University of Alberta

Lexical tone processing by monolingual and bilingual speakers of tone and non-tone languages

by

Donald Kyle Danielson

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

Department of Linguistics

Fall 2012
Edmonton, Alberta

Abstract

This study tests early Chinese-English bilinguals' perception of Thai lexical tone. Lexical tone is a feature that is used contrastively in Chinese but not in English. Chinese-learning infants exhibit native-like treatment of Thai tonal contours, while English learners exhibit non-native perceptual behaviour (Mattock & Burnham, 2006). However, early Chinese-English bilingual adults in the present study do not perform differently on the task than do monolingual English speakers with no tone language experience. Late Chinese-English bilinguals perform more accurately than both of the other groups. Early bilinguals do exhibit evidence of Chinese language experience, as their within-task processing strategies more closely resemble those of late bilingual speakers than those of English monolinguals. Developmental and biographical explanations for the behaviour of the early bilinguals are explored through the use of a language proficiency questionnaire, and the role of bilingual dominance is analyzed in relation to performance on this task.

Acknowledgment

I am grateful for the support of my supervisor, Benjamin Tucker, without whose assistance this project would not have been possible. In no particular order, I also thank Anne-Michelle Tessier, Paula Chesley, Emmanuel Dupoux, Harald Baayen, Santiago Barreda, and the members of the Alberta Phonetics Laboratory for their advice at different stages throughout the experimental and analytical stages of this study. Finally, I am indebted to the generous funding provided to me by various organizations at the University of Alberta: the Department of Linguistics, the Faculty of Graduate Studies and Research, and the Killam Research Fund.

Table of Contents

List of Tables

List of Figures and Illustrations

0 Introduction

Research from the last half-century has demonstrated conclusively that monolingual adult speakers have difficulty processing changes in the speech signal that are not used contrastively in their native language. Although infants are born exhibiting language-general perceptual behaviour, *perceptual reorganization* occurs early in life and results in specialized, language-specific phonological systems by adulthood. This study utilizes the perception of fundamental frequency (F_0), the primary acoustic correlate of lexical tone, to draw conclusions about the ways in which the perceptual framework and developmental trajectory of early bilinguals differs from that of their monolingual peers. The research question is centred upon a large body of investigation that has demonstrated a decline in universal speech cue perception from infancy into adulthood based upon native language input. Drawing upon both infant- and adult-centred literature primarily based on experimentation with monolingual speakers, this work builds upon the argument that the perception of slight fundamental frequency changes in the speech signal constitutes an area of difficulty for listeners that do not use this feature contrastively in their native language, while it presents little challenge for those listeners that do (i.e., speakers of tone languages). This pre-existing claim is tested and validated in the present experimentation. With this assertion in mind, a similar question is raised and tested with early bilingual speakers who have had lifelong experience in both a tone language and a non-tone language. Differences in the behaviour of these early bilinguals with respect to monolinguals point to variations in childhood linguistic input that may significantly affect perception later in life.

In Chapter 1, the work is introduced with a review of the relevant background literature, beginning broadly with studies concerning perception of non-native speech contrasts (Section 1.1). This section outlines seminal research that has investigated the decline of universal speech cue perception as early as infancy, and which has pinpointed various benchmark stages in perceptual development. Special attention is then given to the perception of suprasegmental features, particularly lexical tone, and studies centred primarily upon these speech cues are reviewed (Section 1.2). While most of this research has been conducted with monolingual populations, some more recent research has focussed upon the unique developmental trajectory of early bilingual children, and that work is outlined here (Section 1.3).

Chapter 2 describes the methodology of the current experimentation, comprised of a behavioural study and a linguistic aptitude questionnaire. Chapter 3 presents the major findings of the experiment, and the results are analyzed using a variety of statistical techniques. These methods include traditional tools (e.g., linear regression, analysis of variance, *t*-tests), as well as more specialized procedures (e.g., mixed effects regression, principal components analysis). In the case of the latter methodologies, care is taken to justify their use and briefly describe the relevant mechanisms.

Chapter 4 concludes the study, outlining future directions and arguing that the experimentation outlined in Chapters 2 and 3 contributes significantly to the existing body of literature concerned with acoustic perception and early bilingualism.

1 Background

1.1 Perception of non-native speech contrasts

The fundamental acoustic properties of an individual's native language affect the manner in which he or she processes subtle distinctions in the speech signal. While infants are born with sensitivity to a wide variety of sounds outside their native language(s) (often referred to as *language general* perception), this tendency declines with age and linguistic input. Simply, because the phonemic inventory of linguistic systems varies greatly, an individual language may not utilize certain sounds to form words. When a pair of closely related sounds does not occur in a language, adults have difficulty discriminating between them.

A number of studies have produced findings that demonstrate the robustness of this claim in adult speech perception of non-native consonantal contrasts. For example, English speakers have difficulty distinguishing between the Hindi dental phone [t̪] and its retroflex equivalent [ʈ], as the sounds are not distinctive in English and are rarely realized in natural English speech (Werker, Gilbert, Humphrey, & Tees, 1981). Likewise, English speakers categorize both the unaspirated [p] as well as the aspirated [p^h] of Thai as the same English phoneme, /p/ (Curtin, Goad, & Pater, 1998).

Much attention has also been given to Japanese speakers' perception of the English /r/-/l/ contrast. In English, /r/ is typically realized as the (post)alveolar approximant [ɹ], while in Japanese /r/ is realized as an alveolar tap, more closely related to the North American English [ɾ] in *butter*. Moreover, Japanese does not have a lateral approximant that corresponds to the English /l/ (Best & Strange, 1992). As a result, Japanese speakers have difficulty perceiving the English /r/-/l/ distinction, and do so at a level only slightly better than chance (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). Instead, Japanese speakers often perceive both English sounds as poor realizations of their own phoneme /w/, normally realized in Japanese as a velar approximant (Best & Strange, 1992). English speakers, on the other hand, perceive this native distinction categorically. Some studies (e.g., Goto, 1971) have demonstrated that this effect is evident in Japanese speakers even after years of living and working in an English-speaking environment (but see MacKain, Best, & Strange, 1981).

A similar phenomenon occurs with the perception of non-native vowel contrasts. For example, standard Catalan has a vocalic space consisting of eight contrastive vowels, including the mid-front vowels /e/ and /ɛ/. Catalan speakers, who use the /e/-/ɛ/ distinction to form separate lexical items, easily perceive the two vowels as different. Spanish, on the other hand, has a relatively small vocalic inventory, consisting of only five vowels. The Spanish vowel space includes the mid-front vowel /e/, but not /ɛ/, thus Spanish speakers categorize both vowels as /e/ (Sebastián-Gallés & Bosch, 2005). In other words, the Spanish speakers "map" these two segments to the same conceptual space. This co-perception makes it challenging for Spanish speakers to distinguish between the minimally contrastive sounds, which they perceive as the same phoneme. Even with significant exposure to and experience with Catalan, Spanish-dominant bilinguals show a lack of sensitivity to the /e/-/ɛ/ contrast (Sebastián-Gallés & Soto-Faraco, 1999). Corroborating results have been shown for English speakers' ability to distinguish between oral and nasal vowels, which are not phonemic in English but which are regularly realized as allophones. Hindi speakers, who utilize the nasal-oral contrast phonemically, perceive the distinction categorically, while English speakers tend to perceive the vowels along a continuum (Beddor & Strange, 1982).

Reduction in the categorical perception of non-native phonetic distinctions is so persistent that investigators in some areas have even labelled it as “deafness” (Dupoux, Pallier, Sebastián, & Mehler, 1997, *inter alia*). Importantly, however, researchers have successfully demonstrated that the decreasing perceptual sensitivity described here is likely not due to sensorineural loss, but rather is the result of a shift in attentional focus or processing strategies (Werker & Tees, 1984a). Although speakers are able to accurately perceive phonetic distinctions between minimally contrastive sounds, they only maintain a given distinction in memory for long periods of time when it is phonemic in their native language. If non-native stimulus items are presented with a low interstimulus interval, discrimination is improved because of a lower burden upon short-term memory (Werker & Logan, 1985). This retention may be difficult, however, if the behavioural task is cognitively burdensome and the stimuli have characteristics similar to those of natural, connected speech. Because a decline in demonstrated perceptual abilities is attentional, rather than neurological, intensive training may allow adult speakers to discriminate non-native phonemic contrasts categorically. For example, when Japanese speakers are exposed to intensive immersion training in English, they begin to categorize /r/ and /l/ like English speakers (MacKain, et al., 1981, and see Wang, Spence, Jongman, & Sereno, 1999, below).

This shift in attentional focus, called perceptual reorganization, occurs relatively early in development. Various studies have successfully pinpointed the stages of this process, which affects the perception of different sound classes at various ages. Discrimination of non-native vocalic contrasts begins declining between four and six months of age. By eight months, infants have difficulty differentiating between vowels that are not contrastive in their native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994; Bosch & Sebastián-Gallés, 2003). Later reorganization occurs for consonantal contrasts, and appears to start just before 10 months of age. For example, while younger infants demonstrate language general perception, 10- to 12-month-old English-learning infants exhibit adult-like desensitivity to the Hindi [ʈ] - [t̪] contrast described above (Werker & Tees, 1984b). By one year of age, perceptual reorganization is mostly complete, and infants demonstrate language-specific speech perception of most segments.

1.2 Perception of suprasegmental features

However, perceptual reorganization is not restricted to the segmental properties of the speech signal. Other researchers have demonstrated that a decline in perceptual accuracy is also observed in behavioural tasks testing speakers on non-native suprasegmental contrasts, such as stress and tone, both of which occur across more than one speech syllable.

1.2.1 Lexical stress

Stress is primarily formulated by the manipulation of various degrees of three acoustic properties of the speech signal: duration, fundamental frequency (pitch), and intensity (loudness). Languages like English mainly utilize stress prosodically, for emphasis at the utterance level (e.g., "I didn't say *that* to him" vs. "I didn't say that to *him*"). Stress can also be used in English at the grammatical level, often to distinguish between nominal and verbal forms (e.g., /ˈ**rɛ**.kɚd/ 'record (n.)' vs. /rə.ˈ**kɔrd**/ 'record (v.)'). Note, however, that stress is generally not the only difference between minimal nominal-verbal pairs in English. In both forms of 'record', the unstressed syllable also contains a reduced vowel. Often, however, a language's use of stress is much more widespread. Spanish utilizes stress phonemically to differentiate between many minimal pairs of lexical items (e.g., /ˈ**to**.mo/ 'I take' vs. /to.ˈ**mo**/ 'he/she took'). Because Spanish does not have reduced vowels, the only distinction between these two forms is the placement of lexical stress. Other languages, like French, utilize stress only predictably. Although French makes use of stress for emphasis and phrase-final intonation, it is not used phonemically to distinguish between lexical items.

These cross-linguistically diverse uses of stress have prompted researchers to investigate the perception of this suprasegmental cue by speakers of different languages. Dupoux and colleagues (1997) tested the stress perception of French and Spanish monolinguals using a speeded ABX task. Participants heard two pseudoword stimulus items (A, B) differentiated only by lexical stress placement, and were then prompted to match a third word (X) to one of the first. Spanish speakers performed significantly better on the task than French speakers did. This discovery, that French speakers had difficulty processing differences in lexical stress placement, was replicated and strengthened using a different task (Dupoux, Peperkamp, & Sebastián-Gallés, 2001), and later found to be generalizable even to French speakers with extensive L2 knowledge of Spanish (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). Research with infants indicates that perception of lexical stress also undergoes reorganization, as French-learning infants, but not Spanish-learning ones, lose sensitivity to the placement of stress by nine months of age (Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Gallés, Limissuri & Peperkamp, 2009).

1.2.2 Lexical tone

Many of the world's languages also utilize lexical tone, perceived by the listener as pitch and acoustically correlated to fundamental frequency (F_0). As the majority of tone languages are not contained within the Indo-European language family (Yip, 2002), fewer studies have concentrated on the cross-linguistic perception of this speech cue. However, many Austroasiatic and African languages utilize this feature contrastively, to differentiate between lexical items that are otherwise identical. Mandarin Chinese, for example, has four lexical tones (high, rising, low, and falling), each of which follows a chronologically distinct F_0 trajectory. Tones can be minimally contrastive, differentiating between lexical items that are otherwise identical. The Mandarin syllable /ma/ (meaning 'horse', 'mother', or 'scold', depending on tone) is often cited as an example of minimal tone pairs.

English, on the other hand, uses pitch for certain prosodic purposes, such as contrasting declarative and interrogative sentences, but does not utilize tone to construct minimal pairs of lexical items. As is the case with French speakers' reduced perception of lexical stress, English speakers have difficulty distinguishing minimal pairs differentiated only by tonal contour (Burnham, Francis, Webster, Luksaneeyanawin, Lacerda, & Attapaiboon, 1996; Burnham, Kirkwood, Luksaneeyanawin, & Pansottee, 1992). Mattock and colleagues have demonstrated that perceptual reorganization for tone is complete by about nine months of age. English-learning and Chinese-learning infants discriminate tone types at four and six months, but English-learning infants fail to do so categorically at nine months of age (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008). Not only do English-learning infants have greater difficulty with tone minimal pair differentiation, they are also sensitive to *type* of tone distinction. That is, some pairs of tones are more difficult for English learners to distinguish than others, while Chinese-learning infants do not demonstrate such a bias (Mattock & Burnham, 2006).

However, just as with the segmental contrasts reviewed above, the difficulty in tone perception demonstrated by speakers of non-tone languages appears to be attentional, not neurological. Speakers of English and other non-tone languages are still *able* to distinguish between tone types, but have difficulty doing so under conditions similar to those of natural speech (Burnham, 2000). Wang and colleagues (1999) trained English-speaking university students to consistently categorize Mandarin lexical tones, an effect that persisted in the trained listeners months after the original experimentation. If English speakers' desensitivity to tonal differences were sensorineural, rather than attentional, the result obtained by Wang, et al. would have been unlikely. Participants would not have perceived the differences in lexical tone during the training phase, and could not have demonstrated modified behaviour in the follow-up experimentation.

1.3 Development of bilingual speech perception

While much attention has been given to non-native speech perception and second language phonological development, more recent research has investigated the unique perceptual development of individuals with more than one first language. Bilingual learners, who receive input from at least two distinct phonological systems, undergo perceptual reorganization differently than do their monolingual peers, and exhibit behavioural differences in perception as early as birth (Byers-Heinlein, Burns, & Werker, 2010). For these learners, certain phonetic contrasts may be phonemic in one of their languages, but not in the other. Burns, Yoshida, Hill, & Werker (2007) tested French-English bilingual perception of a voicing contrast (/b/-/p/), the voice onset times (VOT) of which are different in French than in English. Both monolingual and bilingual infants in the 6-8 month range demonstrated language general perception of this VOT distinction. By 10-12 months, however, French and English monolinguals perceived the contrast only at the voicing boundary utilized by their language. Bilinguals, on the other hand, were able to perceive the contrast at both boundaries. In this case, bilinguals exhibit a behaviour that successfully combines the perceptual systems of their two languages.

