NORTHWESTERN UNIVERSITY

Flexibility of Speech Perception and Production Strategies in Mandarin-English Bilinguals

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Laura Ann Burchfield

EVANSTON, ILLINOIS

June 2015

ABSTRACT

Flexibility of Speech Perception and Production Strategies in Mandarin-English Bilinguals

L. Ann Burchfield

This dissertation examines speech perception and production strategies in Mandarin-English bilinguals. We ask what happens when a bilingual's two languages differ in basic perception or production strategies. In particular, we ask whether multiple perception and production strategies are available to bilinguals and how language experience affects a bilingual's choice of strategy. Both of our experiments included English monolingual controls and three bilingual groups: Late Mandarin-English bilinguals (L1 Mandarin L2 English), Early Mandarin-English bilinguals (Mandarin heritage speakers), and Late English-Mandarin bilinguals (L1 English L2 Mandarin). In Experiment 1, we examine speech perception strategies in Mandarin-English bilinguals with a fragment detection task. We find that all groups of bilinguals showed evidence of a syllable-based segmentation strategy in Mandarin, and no evidence of a syllable-based segmentation strategy in English. In Experiment 2, we examine speech production strategies in an implicit priming task. Previous work has suggested that the syllable is a basic unit of speech planning in Mandarin, while the phoneme serves this role in English (O'Seaghdha et al., 2010). We find that the L1 Mandarin L2 English group and the Mandarin heritage speakers differed in production strategies across languages. In particular, the Mandarin Heritage speakers showed L1-like strategies in both Mandarin and English. In both experiments, bilinguals were able to change between L1 and L2 strategies depending on task language. Overall, these results suggest flexibility in bilingual perception and production strategies. We suggest that differences between English and Mandarin, such as differences in

syllable structure and syllable-morpheme correspondence, promote different perception and production strategies and create an environment that allows this bilingual flexibility to surface.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Ann Bradlow for her guidance, insight, and unwavering support. I could not have asked for a better mentor to guide me through this process. I would also like to thank my committee members, Matt Goldrick and Casey Lew-Williams for their valuable input and support of the project.

Many others at Northwestern have also helped to make this project possible. Thank you to Chun-Liang Chan for technical support, Emily Khan for subject running and coding of data, and Victoria Kuritza for subject running. My thanks also go to the participants themselves for their willingness to endure the often tedious experimental tasks. My classmates and members of the Speech Communication Research Group and Sound Lab have also been a valuable source of feedback throughout the course of this project. In particular, I would like to thank my classmate Lauren Ackerman for providing me with encouragement and always being a willing sounding board for ideas.

Funding for this project was provided by a Graduate Research Grant from Northwestern University and by Grant R01-DC005794 from NIH-NIDCD. I thank Northwestern and the NIH for this financial support.

Finally, I would like to thank my family. In particular, my parents, Mitchel and Jan Burchfield and my husband, Laszlo Frazer have provided me with endless support and encouragement throughout the process of completing this dissertation.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

1.1.1 Bilingualism

At first glance, the reasons to study bilinguals might not be obvious. If we assume a bilingual is someone with equal command of two languages and that command is essentially the same as that of a monolingual in each language, there would be little to learn from bilinguals because studying monolinguals would provide the same information. However, research in recent decades suggests that bilinguals do not fit this description.

First, absolutely equal proficiency in two languages is probably not a common state of affairs, and is a difficult notion to define or measure. Even a bilingual exposed to two languages from an early age may use these languages in different social domains, such as speaking one language at school or work and one at home. This could lead to differences in vocabulary or speaking, reading, and writing abilities across the two languages. Probably even more common are individuals with one dominant language, but the ability to communicate effectively in another language as well. The balance between these two languages could vary considerably depending on how often the second language is used, when it was first learned, and how long it has been studied. One way to view bilingualism is as a continuum, with absolute monolinguals on one end and bilinguals with perfectly balanced proficiency between two languages on the other end. In reality, most people would fall somewhere between these two theoretical extremes.

Additionally, the two languages present in a bilingual do not function as completely separate systems. Instead, it appears that a bilingual’s two languages may both be active at any given time and each may influence the other. For example, Spivy and Marian (1999) found that

Russian-English bilinguals look toward distracter objects with Russian names that sound similar to an English target word (for example, looking toward a picture of a marker when hearing the Russian word *marku* “stamp”). This cross-language interference also occurred in the other direction (i.e. with English stimuli and distractor items phonetically overlapping in Russian). In later studies this effect has been replicated even in entirely monolingual environments where the subjects had no indication that the experiment related to bilingualism (Marian and Spivy, 2003).

This between-language influence also extends to the domain of phonemic contrasts. The phoneme /t/ is present in both English and French, but has a longer voice onset time in English than in French. Flege (1986) found that French-English bilinguals produce /t/ with a slightly longer voice onset time in French and a slightly shorter voice onset time in English, compared to monolinguals in each language. This suggests that the realization of the phoneme in each language moved toward the norm of the other. Even brief exposure to a foreign language may influence production of phonemes in the first language. Chang (2012) found that after only a few weeks of intensive Korean study, previously monolingual English speakers showed subtle influences of Korean on productions of English phonemes.

In recent decades, research has also shown that the presence of two language systems in one brain has wide-ranging effects on the cognition of bilingual individuals (see Bialystok, Craik, & Luk (2012) for a review). Bilinguals often outperform monolinguals on tasks requiring selective attention. For example, bilinguals experience less interference than monolinguals in the Simon task, in which participants must attend to the color of stimuli while ignoring location (e.g. Bialystok, Craik, Klien, & Viswanathan, 2004). This may be because constantly juggling the demands of competing languages trains the skills needed to selectively attend to relevant information or ignore irrelevant information. However, monolinguals may outperform bilinguals on some verbal tasks. For example, bilinguals take slightly longer to retrieve names of pictures (Gollan, T. H., Montoya, Fennema-Notestine, & Morris, 2005), and experience tip-of-the tongue states more often (Gollan & Silverberg, 2001; Gollan & Acenas, 2004), as compared to monolinguals. This may be due to interference between a bilingual's two languages and/or the fact that bilinguals have less total exposure to each language than monolinguals.

The complex interactions between a bilingual's languages make bilingualism an important area of psycholinguistic research. Languages vary widely in structure and content, with differing vocabularies, grammars, phoneme inventories, and syllable-structures. Anywhere these differences exist, bilinguals face interactions of their two language systems.

1.1.2 Cross-language Differences in Speech Production and Perception Strategies

1.1.2.1 Perception Strategies

Strategies for segmenting speech into words vary by language. English speakers tend to segment speech at the onset of stressed syllables (Cutler & Norris, 1988). Because a majority of English words begin with a stressed syllable, this is an effective strategy for locating English word boundaries. However, not all languages share this property with English, so other segmentation strategies may be more useful in these languages. One alternative to this strategy is a syllable-based segmentation strategy, in which listeners segment speech at syllable boundaries. French is one example of a language that promotes this strategy (Cutler, Mehler, Norris, & Segui, 1986). Evidence for this comes from a fragment detection task in which participants were asked to respond upon hearing a given sequence of phonemes embedded in a word. French speakers responded more quickly when this phoneme sequence matched the initial syllable of the word, compared to when it corresponded to a partial syllable or a full syllable plus a partial syllable. English speakers however did not show this effect. Given the role of the syllable in Mandarin speech planning and Mandarin's restricted syllable structure and syllable-morpheme correspondence (both discussed in the following section), Mandarin is a likely candidate for a syllable-based segmentation strategy as well.

1.1.2.2 Production Strategies

One area of cross-language difference that has come to recent attention is variability in the basic units of speech production planning. Chen, Chen, & Dell (2002), O'Seaghdha et al. (2010), and Chen & Chen (2013) suggest that the syllable is the basic unit of speech planning in Mandarin, as opposed to the phoneme, which appears to serve this role in English. These studies draw on evidence from an implicit priming task. In this task, subjects were asked to memorize several pairs of words, or word-picture pairs. When the prompt item was presented, they were asked to speak the response word. In English speakers, if the response items all shared the same initial phoneme, response times tended to be faster than if the initial phoneme was not shared. However, in Mandarin, there was no such effect. Instead, reaction time only decreased if the response items all shared an initial syllable.

Evidence from other sources also supports the role of syllables as a basic speech-planning unit in Mandarin. Chen (2000) reports the presence of syllable movement errors in a Mandarin speech corpus. Speech errors of this kind are not commonly reported in English (Bock, 1991; Dell, 1995). Masked priming studies (Chen, Lin, & Ferrand, 2003; You, Zhang, & Verdonschot, 2012) have found that Mandarin speakers produce responses more quickly when they are briefly preceded by a prime that shares an initial syllable with the response, as compared to a prime sharing the same phoneme sequence, but not corresponding to a syllable in the target word (e.g. qi4qiu2 "balloon" was produced more quickly when preceded by qi3 "attempt" than by qing2 "affection"). This is in contrast to English and Dutch speakers, who respond more quickly with increasing phonetic overlap between the target and prime, regardless of syllable structure (Schiller, 1998, 1999).

Several differences between English and Mandarin may contribute to this difference in speech planning strategies. Mandarin has a highly restricted syllable structure, with no consonant clusters and only two possible codas, /n/ and /ŋ/. This results in a very limited number of total syllables (about 400 before tone is taken into account and 1200 after tone is taken into account). This contrasts with English, which allows a wide variety of complex onsets and codas, resulting in over 12,000 syllables (Levelt et al., 1999, based on a count from CELEX) and many more which are legal but not in use. Given this difference, storing a list of all possible syllables and accessing it directly when planning speech might be much more practical in Mandarin than in English. Mandarin's writing system may also play a role in the differences. Mandarin is most commonly written in characters, each of which corresponds to a monosyllabic morpheme. The very nearly one-to-one correspondence between characters and morphemes with syllables may result in (or be a product of) a stronger tendency to plan speech in syllables.

1.2 The Present Study

1.2.1 Goals

This variability in perception and production strategies raises an interesting question in the case of bilinguals. When the two languages of a bilingual vary in basic perception or production strategies, what determines the bilingual's choice of strategy? In this study we set out to answer this question in several specific cases. First, are multiple perception/production strategies available to bilinguals, or are strategies associated with only one language available to each individual? Second, we ask whether bilinguals are more likely to apply L1 (first language) strategies in L2 (second language) environments that more closely resemble L1 (e.g. for words which are phonotactically legal in L1). Third, we ask how degree of experience in L1 and L2 affects a bilingual's choice of strategy. Finally, we ask whether there is a relationship between perception and production strategies within individuals.

We address these questions with two experiments. In the first experiment, we examine speech perception strategies in Mandarin-English bilinguals. In the second, we examine speech production strategies in the same group of bilinguals. Both of these experiments allow us to determine whether bilinguals can change strategies across languages, or are restricted to a single strategy. In the first experiment, we also compare the strategies applied by bilinguals to speech perception in English stimuli with varying degrees of similarity to Mandarin structures. In both experiments, we compare choice of strategies between Mandarin-English bilinguals with varying degrees of proficiency in the two languages. Three groups of bilinguals are included in the study: L1 Mandarin L2 English speakers, Mandarin heritage speakers, and L1 English L2 Mandarin speakers. These groups are compared to an English monolingual control group, which also allows us to ask whether L1 strategies in bilinguals differ from monolingual strategies due to the influence of L2 on L1. Because the studies were conducted at an American university, Mandarin monolingual controls were not available. However, in the case of the production experiment, previous data from Mandarin monolinguals in a nearly identical task was available (Chen & Chen, 2013). In the final chapter, we compare perception and production strategies within individuals who participated in both the first and second experiments.

1.2.2 Findings

To preview the results, we found that bilinguals are able to change strategies depending on the task language in both perception and production. We did not find evidence that an environment more similar to L1 encouraged listeners to apply L1 strategies to L2 input. In the perception experiment, we found evidence that even modest amounts of L2 experience allowed bilingual listeners to adopt L2 strategies. In the production experiment, we found that L1 strategies in bilinguals sometimes differed from those of monolinguals.

