsa/ or /tsʰa/
(half each), and those assigned to the English contrast condition were habituated to either /fa/ or /va/ (half each). The number of trials, but not trial duration, was controlled by infants, such that the habituation phase would terminate if the infant’s looking dropped by 35% or more during a 3-trial window than it had been during their highest looking time to any prior 3-trial fixed window, or by the end of the 24th trial, whichever came first. The test phase consisted of two trials: The same trial comprised the same syllables as used in the habituation trials, and the switch trial comprised the syllables beginning with the contrasting syllable in the same pair, with the same visual display of the checkerboard. Half of the infants heard the same trial first, and half heard the switch first. Finally, the experiment ended with a post-test trial with identical stimuli to the pre-test.

Cognizant that infants learn best when the stimuli match well with their language environment1 (Mattock, Polka, Rvachew & Krehm, 2010), we prepared both the Cantonese and Mandarin subsets of primes and syllables for bilinguals. While monolinguals were exposed to either subset (half-and-half), bilinguals were only shown the faces and phonetic categories of the subset corresponding to their more dominant Chinese language. Thus, Cantonese-English bilingual infants were presented with sound stimuli spoken by the bilingual Cantonese-English speaking adult, and Mandarin-English bilinguals were presented with those spoken by the Mandarin-English speaking adult.

1 Specifically, using the Switch paradigm, Mattock, Polka, Rvachew & Krehm (2010) tested 17-month-old French-English-learning bilingual infants, as well as their French and English monolingual counterparts, and found above-chance performance on associative word learning only if the syllables were produced by the speaker of the same linguistic background as the infants’. In addition, they established that the phonetic categories /g/ and /b/ produced by the French-English bilingual adult speaker were slightly different in VOT from those by French or English monolinguals, and that although measureable, the nuances were not detectable by monolingual and bilingual adults.
After the experiment, families were interviewed for the infant’s language and ethnicity exposure, and asked to fill out the Vancouver Index of Acculturation (Ryder, Alden & Paulhus, 2000) to indicate their affiliation with the North American mainstream culture as well as with their heritage culture.
3. Results

3.1. Pre-test, post-test and last habituation block

A 2 Pre-post-test {Pre-test, post-test} × 2 Group {Between-subject: Bilingual, Monolingual} × 2 Priming {Between-subject: Caucasian, East Asian} × 2 Contrast {Between-subject: English, Chinese} revealed a marginally significant decrease in looking time during the post-test trial (\(M = 15.35\) seconds, \(SD = 1.35\)), compared with the pre-test trial (\(M = 15.57\), \(SD = 1.19\)), as shown in the main effect of Pre-post-test, \(F(1,107) = 3.58, p = .061, \eta^2_p = .032\). There was also a significant 3-way interaction among Pre-post-test × Contrast × Group, \(F(1,107) = 6.77, p = .011, \eta^2_p = .059\), such that bilinguals showed an increase in looking time during the post-test (post-test minus pre-test, \(M = 0.05\), \(SD = 1.79\)) when the contrast was Chinese, and a decrease when the contrast was English (\(M = -0.66\), \(SD = 1.66\)), while monolinguals showed an increase when the contrast was English (\(M = 0.15\), \(SD = 1.02\)) and a decrease when the contrast was Chinese (\(M = -0.52\), \(SD = 1.08\)). However, the differences were too small to be considered evidence for fatigue or general disinterest in the task. This is confirmed by a 2 Recovery × 2 Group × 2 Priming × 2 Contrast ANOVA (following the same logic as in Fennell, Byers-Heinlein and Werker, 2007), showing only the main effect of Recovery across all conditions, \(F(1,106^2) = 1643.28, p < .001, \eta^2_p = .939\), which showed a significant increase in looking time of the post-test on the post-test trial (\(M = 15.35\), \(SD = 1.35\)) compared with the average looking time of the last habituation block (\(M = 6.35\), \(SD = 2.11\)). The fact that the magnitude of Recovery was not a function of Priming, Contrast,

---

\(^2\)There was one missing case; hence, compared with other analyses, there was one less degree of freedom.
Group, or their interactions further helped rule out the possibility that the results in the test phase were influenced by fatigue.