However, some studies have demonstrated that, for certain phonetic contrasts, bilinguals undergo a distinct developmental trajectory that cannot be equated to the combination of their two linguistic systems. For example, as noted above, Spanish monolinguals cease to categorically discriminate the Catalan /e/-/ε/ distinction by eight months of age, while Catalan monolinguals, for whom the distinction is contrastive, continue to easily discriminate the vowels (Bosch & Sebastián-Gallés, 2003). However, the same study demonstrated that Catalan-Spanish bilinguals, when presented with decontextualized stimuli, *also* ceased discriminating between these vowels by eight months. Of course, Catalan-Spanish bilingual adults must be capable of perceiving this particular vocalic distinction, as it is contrastive in one of their languages. In fact, in that study, bilingual infants regained the ability to categorize these sounds by 12 months of age. This "u-shaped" pattern of the bilinguals' behaviour, although task-specific, supports a claim that bilingual perceptual reorganization follows a slightly different path than that of monolinguals (Sebastián-Gallés, 2010).

Although the two studies cited above converge on the idea that simultaneous bilinguals ultimately emerge with functional perceptual systems in both languages at the most basic (segmental) level, investigations at other levels of perception indicate that simultaneous bilinguals may always perform differently from their monolingual counterparts, even into adulthood. In investigations testing the /e/- /ε/ distinction within correct and incorrect pronunciations of Catalan words, Sebastián-Gallés and colleagues have demonstrated that simultaneous bilingual children are less sensitive to mispronunciations on this particular vocalic contrast than are their monolingual

Catalan peers (Ramon-Casas, Swingley, Sebastián-Gallés and Bosch, 2009), an effect that persists into adulthood (Sebastián-Gallés, Echeverria, & Bosch, 2005). It has been proposed that this effect is possibly due to the greater variation in childhood linguistic input experienced by the bilinguals, who in turn learn to accept mispronunciations that monolinguals reject.

Similar effects have been demonstrated for the perception of lexical stress distinctions by simultaneous bilinguals. Dupoux, Peperkamp, & Sebastián-Gallés (2010), whose laboratories have successfully demonstrated the reduced ability of French monolinguals to distinguish between pseudowords differentiated only by stress placement (Dupoux, et al., 1997, 2001, 2008; Skoruppa, et al., 2009), replicated their task with simultaneous French-Spanish bilinguals. Group results indicated that the bilinguals' error rates followed a pattern intermediate to that of the French and Spanish monolingual controls, suggesting that the bilinguals were neither as proficient as the Spanish speakers nor as insensitive to the contrast as the French speakers. However, individual analyses revealed that the bilinguals' error rates were better fit by a bimodal distribution, with some individuals performing like French monolinguals (poorly) and others like Spanish monolinguals (accurately). The bilinguals did not actually exhibit an intermediate behavioural pattern, but rather acted as two groups of monolinguals. Based on a sociolinguistic questionnaire administered to participants, the authors concluded that an individual's belonging to one group or another was the result of differences in input before two years of age.

These findings are consistent with claims by some researchers that the human perceptual system is only capable of processing one language natively (Cutler, Mehler, Norris, & Segui, 1989, 1992). According to this theory, simultaneous bilinguals process one language as well as monolingual speakers do, but exhibit sub-native behaviour in their less dominant language when performing cognitively burdensome tasks (Dupoux, et al., 2010). In addition to the role of input before two years of age reported in that study, researchers have attributed bilingual dominance in one language to a variety of factors, including language of the mother (Sebastián-Gallés, et al., 2005) and subjective measurements, such as forced-choice language preference questions (Cutler, et al., 1992).

1.4 Current research question

Like Dupoux, et al. (2010), the present study investigates the behaviour of the earliest bilinguals in perceiving a contrast that is native in one of their languages and not in the other. However, while Dupoux et al. investigated simultaneous French-Spanish bilinguals on their perception of lexical stress placement, a Spanish-specific linguistic feature, this work is concerned with early simultaneous Chinese-English bilinguals and their perception of lexical tone, a Chinese-specific linguistic feature. Mattock and colleagues (2006, 2008) have successfully demonstrated that infant learners of non-tone languages lose sensitivity to tonal contrasts by nine months of age, and Burnham and colleagues (1992, 1996) have attested to the persistence of this effect into adulthood. In the first place, the present investigation utilizes an alternative methodology to replicate these monolingual-centred findings, confirming a significant difference in the perception of F_0 contours in the speech signal by speakers of tone languages (in this case, Mandarin and Cantonese) and speakers of non-tone languages (in this case, English). The present study is then the first to extend the question of lexical tone perception to a population of early bilingual (Chinese-English) speakers. The results of the investigation are first analyzed between groups, in order to determine where the differences arise in the perceptual behaviour of early bilinguals with regard to their peers in either language control group. If the findings of Dupoux, et al. (2010) are generalizable to lexical tone, it is expected that early bilinguals will follow a pattern intermediate to their peers in the two control groups. On the whole, they will perform neither as well as the Chinese controls nor as poorly as the English controls. However, individual early bilingual participants, when analyzed separately, may cluster with monolinguals in one of the two groups. As this clustering may be the result of childhood linguistic input, a language experience questionnaire is administered to the participants in order to obtain sociobiographical predictors, which are in turn analyzed alongside the behavioural data.

2 Design

The experimental method utilized here is an adaptation of the sequence recall paradigm developed by Dupoux and colleagues, which probed sensitivity to lexical stress position by monolingual French and Spanish speakers (2001, 2008, 2010). This method was developed as an alternative to an earlier speeded ABX task, described in Section 1.2 (Dupoux, et al., 1997). While the ABX task succeeded in better separating Spanish monolingual speakers from French monolingual speakers than did a simpler AX (same-different) task, French speakers still performed significantly better than chance in the ABX paradigm. Moreover, there was considerable overlap between the French and Spanish scores, and a great deal of variability was observed between individual speakers. This high variability and large overlap was not unexpected, as lexical stress employs a number of "noisy" acoustic correlates that are highly salient across populations (Dupoux, et al., 2001).

As reviewed, the developmental decline in sensitivity to non-native segmental contrasts appears to be attentional (Werker & Tees, 1984a). Adults are capable of perceiving these contrasts, but have difficulty doing so in tasks with demands similar to the online processing of natural speech. Because sensorineural perception of F_0 contours is likewise uncontested (e.g., Burnham, 2000; Wang, et al., 1999), it was assumed that there would be similarly minimal results from a same-different or speeded ABX task testing tone perception. In their later experimentation, however, Dupoux and colleagues developed the sequence recall method adopted here, which places even more burden upon short-term memory than does an ABX task and which more closely resembles the cognitive load imposed by speech processing. In fact, this paradigm succeeded in demonstrating a completely nonoverlapping distribution of individual French and individual Spanish performance scores (Dupoux, et al., 2001), and was successfully replicated in later studies (Dupoux, et al., 2008, 2010).

Although the sequence recall method was originally developed for the testing of lexical stress perception, it is utilized here to explore individual differences in the perception of lexical tone. While stimuli in the stress Condition of the Dupoux, et al. experimentation were modified on three acoustic dimensions (F_0, duration, and intensity), stimuli in the tone Condition of the present experiment are only manipulated on the F_0 dimension, the most important perceptual correlate of lexical tone (Gandour & Harshman, 1978). The stimuli in this paradigm are closely controlled acoustically, and any significant behavioural effects are assumed to be derived from the manipulations described below.

2.1 Method

The sequence recall method consists of three phases, repeated for each block of experimentation. For the purposes of this study, these steps are referred to as the *training phase*, the *test phase*, and the *experimental phase*. Appendix 1 is a graphical representation of the presentation method.

2.1.1 Training phase

During this phase, participants are trained to associate novel auditory pseudowords with keys on a standard computer keyboard. Participants are trained on one novel item at a time, using the following procedure:

1. Participants are informed that they will be learning new words. They are told that some of the new words may have familiar sounds, but that they are not words in any known language.
2. Participants are told that they will learn their first new word, and are instructed to press the associated key on the keyboard (e.g., [1]) each time they hear the new word.
3. Participants hear the first word 10 consecutive times, and press the corresponding key after each instance.
4. Participants are then told that they will learn their second new word, and instructed to press the associated key on the keyboard (e.g., [2]) each time they hear this word.
5. Participants hear the second word 10 consecutive times, and press the corresponding key after each instance.

2.1.2 Test phase

In the second phase of each experimental block, participants are tested on their association of the novel words to the keyboard keys. Participants hear 20 instances of the novel words, one at a time. For each instance, they are instructed to press the corresponding key on the computer keyboard. For example, participants must correctly associate the first novel word with key [1] and the second novel word with key [2]. Failure to correctly code the novel words at 90% or higher (two mistakes per block) later excludes a participant's data from analysis.

2.1.3 Experimental phase

In the experimental phase of the procedure, participants are exposed to sequences of the two items from the learned pair of pseudowords. The possible sequences are of five possible lengths ($2 \le n \le 6$), and are randomized within the phase. Participants are exposed to five sequences of each length, resulting in 25 (5 x 5) sequences per phase. After each sequence, participants "recall" the constituent words using the associated keyboard keys. For example, if an experimental sequence is [w_1 w_2 w_2 w_1], the correct keyboard response is [1221]. Only responses that are recalled with 100% accuracy are coded as "correct" for data analysis.

2.2 Stimulus design

Stimulus items were constructed for two Conditions comprising the three blocks of experimentation. Table 1 groups stimulus items by block.

low-rising tone block	[bà] [bǎ]
phoneme block	[mu] [fu]
rising-falling tone block	[bǎ] [bâ]

Table 1: Stimulus items by block

2.2.1 Toneme Condition

The experimental Condition consisted of two blocks of tone-differentiated pseudoword pairs. The three pseudowords in these blocks (ba_{low}, ba_{rising}, and $ba_{falling}$) were constructed to correspond to the F_0 contours of standard Thai tones, as outlined in Mattock & Burnham (2006) (Figure 1). Because the experiment was designed to probe perception of a specific acoustic correlate, it was necessary to test participants using items that were not lexical in any of their languages.

However, due to the relatively low phonemic inventory shared between Mandarin and Cantonese, it was impossible to create a licit stimulus syllable that would have no lexical meaning across all Mandarin and Cantonese tone types. Therefore, in order to maintain lexical ambiguity, Thai tonal contours were utilized.

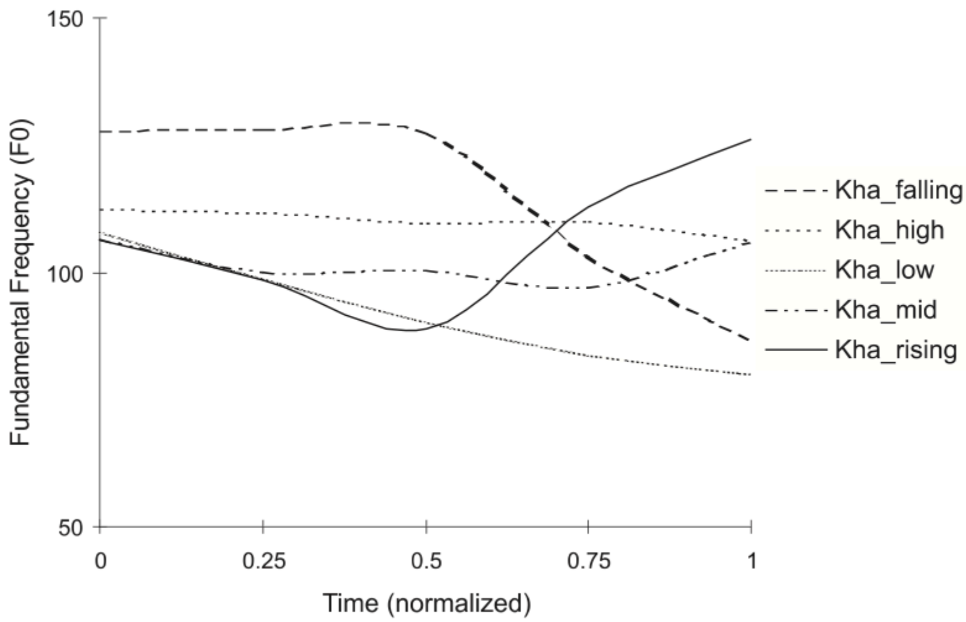


Figure 1: Time normalized Thai tonal contours (Mattock & Burnham, 2006)

The three tonal stimulus items were derived from one instance of the syllable [bɑ]. The original syllable was recorded by a male speaker of English in a sound-attenuated booth at the Alberta Phonetics Laboratory (Department of Linguistics, University of Alberta), using a Countryman E6 Omnidirectional EarSet microphone to a Korg MR-1000 digital recorder. The original syllable consisted of the initial phoneme [b], followed by a long (>1s) vocalic segment [ɑ]. Using Praat (Boersma & Weenink, 2011), the item was first truncated to a length of 500ms. The F_0 of the item was then manipulated using PSOLA with the following procedure in Praat:

1. The pitch was extracted from the stimulus item and converted to a Praat PitchTier ("time-stamped pitch contour") object[1] (Figure 2).
2. The PitchTier was levelled to remove F_0 contour points.
3. Points were added to the PitchTier at five time intervals corresponding to the time normalized values for low, rising, and falling Thai tones as shown in Figure 2, resulting in three new PitchTiers.
4. The original stimulus item was converted to a Manipulation object in Praat, and then merged with each new PitchTier from step 3, resulting in three distinct natural-like stimuli.