1.2.3 Structure of the Dissertation

The structure of this dissertation is as follows: Chapter 2 will describe Experiment 1, which aims to determine the speech segmentation strategies of Mandarin-English bilinguals. Chapter 3 will describe Experiment 2, which aims to determine the speech production strategies of Mandarin-English bilinguals. Chapter 4 will discuss overall findings from the two experiments and compare perception and production strategies within subjects who participated in both experiments.

2 Flexibility of Speech Perception Strategies in Mandarin-English Bilinguals

2.1 Introduction

2.1.1 Background

Speech is a continuous stream of sound, without clear acoustic boundaries separating words. In order to understand speech, listeners must develop strategies to segment the speech stream into useful pieces, such as words and morphemes. Some sources of information available to support segmentation strategies include transitional probabilities between phonemes (Saffran et al., 1996), phonotactic constraints (Weber & Cutler, 2006), and rhythmic properties of the language being spoken (Cutler et al., 1986, Otake et al., 1993; Murty et al., 2007; Kim et al., 2008). Listeners are able to take advantage of all these sources of information, but exactly how they do so varies by language. In the present study, we will focus on rhythm-based segmentation strategies.

English content words tend to begin with stressed syllables (Cutler & Carter 1987) and English listeners tend to segment speech at stressed syllable boundaries. Cutler & Norris (1988) found that English speakers spotted words embedded in non-words more quickly when the non-words consisted of a strong syllable followed by a weak syllable than when they consisted of two strong syllables. Words like "mint" were detected more quickly in non-words like "mintesh" (stress only on the initial syllable) than in "mintayve" (stress on both syllables). This suggests that in the non-words like "mintayve", listeners draw a boundary before the /t/, making it less readily interpretable as part of the word "mint". In a language like English, this strategy will help listeners correctly identify word boundaries most of the time. However, this is not an effective strategy in languages that do not have lexical stress or do not tend to place it at the beginning of words.

An alternative to this stress-based segmentation strategy is a syllable-based segmentation strategy. One language that uses a syllable-based strategy is French (Mehler et al., 1981, Cutler et al., 1986). When asked to press a button after hearing a word starting with a specific sequence of phonemes (a task referred to as a fragment detection task), French speakers responded more quickly when that target sequence corresponded to a full syllable in the word, rather than a partial syllable or a full syllable plus a partial syllable. For example, listeners responded more quickly when asked to detect the sequence bal- in the word "*balcon*" (balcony) than when asked to detect the sequence ba- in the same word because bal- is the initial syllable of "*balcon*" while ba- is only a partial syllable. When presented with English stimuli, they also exhibited a similar response pattern, suggesting that the tendency to segment speech at syllable boundaries was also applied in an unfamiliar language. Conversely, English speakers did not differ in reaction time based on whether the sequence corresponded to a full syllable or not, for both English and French stimuli (e.g. bal- was detected no more quickly than ba- in "balcony" or "*balcon*").

Another rhythm-based segmentation strategy is mora-based segmentation, which is used by speakers of Japanese (Otake et al., 1993) and Telugu (Murty et al., 2007). A mora is sub-syllabic unit, which may correspond to a consonant followed by a vowel, a single vowel, or a coda consonant. For example, the Japanese brand name "*Honda*" has 3 morae: "ho", "n", and "da". Like English and French speakers, Japanese speakers also applied their L1 segmentation strategy to foreign language input. For example Japanese speakers were less accurate when detecting the consonant /n/ in English words when it occurred in the syllable onset position, where it is not moraic (e.g. in the word "enamel"), than in the coda position, where it is moraic (e.g. in the word "inlet").

2.1.2 Hypotheses and Predictions

2.1.2.1 Speech Segmentation Strategies in Bilinguals

These cross-linguistic differences raise the question of what happens when a bilingual's two languages favor different strategies. One possibility is that a listener's dominant language completely determines the speech segmentation strategies available to that listener. Under this view, listeners will apply segmentation strategies from the dominant language when segmenting speech in a non-dominant language. We will call this view the *Fixed Strategy Hypothesis*. Another possibility is that listeners can change speech perception strategies based on task language. Under this view, listeners are able to learn speech segmentation strategies that differ from those of the dominant language and apply them when listening to a non-dominant language. We will call this the *Variable Strategy Hypothesis*.

2.1.2.2 Fixed Strategy Hypothesis

Previous work with bilinguals has supported the Fixed Strategy Hypothesis for rhythm-based segmentation strategies. In the case of French-English bilinguals, even individuals with native or near-native knowledge of both languages apply strategies from the dominant language when listening to the non-dominant language (Cutler et al. 1989, 1992), suggesting a fundamental limitation on learning segmentation strategies other than that of the dominant language. In these experiments, English-dominant participants with native-like proficiency in French did not adopt the syllable-based segmentation strategy characteristic of native French speakers in French (i.e. detecting phoneme sequences that correspond exactly to one syllable of a word more quickly than those that do not). In a second task (the same word-spotting task used in

Cutler & Norris, 1988) English-dominant participants tended to segment speech at stressed syllable onsets, but French-dominant participants with native-like English proficiency did not adopt this stress-based segmentation strategy. This led the authors to conclude that rhythm-based segmentation strategies, including syllable-based and stress-based strategies could not be learned unless encouraged by an individual's dominant language.

Studies in other areas of speech perception have suggested additional limitations on bilingual flexibility. For example, Pallier et al. (1997) found that Spanish-dominant Spanish-Catalan bilinguals had difficulty distinguishing a vowel contrast found in Catalan but not Spanish, even if they had early and continuous exposure to both languages. Dupoux et al. (2010) found that French-dominant French-Spanish bilinguals exhibited stress-deafness, a difficulty in perceiving stress contrasts in Spanish, even with high levels of exposure to Spanish from an early age.

2.1.2.3 Variable Strategy Hypothesis

However, some studies have found evidence that bilinguals can suppress L1 speech segmentation strategies and/or partially adopt strategies characteristic of the non-dominant language. The French-dominant participants with native-like proficiency in English from Cutler et al. (1989, 1992) did not adopt a stressed-based segmentation strategy in English, but they also did not apply a syllable-based segmentation strategy, characteristic of their L1 French, to English. In a fragment detection task with English stimuli, these participants did not differ in response time based on whether the target sequence matched the initial syllable of the context word, suggesting that they did not apply a syllable-based segmentation strategy to English input. However, in a word spotting task, they did not detect embedded words beginning at stressed syllable onsets any more quickly than embedded words beginning at unstressed syllable onsets.

This indicates that they did not adopt the English stress-based segmentation strategy even though they had suppressed the French syllable-based segmentation strategy for English stimuli.

Listeners are also able to apply L2 phonotactic constraints to speech segmentation. Slovak-German bilinguals applied phonotactic constraints from German and lexical knowledge from Slovak to L2 German speech segmentation (Hanulikova et al., 2011). Slovak allows some single consonants words (e.g. /k/ "to") while German does not. Like L1 German speakers, these participants spotted German words (e.g. rose "rose") embedded in non-words more quickly when the words were preceded by a syllable (e.g. suckrose), rather than a single consonant (e.g. krose, trose). This suggests that they favored segmentation outcomes in which the input was segmented into pieces that were phonotactically legal in German. However, in cases where the embedded word was proceeded by a single consonant, subjects responded more quickly if it was a Slovak word than if it was not (e.g. rose was detected more quickly in krose than in trose because /k/ is a Slovak word but /t/ is not), suggesting that Slovak lexical knowledge also influenced German speech segmentation. In another study, German-English bilinguals applied phonotactic constraints from both English and German in segmenting L2 English (Weber & Cutler, 2006). In a word spotting task, L1 German speakers with high English proficiency spotted English words more quickly when they were preceded by a context that forced a syllable boundary in English, but not German (e.g. "lecture" in tharshlecture) or German, but not English (e.g. "lecture" in moycelecture), as compared to when the context did not force a boundary in either language (e.g. lecture in moinlecture).

In some cases, bilingual listeners may be less likely to apply L1 segmentation strategies to L2 input if it violates L1 phonotactics. Japanese speakers in Murty et al. (2007) applied the mora-based strategy characteristic of Japanese to Telegu stimuli, but only those which were phonotactically legal in Japanese. Telugu words containing non-nasal codas elicited a response pattern more closely resembling syllable-based segmentation from Japanese listeners (e.g. god- was detected more quickly and accurately than go- in *godralu*), although the authors note that this result was statistically tenuous. This suggests that listeners may have difficulty applying L1 segmentation strategies to input which does not respect L1 phonotactic constraints. In this case, Japanese listeners did not apply the L1 strategy even though the input should have lent itself to mora-based segmentation because Teleugu also has a mora-based rhythmic structure.

Another possible reason to expect flexibility in perception strategies for Mandarin-English bilinguals is that the evidence for absolute limitations on bilingualism discussed in section 1.2.4 comes from pairs of languages that are quite closely related genetically and through historical contact (e.g. English and French (Cutler et al., 1989, 1992), Spanish and Catalan (Pallier et al., 1997), Spanish and French (Dupoux et al., 2010)). It is possible that speaking two languages with more drastic cross-linguistic differences could necessitate the learning of new strategies that would not be necessary when speaking two more similar languages. In this case, the studies described above would represent specific cases in which there is not sufficient reason for bilinguals to vary strategies across languages, rather than indicating fundamental limitations on bilingualism.

2.1.2.4 English and Mandarin

In this study, we determine the speech perception strategies used by speakers of English and Mandarin Chinese, two languages with highly disparate sound structures. In particular, Mandarin syllable structure is highly restricted when compared with English. In English, a single syllable may have clusters of several consonants at both the beginning and end of a syllable. For example the single-syllable word “strengths” (/strɛŋθs/) consists of a cluster of three consonants followed by a vowel, followed by another cluster of three consonants. In Mandarin, syllables contain a vowel preceded by at most one consonant and followed by at most one consonant. Additionally, only two consonants, /n/ and /ŋ/, are possible in the final position (Tseng, 2005). One consequence of this structure is that syllable boundaries are more predictable in Mandarin than in English. Any consonant other than /n/ or /ŋ/ can automatically be assumed to be a syllable onset. Chen et al. (2007) found that a computational model taking into account only the sequence of phonemes could predict syllable boundaries more accurately in Mandarin than in English.

In the Chinese writing system, each written character represents a syllable. This writing system may exist because of a language that lends itself well to segmentation into syllable-sized chunks and/or it may promote a tendency to segment the language into syllables. Furthermore, each Chinese syllable (with very few exceptions) represents a morpheme. This tight correspondence between syllables and morphemes could further promote a syllable-based segmentation strategy because a strategy for segmenting speech into morphemes is presumably necessary and would imply a mechanism for segmenting syllables.

Additionally, Mandarin speakers may plan speech in whole syllables, rather than individual phonemes, when speaking (O'Seaghdha et al., 2010, Chen, Chen, and Dell, 2002, Chen & Chen, 2013). These studies draw on evidence from an implicit priming task. In this task, subjects were asked to memorize several pairs of words, or word-picture pairs. When the prompt item was presented, they were asked to speak the response word. In English speakers, if the response items all share the same initial phoneme, response times tend to be faster than if the initial phoneme is not shared. However, in Mandarin, there is no such effect. Instead, reaction time is only decreased if the response items all share an initial syllable. These results suggest that syllables may be the basic unit of speech planning in Mandarin.

All of these factors together (predictable structure, correspondence to written characters, correspondence to morphemes, role in speech planning) suggest that syllables play a highly fundamental role in Mandarin. Because of this, Mandarin may be even better suited to syllable-based segmentation than French, which has previously been shown to promote a syllable-based segmentation strategy. It is possible that these factors might be enough to promote a syllable-based segmentation strategy in Mandarin, even when it is not a listener's dominant language. However, this could also mean that Mandarin-dominant speakers might have difficulty switching away from a syllable-based segmentation strategy when listening to other languages.