3.2. Habituation & Priming phase

To rule out the possibility that the different groups of infants had non-equivalent opportunities to process the primes or to learn the speech syllable presented in the habituation phase, we also analyzed the habituation and priming phase. A 2 Group \{Between-subject: Bilingual, Monolingual\} × 2 Priming \{Between-subject: Caucasian, East Asian\} × 2 Contrast \{Between-subject: English, Chinese\} ANOVA testing the numbers of habituation trials needed before the habituation criterion revealed a significant 2-way interaction between Group and Priming, $F(1,107) = 5.55, p = .020, \eta^2_p = .069$. It took monolinguals more trials to habituate when they had been primed with East Asian faces ($M = 11.18, SD = 4.39$) than Caucasian faces ($M = 9.09, SD = 2.61$), while the reverse was true for bilinguals (East Asian: $M = 10.50, SD = 4.15$; Caucasian: $M = 11.88, SD = 4.55$). Since bilinguals were predominantly exposed to East Asian faces, and monolinguals to Caucasian faces, this finding can be interpreted as evidence that both groups were discriminating between the ethnicities of the faces and habituated faster to the familiar ethnicity – a finding interesting unto itself. However, there was no evidence that infants were habituating differently across the four priming-contrast combinations (Group × Priming × Contrast interaction, $F(1,107) = 2.46, p = .12, \eta^2_p = .022$), thus ruling out the possibility that differential processing opportunities influenced the main results.
In the priming phase, infants in both groups uniformly looked more at the East Asian primes ($M = 12.27$, $SD = 2.21$) than the Caucasian primes ($M = 10.82$, $SD = 2.38$), as shown in the main effect of Priming, $F(1,107) = 10.96$, $p = .001$, $\eta^2_p = .093$. This again suggests that infants have the ability to discriminate between the two ethnicities. This is also consistent with the work of May, 2015, wherein even Vancouver infants from English monolingual, mono-cultural (Caucasian) homes showed a preference for East Asian faces.

3.3. Test phase

The main analyses tested whether infants’ ability to discriminate contrastive sounds in Chinese and in English is a function of the priming condition. A 2 Trial Type {Within-subject: Switch, Same} × 2 Group {Between-subject: Bilingual, Monolingual} × 2 Priming {Between-subject: Caucasian, East Asian} × 2 Contrast {Between-subject: English, Chinese} ANOVA revealed a marginally significant 3-way interaction among Trial Type × Group × Contrast, $F(1,107) = 6.88$, $p = .056$, $\eta^2_p = .034$, which showed a group difference between monolinguals and bilinguals in the discriminability of the contrasts. The three-way interaction was followed by two separate three-way ANOVAs for bilinguals and monolinguals, each looking at Trial Type {Within-subject: Switch, Same} × 2 Priming {Between-subject: Caucasian, East Asian} × 2 Contrast {Between-subject: English, Chinese}. For bilinguals, there was a marginal Trial Type × Contrast interaction, $F(1,41) = 5.12$, $p = .10$, $\eta^2_p = .059$. Follow-up one-sample t-tests revealed that discrimination of the Chinese contrast was above chance when the infants were primed with East Asian faces, $t(11) = 2.13$, $p = .03$, Cohen's $d = .61$, but not different
from chance for the other three prime-contrast combinations. Monolingual infants’ discrimination was not significantly different from chance in any of the four conditions. Further testing with pair-wise paired t-tests confirmed this pattern of results, as summarized in Table 2, which also provides the descriptive statistics of each prime-contrast combination of both groups of infants. Finally, no between-subject effects, which ignored the looking time difference between the Same and Switch trial by lumping them together, were found. This means that no particular ethnicity-language combinations made infants pay more attention during the test phase. Thus, the data provided support for the hypothesis that bilinguals better discriminated the Chinese contrast when primed with East Asian faces than when primed with Caucasian faces, where discrimination was not above chance. Conversely, it was not found that bilinguals’ English discrimination was better when primed with Caucasian faces than when primed with East Asian faces, as discrimination was not above chance in either priming conditions.