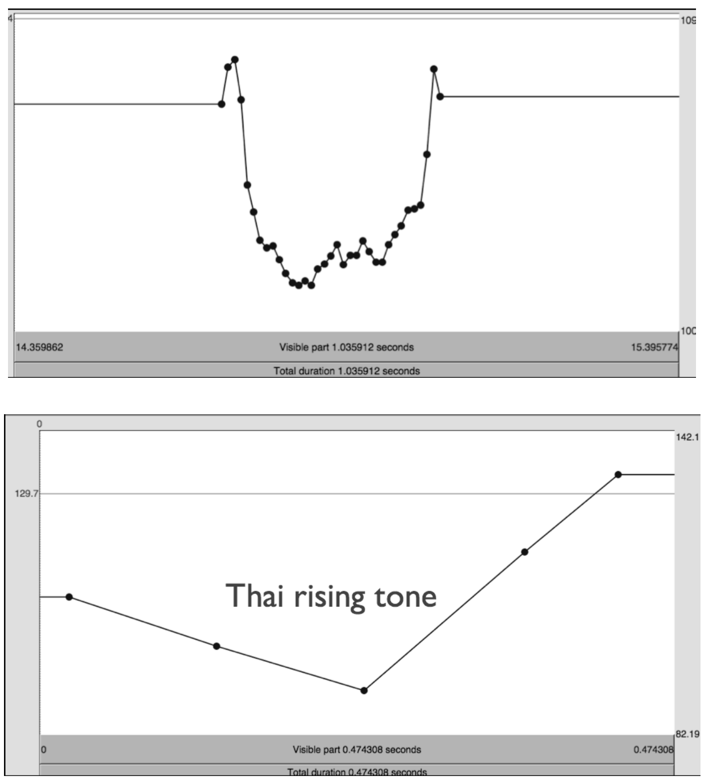


Figure 2: Original stimulus PitchTier and modified PitchTier with rising tonal contour

Since the three resulting items were created using the same input, they were consistent across all acoustic dimensions with the exception of F_0. Any resulting differences in perception of the stimuli are assumed to derive from differences in fundamental frequency alone.

Although some similarities exist between the tonal contours of Thai, Mandarin, and Cantonese, and although [bɑ] is a potentially lexical syllable in both Chinese languages, stimulus items were tested by one male native speaker of Cantonese and one female native speaker of Mandarin, both trained linguists, who certified that, because of the Thai tone contours, the items did not convey lexical meaning in either language.

2.2.2 Phoneme Condition

In order to test participants for online perception of the specific acoustic cue in question (F_0) and not on individual differences in task-specific abilities, it was necessary to create a control condition in which performance would not be expected to improve or deteriorate based on the specific language experience of any of the participants. In this case, a simple phonemic distinction ([m] vs. [f]) was chosen. These bilabial phonemes differ in manner of articulation, and are contrastive in the languages of all participants tested. Hypothetically, then, participants would perform roughly equally in this phoneme Condition. By subtracting their performance on this Condition from their performance in the tone Condition, a "difference score" can be obtained that partially controls for non-linguistic individual differences, such as memory. The usage of this control condition and calculation of this score is based upon a similar procedure in Dupoux, et al. (2010), and is described in greater detail in Chapter 3.

The two phoneme Condition stimuli were also created using Praat (Boersma & Weenink, 2011), and the same recording hardware described above (Section 2.2.1). One instance of the stimulus item [mu] and one instance of the item [fu] were recorded by the same male speaker that recorded the tonal stimuli. The item [mu] was 502ms in duration (compared to 500ms for the tonal stimuli) and was used in experimentation. The second item was formulated by splicing the vocalic portion of [mu] to the fricative generated in the production of [fu]. The resulting stimulus item [fu] was 565ms in duration. Because the items were constructed with the initial phoneme as the minimal distinction, differences in duration are inherent in the phoneme of interest.

2.3 Presentation

The three phases of experimentation were presented to participants in one of three sound-attenuated booths at the Alberta Phonetics Laboratory. Participants were instructed that they would be watching a monitor and hearing items through headphones attached to a computer, and would be using only the number keys and "Enter" key on a standard computer keyboard. Participants were instructed to follow the onscreen instructions, and were given no additional information prior to completing the task.

The entire experimental presentation was completed using ACTUATE (Westbury, 2007). The flexibility of the software allowed for the administration of the sequence recall method by randomizing pre-set sequences of the stimulus items within block conditions. As described above, the sequences ranged in length from n=2 to n=6. Each of these five lengths was represented by five unique sequences, resulting in 25 multi-length sequences per experimental block. The interstimulus interval between syllables in a sequence was 1000ms. Participants completed all three blocks of experimentation, consisting of three phases each, in the same order. Data from participants that experienced technical difficulties during the experiment were excluded from analysis.

2.4 Questionnaire

Each early Chinese-English bilingual participant was administered the Language Experience and Proficiency Questionnaire (LEAP-Q) post-experimentally (Marian, Blumenfeld, & Kaushanskaya, 2007)[2]. This questionnaire surveys participants on a number of linguistic and sociolinguistic criteria, including subjective measures of language proficiency in different modalities, age of acquisition, time of residence in distinct linguistic environments, methods of acquisition, and modes of daily language use. The questionnaire has been demonstrated to be internally robust, with factor analysis revealing distinct clusters of individual questions corresponding to L1 and L2 proficiency across speakers. For example, Marian, et al. (2007) found that individual questions concerning L1 competence clustered together naturally as the first factor in a principal components analysis. Moreover, they demonstrated that the method is a reliable predictor of performance on external, objective language proficiency examinations, even though the questionnaire is based entirely on participant responses.

Because of the accessibility of the questionnaire, its ease of administration, and its previously demonstrated reliability, it is used in this experimentation to collect sociobiographical and linguistic data about the bilingual participants. These data are then used in multiple analyses of the early bilinguals in order to determine whether proficiency in the tone-differentiated sequence recall task is affected by such sociolinguistic measurements. Correlation of behavioural task results with one or more of the factors in the questionnaire may demonstrate the importance of subtle biographical and developmental differences in the acquisition of native-like adult phonology.

2.5 Recruitment and participants[3]

Participants were recruited on a volunteer basis or using an online recruitment system from the Department of Linguistics subject pool at the University of Alberta. The latter group of participants received undergraduate course credit for their participation. Recruitment and experimentation took place in three phases, in the Winter, Spring/Summer, and Fall terms of 2011.

Data from 77 adults (ranging in age from 18 to 32 years, median age = 21) were included in the analysis. An additional 14 adults participated in the experiment, but were excluded from the data analyses because they did not accurately complete the training phase (6), because they made ‘complete reversals’ during the test phase by associating the assigned keyboard key with the opposite stimulus item (5), because they experienced technical difficulties with the presentation software and did not complete the experiment (2), or because they had non-Chinese tone language experience (1). All of the participants reported having normal hearing. No participant’s data was excluded from analysis for differing from the Group difference score mean by two or more standard deviations in either direction.

Of the participants included in analysis, 44 were monolingual speakers of Canadian English and had no L1 or L2 experience with a tone language. 18 additional speakers were born in Taiwan or Mainland China and immigrated to Canada as adults (late bilingual speakers). Each of these speakers considered his or her dominant language to be Mandarin or Cantonese Chinese, and reported sub-native abilities in English. Finally, 15 of the participants were either born in Canada to Chinese-speaking parents or immigrated to Canada from a Chinese-speaking country before age 8. Each of these early bilingual speakers reported lifelong exposure to and native-like abilities in both English and Mandarin or Cantonese.

3 Results and discussion

The results that follow are organized into multiple sections. First, group results are provided as mean accuracy rates in both tone and phoneme conditions. These results are initially analyzed between subject groups, using factorial analyses of variance (ANOVA) and *t*-tests to highlight differences in the mean accuracies between groups. Next, accuracy results are presented as *difference scores*, partially controlling for individual variations in task-specific processing strategies. These scores are, in turn, analyzed using *t*-tests. With these main findings presented using traditional statistical tools, the next portion of the analysis ultimately converges upon similar conclusions, instead using multiple logistic regression models with random effects.

The second section of results focuses upon differences in individuals' treatment of different stimulus types (contour-contour and contour-level minimal pairs) within the tone condition of experimentation. Again, differences across groups are analyzed using *t*-tests.

In the final portion of this chapter, the analysis centres on individual results in the early bilingual group of participants. As noted in the introduction, the performance of participants in this group on a task specific to one (but not the other) of their native languages may be affected by questions of language dominance and developmental factors. Moreover, the participants in this group demonstrate a wider distribution of difference scores (greater variance in responses) (Section 3.1.2), which may be the result of individual biographical variation. The effects of some of these factors upon the behavioural data are explored in this last section.

3.1 Group accuracy results

Main results are presented in Table 2 and Figures 3-4 as accuracy rates in both tone and phoneme Conditions.

Group	*tone accuracy*	*phoneme accuracy*
English	75.02 *(0.96)*	91.02 *(0.87)*
Early bilinguals	68.93 *(1.82)*	84.82 *(1.96)*
Late bilnguals	82.76 *(1.32)*	89.35 *(1.49)*

Table 2: Tone and phoneme accuracies by Group (standard error in italics)

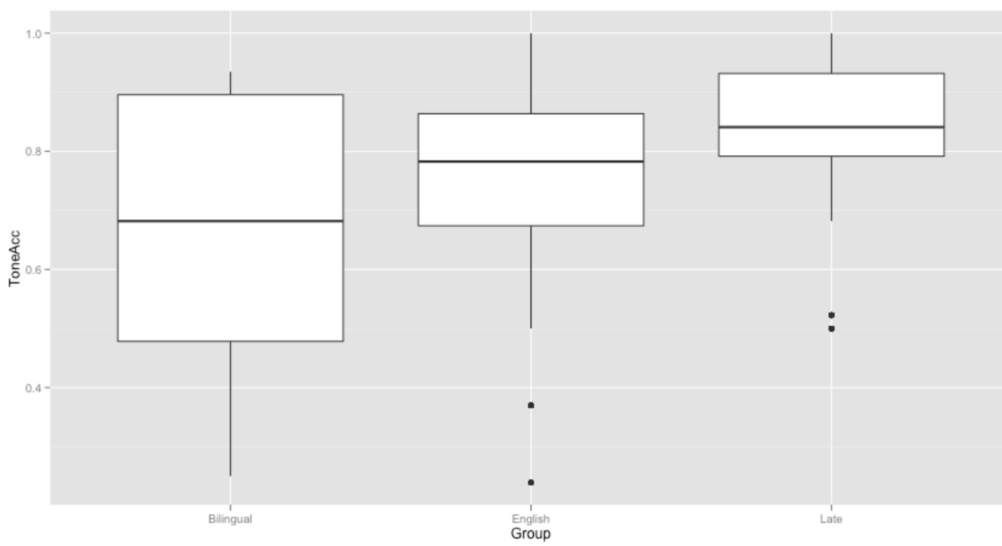


Figure 3: Tone Accuracy by Group

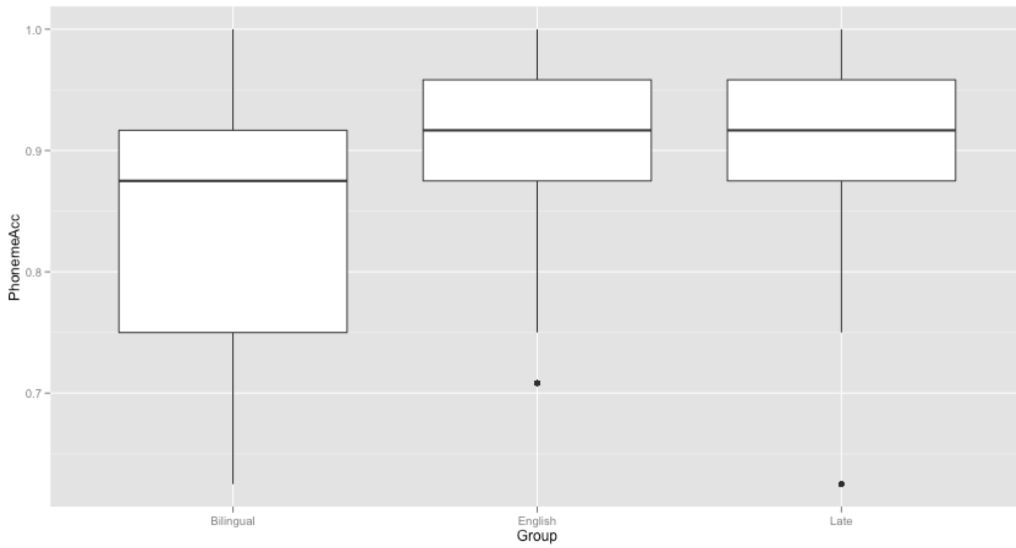


Figure 4: Phoneme Accuracy by Group

3.1.1 Analysis of variance and t*-tests*

A two factor by-subjects ANOVA was run upon the raw accuracy scores, with factors Group (levels English, Late, and Early) and Condition (levels Phoneme and Tone). Significant main effects were observed for Group ($F(2,5325)=20.75$, $p < .001$) and Condition ($F(1,5325)=151.29$, $p < .001$), with a significant interaction between the two ($F(2,5325)=6.29$, $p = .002$). Post-hoc analyses using *t*-tests were run pairwise on the tone accuracy scores to determine the direction of the significant effects in that Condition. As illustrated in Figure 3, late bilinguals performed significantly better (M = 82.76) than English speakers (M = 75.02) in the tone Condition ($t(1720.33) = -4.73$, $p < .001$). Late bilinguals also performed significantly better in the tone Condition than did early bilinguals (M = 68.93) ($t(1234.87) = -6.14$, $p < .001$). Surprisingly, however, English speakers with no tone language experience also performed significantly better than did the early bilinguals in that Condition ($t(1032.81) = -2.96$, $p = .003$).

3.1.2 Difference score results

However, raw accuracies by Condition only give a partial indication of the three Groups' performance. Although a significant difference was found between English and early bilingual speakers' performance in the tone Condition, a parallel significant difference was also found between the two Groups' performance in the *phoneme* Condition ($M_{english}$ = 91.02, $M_{bilingual}$ = 84.82) (t(474.38) = -2.89, p = .004) indicating that an increased error rate by early bilingual speakers may be the result of task-specific, rather than linguistic, effects. That is, an increased error rate in the tone Condition by early bilingual speakers may indicate idiosyncratic difficulty with the sequence recall task by the sample of participants, and may not demonstrate reduced sensitivity to tonal contrasts by that Group. This effect can be partially conceptualized using *difference scores,* also utilized in Dupoux, et al. (2010). Difference scores are calculated by subtracting a participant's accuracy in the tone Condition from his or her accuracy in the phoneme Condition. The result of this simple calculation is a score that minimizes the effects of individual differences in memorization abilities or task-specific strategies, while statistics based on raw Group means by Condition (as above) might be influenced by such effects. Higher difference scores indicate a more distant relationship between a participant's proficiency in the tone Condition and his or her proficiency in the phoneme Condition, and demonstrate a relatively reduced sensitivity to tone changes. Visual examination of the difference scores for each Group (Table 3, Figure 5) supports the finding that, although English speakers perform more accurately in the tone Condition than do early bilinguals, they also perform more accurately in the phoneme Condition, possibly negating this distinction.