2.1.2.5 The Present Study

The current study examines speech segmentation strategies in Mandarin English bilinguals with a fragment detection task, based closely on that on Cutler et al. (1986, 1989, 1992). Groups under investigation include Mandarin-dominant speakers, English-dominant speakers, and Mandarin heritage speakers (English-dominant, but with early Mandarin experience). These groups perform a fragment detection task in both English and Mandarin. Additionally, in the English condition, we include stimuli that are both legal and non-legal according to Mandarin phonotactics in order to test whether Mandarin speakers are more likely apply an L1 strategy to Mandarin-like stimuli. In previous work, listeners may apply L1 segmentation strategies to other languages, but in most cases the non-L1 stimuli were phonotactically legal in the L1 (Cutler et al., 1986, 1989, 1992; Cutler & Otake, 1994). This also allows us to test whether syllable-based segmentation strategies are more likely for syllables ending in certain phonemes. Content, Meunier, Kearns, & Frauenfelder (2001) found that French speakers show evidence of a syllable-based segmentation strategy in a fragment detection task for non-word stimuli when the syllable boundary occurs directly before or after a liquid consonant, but not for other consonant types. This may be due to differing acoustic cues to syllable boundaries for these consonants. Varying our syllable types allows us to test whether this is also true for Mandarin and/or English speakers.

2.1.2.6 Predictions

The fixed strategy hypothesis predicts that even the pressures for syllable-based segmentation strategies in Mandarin would not be enough to cause English-dominant listeners to apply a syllable-based segmentation strategy to Mandarin stimuli. It also predicts that Mandarin-dominant speakers would apply a syllable-based segmentation strategy in English, regardless of whether the syllables are phonotactically legal in Mandarin or not. For the Mandarin heritage speakers, this hypothesis would predict that if they apply a syllable-based segmentation strategy to stimuli in either English or Mandarin, they would apply it in both.

The Variable Strategy Hypothesis predicts the possibility of English-dominant listeners applying a syllable-based segmentation strategy to Mandarin stimuli. For Mandarin-dominant listeners, it predicts that they may abandon a syllable-based strategy when presented with English stimuli. This might occur for all syllable types, or only for syllable-types not found in Mandarin. Mandarin heritage speakers would be predicted to apply a syllable-based segmentation strategy to Mandarin input, but suppress it for English stimuli.

2.2 Materials and Methods

2.2.1 Participants

Three groups of bilinguals participated in the experiment: L1 Mandarin L2 English speakers (strongly Mandarin dominant), Mandarin heritage speakers (L1 Mandarin, English dominant), and L1 English L2 Mandarin speakers (strongly English dominant). English monolinguals also participated, allowing us to determine baseline response patterns for native English speakers and for listeners completely unfamiliar with Mandarin. Table 1 summarizes relevant demographic information for each of the four participant groups.

Table 1: Demographic information for participant groups

Group	**Number**	**Mean Years of English Study**	**Mean Years of Mandarin Study**	**Mean Age**	**Sex**
L1 Mandarin L2 English	20	12.0	From birth	22.9	8 female
Mandarin Heritage Speakers	20	16.6	From birth	18.85	16 female
L1 English L2 Mandarin	11	From birth	4.36	20.55	7 female
English Monolinguals (English Task)	20	From birth	None	19.3	10 female
English Monolinguals (Mandarin Task)	20	From birth	None	19.2	16 female

2.2.2 Materials

Materials for each condition consisted of seven pairs of words. The two words in each pair shared the same first three initial phonemes, but differed as to whether the final phoneme was followed by a consonant or a vowel, manipulating the placement of the syllable boundary. In CVC- initial stimuli, this "pivotal consonant" was the final phoneme of the first syllable while in CV- initial stimuli it was the initial phoneme of the second syllable[1].

We included stimuli with three different pivotal consonants in the English portion of our experiment. These included /l/, /s/, and /n/. Stimuli in previous experiments that established the differences in segmentation strategies between English and French (Cutler et al, 1986) had the pivotal consonant /l/, so it is included here for comparability to previous results. The only consonant allowed at both the beginning and end of a syllable in Mandarin is /n/, so it was included in order to allow for a comparison between perception strategies in Mandarin-like and non Mandarin-like syllable structures. To further facilitate this comparison, we divided these stimuli into two subcategories; mono-morphemic and bi-morphemic. We also included /s/ in order to have an additional final consonant that was not legal in Mandarin. Mandarin stimuli were all bimorphemic (as necessitated by the structure of Mandarin) and shared the consonant /n/ as their third phoneme[2].

The stimuli for each condition are listed in tables 2-6.

Table 2: English: Monomorphemic nasal stimuli

CV target	**CVC target**	**CV word**	**CVC word**
pe-	pen-	penny	pencil
ba-	ban-	banner	banter
va-	van-	vanish	vantage
ma-	man-	manner	mantis
te-	ten-	tennis	tendon
bu-	bun-	bunny	bungee
pa-	pan-	panel	panda

Table 3: English: Bimorphemic nasal stimuli

CV target	CVC target	CV word	CVC word
su-	sun-	sunny	Sunday
mi-	min-	mining	mineshaft
si-	sin-	sinner	sinless
rai-	rain-	rainy	raindrop
gu-	gun-	gunner	gunsmith
to-	ton-	tonal	toneless
dow-	down-	downy	downcast

Table 4: English: Monomorphemic fricative stimuli

CV target	CVC target	CV word	CVC word
ca-	cas-	castle	casket
mu-	mus-	muscle	muster
mi-	mis-	missile	mister
ma-	mas-	massif	master
fo-	fos-	fossil	foster
le-	les-	lesson	Lester
bu-	bus-	bustle	buskin

Table 5: English: Liquid (stimuli from Cutler et al., 1986)

CV target	CVC target	CV word	CVC word
ba-	bal-	balance	balcony
ca-	cal-	calorie	calculate
ga-	gal-	galaxy	galvanize
ma-	mal-	malady	malcontent
pa-	pal-	palace	palpitate
sa-	sal-	salad	salvage
ta-	tal-	talon	talcum

Table 6: Mandarin stimuli

CV target	CVC target	CV word	CVC word
xi-	xin-	xi1nan2 (southwest) 西南	xin1sheng1 (newborn) 新生
fa-	fan-	fa1nu4 (get angry) 发怒	fan1xin1 (refurbish) 翻新
ju-	jun-	ju1na2 (arrest) 拘拿	jun1dui4 (army) 军队
da-	dan-	da4nao3 (brain) 大脑	dan4shi4 (however) 但是
hua-	huan-	hua2ni4 (satiny) 滑腻	huan2zui3 (to retort) 还嘴
ke-	ken-	ke3neng2 (possible) 可能	ken3ding4 (to be certain) 肯定
qi-	qin-	qi4nei3 (be discouraged) 其内	qin4ding1 (thumbtack) 揿钉

2.2.3 Procedure

Procedures were based closely on those of Cutler et al. (1986, 1989, 1992), but differed in that the experimental stimuli were presented on a computer with the SuperLab software package, rather than played from a cassette tape with orthographic targets displayed on separate monitor. Participants saw target phoneme sequences on a computer screen, and were instructed to press a button when a word beginning with the sequence was heard. Before beginning the task, participants were familiarized with each of the target sequences, in order to eliminate any ambiguity in pronunciation. The phoneme sequences were displayed to participants in pinyin for Mandarin targets and standard English orthography for English targets. The phoneme sequences were either CV or CVC. These were spread across 28 trials, such that each of the fourteen target words was presented twice, once with a CV target syllable and once with a CVC target syllable.

In each trial, a sequence of words (recorded by a male native speaker of the target language) was presented. The target occurred as either the second, third, fourth, or fifth word in the sequence. Additionally, 28 filler trials were included. Each of the target sequences occurred in two filler trials, one in which a word with the target appeared and one in which it did not. For English, this procedure was repeated four times with subjects receiving the four conditions in random order. Reaction times were recorded for each button press.

The Mandarin and the English conditions occurred on separate days, with the non-dominant language occurring on the first day and the dominant language occurring on the later day. This was to minimize influence of the dominant language in the non-dominant language task. English learners carried out the English conditions on the first day while Mandarin learners and Mandarin Heritage speakers carried out the Mandarin condition on the first day.

Some modifications to the procedure were made for the English monolingual group. Each English monolingual participated only in the English conditions or the Mandarin condition. In the Mandarin condition, target sequences were not presented in standard pinyin, but rather in a transliteration scheme more intuitive for those with no Mandarin experience. Additionally, participants were given examples of similar sounds occurring in English words to clarify the phonemic content (e.g. “fah” rhymes with “blah”), with the caveat that Mandarin sounds may be somewhat different from their closest equivalents in English. The target sequences are listed in table 7.

Table 7: English transliteration of pinyin stimuli

Pinyin	English transliteration
xi-	shee-
xin-	sheen-
fa-	fah-
fan-	fahn-
ju-	joo-
jun-	joon-
da-	dah-
dan-	dahn-
hua-	hwah-
huan-	hwahn-
ke-	kuh-
ken-	kuhn-
qi-	chee-
qin-	cheen-

2.2.4 Analysis

We constructed a linear mixed-effects model for each of the five conditions (four English conditions and one Mandarin condition). Because four language groups participated in the experiment, group was expressed as three different fixed effects. These were: 1) Monolingual participants vs. bilingual participants (English monolingual baseline compared to all other groups), 2) L2 Mandarin participants vs. L1 Mandarin participants (L1 English L2 Mandarin group as a baseline compared to Mandarin heritage speakers and the L1 Mandarin L2 English group), and 3) Mandarin heritage speakers vs. the L1 Mandarin L2 English group. The models included fixed effects of language group (English monolinguals vs. all other groups, L1 English (L2 Mandarin) vs. L1 Mandarin (L2 English) and Mandarin heritage speakers, and Mandarin heritage speakers vs. L1 Mandarin (L2 English)), target (CV vs. CVC), context (CV vs. CVC), and all interactions of these three variables. Models also included a maximal random effects structure (Barr et. al, 2013), which consisted of random intercepts and random slopes of target, context, and their interaction by subject and random intercepts and random slopes of target, context, and each of the three group contrasts by item[3].

2.3 Results

2.3.1 English Condition

2.3.1.1 Monomorphemic Nasal Condition

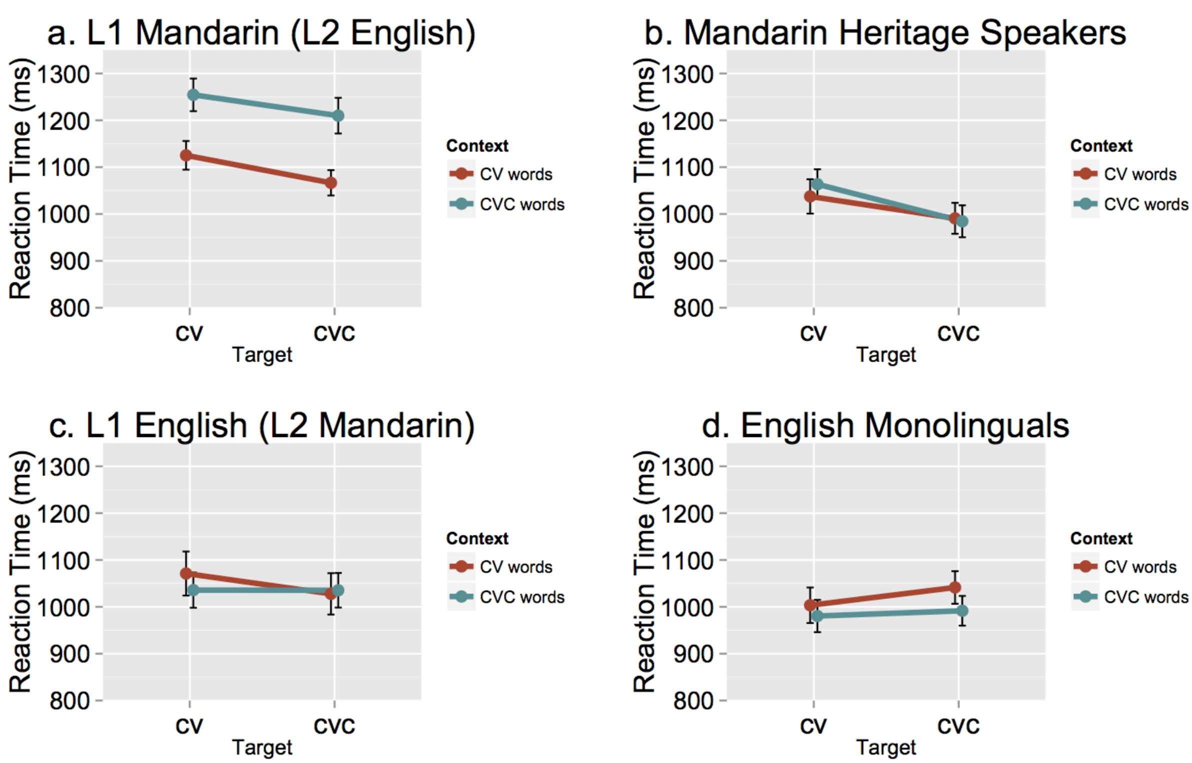

Figure 1: Reaction time in the Monomorphemic Nasal condition as a function of target and context for each language group

Results for this condition are shown in Figure 1. A significant group (L1 Mandarin L2 English vs. Mandarin heritage speakers) by context interaction ($\beta = -112.464$, $t = -2.91$, $p <.01$) indicates that the L1 Mandarin L2 English group responded more quickly to CV words than to CVC words while the Mandarin heritage speakers did not. A second group by context interaction (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English) was significant ($\beta = 57.145$, $t = 1.95$, $p <.05$), which also reflects the faster response to CV words in the L1 Mandarin L2 English group. The tendency to respond more quickly to CV than CVC words is found in several groups in the other English conditions as well and most likely reflects the fact that the CVC words are longer (due to containing an additional consonant).