3.4. Additional analyses to rule out alternative explanations for the findings

Given the potential influence of language exposure on phonetic discrimination, exploratory bivariate correlations were calculated between discrimination scores (the difference between the looking time to the switch trial from the same trial) and each of the two language exposure indices – the percentage of exposure to Chinese, and the weaker language of the two. The percentage was assessed using a variant, as described in Methods, of the Language Exposure Questionnaire (Bosch & Sebastián-Gallés, 1997), adapted to include an ethnicity exposure section. The percentage of exposure to Chinese was calculated as the total lifetime exposure, in terms of hours, to Chinese languages
reported by the caregiver to be directly spoken to the infant, divided by the total hours of
direct exposure to all languages. Since the bilinguals were almost exclusively exposed to
English or Chinese languages, knowing the percentage of Chinese provides sufficient
information for their English exposure as well (for bilinguals, the correlation between
Chinese and English exposure was -.97, \( p < .001 \)). The percentage of the weaker
language of the two was calculated to index language balance (as in Byers-Heinlein,
2013). The higher the percentage, the greater the balance between the two languages.
Since there was not enough variability in language exposure among monolinguals, the
bivariate correlations were calculated only for bilinguals. The results showed that in the
Caucasian priming condition, the percentage of exposure to Chinese negatively predicted
discrimination (\( r(22) = -.45, \ p = .026 \)), indicating that the more exposure to English (i.e.
the less exposure to Chinese), the better the discrimination for both the Chinese and the
English contrast. None of these indices significantly predicted phonetic discrimination in
any other conditions. Finally, a series of 2 Priming × 2 Contrast ANOVAs with the
language exposure indices as the D.V. confirmed that language exposure did not differ as
a function of priming, contrast, or priming by contrast.

As predicted, for both the monolingual and bilingual groups, phonetic
discrimination could not be explained by the gender of the infant or the version of the
stimuli. Bivariate correlations between any of two of the six trained coders revealed high
inter-rater reliabilities. On the same trial, the correlations were \( r(17) = .93, \ r(44) = .91,\)
\( r(26) = .99, \) and \( r(12) = .92, \) respectively for Coder 1 & 2, Coder 1 & 4, Coder 2 & 3, and
Coder 5 & 6. On the switch trial, the correlations were $r(17) = .97$, $r(44) = .97$, $r(26) = .98$, and $r(12) = .93$, $p < .001$ for all the rs.

Table 2. Infants’ looking times and pairwise t values

<table>
<thead>
<tr>
<th>Prime</th>
<th>Contrast</th>
<th>Same (sec)</th>
<th>Switch (sec)</th>
<th>Overall looking time</th>
<th>t</th>
<th>Cohen’s $d$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiCH East Asian</td>
<td>Chinese</td>
<td>7.21 (3.13)</td>
<td>9.50 (3.54)</td>
<td>16.71 (5.55)</td>
<td>2.13</td>
<td>0.61</td>
<td>.03*</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>8.09 (3.73)</td>
<td>7.54 (3.63)</td>
<td>15.63 (5.55)</td>
<td>-0.39</td>
<td>-0.11</td>
<td>.35</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Chinese</td>
<td>9.19 (4.61)</td>
<td>10.16 (4.72)</td>
<td>19.35 (8.86)</td>
<td>1.15</td>
<td>0.33</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>9.14 (4.09)</td>
<td>9.20 (3.82)</td>
<td>18.33 (7.25)</td>
<td>0.06</td>
<td>0.02</td>
<td>.48</td>
</tr>
<tr>
<td>Mono E East Asian</td>
<td>Chinese</td>
<td>7.33 (3.28)</td>
<td>7.78 (2.92)</td>
<td>15.11 (4.98)</td>
<td>0.50</td>
<td>0.12</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>9.34 (3.29)</td>
<td>9.29 (3.31)</td>
<td>18.63 (4.89)</td>
<td>-0.04</td>
<td>-0.01</td>
<td>.48</td>
</tr>
<tr>
<td>Caucasian</td>
<td>Chinese</td>
<td>9.95 (4.01)</td>
<td>7.66 (3.23)</td>
<td>17.62 (4.90)</td>
<td>-1.59</td>
<td>-0.42</td>
<td>.07a</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>8.44 (3.59)</td>
<td>8.77 (4.17)</td>
<td>17.21 (7.46)</td>
<td>0.68</td>
<td>0.15</td>
<td>.25</td>
</tr>
</tbody>
</table>