Group	*difference score*
English	16.05
Early bilingual	15.81
Late bilingual	6.95

Table 3: Difference scores (Phoneme Accuracy – Tone Accuracy) by Group

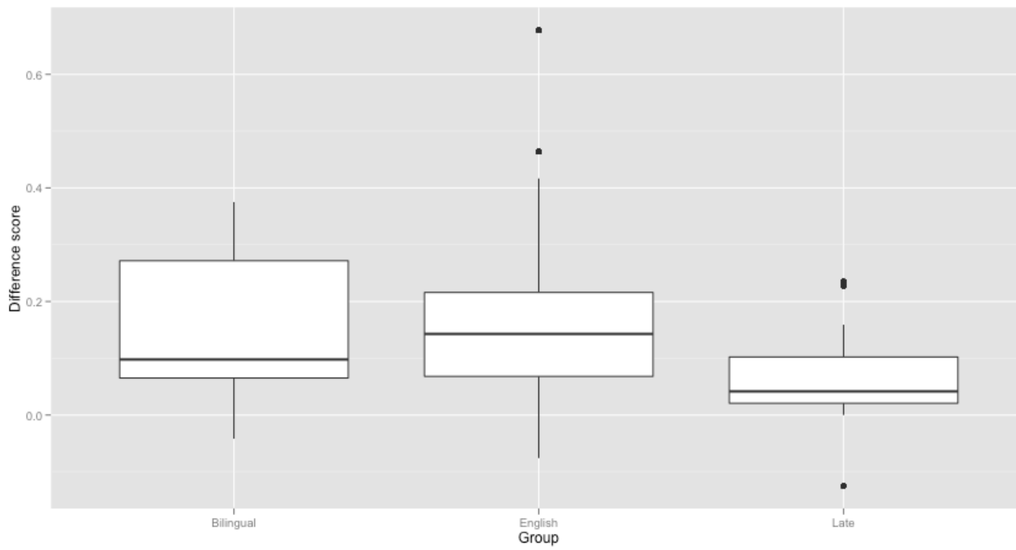


Figure 5: Difference scores (Phoneme Accuracy – Tone Accuracy) by Group

3.1.3 Difference score t*-tests*

T-tests confirm a significant discrepancy between the difference scores of the late bilinguals and those of the early bilinguals ($t(19.69) = -2.11$, $p = .05$), as well as a significant difference between those of late bilinguals and those of English monolinguals ($t(57.147) = 3.07$, $p = .003$). However, unlike the raw scores analysis, these tests reveal no significant difference between English monolinguals and early bilinguals ($t(23.42) = 0.06$, $p = 0.96$). Although English monolinguals performed surprisingly better than early bilinguals in the tone Condition, analyses conducted on the difference scores indicate that this discrepancy is task-specific, rather than linguistic.

3.1.4 *Multiple logistic regression models with mixed effects*

The difference score analysis, though, compromises statistical power by reducing the number of data points for each participant to one mean value, ignoring the underlying distribution of participants' accurate and inaccurate responses. More importantly, perhaps, the statistical tools used in the difference score analysis do not allow for the possibility that task-specific characteristics are more challenging or facilitatory for some participants than for others.

To correct for the reduction in power, while still acknowledging participants' individual task-specific differences, performances in the tone Condition are subsequently modelled using mixed effects logistic regression (Bates & Maechler, 2011). These models, like ANOVA, are used to predict accuracy using fixed effect factors, such as Group. Additionally, though, these tools allow the user to specify random effect factors, such as Subject. Because individual participants may perform differently on a given task (in this case, sequence recall), this tool does not assume that all individuals' baseline task accuracies are equal. *Random intercepts* by Subject account for this variance. Moreover, mixed effects models allow the user to include *random slopes* by Subject for individual variables, thus recognizing that some independent variables may have greater effects on some participants' performance than on others'.

Therefore, in the analyses that follow, we indicate in the random effects structure of the linear mixed effects model that tone Accuracy Score intercepts may vary randomly by Subject (Baayen, 2008).[4] Furthermore, we add two numeric predictors into the model to further isolate the effects of our main factor of interest (Group). These numeric predictors are Phoneme Accuracy (an individual's overall performance in the phoneme Condition, replacing the need for a difference score) and Sequence Length (the numeric length of an individual sequence). As might be expected, the length of a sequence is responsible for a great deal of variance in the accuracy scores of participants (see Figure 6). Moreover, we can allow the slopes of the partial effects for Sequence Length to vary by Subject, recognizing the possibility that this numeric factor is more important for some participants than for others.

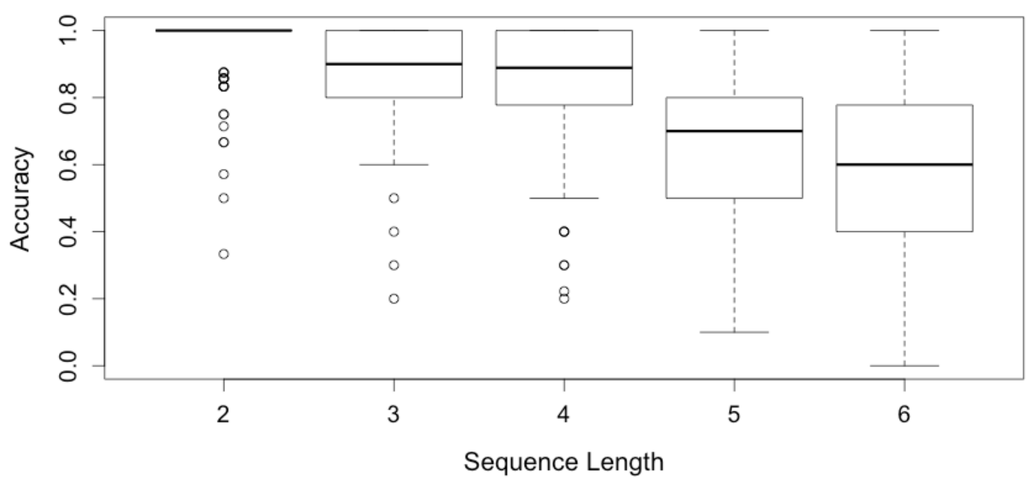


Figure 6: Accuracy scores by Sequence Length across Subjects and Conditions

The addition of each predictor to the model was justified by performing stepwise likelihood ratio comparisons of increasingly complex models (additive model comparison). That is, the inclusion of each new factor produced a significantly better fitting model than a simpler model excluding the new factor. Inclusion of the factor Language, which would have accounted for a speaker's knowledge of Mandarin versus Cantonese, did not improve the model and did not significantly predict a participant's response accuracy. Therefore, in all analyses that follow, participants are analyzed across Language. Additionally, a mixed model with interactions between factors was also tested, but an ANOVA revealed that the models with only main effects fit the data more closely (AIC = 3099.0 vs $AIC_{interactions}$ = 3103.7). Finally, the addition of random slopes by Subject for Group and Phoneme Accuracy did not improve the model's fit. The final mixed effects model, then, predicts Accuracy in the tone Condition with the factor Group and numeric predictors Phoneme Accuracy and Sequence Length. Random intercepts are specified for Subject, and random slopes for Sequence Length by Subject (Table 4).

3.1.5 Multiple logistic regression results

Ref. Level: English	Regression coefficient	Standard error	z-value	Pr(>|z|)
(Intercept)	-2.44	1.16	-2.11	0.035 *
SeqLength	-0.77	0.04	-17.22	<0.001 *
PhonemeAcc	7.92	1.24	6.37	<0.001 *
GroupEarly	**0.13**	**0.27**	**0.48**	**0.630 n.s.**
GroupLate	**0.65**	**0.25**	**2.63**	**0.009 ***

Ref. Level: Early	Regression coefficient	Standard error	z-value	Pr(>|z|)
(Intercept)	-2.93	1.12	-2.61	0.009 *
SeqLength	-0.71	0.04	-18.55	<0.001 *
PhonemeAcc	8.26	1.30	6.35	<0.001 *
GroupEnglish	**-0.19**	**0.29**	**-0.64**	**0.523 n.s.**
GroupLate	**0.57**	**0.33**	**1.71**	**0.088 n.s.**

Table 4: Mixed effects model 1

Two final binomial mixed models were examined by alternating the Group reference level. That is, the two models were constructed such that comparisons could be made between each Group of interest with the other two Groups. All of the models revealed significant effects for Sequence Length (Markov-chain

Monte Carlo estimated $p < .001$) and Phoneme Accuracy (MCMCp $< .001$). As expected, Sequence Length was negatively correlated with Tone Accuracy, as it is across Conditions and noted in Figure 6. That is, in general, longer sequences were more difficult for participants to recall than shorter sequences. Figure 7 plots the partial effect of Sequence Length on Tone Accuracy (Baayen, 2011). Phoneme Accuracy was positively correlated with Tone Accuracy, indicating that higher individual performance in the Phoneme Condition resulted in higher performance in the Tone Condition. This finding indicates, once again, that some of the individual variation in performance is a result of task-specific processing strategies, not linguistic effects. Figure 8 plots the partial effect of Phoneme Accuracy on Tone Accuracy in the model.

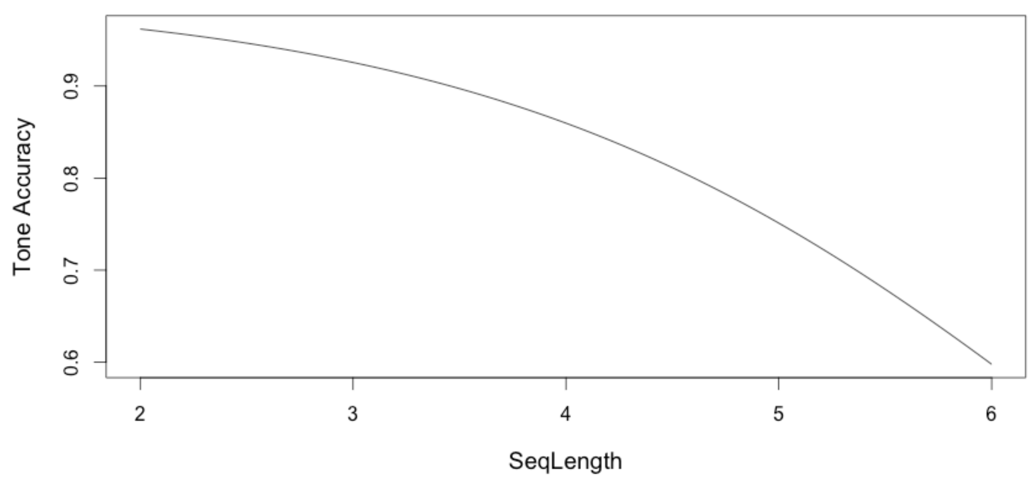


Figure 7: Partial effect of Sequence Length on estimated Tone Accuracy in mixed model

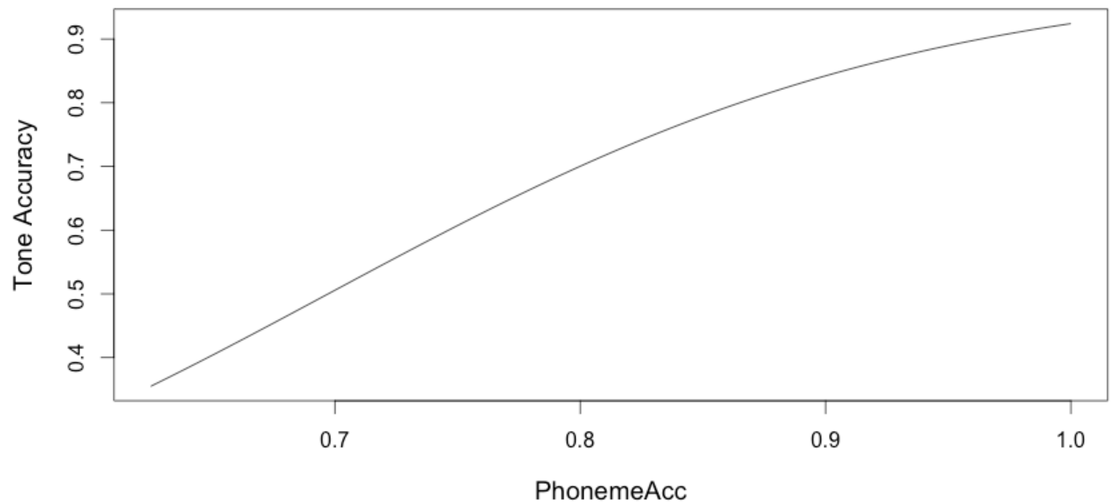


Figure 8: Partial effect of Phoneme Accuracy on estimated Tone Accuracy in mixed model

With regard to Group effects, in the first model, English monolingual speakers were contrasted with early and late bilingual speakers. As indicated in the initial raw score ANOVAs and difference score analyses above, the mixed model confirmed that individual late bilingual speakers performed significantly better than did English speakers in the tone Condition (MCMCp = .009). However, no significant difference is observed between English monolingual speakers and early bilinguals in this Condition (MCMCp = 0.630), also confirming the results of the difference score analyses. In the second model, early bilingual speakers were contrasted with the two other Groups. Contrary to the results obtained in the difference score analysis, early bilinguals in this group did not differ significantly from late bilinguals (MCMCp = 0.088). In other words, as a group, early bilinguals exhibited a pattern intermediate to the English monolingual and Chinese-dominant speakers and did not exhibit statistically divergent accuracies from either Group, partially opposing the *t*-test analyses above.

3.1.6 Data exclusion and new models

Although the mixed models above fail to converge on the same conclusion as the *t*-tests in isolating a significant difference between the two groups of bilingual speakers, the *p*-value obtained is relatively low (0.088). Furthermore, recall that the early bilingual test Group consisted not only of English-Chinese bilinguals that were born in Canada, but also of speakers who immigrated to Canada before eight years of age. These bilinguals, who were not regularly exposed to English before moving, learned Chinese before English and might be expected to perform differently than the earliest bilinguals regularly exposed to both languages throughout early development. Previous studies have suggested that early *sequential* bilinguals, even those who appear to be native speakers of both languages, in fact perform significantly differently than early *simultaneous* bilingual speakers of the same languages (Sebastián-Gallés, et al., 2005, *inter alia*.). Including these sequential bilinguals in the analysis may have biased the results such that the behaviour of true simultaneous bilinguals was masked. In order to investigate this possibility, data from participants that moved to Canada as children were excluded. Because there were only four such sequential bilinguals, separate analyses for this Group would have been statistically weak and were not conducted. Rather, new models were fit to the data comparing early simultaneous bilinguals (age of arrival = 0) to English monolinguals and late bilinguals (age of arrival > 17). These models followed the same format as the above models that included all speakers (response variable Tone Accuracy; numeric predictors Sequence Length, Phoneme Accuracy; fixed-effect factor Group; random-effect factor Subject, random slopes for Sequence Length by Subject).