A marginally significant main effect of target indicated an overall tendency for participants to respond more quickly to CVC than CV targets ($\beta = -33.548$, $t = -1.74$, $p = .072$). A group (English monolinguals vs. all other groups) by target interaction was marginally significant ($\beta = -55.947$, $t = -1.64$, $p = .097$), indicating that groups other than English monolinguals tended to respond more quickly to CVC targets than to CV targets. A significant main effect of group (L1 Mandarin L2 English vs. Mandarin heritage speakers, $\beta = -142.593$, $t = -1.94$, $p <.05$) indicates that the L1 Mandarin L2 English group responded more slowly than the heritage speakers, probably due to lower levels of English experience. No other main effects or interactions reached significance.

The lack of target by context interaction suggests no difference between trials with matched or mismatched targets and contexts. Furthermore, the lack of target by context by group interactions suggests that this lack of interaction is consistent across groups. These results show no evidence of a syllable-based segmentation strategy for any of the listener groups, similar to results for English-dominant listeners from previous studies.

2.3.1.2 Bimorphemic Nasal Condition

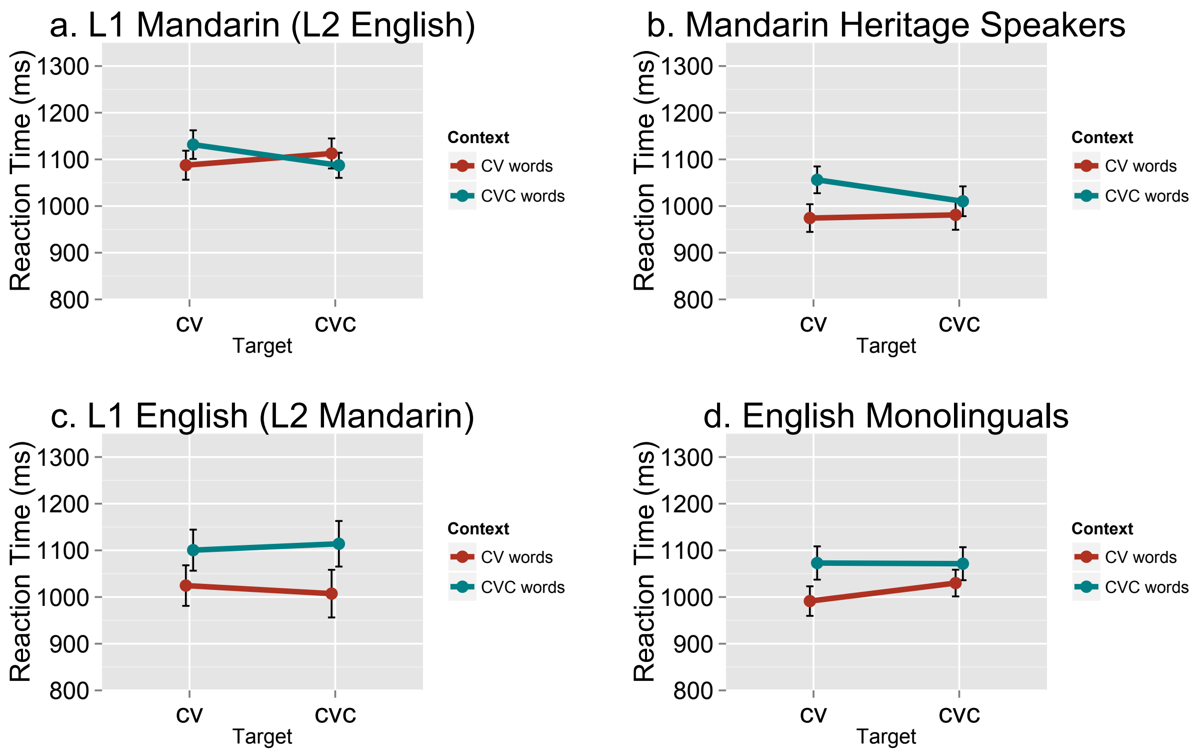

Figure 2: Reaction time in the Bimorphemic Nasal condition as a function of target and context for each language group

Results for this condition are shown in Figure 2. Similar to the monomorphemic nasal condition, these results show no evidence of a syllable-based segmentation strategy. No significant target by context interaction or target by context by group interactions were found. A main effect of context ($\beta = 55.607$, $t = 2.55$, $p < .05$) indicated that subjects responded to CV words more quickly than CVC words. A marginally significant main effect of group ($\beta = -101.839$, $t = -1.68$, $p = .087$) (L1 Mandarin L2 English vs. Mandarin heritage speakers) indicated that L1 Mandarin L2 English speakers responded more slowly than Mandarin Heritage speakers.

2.3.1.3 Monomorphemic Fricative Condition

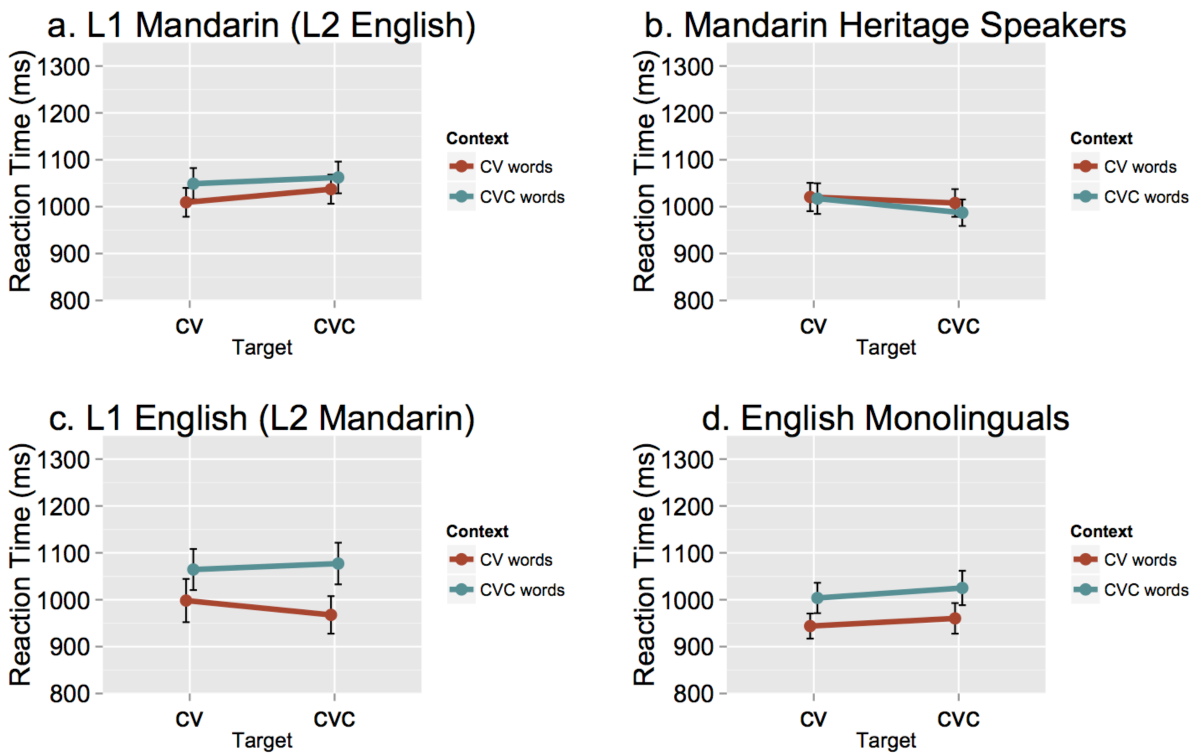

Figure 3: Reaction time in the Monomorphemic Fricative condition as a function of target and context for each language group

Results for this condition are shown in Figure 3. Similar to the other English conditions, no significant target by context interaction or target by context by group interactions were found. A significant context by group (L1 English L2 Mandarin vs. Mandarin Heritage speakers and L1 Mandarin L2 English) interaction ($\beta = -58.575$, $t = -2.04$, $p < .05$) indicated that the subjects in the L1 English L2 Mandarin group responded more quickly to CV words than CVC words, while the Mandarin heritage speakers and L1 Mandarin L2 English group did not.

2.3.1.4 Liquid Condition

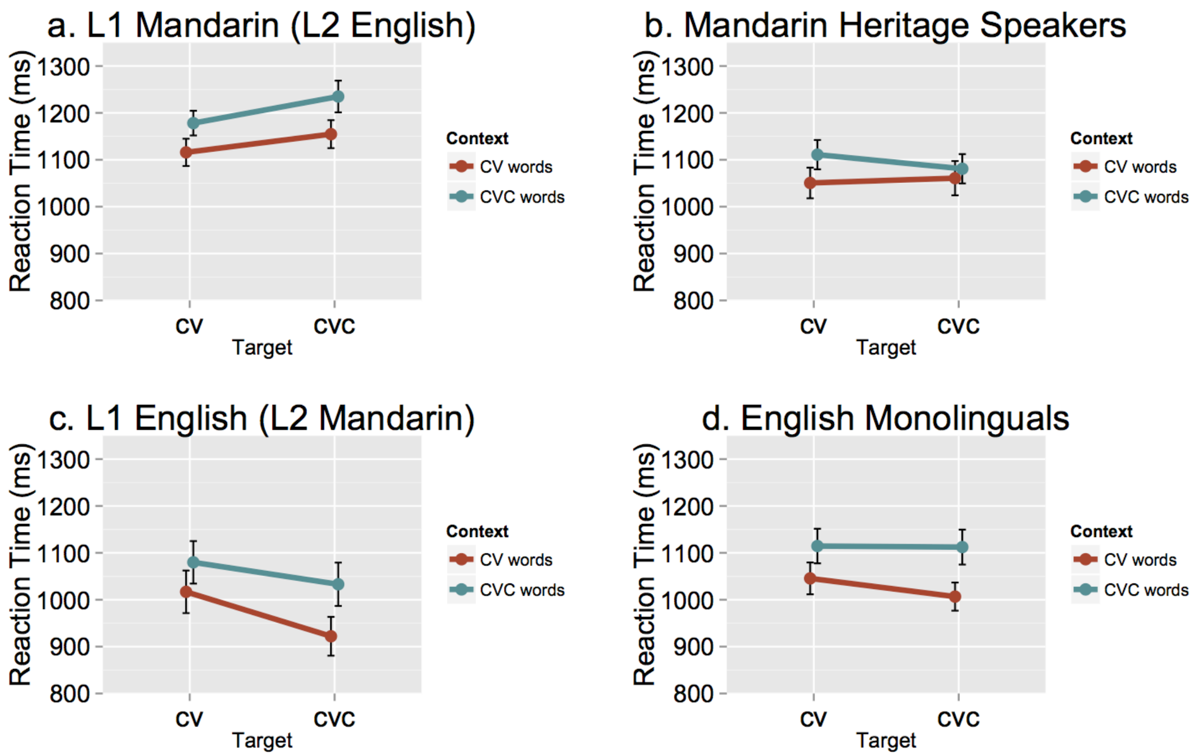

Figure 4: Reaction time in the Liquid condition as a function of target and context for each language group

Results for this condition are shown in Figure 4. Similar to the other English conditions, No significant target by context interaction or target by context by group interactions were found. A significant main effect of context ($\beta = 74.060$, $t = 3.17$, $p < .01$) indicated that subjects responded more quickly to CV words than CVC words. A significant target by group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English) interaction ($\beta = 72.491$, $t = 2.15$, $p < .05$) indicates a tendency toward slower responses for CV targets in the L1 English L2 Mandarin group and a tendency toward faster responses to CV targets in the L1 Mandarin L2 English group.