Infants’ looking times (sec) to the Same, Switch, and both test trials (in the column “Overall”), pairwise $t$ values (switch minus same) and standardized effect sizes, Cohen’s $d$s, based on the difference in looking times to the same and the switch test trials in each of the experimental conditions.

Standard deviations are given in parentheses.

a. Negative differences signify a preference for the same trial that is not interpretable.

**Figure 4. Looking times data**
The looking times (in sec) to the Switch and the Same trial of bilingual and monolingual infants in the each of the priming-contrast combinations. For example, the leftmost column, “Chinese – East Asian” shows the infants who were primed with East Asian faces, and heard the Chinese contrast during the habituation and priming phase. The error bars represent the .90 C.I.s of paired comparisons between the Switch and the Same trial.
4. Discussion

It was predicted that bilinguals primed with Caucasian faces would better discriminate the English contrast than those primed with East Asian faces, and that bilinguals primed with East Asian faces would discriminate Chinese better than those in the Caucasian priming condition. Bilinguals were hypothesized to be able to discriminate both the Chinese and the English contrasts better than chance, and monolinguals were expected to discriminate the English contrast only. Finally, bilinguals were assumed to be tracking language-ethnicity links. The results provide support for the hypothesis that brief exposure to East Asian faces made 10-month-old bilingual Chinese-English learning infants’ look more during the Switch than the Same test trial to a greater extent than infants in the Caucasian priming condition, in the sense that the East-Asian-primed infants looked significantly longer at the switch trial while the Caucasian-primed group did not differ from chance. However, brief exposure to Caucasian faces did not make the infants discriminate the English (non-Chinese) phonetic categories better than those in the East Asian priming condition. In addition, the results appeared to reflect the difficulty of the /fa/ - /va/ contrast, as neither bilinguals nor monolinguals discriminated the English contrast at all.

Given that the processing advantage for Chinese sounds was selective for East Asian faces, the result can be taken as confirming the assumption that bilingual infants were tracking the link between language and ethnicity in their environment, as Kandhadai, Danielson and Werker (2014) conjectured. Furthermore, given that the ethnicity primes were not coupled with language, bilinguals’ selective activation of the
Chinese sound discrimination hypothetically reflects that language and ethnicity are in the same package of the same language mode (Grosjean, 2001), or functional category of language (Byers-Heinlein, 2014). A relevant point is that bilinguals seem to register that the /tsa/ and / tsʰa / are components of Chinese languages, indicating their phonological awareness (as mentioned in e.g. Sundara, Polka and Molnar, 2008; Werker & Byers-Heinlein, 2008; Curtin, Byers-Heinlein & Werker, 2011; Byers-Heinlein & Fennell, 2014). Most importantly, the result supports the overall thesis that bilinguals indeed use the ability to track language-ethnicity links to selectively activate the relevant language mode, in this case the link between East Asian ethnicity and Chinese language.

Strengthening this point is the fact that lower-level accounts seem not supported: the selective processing advantage for the discrimination of Chinese sounds is unlikely due to different levels of attention as a function of any prime-contrast combinations, because if the looking times to the same and the switch trials were lumped together as an index of total amount of attention, none of the factors – Priming, Contrast, nor the interaction between them – were significant. There is also no strong enough evidence to conclude different levels of fatigue, making the attention account less plausible. Last but not least, the differential looking times to the East Asian and Caucasian faces during the priming phase confirms that both bilinguals and monolinguals are able to tell the ethnicities apart. Thus, the result does seem to indicate that it was the priming with East-Asian faces that resulted in an improvement in discrimination of the Chinese contrast.