Ref. Level: Early	Regression coefficient	Standard error	z-value	Pr(>|z|)
(Intercept)	-2.19	1.09	-2.00	0.046 *
SeqLength	-0.76	0.05	-16.78	<0.001 *
PhonemeAcc	7.30	1.26	5.79	<0.001 *
GroupEnglish	**0.24**	**0.31**	**0.76**	**0.448 n.s.**
GroupLate	**0.88**	**0.35**	**2.54**	**0.011 ***

Table 5: Mixed effects model 2

With the removal of the middle group of bilinguals, the distinction in tone accuracies between early bilinguals and late bilinguals reached significance (MCMCp = 0.011), now echoing the difference score analyses above. That is, even while accounting for individual task-specific differences, late bilinguals performed significantly more accurately in the tone Condition than did early bilingual speakers. The difference between these early bilinguals and English monolinguals, however, is still negligible (MCMCp = 0.448). Figure 9 plots the partial effect of Group upon Tone Accuracy with sequential bilinguals removed from analysis.

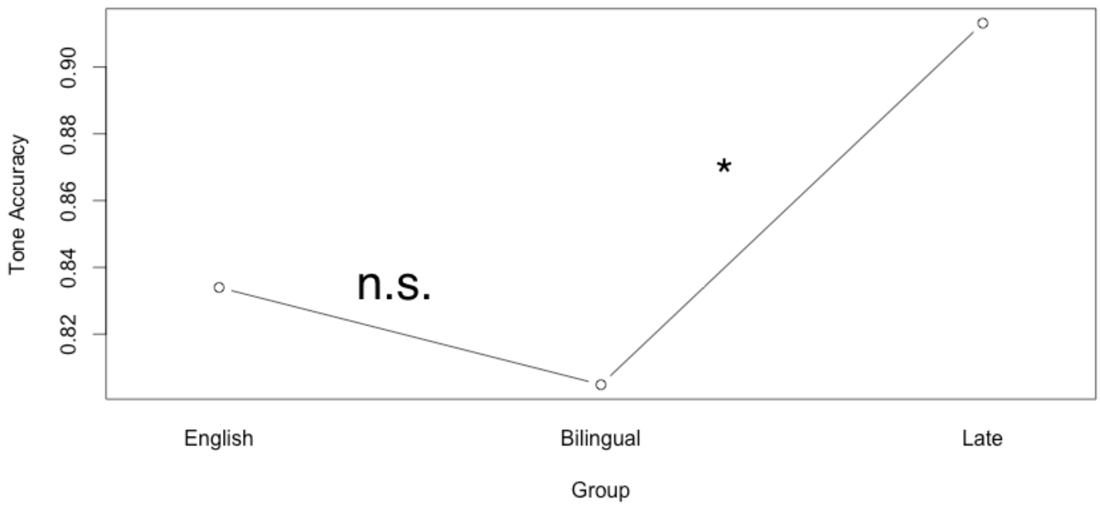


Figure 9: Partial effect of Group on estimated Tone Accuracy in final mixed model

3.1.7 General discussion of group results

The final mixed effects models confirm the earlier ANOVAs and *t*-tests in suggesting that late bilinguals indeed process minimal F_0 distinctions differently than English monolinguals do. Although all Groups discriminate tone differences in this task at levels significantly better than chance, late bilingual speakers perform more consistently than English speakers do, indicating a processing or attentional advantage for this specific acoustic dimension. This isolated finding validates the sequence recall methodology for tone perception, as these results for English and Chinese speakers parallel earlier findings by Burnham and colleagues (1992; 1996) and those regarding the stress perception of Spanish and French speakers in Dupoux, et al. (2001; 2008).

Interestingly, late bilinguals also perform significantly more accurately than do early bilingual speakers with lifelong exposure to both English and Chinese. This finding is confirmed both with *t*-tests conducted on individual difference scores, as well as with mixed effects models. Although these early bilinguals use tonal distinctions to navigate linguistic subtleties in one of their native languages, these results suggest that they do so differently than speakers exposed primarily to Chinese. Importantly, difference score *t*-tests and mixed models converge to indicate that early bilinguals do not perform significantly better than do English monolinguals with no tone language exposure. These findings support claims that even simultaneous bilinguals have underlying language dominance based on a number of early developmental factors (Cutler, et al., 1989, 1992; Sebastián-Gallés, et al., 2005). Section 3.3 explores a few of these developmental factors.

3.2 Effects of tone type

Although early bilinguals do not perform more accurately in the tone condition than do English monolinguals, it is possible that the two groups are still processing the tonal differences in distinct ways. Mattock & Burnham (2006) demonstrated that infant learners of English, who generally categorize tones less accurately than Chinese learners, are better at distinguishing between two contour tones (e.g., rising and falling) than between a level tone and a contour tone. Chinese-learning infants do not demonstrate this bias. The stimulus items in the present experiment are also paired in contour-contour and level-contour blocks, allowing for analysis of participants' perception at this level.

3.2.1 Tone type results

Accuracies by Group for both Tone Types are visualized in Figure 10. Pairwise *t*-tests were run on the tone accuracy scores in order to determine the significance of Tone Type for each Group. Tone Type does not emerge as a significant predictor of performance for late bilinguals ($M_{contour\text{-}contour} = 82.05$, $M_{level\text{-}contour} = 83.41$; $t(803.58) = -0.51$, $p = 0.61$) or for early bilinguals ($M_{contour\text{-}contour} = 61.64$, $M_{level\text{-}contour} = 59.74$; $t(460.92) = 0.4172$, $p = 0.68$), but it does emerge as significant for English monolinguals ($M_{contour\text{-}contour} = 72.24$, $M_{level\text{-}contour} = 77.56$; $t(1963.41) = -2.75$, $p = 0.006$).

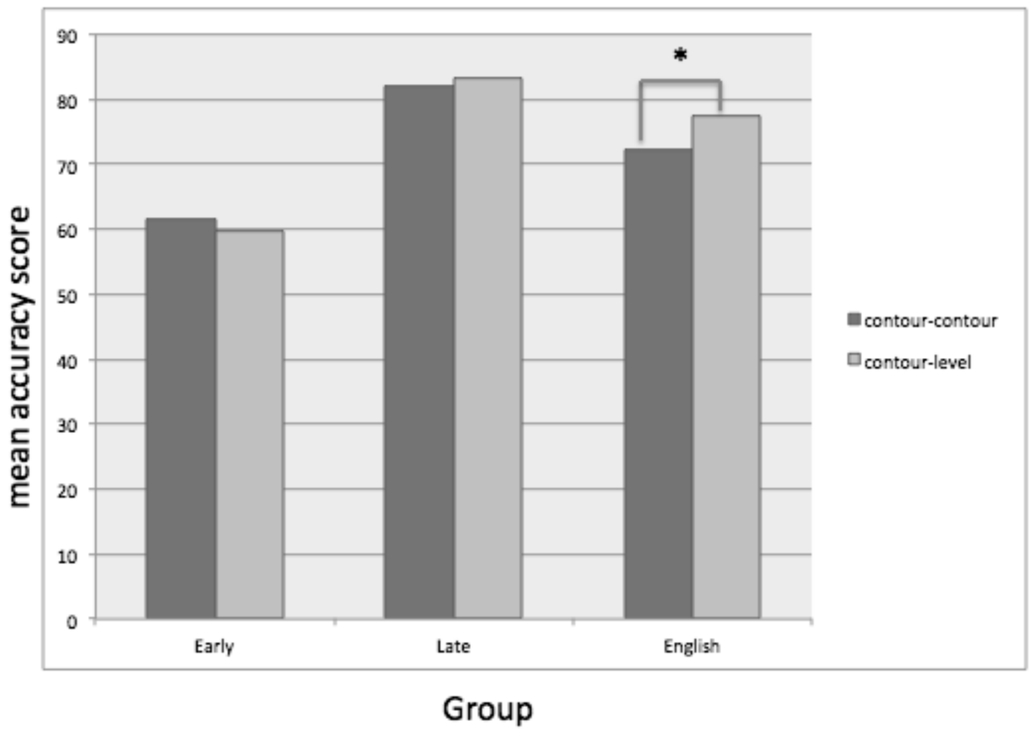


Figure 10: Accuracies for two Tone Types by Group

3.2.2 Discussion of tone type results

In this case, English monolinguals more accurately recall sequences with a level-contour distinction than with a contour-contour distinction. Early and late bilinguals, on the other hand, do not favour either tonal pair. This result is similar to those found in studies investigating hemispheric processing of tone, in which Chinese-English bilinguals do not differ from Chinese monolinguals (Sereno &

Wang, 2006). On the other hand, the current finding should be the source of greater investigation, as it is distinct from the result obtained in Mattock & Burnham with infant learners (2006). Although English monolingual infants in that study did treat contour-contour and level-contour pairs differently while Chinese infants did not, they were more accurate in the contour-contour condition than in the level-contour condition. In the present adult-centred study, the bias was observed in the English speakers' data, but in the opposite direction. Regardless of the direction of English speakers' bias, the results of the present study indicate that, although English speakers and early bilinguals do not perform significantly differently in their overall recall of tone differentiated stimuli, early bilinguals behave more like late bilinguals in their equal treatment of tone types.

3.3 Early bilingual individual results

Group analyses have demonstrated that the early bilingual speakers in this experiment did not perform significantly better in the tone Condition of the sequence recall task than did their English monolingual counterparts. Despite lifelong exposure to and use of a Chinese language, early bilingual speakers may exhibit a reduced attentional behaviour to tone differences. Nevertheless, further investigation of the bilinguals' interaction with specific tone types reveals that these speakers are not processing tonal differences in precisely the same manner as the English monolinguals. Furthermore, it is evident from examination of raw by-Group results that the tone accuracies (Figure 3) and difference scores (Figure 5) of the early bilinguals have a wider distribution than those of the English and late bilingual speakers. Because of the somewhat surprising by-Group analysis and this wide distribution, further investigation of individual bilingual performance is warranted.

In the following analyses, measurements from the LEAP-Q post-experimental survey are analyzed alongside accuracy scores from the behavioural task. The responses chosen from the LEAP-Q consist of 22 variables, many of which (such as the frequencies with which a participant chooses to speak Chinese and with which she chooses to speak English) have absolute correlation values higher than 90%. Because this collinearity renders traditional regression analysis unsatisfactory and the large number of variables makes such a process cumbersome, a few exploratory statistical tools are utilized here. The primary tool utilized is principal components analysis (PCA) (Pearson, 1901). From a number of interrelated and collinear factors (like the LEAP-Q measurements collected here), this tool constructs orthogonal vectors called principal components. These components, consisting of clusters of individual variables, are then used in regression models as predictors of performance (in this case, accuracy in the tone Condition of the sequence recall experiment). As in Marian, et al. (2007), principal components derived from these data serve to highlight specific areas of language proficiency and experience that, in this case, may affect F_0 perception. The PCA for these data is conducted in R (R Development Core Team, 2012) according to the procedure specified by Baayen (2008), with subsequent linear models constructed using the resulting components as predictors.

3.3.1 Principal components analysis of LEAP-Q data

The present PCA is conducted for early bilingual participants (age of arrival = 0), using the numeric predictors obtained from the LEAP-Q survey. The complete set of these predictors is listed in Appendix 2.

Six principal components emerged as accounting for 5% or more of the data variation between subjects' LEAP-Q responses. Figure 11 visualizes all of the principal components derived from this initial procedure, with the significant components highlighted in black. For illustration purposes, Table 6 lists the two most important principal components, along with the proportional loadings of each LEAP-Q question. These loadings indicate the positive or negative correlation of each original LEAP-Q item to the principal components, generalizing the importance of each survey question to each component (Baayen, 2008). Proportional loadings for all six significant principal components are given in Appendix 3.

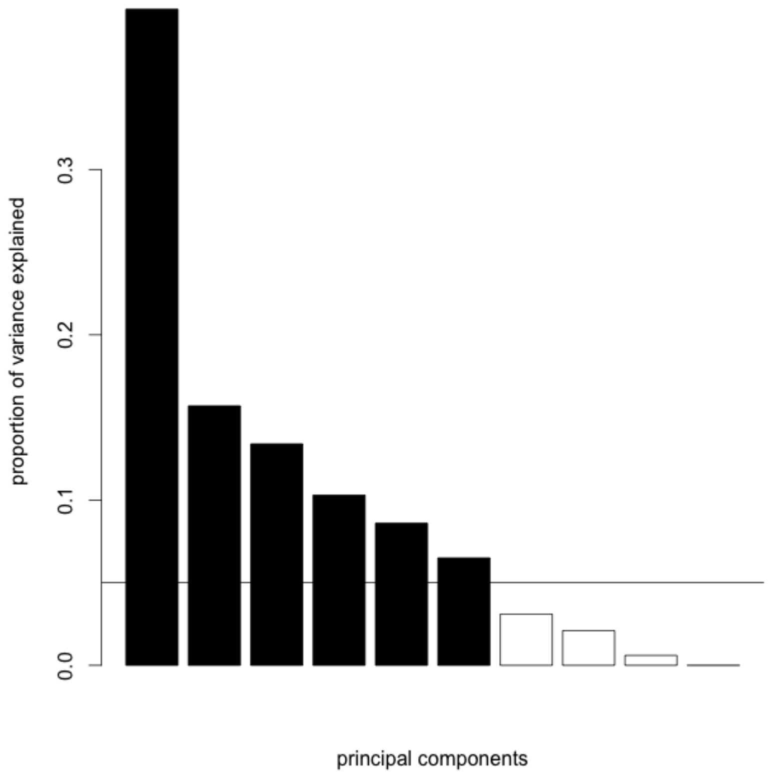


Figure 11: Proportions of variance explained by principal components 1-10

PC1 (39.7%)		PC2 (15.7%)	
ChooseReadChinese	0.29	UnderstandEnglish	0.5
ChooseSpeakChinese	0.28	ReadEnglish	0.48
BeginEnglish	0.27	SpeakEnglish	0.37
SpeakChinese	0.26	EnglishSchool	0.2
UnderstandChinese	0.26	ChooseReadChinese	0.16
BeginReadEnglish	0.25	SpeakChinese	0.14
ChineseExpose	0.24	UnderstandChinese	0.13
ReadChinese	0.22	ChooseSpeakChinese	0.12
ChineseHome	0.19	ReadChinese	0.05
ForeignEnglish	0.17	BeginChinese	0.04
BeginReadChinese	0.05	ChooseSpeakEnglish	0.03
UnderstandEnglish	-0.01	EnglishExpose	0.02
EnglishHome	-0.03	BeginReadChinese	0.01
EnglishSchool	-0.04	ChineseExpose	-0.04
ReadEnglish	-0.11	ChineseHome	-0.06
ChineseSchool	-0.13	BeginEnglish	-0.08
SpeakEnglish	-0.16	EnglishHome	-0.1
BeginChinese	-0.2	ForeignChinese	-0.17

ChooseReadEnglish	-0.26	ChineseSchool	-0.19
ChooseSpeakEnglish	-0.27	ForeignEnglish	-0.21
EnglishExpose	-0.28	ChooseReadEnglish	-0.23
ForeignChinese	-0.29	BeginReadEnglish	-0.28

Table 6: Principal components 1 and 2 with proportion of variance explained and LEAP-Q item loadings

Principal component 1 (PC1), which explains nearly 40% of the variance between early bilingual participants' LEAP-Q responses, is most heavily influenced by factors concerning the participant's Chinese language experience and proficiency. These factors include language choice, age of acquisition, and current self-rated proficiency in Chinese. Loadings within this principal component are relatively evenly distributed: 16 proficiency measures have absolute loadings greater than 0.15, and none have loadings at 0.3 or higher. PC2, which accounts for roughly 15% of the total LEAP-Q variance, is dominated by factors concerning a participant's English use. This principal component is more narrowly distributed, as only 10 loadings have absolute values greater than 0.15. Additionally, some loadings in this component have much more significant values. Self-rated proficiencies in English understanding, reading, and speaking hold the highest loadings in PC2 at 0.5, 0.48, and 0.37, respectively. The age at which an individual began to read English as a child correlates negatively with PC2 (-0.28), consistently indicating that a later age of literacy has a negative effect upon the positively correlated self-rated measurements. The clusters obtained here by principal components analysis are immediately satisfying, as they appear to naturally group variables together by language. This internal validation of the LEAP-Q method is a partial replication of the findings produced in the original study introducing the questionnaire (Marian, et al., 2007). Furthermore, the greater variability in PC1 (a measure of Chinese dominance) is unsurprising, as early bilinguals in this sample have lived exclusively in an English-speaking country. It seems reasonable to expect, therefore, that greater variance would exist in aptitude measurements of the minority language.