2.3.2 Mandarin Condition

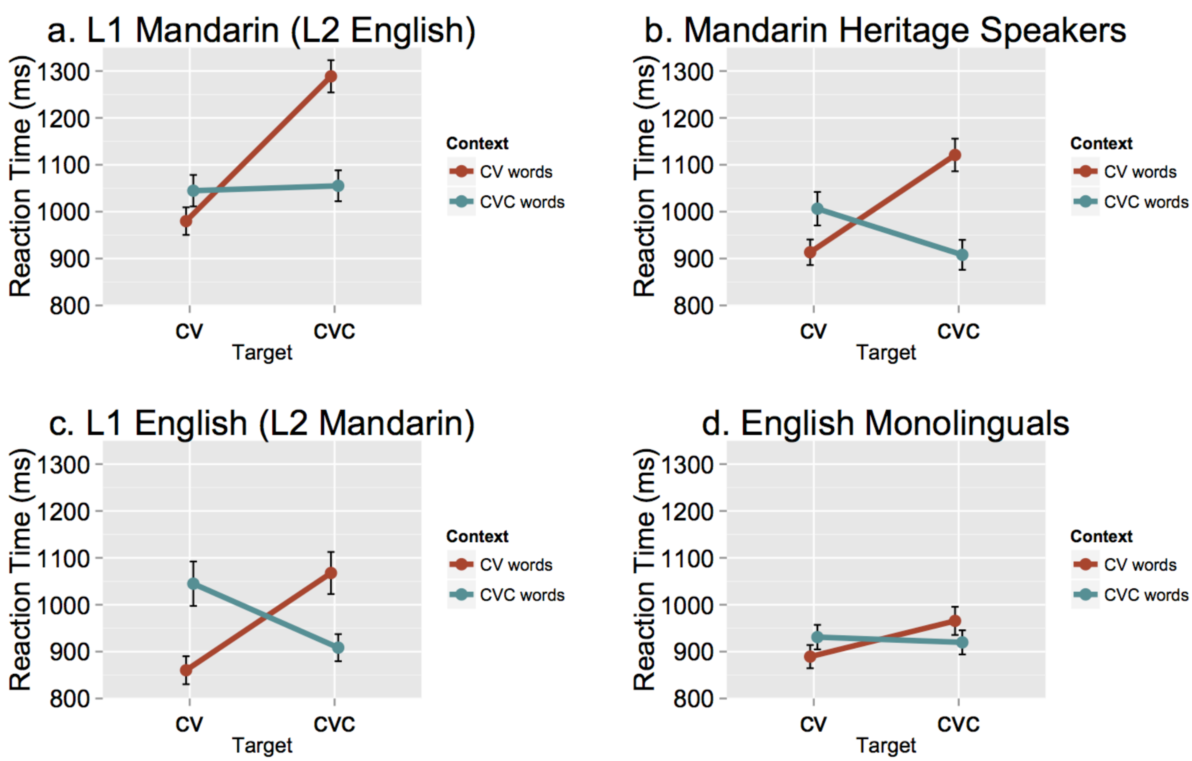

Figure 5: Reaction time as a function of target and context in the Mandarin condition for each language group

The results for this condition can be seen in Figure 5. A significant main effect of target ($\beta = 73.35$, $t = 3.59$, $p <.01$) indicated that CV targets were found more quickly overall than CVC targets. A main effect of language group (L1 Mandarin L2 English vs. Mandarin heritage speakers) was marginally significant ($\beta = -99.96$, $t = -1.90$, $p = .054$), indicating that the L1 Mandarin L2 English group responded more slowly than the Mandarin heritage speakers.

Several two-way interactions also reached significance. The interaction of target and context ($\beta = -276.42$, $t = -9.39$, $p <.001$) indicated that overall, speakers responded more quickly when the target and context matched (e.g. faster responses for the CV target in a CV context). This suggests that subjects applied a syllable-based segmentation strategy to these stimuli. Also significant were a group by target interaction (L1 Mandarin L2 English vs. Mandarin heritage speakers, $\beta = -104.56$, $t = -2.47$, $p < .05$) and a group by context interaction (English monolinguals vs. all other groups, $\beta = -69.82$, $t = -2.33$, $p < .05$). An additional group by target interaction (English monolinguals vs. all other groups, $\beta = 56.41$, $t = 1.64$, $p = .096$) and group by context interaction (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English, $\beta = -46.98$, $t = -1.77$, $p = .072$) were marginally significant.

A significant three-way interaction of target, context, and group (English monolinguals vs. the three bilingual groups, $\beta = -238.47$, $t = -4.01$, $p < .001$) indicated that the target by context interaction was larger in the three bilingual groups than in the English monolingual group.

In order to examine these interactions in more detail, individual linear mixed-effects models were constructed for each group in addition to the full model for this condition. These models included fixed effects of target and context as well as the interaction of these two variables. They also included a maximal random effects structure, which consisted of random intercepts by subject and item as well as random slopes of target, context, and their interaction by subject and random slopes of target and context by item.

The three-way target by context by group interaction indicated that the interaction of target and context differed across groups. In the individual group linear mixed-effects models, all three groups with Mandarin experience (L1 Mandarin L2 English ($\beta = -293.48$, $t = -5.16$, $p < .001$), Mandarin Heritage Speakers ($\beta = -316.19$, $t = -5.60$, $p < .001$), and L1 English L2 Mandarin ($\beta = -391.46$, $t = -4.16$, $p < .01$)) showed a significant interaction of target and context (figures 5a, 5b, and 5c). A trend toward a similar interaction is visible in the English monolingual data (figure 5d), but this was not significant ($\beta = -84.15$, $t = -1.660$ $p = .1062$). These interactions indicate that groups with Mandarin experience tend to segment speech at syllable boundaries. The non-significant effect in English monolinguals suggests that acoustic cues to syllable boundaries may be present, but that experience increases sensitivity to these cues.

The L1 Mandarin L2 English group ($\beta = 157.54$, $t = 3.989$, $p < .01$) and the Mandarin heritage speakers ($\beta = 54.02$, $t = 1.832$, $p = .080$) responded more slowly to CVC targets than CV targets, although the effect was marginally significant in the Mandarin heritage speakers. The fact that this effect was larger in the L1 Mandarin L2 English group accounts for the significant group by target interaction in the full model. The L1 Mandarin L2 English group ($\beta = -86.01$, $t = -1.782$, $p < .05$) and the Mandarin heritage speakers ($\beta = -56.61$, $t = -2.08$, $p = .054$) also both responded more slowly to CV words than CVC words while the other two groups did not. This can account for the group by context interactions found in the main model.

In the L1 Mandarin L2 English group, the target by context interaction, along with the large main effect of target, reflects a pattern of responses in which a target-context mismatch slows down response times only for CV words (figure 4a). This contrasts with the L1 English L2 Mandarin group, which shows a significant target by context interaction, but no main effects, suggesting a similar effect of target-context mismatch in both CV and CVC words (figure 4c). The smaller target effect in the Mandarin heritage speakers suggests a pattern intermediate between the other two groups (figure 4b).

2.4 Discussion

2.4.1 Overview

One major finding of this study is that all groups other than English monolinguals showed evidence of syllable-based segmentation strategies in Mandarin while no groups showed evidence of syllable-based segmentation strategies in English. This supports the variable strategy hypothesis. Even for syllables with Mandarin-like phonotactics, L1 Mandarin speakers did not apply a syllable-based segmentation strategy to English stimuli. Additionally, L1 Mandarin speakers did not apply a purely syllable-based strategy to Mandarin stimuli. Rather, they applied a strategy that may take into account knowledge of the Mandarin lexicon and/or relative frequencies of open and closed syllable types in Mandarin. L2 Mandarin speakers did not fully adopt this strategy, but instead adopted a more purely syllable-based segmentation strategy. Each of these points is discussed in more detail below.

2.4.2 Dominant language strategies in English and Mandarin

These results indicate that English segmentation strategies by L1 English speakers and Mandarin segmentation strategies by L1 Mandarin speakers are very different. Figure 5a shows that Mandarin speakers are highly sensitive to syllable structure when hearing CV words, indicated by the strong positive slope of the red line. When hearing a word like “xi1-nan2”, Mandarin speakers were much faster to detect the sequence “xi” than “xin”. This difference in response times between CV and CVC targets in the CV context is quite large, approximately 300ms (compared to a difference of approximately 20-30ms found in French in Mehler et al. (1981)). One possible explanation for this difference in effect sizes is that Mandarin syllables typically correspond to a morpheme, so speakers may have a reason in addition to acoustic cues or predictable syllable shapes to draw a strong boundary (possibly similar to a word boundary) between syllables that is not present in languages without such a strong syllable-morpheme correspondence (e.g. French).

Figures 1-4d show that native English speakers were not sensitive to correspondence between syllable structure in the target and context when hearing words with a similar structure in English. When hearing a word like “pe-nny” English speakers detected the sequence “pe-” no faster than "pen-". This is in line with results for liquid consonants in previous work (Cutler et al., 1986, 1992) and expands this finding to syllables ending with several different consonant types.

However, in words with initial CVC syllables (e.g. "xin1-sheng1" or "pen-cil"), neither group differed in reaction time across target sequences. Speakers of both languages were just as fast to detect "xi-" or "pe-" as "xin-" or "pen-" in these words. This is unexpected for the Mandarin speakers because a purely syllable-based segmentation strategy would predict that extracting a CV sequence from a CVC syllable should take longer than extracting a CVC sequence (e.g. extracting "xi-" from xin1sheng1 should take longer than extracting /xin/ from xin1sheng1). Possible reasons for this response pattern are discussed in the following section.

2.4.3 Possible reasons for lack of syllable effect in L1 Mandarin speakers for CVC context

One possible explanation is that Mandarin contains substantially more CV than CVC syllables (about 75% CV syllables, according to Tseng (2005)), which might make a segmentation strategy that first looks for CV syllables advantageous. If this is the case, we would expect a similar result in other languages with a high proportion of open syllables. Under 10% of Japanese syllables have a coda (Tamaoka & Makioka, 2004), making it a possible candidate to test this hypothesis. However, previous work with the fragment detection task in Japanese found a different pattern of results indicating a mora-based segmentation strategy (see Otake et al. (1993) for a full explanation of these results). A language with a high prevalence of open syllables, but without such a fundamental role played by the mora would be a good test case for the hypothesis that listeners will tend to segment speech into CV sequences when CV is the predominant syllable type.

On a related note, the CV sequences embedded in the CVC syllables all correspond to existing morphemes, and often several morphemes (e.g. xi- from xin1 (*new*) can mean xi1 (*west*), xi1 (*to inhale*), xi1 (*stream*), xi3 (*to wash*), xi4 (*department*), and many others). These competing morphemes may be highly active, causing the corresponding phoneme sequence to be easily noticed. L2 Mandarin speakers with smaller vocabularies may not necessarily experience the same degree of activation, which may explain why other groups did not show this pattern in Mandarin. A systematic examination of this phenomenon would need to take into account the number of morphemes corresponding to each syllable, as well as their frequency. Additionally, whether or not these morphemes match the target in tone and whether they correspond to independent words may affect the degree of activation. Such an analysis is beyond the scope of the present study, which only includes items beginning with seven CVC syllables in the Mandarin condition.