One important question left to answer is why the /fa/ - /va/ contrast was not discriminated by bilinguals primed with Caucasian faces, and importantly, if it suggests
that bilinguals did not activate the English mode. One simple explanation is that bilinguals did not activate the English mode in response to Caucasian faces due to the scarcity of exposure to Caucasian faces (9% only out of their total life time). If bilinguals are tracking language-ethnicity co-occurrences, it can be hypothesized that insufficient experience with Caucasian faces prevented bilinguals from regarding Caucasian faces as suggestive of the English language. In fact, just as monolinguals in Vancouver of a similar age learned the specific pairing of East Asian and Cantonese instead of merely expecting unfamiliar language to co-occur with unfamiliar ethnicity (May, 2015), it is possible that bilinguals in the same area also had to learn the specific combination between Caucasian faces and English and the 9% life exposure was insufficient to allow for this. However, one limitation of this account is that it does not explain why, in the same study by May (2015), the scarce input of the East Asian faces from the neighbourhood sufficed for monolinguals to learn the specific link between East Asian and Cantonese. Unless there are strong reasons to suspect that the possible privileged status of language-ethnicity links would be missing in the bilingual population, this account is less likely.

Alternatively, bilinguals could be already in the English mode in the face of Caucasian primes, but the /fa/ - /va/ contrast chosen to represent the English language was inherently difficult for 10-month-olds. Given that phonetic discrimination very likely depends not only on the developmental level, but also the perceptual properties of the contrast itself (see Werker & Curtin, 2005, for a review; and see Narayan, Werker, & Beddor, 2010 for empirical evidence of the difficulty of some acoustically similar
contrasts), the developmental-level readiness among 10-month-olds for their native language in general (e.g. Werker & Tees, 1984; Kuhl et. al., 2006; Albareda-Castellot, Pons & Sebastián- Gallés, 2011; Sundara, & Scutellaro, 2011; Burns, Yoshida, Hill & Werker, 2007; Sundara, Polka & Molnar, 2008; Burns et al., 2007) might not be enough to enable the discrimination. This is consistent with the lack of findings that infants at 10 months of age could discriminate fricative sounds, to which /f/ and /v/ belong, in the literature. Indeed, the voicing distinction in another fricative contrast, /s/ - /z/ was shown to be not discriminable even in 12-month-old Catalan-learning monolinguals, who were exposed to the contrast (Bosch & Sebastián-Gallés, 2003b). In support of this explanation is the finding that even monolinguals, typically found to have specialized in their native language discrimination by 10 months of age, did not discriminate this English contrast. Thus, it remains a strong possibility that, due to the likely privileged status of ethnicity-language links, the minimal exposure among bilinguals to Caucasian faces was enough for bilinguals to establish the Caucasian-English link and to activate the English mode, but the effects were attenuated by the difficulty of the English contrast, ultimately yielding chance performance.

A related issue is whether the English mode was also activated after East Asian primes in bilingual infants. Although in the current sample East Asian faces negatively predicted the amount of English spoken to the bilingual infant, there was still a 64% chance for bilinguals to receive English input from caregivers of East Asian ethnicity. Thus, if bilinguals were tracking language and ethnicity co-occurrence, it can still be hypothesized that East Asian faces may activate the bilingual mode (Grosjean, 2001)
where both English and Chinese were activated instead of Chinese alone; it is due to the difficulty of the English contrast that the current study failed to provide evidence for it. Along the same line of reasoning, the fact that most infants had been tested by East Asian experimenters, who, while explaining the research to the caregivers, might have already primed the infant towards the “Chinese end”, may explain the higher performance on the Chinese contrast. Again, it is also possible that the East Asian experimenters instead primed infants into the bilingual mode, but the difficulty of the English contrast made the activation of English discrimination more difficult to show.