Additionally, PCA outputs coordinates for each Subject ID, representing the relationship of each participant to each principal component. These coordinates are useful for determining which PCs are accurate clusters of individuals' responses. Figure 12 plots individual sociolinguistic variable loadings for PC1 against corresponding loadings for PC2, and adds the PCA coordinates for individual Subject IDs to visualize the individual participants' relationships to component variables. Variables that are very similar in their relative correlations to PC1 and PC2 show overlapping patterns in the plot. A few further generalizations can be drawn from examination of this figure. As noted, language choice and proficiency in Chinese dominate the positive effects in PC1, which in turn appears to be particularly important for Subjects "bi4", "11339", and "11684". Manual examination of these participants' LEAP-Q data, which contain relatively high Chinese language competency ratings, confirms their close relationship to PC1. On the other hand, no Subject ID is found to be particularly attracted to PC2, roughly a measure of English competence. This lack of polarity with regard to PC2 is likely due to ceiling effects: examination of the LEAP-Q data for early bilinguals reveals very high self ratings in English proficiency, again likely due to the linguistic context from which members of this Group are sampled.

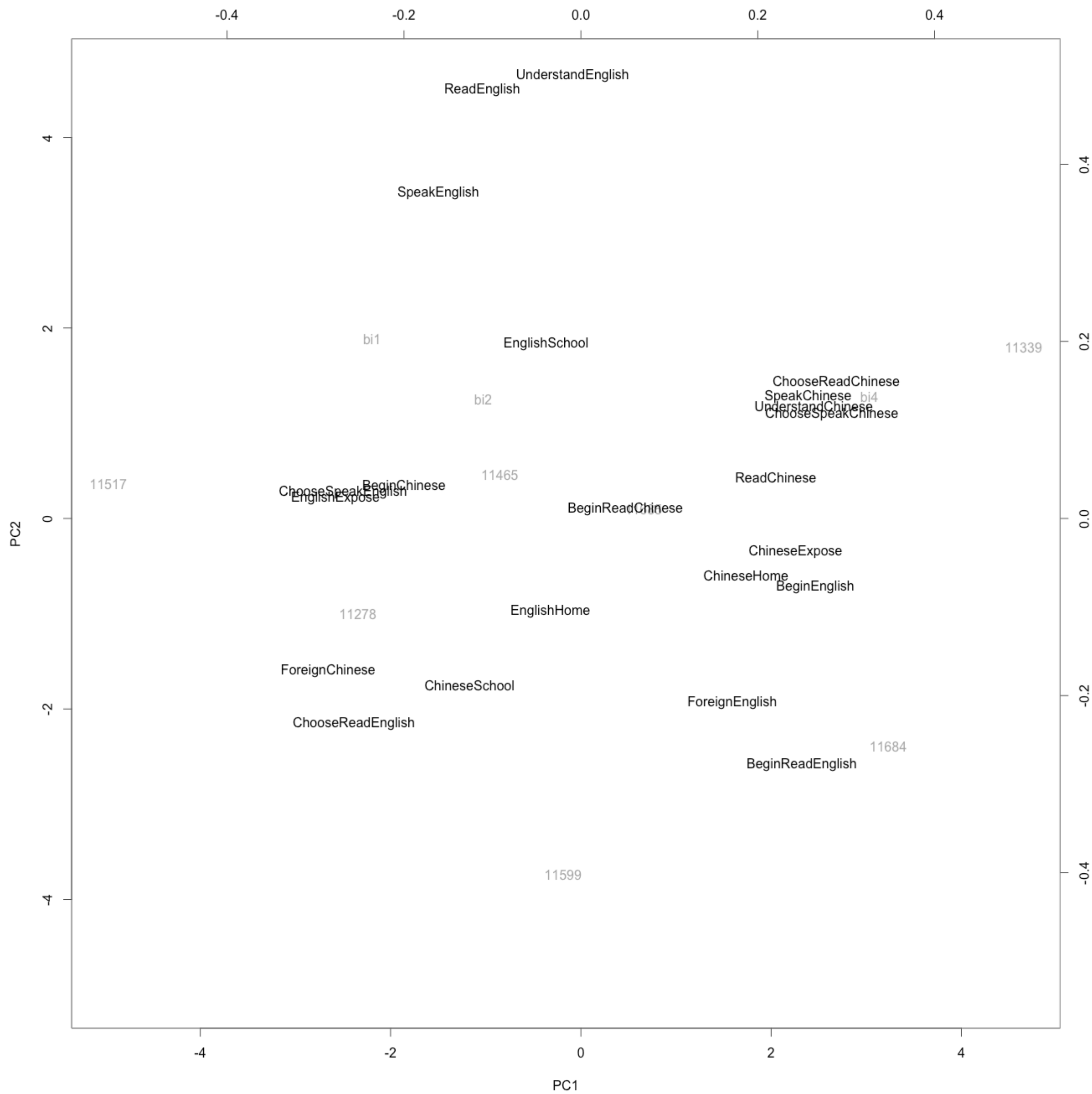


Figure 12: LEAP-Q variable loadings (black) and Subject ID coordinates (gray) against Principal Components 1 and 2

The observations made by examining the variable loadings and Subject-specific coordinate plots indicate that inter-Subject variability within the early bilingual Group is primarily on the PC1 axis, and consists mostly of proficiency levels in Chinese. However, this observation does not account for the importance of either principal component upon the participants' performance in the behavioural experimentation of interest here. In the by-Groups analysis above, late bilingual speakers performed significantly better on the task than did English monolinguals and early bilinguals on the whole. However, as noted, early bilinguals' difference scores had a wider distribution than those of the English speakers and late bilinguals. That is, some early bilinguals performed more like late bilingual speakers than others. Furthermore, as a group, early bilinguals seemed to process tone type distinctions similarly to Chinese-dominant speakers, insofar as they were not more sensitive to one tonal distinction than to the other. It is therefore reasonable to hypothesize that a early bilingual's self-perceived proficiency in Chinese (as approximated by PC1) would have a significant effect upon his or her accuracy in the tone Condition of the sequence recall task.

3.3.2 Principal components regression

To test the effect of the two principal components upon tone perception accuracy, a principal components logistic regression analysis was performed. The Subject-specific coordinates of the six significant PCs were converted into numeric vectors in R and added as variables to be associated with each participant. These vectors, along with numeric variables Sequence Length and Phoneme Accuracy and the random effect variable Subject, were used as predictors in a mixed model. This model fit the data significantly better (AIC = 491.49) than did a model without the six PCs but still including Sequence Length, Phoneme Accuracy, and random effects for Subjects (AIC = 497.01) ($p = 0.007$). It also fit the data better than did a model only including the random effect of Subjects (AIC = 578.75) ($p < 0.001$). That is, the inclusion of the six new principal components calculated a mixed-effects regression model more accurately fitted to the data than models without these components, confirming the importance of the sociolinguistic information gained from the LEAP-Q survey. Moreover, the mixed model analysis revealed that none of the effects of the principal components were correlated at a level higher than 20%, a significant improvement over the highly correlated data initially gleaned from the questionnaire. However, subsequent removal of the four PCs with the least contribution to overall variance did not significantly affect the fit of the model. ($AIC_{6PC} = 489.72$ vs. $AIC_{2PC} = 491.49$, $p = 0.18$), and these four components were removed. Likewise, models including random slopes by Subject for Sequence Length, Phoneme Accuracy, PC1, and PC2, as well as models with factor interactions were calculated. None of these models more closely fit the data, and were rejected. The final model, therefore, predicted Tone Accuracy by early bilinguals using Sequence Length, Phoneme Accuracy, PC1, and PC2, with random intercepts for individual Subjects.

	Regression coefficient	Standard error	z-value	Pr(>|z|)
(Intercept)	-5.33	1.12	-4.78	<0.001 *
PC1 (Chinese)	**-0.12**	**0.04**	**-3.11**	**0.002 ***
PC2 (English)	**-0.17**	**0.06**	**-2.85**	**0.004 ***
PhonemeAcc	10.74	1.44	7.46	<0.001 *
SeqLength	-0.70	0.09	-7.68	<0.001 *

Table 7: Mixed effects model with principal component predictors

This final model again indicated significant main effects of Sequence Length (MCMCp < .001) and Phoneme Accuracy (MCMCp < .001) on Tone Accuracy, as observed in the prior by-Groups analysis. As in previous analysis, Sequence Length had an inhibitory effect upon participants' tone processing, while their accuracy in the phoneme Condition had a positive correlation to their tone Condition performance. Plots of the partial effects for these factors are similar to those in the by-Groups analysis (Figures 7 and 8). As might be expected, the model also indicated a negative main effect of PC2 (roughly interpreted as English proficiency) upon accuracy in the tone Condition (MCMCp = 0.004). However, the model also produced a surprising negative effect of PC1 (interpreted as Chinese proficiency) in the tone Condition (MCMCp = 0.002). Figure 13 shows the partial effects of PC1 upon Tone Accuracy in this mixed model. Values on the *x*-axis indicate participants' component-specific coordinates. The higher a participant's coordinates for the component, the more significant that component is within his or her LEAP-Q responses. In this case, higher coordinates in PC1 indicate a participant's higher self-assessment in and exposure to his or her Chinese language.

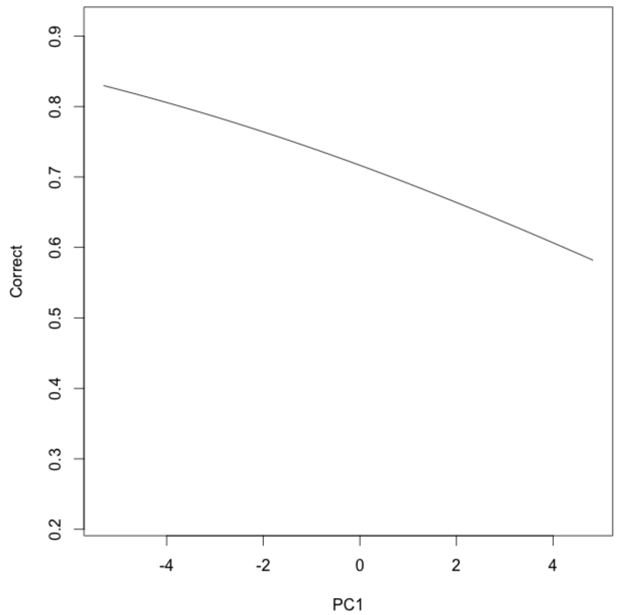


Figure 13: Partial effect of PC1 upon the estimated Tone Accuracy of early bilinguals

As can be visualized here, high PC1 coordinates are negatively correlated to performance in the tone Condition.

3.3.3 Discussion of principal components regression

The findings of the principal components regression analysis were unexpected, as they suggest that high self-assessment in Chinese does not necessarily contribute to increased tonal perception accuracy. While the analysis suggests an expected negative effect of English dominance on tone Condition scores, competing negative effects of Chinese dominance seem to negate this finding. In the case of these particular participants, high LEAP-Q values in either language seem to negatively affect F_0 contour identification. Possible explanations for this contradictory finding are explored in Chapter 4.

4 General discussion and conclusion

This study tested three groups of adult speakers on their general perception of lexical tone, using a sequence recall method adapted from Dupoux, et al. (2010). Of the three groups, two had significant lifelong experience in a tone language: late Chinese-English bilinguals and early Chinese-English bilinguals. Based on previous findings with monolinguals (Burnham, et al., 1992, 1996), the late bilingual group's experience with lexical tone in their own language was expected to positively affect their performance in this experimentation using non-native tones. The early bilingual participants, on the other hand, were expected to produce accuracy scores intermediate to those of the late bilinguals and the third group of participants, English speakers, with no tone language experience. Additionally, the early bilingual participants were administered the Language Experience and Proficiency Questionnaire (Marian, et al., 2007), the results of which were expected to partially predict their accuracy scores in the tonal condition.

Overall, English speakers with no tone language experience exhibited reduced accuracy in this task, which tested rapid, online perception of natural-like stimuli differentiated only by F_0 contour. Late Chinese-English bilinguals, whose L1 and dominant language is tonal, exhibited greater accuracy in this task. These between-groups results emerged significantly using more traditional statistical tools (ANOVA, *t*-tests), without accounting for individual task-specific differences using mixed effects models. When these differences were recognized as random by-subjects effects using mixed models, the findings were even clearer.

These results replicate the findings of previous studies testing tone perception by monolingual speakers. Infant-centred research has suggested that learners of non-tone languages undergo perceptual reorganization and become less sensitive to F_0 contour distinctions by nine months of age (Mattock & Burnham, 2006; Mattock, et al., 2008). By adulthood, these differences are robust (Burnham, et al., 1992, 1996). This experimentation also produces this result, using a modified methodology that simulates the cognitive burdens of natural speech processing.