2.4.4 Flexibility of strategies across task language

These results also suggest that bilinguals with varying levels of experience in English and Mandarin are able to adopt segmentation strategies appropriate for the task language, even if it is not their dominant language. In English conditions, Mandarin-dominant speakers did not apply the dominant-language strategy of syllable-based segmentation (figures 1-4a). French-dominant listeners with very high English proficiency were able to suppress the French syllable-based strategy when listening to English stimuli in Cutler et. al (1992), but listeners from Japanese and French language backgrounds with at least some degree of English experience have been shown to apply dominant-language segmentation strategies when segmenting English speech (Cutler et al., 1986, Cutler & Otake, 1994), so this result is surprising because these strongly L1 dominant Mandarin speakers apparently abandoned their dominant language strategy when presented with English stimuli. Additionally, this occurred for stimuli in all four English conditions, including those designed to be more similar to Mandarin. This suggests that it was the fact that the stimuli were in English, rather than the presence of less familiar syllable structures that led L1 Mandarin speakers away from a syllable-based segmentation strategy in the English conditions.

Native English speakers who have studied Mandarin readily switched to a syllable-based strategy when listening to Mandarin. However, this switch in strategies does not occur without some familiarity with Mandarin. This suggests that sensitivity to the acoustic cues responsible for this pattern only develops with language experience. The amount of experience required appears to be quite modest, as Mandarin learners in this group only had an average of 4.4 years of experience. This is in sharp contrast to findings with English-dominant speakers of French, who did not adopt a syllable-based segmentation strategy, even when they had native or near-native knowledge of French (Cutler et. al., 1992).

Why would L1 English speakers switch to a syllable-based segmentation strategy so easily in Mandarin when Cutler et al. (1992) found that they do not in French? As suggested in the introduction, the highly restricted syllable structure of Mandarin may make the location of syllable boundaries an easier task than it would be in French. A limited inventory of codas and lack of consonant clusters means that syllable boundaries may be located quite accurately, even with limited lexical knowledge. If this leads L1 English speakers to a syllable-based segmentation strategy in Mandarin, we would expect them to adopt a syllable-based segmentation strategy in any language with similarly restricted syllable structure. In Otake et al. (1993) monolingual English speakers performed a similar fragment detection task in Japanese, which (like Mandarin) does not allow consonant clusters or non-nasal codas. These English speakers showed no trace of syllable-based segmentation in Japanese. However, this study did not include data from L1 English speakers with Japanese experience, so we do not know if

English speakers might adopt a syllable-based segmentation strategy in Japanese with sufficient experience.

Another possibility is that the L1 English speakers may have interpreted the Mandarin stimuli as containing two stressed syllables. Mandarin does not have lexical stress as English does, but syllables with neutral tone have some properties in common with unstressed syllables, including shorter duration (Chen & Xu, 2006). If L1 English speakers interpret all non-neutral tone syllables as stressed, applying the stress-based segmentation strategy characteristic of English would result in a pattern of responses much like that of a syllable-based strategy. However, in this case, English speakers would not be expected to respond more quickly to targets that match the initial syllable of a word if that word's second syllable had a neutral tone. An experiment investigating segmentation of words with a neutral tone on the second syllable by L1 English speakers of L2 Mandarin could determine if this is the case.

2.4.5 Concluding remarks

The results of this experiment demonstrate that bilinguals who speak both Mandarin and English are able to change segmentation strategies when speaking their non-dominant language. This suggests a flexibility in speech segmentation strategies that has not been observed in previous studies. Possible factors driving this result include the differences in syllable structure and syllable-morpheme correspondence between English and Mandarin as well as listeners' mapping between the tone and stress systems of these two languages. These differences appear to create an environment in which it is highly useful for listeners to segment speech with different strategies and listeners' perceptual systems are able to adjust accordingly.

However, L1 English speakers do not adopt exactly the same strategy in Mandarin as L1 Mandarin speakers. The L1 Mandarin strategy is somewhat different than a purely syllable-based segmentation strategy. It may take into account additional properties of Mandarin such as syllable-type frequency and/or the prevalence of homophony that occurs due to Mandarin's monosyllabic morphemes and restricted syllable inventory. While these results support the variable strategy hypothesis, they also underscore the fact that speech segmentation is a complex problem with multiple possible solutions and that L1 segmentation strategies may be finely tuned to properties of the language.

Endnotes:

1. Consonants located between two vowels in English may be classified as ambisyllabic (potentially belonging to two syllables) under some analyses. For the sake of simplicity, we describe words such as "penny" as having the structure CV-CV, in contrast to words such as "pencil", which have the structure CVC-CV.
2. The bimorphemic nasal conditions in English and Mandarin differed somewhat in structure due to unavoidable differences between the languages. In English, the initial morpheme consisted of a CVC sequence and was held constant while the second morpheme started with either a consonant (CVC context) or a vowel (CV context) (e.g. su-nny, Sun-day). In Mandarin, the words consisted of an initial CV morpheme followed by a second morpheme beginning with a consonant or an initial CVC morpheme followed by a second morpheme beginning with a consonant.
3. Each pair of words beginning with the same initial phoneme were considered one item. For example, words beginning with pen- (penny and pencil) were coded as the same item.

3 Phonological Units of Speech Production in Mandarin-English Bilinguals

3.1 Introduction

3.1.1 Phonological Encoding and Proximate Units

Phonological encoding is an important step in the planning of speech. It may be defined as the process by which a speaker retrieves the sound form of a word. In accounts of phonological encoding based on Indo-European languages, the selection of phonemes is a necessary step in the selection syllables. For example, in the well-known model of speech production proposed by Levelt, Roelofs, and Myer (1999), activation of phonemes precedes the activation of syllables. However, more recent research suggests that in some languages, units other than phonemes may fill this role. In particular, the syllable has been proposed as a basic unit in Mandarin phonological encoding by O'Seaghdha, Chen, & Chen (2010). O'Seaghdha et al. (2010), refer to the first selectable phonological units below the level of the word as *proximate units* and propose that phonemes are the proximate units in English while syllables are the proximate units in Mandarin.

3.1.2 Implicit Priming

One important line of evidence for these proposals comes from implicit priming experiments. In these experiments, speakers produce a spoken word in response to a prompt, which may take a variety of forms, such as a written word, a picture, or a symbol. These responses are typically produced in sets of three or four, which may either share some phonetic material (homogeneous context) or not (heterogeneous context). If subjects respond more quickly in homogeneous sets than heterogeneous sets, this may be considered evidence that speakers were able to "pre-plan" the shared material, which is predictable within the set. This might suggest that the speakers' proximate unit is contained within the shared material. For example, if speakers produce responses more quickly in a homogeneous set in which all items begin with the phoneme /t/ than in a heterogeneous set, this might suggest that phonemes serve as a proximate unit for that group.

English (O'Seaghdha et al., 2010) and Dutch (Myer, 1991) speakers do respond more quickly in sets that share an initial phoneme than in sets that share no initial phonetic material. However, for Mandarin speakers, response times to sets with a shared initial phoneme are no faster than for sets with no shared phonetic material (O'Seaghdha et al., 2010; Chen, Chen, & Dell, 2002; Chen & Chen, 2013). However, in these experiments, Mandarin speakers did respond more quickly to sets with a shared initial syllable as compared to heterogeneous sets. This suggests that syllables, rather than phonemes, may be the proximate units for Mandarin.

3.1.3 Points of Uncertainty

Complicating this picture slightly are studies which have found effects of shared sub-syllabic material in Mandarin production. One study found that a shared CV sequence was sufficient to speed reaction times for native Mandarin speakers in a masked priming task (Verdonschot, Lai, Chen, Tamaoka, & Schiller, 2015). In this study, participants produced responses more quickly when reading characters for words preceded by a prime sharing an initial CV sequence, even when the initial syllable was not the same (e.g. *ba1* (eight) might be produced more quickly when primed by *ban1* (class) than when proceeded by an phonologically unrelated prime). In another study, native Mandarin speakers exhibited a different ERP response, but no difference in reaction time, when preparing to produce adjective-noun pairs that shared an initial phoneme versus pairs that did not share initial phonetic material (Qu, Damian, & Kazanina, 2012). However, both of these studies replicated the finding (from Chen, Chen, & Dell, 2002 and O'Seaghdha et al., 2010) that a shared initial phoneme did not speed reaction times for Mandarin speakers. These studies suggest that phonemes do play a role in Mandarin phonological encoding, but it that it may be different than their role in English. For example, they might be specified at a different stage in the process of phonological encoding.

3.1.4 Proximate Units in Bilinguals

The observed cross-linguistic differences in phonological encoding processes raise an important question regarding bilinguals. Do bilinguals whose languages have different proximate units plan speech in L1 units for both languages, or do they switch strategies between languages? Under one view, proximate units are something that is set at the level of the speaker and cannot be changed across languages. We will call this the fixed-strategy hypothesis. This would predict that speakers will always prepare speech with proximate units from the L1, even if proficiency in the L2 is very high. Under a different view, proximate units could be adopted by a speaker based on the demands of the task language. We will call this the variable-strategy hypothesis. This would predict that, given sufficient proficiency in L2, speakers may adopt the proximate units of the L2.

In the domain of speech segmentation, previous work has suggested that bilinguals do not easily switch strategies across languages. French listeners tend to segment speech at syllable boundaries (Cutler, Mehler, Norris, & Segui, 1986) while English listeners tend to segment speech at the onset of stressed syllables (Cutler & Norris, 1988). Cutler, Mehler, Norris, & Segui (1992) investigated speech segmentation strategies in French-English bilinguals. Despite very high proficiency in both English and French, these speakers tended to pattern into a French-dominant group, who favored a syllable-based segmentation strategy and an English dominant group who favored a stress-based segmentation strategy. These results suggest that bilinguals may have access to segmentation strategies from only one of their languages. If the restricted strategy hypothesis applies generally to both perception and production strategies, this would predict that individuals would apply the same production strategy to both Mandarin and English words.

In contrast, one recent study provided preliminary support for the variable strategy hypothesis for speech production in Mandarin English bilinguals (Verdonschot, Nakayama, Zhang, Tamaoka, & Schiller, 2013). In this study, L1 Mandarin speakers with a high level of English proficiency read aloud words which were preceded by a masked prime. The prime was presented briefly, followed by a masker (in this case the symbols ##), followed by the target word. In English, these speakers responded more quickly when the masker shared a single initial phoneme with the target word. Onset priming of this kind was also observed for some Mandarin stimuli, but only when the prime and target shared the same syllable structure (e.g. *bi1* primed *ba1*, but not *ban1*). The reason syllable structure affected this result is unclear. This could be taken as evidence that the speakers were able to adopt the phoneme as the proximate unit in English and that this may have even influenced the choice of proximate unit in their L1 Mandarin. However, this study did have some limitations, particularly the use of orthographic stimuli. In the English stimuli, the primes with a shared initial phoneme also shared orthographic material (the initial letter). This study also lacked comparison to an English monolingual control group and comparison between groups with differing levels of English proficiency.

3.1.5 The Present Study

The present study avoids the confound of orthographic similarity in English primes by presenting stimuli as auditory recordings, rather than in orthographic form. Moreover, a critical innovation of this study is the inclusion of groups with a range of experience in Mandarin and

English. These groups include English monolingual controls (English-dominant, no Mandarin experience), L1 English learners of Mandarin (L1 English, L2 Mandarin, English-dominant), Mandarin Heritage speakers (L1 Mandarin, L2 English, English-dominant), and L1 Mandarin learners of English (L1 Mandarin, L2 English, Mandarin-dominant). For a fully crossed design, we would also need Mandarin monolinguals and English heritage speakers (dominant in L2 Mandarin). However, these groups were not available in the participant population because the study was conducted at an American university. Table 8 depicts the groups included in this study, with groups not included in this study in grey.