Of interest is also why the bilingual discrimination of Chinese was at chance given the Caucasian primes. To answer this question, it may be useful to consider the baseline discriminability of bilinguals on the /tsa/ - /tsʰa/ contrast and the nature of language mode activation (see Byers-Heinlein, 2014 for a summary of these issues). If, be it due to the developmental level or perceptual properties of the contrast, /tsa/ - /tsʰa/ was too difficult to be discriminated, the apparent absence of Chinese discrimination in the Caucasian priming condition would be a natural consequence, especially when the bilingual infant was not in the relevant language mode. Under this account, the activation of the relevant language mode has to be conceptualized as facilitative enough to have boosted the performance of the /tsa/ - /tsʰa/ contrast to reconcile with the finding that /tsa/ - /tsʰa/ could be discriminated in the East Asian priming condition. Given that affricates, including /ts/ and /tsʰ/, are understudied, this account is possible. As a second account, if the activation of the relevant language mode entails the suppression of the irrelevant one, the Caucasian primes, which activated the English mode, could be
hypothesized to have impeded the /tsa/ - / tsʰa / distinction among bilinguals, whatever is
the baseline performance of this contrast. This account is consistent with the line of
reasoning that bilinguals must, and will become adept at, suppressing the language not
demanded by the situation. Scholars who hold this view also tend to report evidence for
bilingual cognitive advantages in conflict resolution or executive function tasks (e.g.
Bialystok, Craik & Luk, 2012; Costa, Hernández, & Sebastián-Gallés, 2008). Although
the current results cannot tease apart these two accounts, which are not mutually
exclusive, they concur on the overall thesis that bilinguals were able to use their
knowledge of language-ethnicity links to activate relevant language modes. Finally, a less
favourable explanation is that the baseline discriminability among bilinguals is present
for the Chinese but not the English contrast. Without the ability to selectively activate
language modes, bilinguals were simply distracted less by the East Asian faces than the
unfamiliar Caucasian faces, contributing to the above chance performance on the Chinese
contrast discrimination. This final account is, however, unlikely, because bilinguals, like
monolinguals, actually looked longer at the East Asian primes than the Caucasian primes.
If distraction was key, it would be expected that the East Asian faces would interfere with
the performance more, which is not supported by the result pattern. Therefore, the more
likely explanations still favour the presence of bilingual sociolinguistic sensitivity.

Given that monolinguals are sensitive to the links between language and ethnicity
(e.g. May, 2015; Uttley et al., 2013), it is possible that all the factors discussed above –
the relative status of activation of the English mode and the baseline performance on the
discrimination of both contrasts – are also relevant. For example, the lack of
discrimination of the Chinese contrast in the Caucasian condition by monolinguals could have been due both to the lack of a Chinese mode and perceptual attunement that has pruned out the Chinese distinction, or it could be due to the Caucasian-incurred suppression of the Chinese distinction, which could be so perceptually salient that the baseline performance would have been above chance with no primes. Lower-level explanations, such as not discriminating the ethnicities, are unlikely (see Kelly et al., 2005 for evidence in Caucasian infants; Bar-Haim, Ziv, Lamy, & Hodes, 2006, for African infants). Moreover, monolinguals in the current study showed a preference for East Asian faces (consistent with May, 2015 and Liu et al., 2015), indicating successful discrimination.

Taken together, the current study provides evidence that 10-month-old bilingual Chinese-English learning infants were able to apply at least part of their knowledge of ethnicity-language links to activate the relevant language mode, as they discriminated the Chinese contrast better in the East Asian priming condition than in the Caucasian priming condition. Thus, after considering factors such as the difficulty of the contrast and whether the activation of one language mode suppresses the other, among others, I argue that it is more likely than not that bilinguals were capable of selective activation of phonetic distinctions in response to ethnicity as the contextual cue. This capability, given their ability to tell ethnicities apart in the first place, is a sufficient condition for Kandhadai, Danielson & Werker (2014)’s proposal that bilinguals are able to track language-ethnicity co-occurrences, possibly with biological readiness. Lower-level
attentional accounts are unlikely. The monolingual result is not conclusive.

Future research is needed to find a phonetic contrast to operationally represent the

English mode with comparable difficulty to the Chinese one. Also, direct
information of the baseline performance on the phonetic discrimination of the two
language contrasts – without any priming – could be helpful in understanding the nature
of the activation of language modes, whether it is facilitative, suppressive, or both.