Additionally, this experimentation is the first to probe early simultaneous bilinguals of tone and non-tone languages on their online perception of lexical tone. Previous research with lexical stress recall might predict that the early bilingual speakers in this study would present tonal accuracy scores intermediate to their Chinese-dominant and English monolingual peers (Dupoux, et al., 2010). This pattern was not observed in the ANOVA and *t*-test portion of the analysis, nor was it demonstrated using mixed effects models. When only true early simultaneous bilinguals (age of arrival = 0) are considered, a significant difference emerges between them and their Chinese-dominant counterparts. In fact, these bilinguals, who have had continuous lifelong exposure to a tone language, do not produce more accurate scores than English speakers with no tone language experience. This finding supports the notion that bilinguals only truly process one language natively (Cutler, et al., 1989, 1992), and indicates that these bilinguals, as a group, natively process English and not Chinese.

Despite this finding, the early bilinguals' experience with Chinese is not negligible even in this particular task. Although their accuracy scores align with those of the English monolinguals, they exhibit a wider distribution of difference scores. Moreover, their treatment of tonal contour pairs is more similar to the late bilingual speakers. That is, while English speakers appear to find the perception of level-contour pairs easier than contour-contour pairs, late bilingual speakers do not exhibit such a preference. This finding is somewhat similar to that obtained by Mattock & Burnham (2006), who suggested that Chinese-learning infants did not find either condition significantly easier. In this present task, early bilinguals also did not demonstrate a type preference. This point of similarity with the late bilingual speakers perhaps indicates that, despite lower overall accuracy scores, early bilinguals still utilize their tone language experience when completing this task. This result is, in part, a behavioural replication of neurological studies indicating that Chinese-English bilinguals do not perform differently from Chinese monolinguals in hemispheric processing of tone (Sereno & Wang, 2006).

These findings raise a number of new questions about language dominance. First, if early bilingual speakers in this population are consistently exposed to and are required to use Chinese, and if their task-specific behaviour in the tone Condition is similar to Chinese-dominant participants, why are their tonal accuracy scores lower? The wider distribution of scores in the early bilingual group may point to individual biographical differences causing perceptual variation on this task. Explanation of these effects was attempted using the LEAP-Q method (Marian, et al., 2007). The questionnaire was validated using principal components analysis, as it was by Marian and colleagues when introducing the method. That is, questions addressing experience and proficiency in an individual language appeared to cluster together for all participants, indicating, at least, that the survey produces consistent measurements. On the other hand, these clusters were not particularly satisfying predictors of accuracy in the task of interest here. Using principal components regression, we saw that high LEAP-Q correlates for English proficiency and exposure were negatively associated with high tone accuracy scores, a result that is consistent with the hypothesis that dominance in English would adversely affect tone perception. However, a negative effect is *also* observed for Chinese proficiency and exposure, which is surprising. The inconsistency of these results may support claims by some that subjective measurements of language proficiency are irrelevant to determining bilingual dominance (Dupoux, et al., 2010; Flege, MacKay, & Piske, 2002). On the other hand, these results may simply point to a methodological flaw in using the LEAP-Q for determination of dominance in simultaneous bilinguals. Although the survey successfully probes questions of dominance for adult L2 learners (Marian, et al., 2007), it may be inappropriate for analyzing speakers with two first languages.

Although their behavioural scores are significantly different, speakers in both Chinese-speaking groups do not differ on their age of (Chinese) acquisition, the native language of both parents, or the predominant language in their homes. Instead, the greatest difference between the groups appears to be the language of exposure outside the home. Recent studies have demonstrated that home-external factors, which contribute to the “richness” of the input, are more influential in the morphosyntactic acquisition of early bilingual children than are home-internal factors (see Paradis, 2011 for a review). At that linguistic level, then, early bilinguals are particularly affected by the status of their two first languages in their communities. In the case of participants residing in Alberta, Cantonese- and Mandarin-English bilinguals are primarily exposed to English outside the home. The influence of an external ambient language (in this case, English), even when an individual does not experience it until school age, may overshadow traditional notions of bilingual language development, such as maternal language (in this case, Cantonese or Mandarin). As it is currently implemented, the LEAP-Q survey does not easily account for such early childhood factors. Research with pre-school and early school age children in these bilingual communities may shed more light on questions of home-internal and home-external developmental factors affecting phonetic perception. Additionally, recruitment of participants in more entrenched bilingual communities, such as those in British Columbia, California, and Hawaii, may yield different results than those obtained here.

The use of non-native tonal contours may be another source of complication in the present experimental design. Non-native tonal contours are more difficult for speakers than native tonal contours are (Lee, Vakoch, & Wurm, 1996). Although late bilingual speakers did not seem to be significantly challenged by the use of Thai tonal contours in this experiment, early bilingual speakers were. It is possible that if tested using Chinese tonal contours, early bilingual speakers would approach the accuracy levels of their late bilingual peers, perhaps demonstrating an intermediate pattern more similar to that observed by Dupoux, et al. (2010). It may also be the case that certain aspects of the experimental task were too easy for certain speakers, allowing the scores of the English monolingual Group to equal those of the early bilinguals. In their original experimentation developing the sequence recall task, Dupoux and colleagues observed ceiling effects that rendered their results insignificant. Only when the within-sequence ISI was reduced significantly and intra-token variability was introduced did differences between Groups emerge clearly (E. Dupoux, personal communication, 14 December 2011). Future implementations of the present experiment (in progress) account for this possibility, reducing the ISI from a long 1000ms and introducing changes in the base F_0 used to formulate stimulus tokens.

Overall, this research program constitutes progress in extending questions of speech perception to the area of lexical tone. The present study adds to the body of literature concerned with perception of non-native speech cues, replicating work conducted with monolingual speakers of tone and non-tone languages by adapting a new behavioural method. Moreover, this study is the first to extend these questions to investigate dominance in a population of early simultaneous bilinguals, for whom the contrast of interest is native in one language but not in the other. Findings with this test group underscore previous claims regarding the relationship of a bilingual's two languages within his or her individual perceptual system, indicating that even simultaneous bilinguals may only process one of their first languages with complete native-like proficiency. In this case, early bilingual speakers of a tone and a non-tone language exhibit accuracies similarly to speakers with no tone language experience at all, while subtle differences in their processing strategies highlight underlying evidence of their tone language experience. These findings necessitate the closer investigation of biographical and developmental factors that might give rise to such patterns.

---

[1] Praat PitchTiers represent “time-stamped pitch contours”. Pitch points within the Tier are controlled for time and F_0, while space between points is “linearly interpolated” to calculate contour frequencies (Boersma & Weenink, 2012).

[2] The Canadian version of this questionnaire was retrieved 10 January 2011 from <http://comm.soc.northwestern.edu/bilingualism-psycholinguistics/files/LEAPQ-CANADA2006.doc>.

[3] Recruitment of participants, data collection, and data retention were completed with approval from the Research Ethics Board of the University of Alberta (Approval #00021378, expiring 13 February 2013).

[4] In many linguistic experiments involving words and pseudowords, it is also advisable to include random intercepts for stimulus items, as these items are only a sample of a larger population and may be affected by variables outside the scope of experimental interest (e.g., frequency). However, in this case, sequences were random formations of highly controlled, non-word syllables, and were therefore not included in the random effects structure.

Bibliography

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2011). Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics" (Version 1.3) [R package]. Available from http://cran.r-project.org.

Bates, D. & Maechler, M. (2011). Linear mixed-effects models using S4 classes (Version 0.999375-42) [R package]. Available from http://cran.r-project.org.

Beddor, P. S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, *71*, 1551-1561.

Best, C. & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305-330.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.18, retrieved 10 March 2011 from http://www.praat.org/.

Boersma, P. & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.04, retrieved 12 January 2012 from http://www.praat.org/.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language specific vowel contrast in the first year of life. *Language and Speech*, *46*, 217–244.

Burnham, D. (2000). Excavations in language development: Cross-linguistic studies of consonant and tone perception. In D. Burnham, S. Luksaneeyanawin, C. Davis, & M. Lafourcade (Eds.), *Interdisciplinary approaches to language processing: The International Conference on Human and Machine Processing of Language and Speech* (pp. 44-69). Bangkok, Thailand: Chulalongkorn University Press.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Lacerda, F., & Attapaiboon, C. (1996). Facilitation or attenuation in the development of speech mode processing? Tone perception over linguistic contexts. In P. McCormack & A. Russell (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 587-592). Canberra: Australian Speech Science and Technology Association.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. In *Pan-Asiatic Linguistics: Proceedings of the 3rd International Symposium of Language and Linguistics*. Bangkok: Chulalongkorn University Press, 546-560.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science, 21,* 343–348.

Curtin, S., Goad, H., & Pater, J. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*, 89-405.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilingualism. *Nature*, *340*, 229-230.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation in bilinguals. *Cognitive Psychology*, *24*, 381-410.

Dupoux, E., Pallier, C., Sebastián, N., & Mehler, J. (1997). A destressing ‘‘deafness” in French? *Journal of Memory and Language*, *36*, 406– 421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress ‘deafness’. *Journal of the Acoustical Society of America*, *110*, 1608–1618.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress ‘deafness’ in simultaneous French–Spanish bilinguals. *Cognition*, *114*, 266-275.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: the case of French learners of Spanish. *Cognition*, *106*, 682–706.

Flege, J., MacKay, I., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567–598.

Gandour, J., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*, 1-33.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sound “L” and “R”. *Neuropsychologia, 9*, 317-323.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science, 255,* 606-608.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25,* 527-542.

MacKain, K., Best, C., Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940-967.

Mattock, K. & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, *10*, 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*, 1367-1381.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception* & *Psychophysics, 18*, 331-340.

Paradis, J. (2011). The impact of input factors on bilingual development: Quantity versus quality. Peer commentary on A. Sorace, Pinning down the concept of interface in bilingualism. *Linguistic Approaches to Bilingualism*, *1*, 67-70.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *2*, 559-572.

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421-435.

R Development Core Team. (2012). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology, 59,* 96–121.

Sebastián-Gallés, N. (2010). Bilingual language acquisition: Where does the difference lie? *Human Development*, *53*, 245-255.

Sebastián-Gallés, N. & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111-123.

Sebastián-Gallés, N., Bosch, L. (2005). Phonology and bilingualism. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches.* Oxford: Oxford University Press, 68-87.

Sebastián-Gallés, N., Echeverria, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*, 240–255.

Sereno, J. A. & Wang, Y. (2006). Behavioral and cortical effects of learning a second language: The acquisition of tone. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins, 239-258.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*, 914-919.

Wang, Y., Spence, M., Jongman, A., and Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649 – 3658.

Werker, J. F. & Tees, R. C. (1984a). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*, 1866-1878.

Werker, J. F. & Tees, R. C. (1984b). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour & Development*, *7*, 49-63.

Werker, J. F., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, *37*, 35–44.

Westbury, C. (2007). ACTUATE: Assessing Cases: The University of Alberta Testing Environment [Computer program]. Version 1.0.0.0, from http://www.psych.ualberta.ca/~westburylab/downloads/actuate.download.html/.

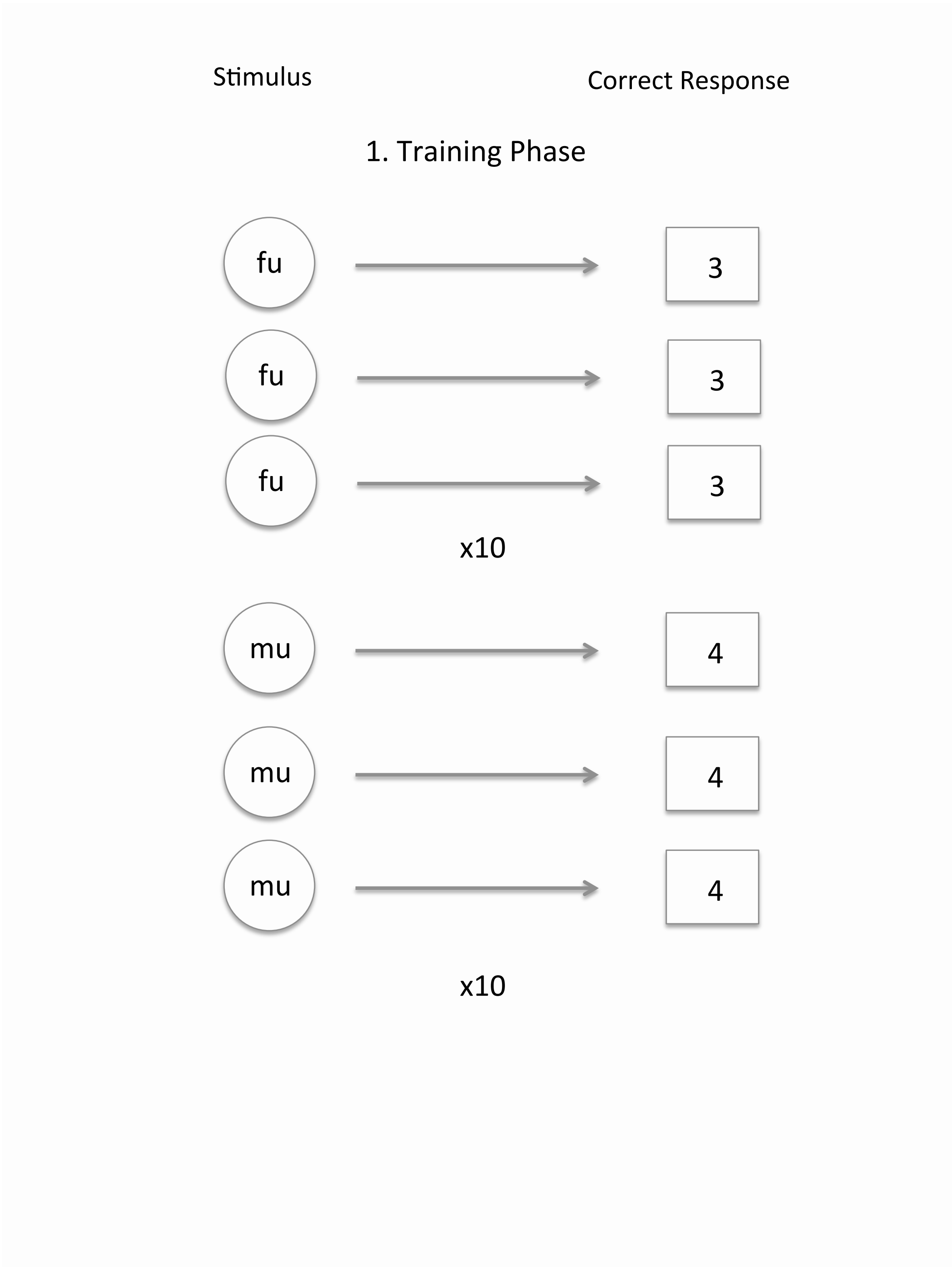
Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Dordrecht, The Netherlands: Springer.