Table 8: L1, L2, and dominant language for each group included in this study

Group	**L1**	**L2**	**Dominant Language**
L1 English L2 Mandarin	English	Mandarin	English
(English heritage speakers)	English	Mandarin	Mandarin
English monolinguals	English	None	English
L1 Mandarin L2 English	Mandarin	English	Mandarin
Mandarin heritage speakers	Mandarin	English	English
(Mandarin monolinguals)	Mandarin	None	Mandarin

If the restricted strategy hypothesis applies to speech production strategies, a speaker's production strategy could be determined by L1 or by dominant language. If L1 determines a speaker's strategy, we would expect any group with Mandarin as the L1 (L1 Mandarin L2 English group and Mandarin heritage speakers) to produce speech with a syllable-based strategy and any group with L1 English (English monolinguals, L1 English L2 Mandarin group) to produce speech with a phoneme based strategy. Similarly, if dominant language determines a speaker's strategy, we expect a syllable-based strategy for all Mandarin-dominant groups (L1 Mandarin L2 English group) and a phoneme-based strategy for all English dominant groups (English monolinguals, L1 English L2 Mandarin group, Mandarin heritage speakers).

If the variable strategy hypothesis applies to speech production strategies, we would expect that some groups might plan speech with a phoneme-based strategy in English and a syllable-based strategy in Mandarin. Which groups would show this pattern depends on the degree of L2 experience required to adopt a new speech planning strategy. If relatively little L2 experience is required, all bilingual groups (L1 Mandarin L2 English, L1 English L2 Mandarin, Mandarin heritage speakers) might adopt a phoneme-based strategy in English and a syllable-based strategy in Mandarin. If a very high degree of L2 experience is required, we might expect that only the Mandarin heritage speakers would show this pattern of results.

3.2 Experiment

3.2.1 Participants

Participants were divided into four groups: L1 Mandarin L2 English, L1 English L2 Mandarin, Mandarin heritage speakers, and English monolingual controls. Demographic information for each group is listed in Table 9. All participants gave informed consent and were paid $10 per hour or given course credit.

Table 9: Demographic information for participants

Participant groups	**Age**	**Sex**	**Years of English Study**	**Years of Mandarin Study**	**number**
L1 Mandarin L2 English	22.88	7 female 10 male	10.53	From birth	English condition: 15 Mandarin condition: 11
Mandarin heritage speakers	18.85	16 female 3 male	16.6	From birth	English condition: 14 Mandarin condition: 15
L1 English L2 Mandarin	20.55	7 female 4 male	From birth	4.36	English condition: 9 Mandarin condition: 10
English monolinguals	19.25	7 female 5 male	From birth	none	English condition: 12

3.2.2 Materials

Materials consisted of 36 picture-word pairs, 18 in English and 18 in Mandarin. The 18 words in each language were further divided into 6 groups of 3. Three of these groups shared an initial phoneme and three shared an initial syllable. These groups of three formed the homogeneous phoneme and homogeneous syllable conditions. Heterogeneous sets were formed by taking one item from each of the homogeneous phoneme sets or the homogeneous syllable sets. The items are all listed in Table 10 (English) and Table 11 (Mandarin).

Table 10: English stimuli

	Phoneme			**Syllable**		
	Homogeneous			Homogeneous		
	/t/	/m/	/h/	tur-	mus-	ham-
heterogeneous	tiger	mushroom	horseshoe	turnip	musket	hamlet
	table	mountain	helmet	turkey	mustang	hamper
	teapot	music	hippo	turtle	mustard	hamster

Table 11: Mandarin stimuli (from Chen & Chen (2013))

	Phoneme			**Syllable**		
	Homogeneous			Homogeneous		
	/t/	/m/	/ʂ/	tu-	mao-	shan-
heterogeneous	tuo1ba3 "mop" 拖把	mei2gui1 "rose" 玫瑰	shu4ye4 "leaf" 樹葉	tu1ying1 "vulture" 禿鷹	mao2yi1 "sweater" 毛衣	shan4zi "hand fan" 扇子
	tie3chui2 "hammer" 鐵鎚	mi4feng1 "bee" 蜜蜂	sha1fa1 "sofa" 沙發	tu3si1 "toast" 吐司	mao4zi5 "hat" 帽子	shan1yang2 "goat" 山羊
	tai4yang2 "sun" 太陽	ma1zu3 "Mazu" 媽祖	shou3tao4 "gloves" 手套	tu4zi "rabbit" 兔子	mao1mi1 "cat" 貓咪	shan3dian4 "lightning" 閃電

3.2.3 Procedure

This experiment used an implicit priming paradigm, based on that of Chen & Chen (2013). Each subject completed the task in English and Mandarin on separate days, with the non-dominant language occurring on the first day. This was done to minimize the influence of the dominant language on performance in the non-dominant language. The L1 Mandarin L2 English group completed the task in English on the first day while the Mandarin Heritage speakers and the L1 English L2 Mandarin group completed the Mandarin task on the first day. English monolinguals completed only the English task.

Following the design of Chen & Chen (2013), the task for each language was divided into two halves, one composed of items from the phoneme condition and one composed of items from the syllable condition. Half of subjects completed the phoneme condition first and half completed the syllable condition first. Each half of the experiment was further subdivided into three blocks, with each block consisting of the six sets in that condition (three homogeneous and the three heterogeneous sets) in random order. Within each set, the items were repeated six times in random order. This resulted in a total of 648 trials per subject in each language, consisting of 2 conditions (syllable, segment) × 2 contexts (homogeneous, heterogeneous) × 3 sets × 3 blocks × 3 items × 6 repetitions.

Before each block, participants were presented with a training phase, the purpose of which was to familiarize participants with the task. For this training, the three pictures for that block were presented along with recordings of the associated word. After the presentation of the picture-word pairs, participants were given the option to repeat the training phase or to continue to the test phase. This choice was indicated with a button press. During the test phase, each trial began with a 1000Hz warning tone lasting for 200ms during which dashed lines flanking a blank space (-- --) were displayed in the middle of the screen followed by 800ms of silence and a blank screen. Next, a picture was displayed for 2000ms, during which time participants voices were recorded as they named the picture. After a blank interval of 200ms, the next trail began.

Participants completed one practice block before the trail, which contained three items not used in the main experiment. Participant responses were recorded with a head-mounted microphone in a sound attenuated booth. The experiment was presented with the SuperLab software package. Each experimental session lasted approximately 1 hour.

3.2.4 Analysis

Response time was measured as the time from the appearance of the picture to the onset of speech. This measurement was performed in PRAAT (Boersma & Weenik, 2014) with a script that marked the onset of speech as the point where the recording exceeded a volume threshold.

This threshold was varied by subject and ranged from 40 to 50 db. Output from the script was corrected when the onset was marked as a non-speech sound preceding actual speech (e.g. background noise or a lip smack), or when the onset was marked at a point more than halfway through the initial fricative (for fricative onsets) or burst (for stop onsets). When corrected by hand, the onsets were placed at the beginning of frication for fricative onsets, the beginning of the burst for stop onsets, and at the onset of voicing for nasal onsets.

Trials in which the subject did not respond, the response was incorrect, or the response was disrupted by non-speech sounds (e.g. coughs or yawns) as well as false starts (in which a participant produced a partial response preceding a correct response) were excluded from analysis. For all groups, fewer than 10% of responses were excluded for any of these reasons. For L2 speakers, a response was considered correct if it began with the correct phoneme (for the phoneme condition) or syllable (for the syllable condition) and was not one of the other items in the experiment. For example an L2 English L1 Mandarin speaker who produced “musket” as /mʌski/ would be considered correct.

3.3 Results

3.3.1 Statistical models

Data for the three bilingual groups (L1 Mandarin L2 English, Mandarin heritage speakers, and L1 English L2 Mandarin) was submitted to a linear mixed-effects analysis. Fixed effects included task language (English vs. Mandarin), condition (phoneme vs. syllable), context (homogeneous vs. heterogeneous), and group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English, Mandarin Heritage speakers vs. L1 Mandarin L1 English) as well as interactions of these variables. A maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) was included in the model. This consisted of random intercepts for subjects and items as well as random slopes of condition, context and their interaction by subject and random slopes of context by item. Significant and marginally significant effects in this model are indicated in Table 12.

To further investigate the interactions from the full model, data for each language group was submitted to a separate linear mixed-effects analyses. Fixed effects included task language (English, Mandarin), condition (phoneme, syllable), and context (homogeneous, heterogeneous) as well as the interactions of these three variables. A maximal random effects structure was included in the model. This consisted of random intercepts for subjects and items as well as random slopes of context by item and by subject.

For groups in which significant interactions involving language were present, the data was submitted to separate linear mixed-effects models for each task language (English and Mandarin). These included fixed effects of condition (phoneme, syllable), and context (homogeneous, heterogeneous) as well as the interactions of these two variables. A maximal random effects structure was also included. This consisted of random intercepts for subjects and items as well as random slopes of context by item and subject. Because the English monolingual group completed only the English task, only this analysis was performed for this group. All reported p-values were obtained through model comparison.

3.3.2 Full Model

Table 12: Significant and marginally significant effects in the full model including condition, context, language, and group

Effect	**β**	**t**	**p**
Context	0.008	1.90	< .05
Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	-0.040	-2.17	< .05
Condition x Context	0.027	2.99	< .01
Context x Language	0.033	4.44	< .0001
Condition x Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	0.025	1.96	<.05
Language x Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	0.020	6.11	< .0001
Language x Group (Mandarin heritage speakers vs. L1 Mandarin L2 English)	0.065	11.87	< .0001
Condition x Context x Language	0.033	2.22	< .05
Condition x Language x Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	0.027	4.17	< .0001
Condition x Language x Group (Mandarin heritage speakers vs. L1 Mandarin L2 English)	0.095	8.96	< .0001
Context x Language x Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	0.018	3.01	< .01
Condition x Context x Group (Mandarin heritage speakers vs. L1 Mandarin L2 English)	0.025	1.70	= .084
Condition x Context x Language x Group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English)	0.024	1.96	= .054

3.3.2.1 Main effects

The main effect of context indicates that, in general, words in the homogeneous context were produced more quickly than those in the heterogeneous context. Given that the context manipulation was included to facilitate faster production times through shared phonetic material, this result is expected. The main effect of group indicates that the L1 English L2 Mandarin group was slower overall than the Mandarin heritage speakers or the L1 Mandarin L2 English group.

3.3.2.2 Two-way interactions

The condition by context interaction indicates that the benefit to reaction time for the homogeneous context was greater for the syllable condition than the phoneme condition. Given that sets the homogeneous syllable condition have more phonetic overlap than sets in the homogeneous phoneme condition and that Mandarin speakers in previous work benefited from a shared syllable but not a shared phoneme, this is not surprising. The context by language interaction indicates that benefit to reaction times from homogenous sets is greater in English than in Mandarin. As we will discuss further in the following sections, reaction times in the homogeneous context in Mandarin were actually sometimes slower than reaction times in the heterogeneous context. The condition by group interaction (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English) indicates that the L1 English L2 Mandarin group showed a smaller difference in reaction time between the syllable and phoneme conditions than the other two groups. The two significant language by group interactions indicate that 1) the L1 English L2 Mandarin group differed less in reaction time across languages than the other two groups and 2) Mandarin heritage speakers differed less in reaction time across languages than the L1 Mandarin L2 English group.

3.3.2.3 Three-way interactions

The condition by context by language interaction indicates that the interaction of condition and context differs across languages. The two condition by language by group interactions indicate that the difference in reaction times between the syllable and phoneme conditions was 1) less different between English and Mandarin in the L1 English L2 Mandarin group than the other two groups and 2) less different between English and Mandarin in Mandarin Heritage speakers than in the L1 Mandarin L2 English group. The context by language by group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English) interaction indicates that the L1 English L2 Mandarin group benefited from the homogeneous context in both languages while the Mandarin heritage speakers and L1 Mandarin L2 English group benefited from the homogeneous context only in English. The marginally significant condition by context by group (Mandarin heritage speakers vs. L1 Mandarin L2 English) interaction suggests that the benefit of the homogenous context in the syllable condition over that of the homogenous context in the phoneme condition may be greater for the L1 Mandarin L2 English group than for the Mandarin heritage speakers. The two-way interactions that contribute to the interactions in this section are discussed in more detail in the individual group sections.