Finally, broader implications of the symphony among biological endowment, language-
specific input, and general learning mechanisms on language-related perception beyond
the purely linguistic domain can be further investigated.
References


Byers-Heinlein, K. (2014). Languages as categories: Reframing the “one language or two” question in early bilingual development. *Language Learning, 64*(s2), 184-201.


Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants’ selective attention to the mouth of a talking face. Psychological science, 0956797614568320.


Appendices

Appendix A: Language and Ethnicity Questionnaire

Language and Ethnicity Questionnaire

Personal Data

BABY ID: ________________________________

DATE OF BIRTH: ________________________________

DATE OF EXPERIMENT: ________________________________

LANGUAGES SPOKEN BY FAMILY:

<table>
<thead>
<tr>
<th>Parent A</th>
<th>Parent B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side</td>
<td>Side</td>
</tr>
</tbody>
</table>

ETHNICITY OF FAMILY MEMBERS:

<table>
<thead>
<tr>
<th>Parent A</th>
<th>Parent B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Side</td>
<td>Side</td>
</tr>
</tbody>
</table>

NOTES:

TYPICAL DAY LENGTH: Wake up time: ______ Bed time: _______ Typical day length (max 24): ______

PARENTS/CARETAKERS (e.g., parents, grandparents, babysitters, etc.):

<table>
<thead>
<tr>
<th>Who?</th>
<th>Language Spoken</th>
<th>Ethnicity</th>
<th>What ages?</th>
<th>More than 1 hour per week</th>
<th>Hours/Week</th>
<th>Since When?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FAMILY (e.g., grandparents, siblings, aunt, uncles etc.):

Who spends time with the baby and what language do they speak

<table>
<thead>
<tr>
<th>Who?</th>
<th>Language Spoken</th>
<th>Ethnicity</th>
<th>What ages?</th>
<th>More than 1 hour per week</th>
<th>Hours/Week</th>
<th>Since When?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VIDEO CONFERENCING/ TV
Does the baby participate in video conferencing (e.g. Skype)? Y / N Watch TV? Y / N

<table>
<thead>
<tr>
<th>Who/Shows</th>
<th>Language</th>
<th>Ethnicity</th>
<th>What ages?</th>
<th>&gt; 1 hour a week</th>
<th>Hours/Week</th>
<th>Since when?</th>
</tr>
</thead>
</table>

FRIENDS
Who spends time with the baby and what language do they speak

<table>
<thead>
<tr>
<th>Who?</th>
<th>Language Spoken</th>
<th>Ethnicity</th>
<th>What ages?</th>
<th>More than 1 hour per week</th>
<th>Hours/Week</th>
<th>Since When?</th>
</tr>
</thead>
</table>

BABY GROUPS (play groups, story time, mom/baby classes, etc)
Do you attend baby groups?

<table>
<thead>
<tr>
<th>Which?</th>
<th>Language Spoken</th>
<th>Ethnicity</th>
<th>What ages?</th>
<th>More than 1 hour per week</th>
<th>Hours/Week</th>
<th>Since When?</th>
</tr>
</thead>
</table>

DAYCARE
Does your child attend daycare?

<table>
<thead>
<tr>
<th>Since when?</th>
<th>Language Spoken</th>
<th>Ethnicity of daycare provider</th>
<th>Hours/Week</th>
</tr>
</thead>
</table>

NEIGHBORHOOD AND PUBLIC AREA
How often do you go outside with your baby?

<table>
<thead>
<tr>
<th>Where?</th>
<th>Languages Spoken</th>
<th>Ethnicity</th>
<th>Days/Week</th>
<th>Hours/Day</th>
<th>Since When?</th>
</tr>
</thead>
</table>


**TRAVEL**

Has the child lived/vacationed in any country where s/he would hear a language other than English?

If yes, Where? _____________________________________________________________

When? ________________________________________________________________

And for How long? _______________________________________________________

**TOTAL ESTIMATE:** ............... % L1/ ............... % L2/ ...............% other

............... % E1/............... % E2/...............% other