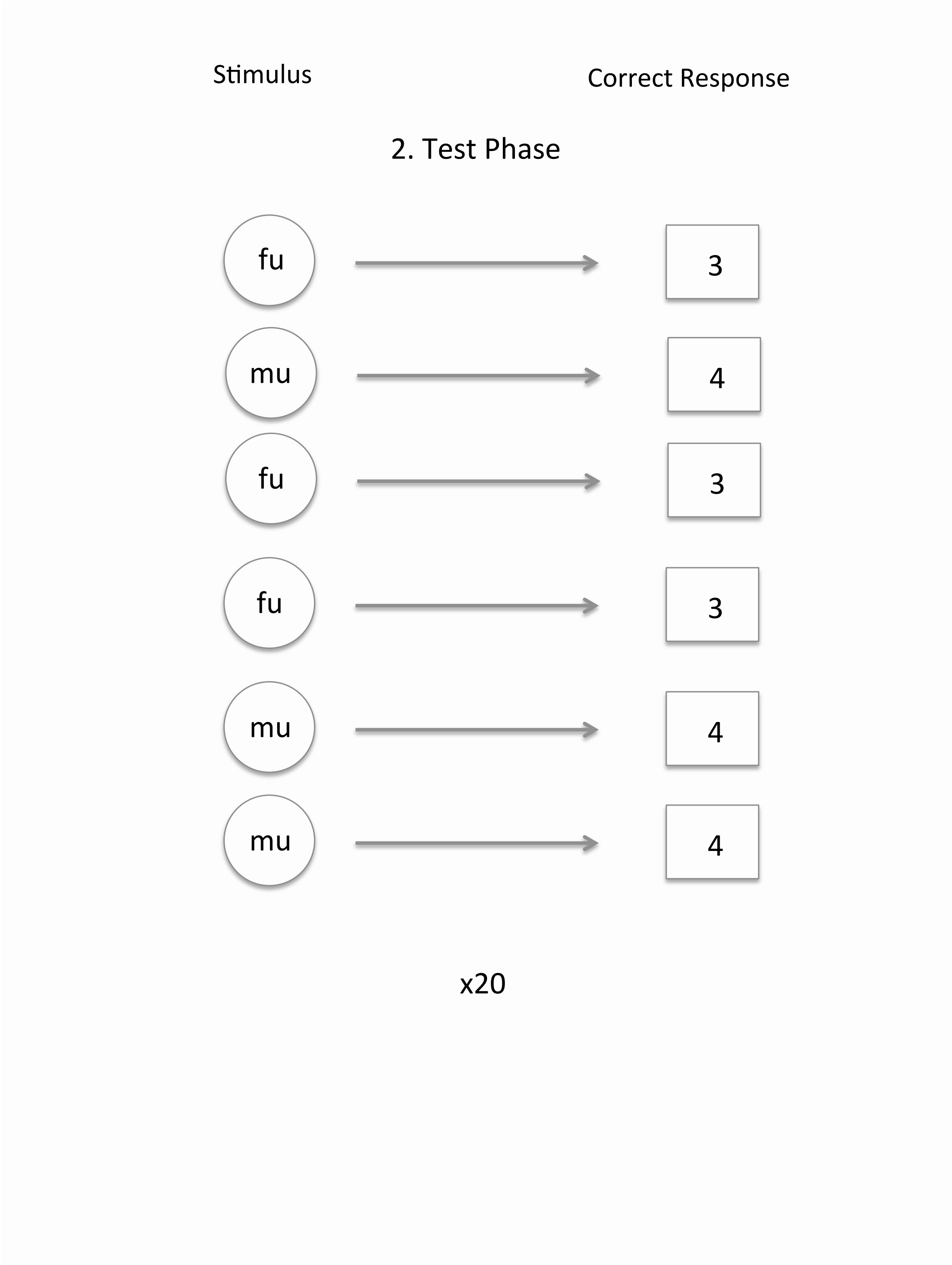
Wickham, H. (2012). An implementation of the Grammar of Graphics (Version 0.9.0) [R package]. Available from http://cran.r-project.org.

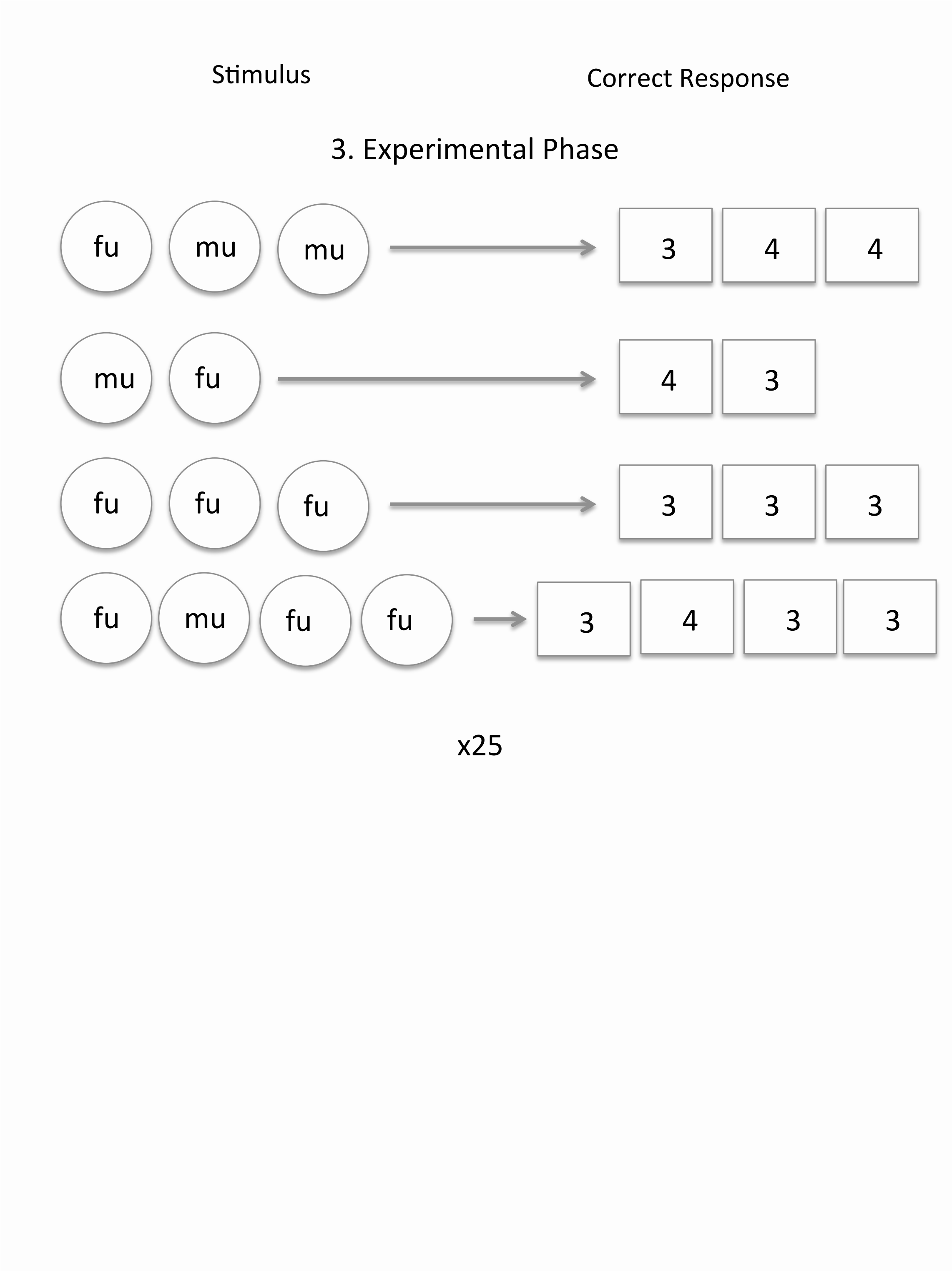
Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Appendix 1

Graphical representation of stimulus presentation




Stimulus
Correct Response
2. Test Phase
fu
3
mu
4
fu
3
fu
3
mu
4
mu
4
x20


Stimulus
Correct Response
3. Experimental Phase
fu mu mu
3 4 4
mu fu
4 3
fu fu fu
3 3 3
fu mu fu fu
3 4 3 3
x25

Appendix 2
LEAP-Q numeric predictors

Current language exposure

1. ChineseExpose (Percent of time participant is currently exposed to Chinese)
2. EnglishExpose (Percent of time participant is currently exposed to English)

Current language choice

3. ChooseReadChinese (Percent of time participant chooses to read Chinese)
4. ChooseReadEnglish (Percent of time participant chooses to read English)
5. ChooseSpeakChinese (Percent of time participant chooses to speak Chinese)
6. ChooseSpeakEnglish (Percent of time participant chooses to speak English)

Acquisition

7. BeginChinese (Age at which Chinese learning began)
8. BeginEnglish (Age at which English learning began)
9. BeginReadChinese (Age at which Chinese reading began)
10. BeginReadEnglish (Age at which English reading began)
11. ChineseHome (Years spent in a Chinese-speaking living environment)
12. EnglishHome (Years spent in a English-speaking living environment)
13. ChineseSchool (Years spent in a Chinese-speaking instructional environment)
14. EnglishSchool (Years spent in a English-speaking instructional environment)

Language-specific proficiency

15. SpeakChinese (1-10 scale of speaking proficiency in Chinese)
16. SpeakEnglish (1-10 scale of speaking proficiency in English)
17. ReadChinese (1-10 scale of reading proficiency in Chinese)
18. ReadEnglish (1-10 scale of reading proficiency in English)
19. UnderstandChinese (1-10 scale of comprehension proficiency in Chinese)
20. UnderstandEnglish (1-10 scale of comprehension proficiency in English)
21. ForeignChinese (1-10 scale of self-rated foreign accent in Chinese)
22. ForeignEnglish (1-10 scale of self-rated foreign accent in English)

Appendix 3

Proportional loadings of LEAP-Q predictors on principal components

PC1		PC2		PC3	
ChooseReadChinese	0.29	UnderstandEnglish	0.5	ChineseSchool	0.41
ChooseSpeakChinese	0.28	ReadEnglish	0.48	ReadChinese	0.36
BeginEnglish	0.27	SpeakEnglish	0.37	ForeignEnglish	0.35
SpeakChinese	0.26	EnglishSchool	0.2	BeginChinese	0.27
UnderstandChinese	0.26	ChooseReadChinese	0.16	SpeakEnglish	0.18
BeginReadEnglish	0.25	SpeakChinese	0.14	EnglishExpose	0.14
ChineseExpose	0.24	UnderstandChinese	0.13	UnderstandEnglish	0.14
ReadChinese	0.22	ChooseSpeakChin	0.12	ChooseReadChinese	0.11
ChineseHome	0.19	ReadChinese	0.05	ChooseSpeakChin	0.11
ForeignEnglish	0.17	BeginChinese	0.04	BeginReadEnglish	0.06
BeginReadChinese	0.05	ChooseSpeakEng	0.03	EnglishHome	0.06
UnderstandEnglish	-0.01	EnglishExpose	0.02	ChooseReadEnglish	0
EnglishHome	-0.03	BeginReadChinese	0.01	SpeakChinese	-0.03
EnglishSchool	-0.04	ChineseExpose	-0.04	BeginEnglish	-0.04
ReadEnglish	-0.11	ChineseHome	-0.06	EnglishSchool	-0.06
ChineseSchool	-0.13	BeginEnglish	-0.08	ReadEnglish	-0.06
SpeakEnglish	-0.16	EnglishHome	-0.1	UnderstandChinese	-0.08
BeginChinese	-0.2	ForeignChinese	-0.17	ForeignChinese	-0.1
ChooseReadEnglish	-0.26	ChineseSchool	-0.19	ChineseHome	-0.2
ChooseSpeakEnglish	-0.27	ForeignEnglish	-0.21	ChineseExpose	-0.23
EnglishExpose	-0.28	ChooseReadEnglish	-0.23	ChooseSpeakEng	-0.29
ForeignChinese	-0.29	BeginReadEnglish	-0.28	BeginReadChinese	-0.45

PC4		PC5		PC6	
ChineseExpose	0.37	BeginEnglish	0.32	EnglishSchool	0.57
BeginChinese	0.36	EnglishSchool	0.3	EnglishHome	0.31
ChooseSpeakChin	0.18	ForeignChinese	0.21	ForeignEnglish	0.16
ChooseReadChinese	0.17	ReadChinese	0.17	SpeakEnglish	0.11
ForeignChinese	0.15	ChooseReadEnglish	0.12	BeginEnglish	0.08
ChooseSpeakEnglish	0.07	ChineseHome	0.12	BeginChinese	0.07
ReadEnglish	0.05	ChooseSpeakChin	0.09	ChineseExpose	0.05
BeginReadChinese	0.05	UnderstandEnglish	0.05	ChooseReadChinese	0.04
ReadChinese	0.05	ChineseExpose	0.03	ChooseSpeakChinese	0
BeginReadEnglish	0.04	EnglishExpose	0.03	BeginReadChinese	0
UnderstandEnglish	0.03	ChineseSchool	-0.03	BeginReadEnglish	-0.02
ChineseSchool	0.03	ChooseReadChinese	-0.06	ReadEnglish	-0.06
BeginEnglish	-0.01	ChooseSpeakEnglish	-0.07	ForeignChinese	-0.06
ChooseReadEnglish	-0.16	ReadEnglish	-0.07	ChineseHome	-0.09
SpeakEnglish	-0.17	ForeignEnglish	-0.1	UnderstandEnglish	-0.14
SpeakChinese	-0.17	UnderstandChinese	-0.11	ChineseSchool	-0.18
EnglishHome	-0.22	SpeakEnglish	-0.17	ReadChinese	-0.18
ForeignEnglish	-0.23	SpeakChinese	-0.19	ChooseSpeakEnglish	-0.2
UnderstandChinese	-0.26	BeginChinese	-0.23	EnglishExpose	-0.22
EnglishExpose	-0.28	BeginReadEnglish	-0.24	ChooseReadEnglish	-0.29
EnglishSchool	-0.3	BeginReadChinese	-0.35	UnderstandChinese	-0.31
ChineseHome	-0.46	EnglishHome	-0.6	SpeakChinese	-0.39