3.3.2.4 Four-way interactions

A four-way interaction between condition, context, language, and group (L1 English L2 Mandarin vs. Mandarin heritage speakers and L1 Mandarin L2 English) approached significance. This reflects the fact that the Mandarin heritage speakers and the L1 Mandarin L2 English group showed three-way interactions of context, condition, and language while the L1 English L2 Mandarin group did not. These three-way interactions and their interpretation are discussed in the individual group sections.

3.3.3 English Monolinguals

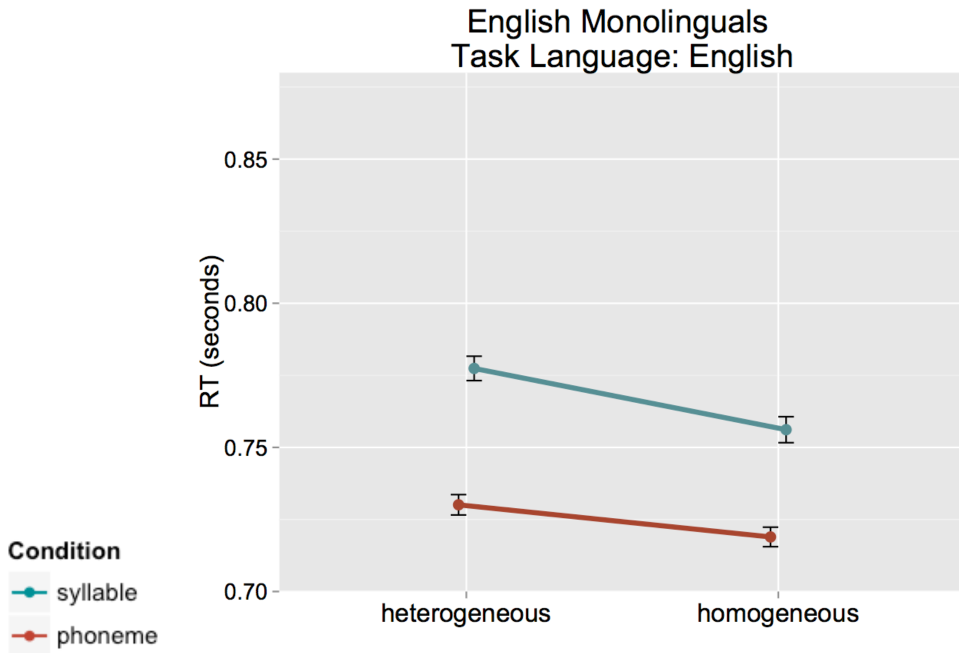

Figure 6: Response times in the English task for English monolinguals in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

The overall pattern of this data indicates a faster response time in the homogeneous context for both the syllable and phoneme conditions, which agrees with previous work showing that a shared initial phoneme is sufficient to speed responses in English (e.g. O'Seaghdha et al., 2010). In the linear mixed-effects model for this group, this main effect of context was marginally significant (β=.0163, t=1.731, p = .0850). Additionally, we found a main effect of condition, with responses in the syllable condition significantly slower than those in the phoneme condition (β=.0421, t=6.497, p<.001). This is most likely because the words in the syllable condition were of lower frequency than the words in the phoneme condition. The interaction of context and condition was not significant. These results are pictured in Figure 6.

3.3.4 L1 Mandarin L2 English

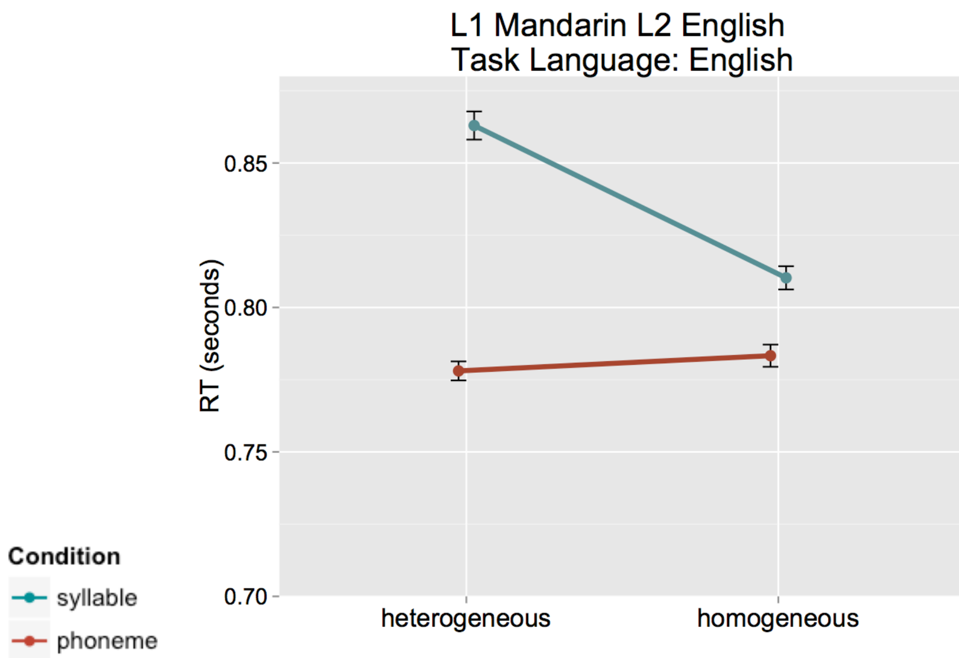

Figure 7: Response times in the English task for the L1 Mandarin L2 English group in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

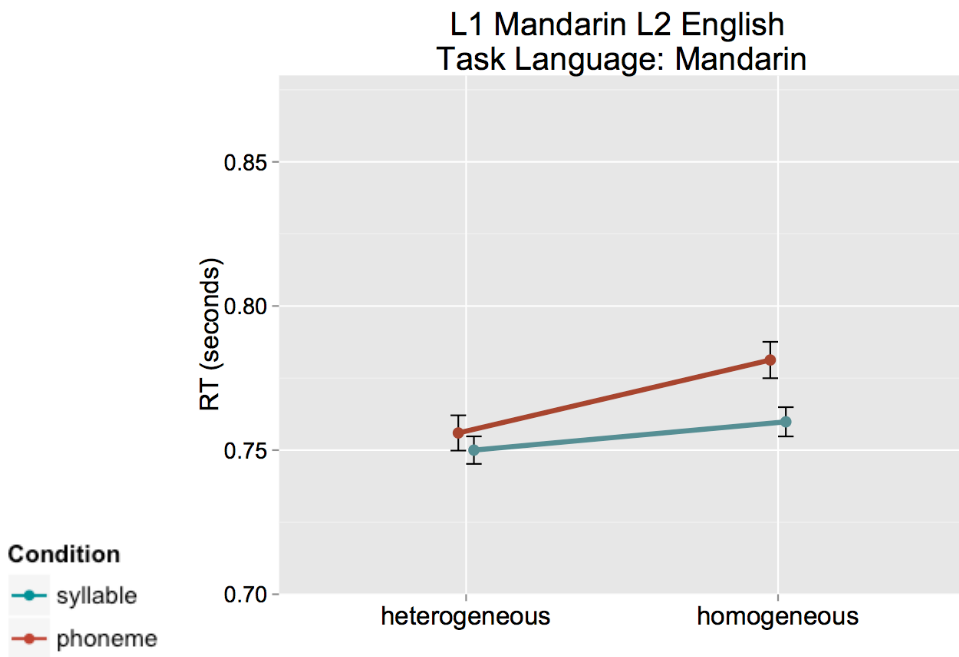

Figure 8: Response times in the Mandarin task for the L1 Mandarin L2 English group in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

3.3.4.1 Overall Results

In the full model for this group, the three-way interaction between condition, context, and language was significant (β=.045, t=2.15, p<.05), indicating that the pattern of interaction between condition and context differed in the two task languages. This result suggests a difference in production strategies across languages, which supports the variable strategy hypothesis. These strategies are examined in the individual models in the following section.

Additional significant effects in this model included main effects of condition (β=.022, t=2.00, p <.05), and language (β=.041, t=3.65, p<.001) as well as interactions of context by language (β=.041, t=3.97, p >.001), context by condition (β=.037, t=3.57, p <.001), and condition by language (β=.070, t=3.17, p<.01). Participants responded more quickly in Mandarin than in English, probably due to higher proficiency in Mandarin than in English. Responses were slower overall for the syllable condition than the phoneme condition in English, leading to the condition by language interaction and the main effect of condition. As with the previous group, this was most likely due to the lower frequency of the English words in the syllable condition. The context by condition interaction reflects the fact that these speakers responded more quickly for sets with a shared initial syllable (at least when the task language was English), but not for those that shared only an onset. These results are pictured in Figures 7 and 8.

3.3.4.2 Results in English

The model for this group in English showed significant main effects of context (β=.0256, t=3.33, p<.01) and condition (β=.0575, t=3.95, p<.001) as well as a significant interaction of these two variables (β=.0610, t=4.66, p<.001). Responses were faster in the phoneme condition and the homogeneous context. The interaction indicates that the benefit of the homogeneous context only occurred for the syllable condition. This pattern is consistent with previous results from this task for Mandarin speakers in Mandarin (O'Seaghdha et al, 2010; Chen, Chen, & Dell, 2002, Chen & Chen, 2013). It suggests that L1 Mandarin speakers may also use a syllable-based planning strategy in English.

3.3.4.3 Results in Mandarin

However, the pattern of results for this group in Mandarin is different from that observed for L1 Mandarin speakers in previous studies. In this group, response times were longer in the homogeneous context than the heterogeneous context, although this trend did not reach significance (β=-.016, t=-1.53, p = .120). This indicates that shared phonetic material did not lead to faster responses. If anything, it may have led to slower responses. Possible reasons for this result will be explored in section 3.6. The effects of condition and its interaction with context were not significant.

3.3.5 Mandarin Heritage Speakers

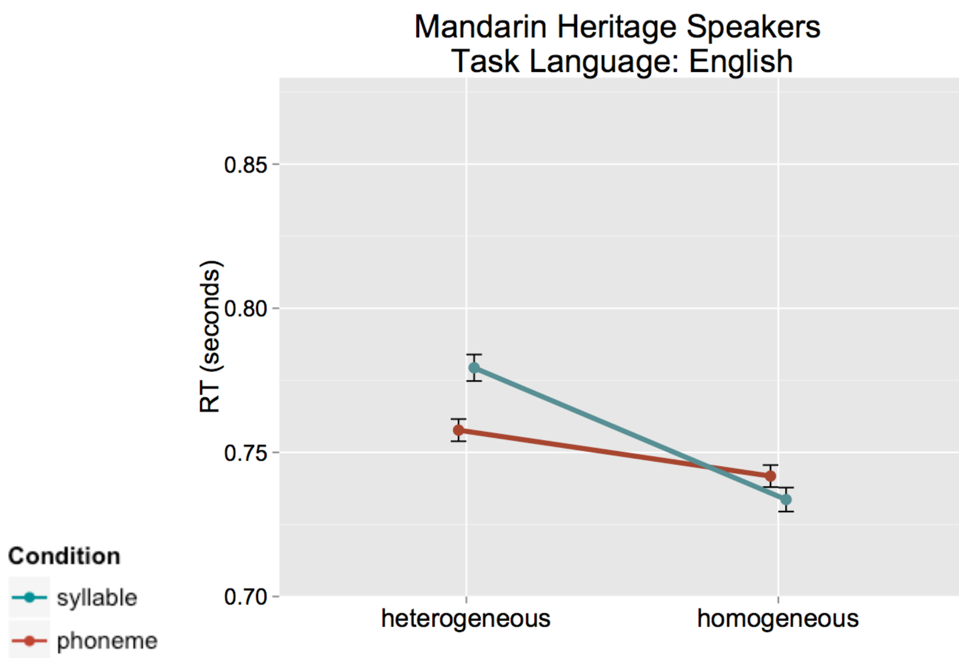

Figure 9: Response times in the English task for the Mandarin heritage speakers in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

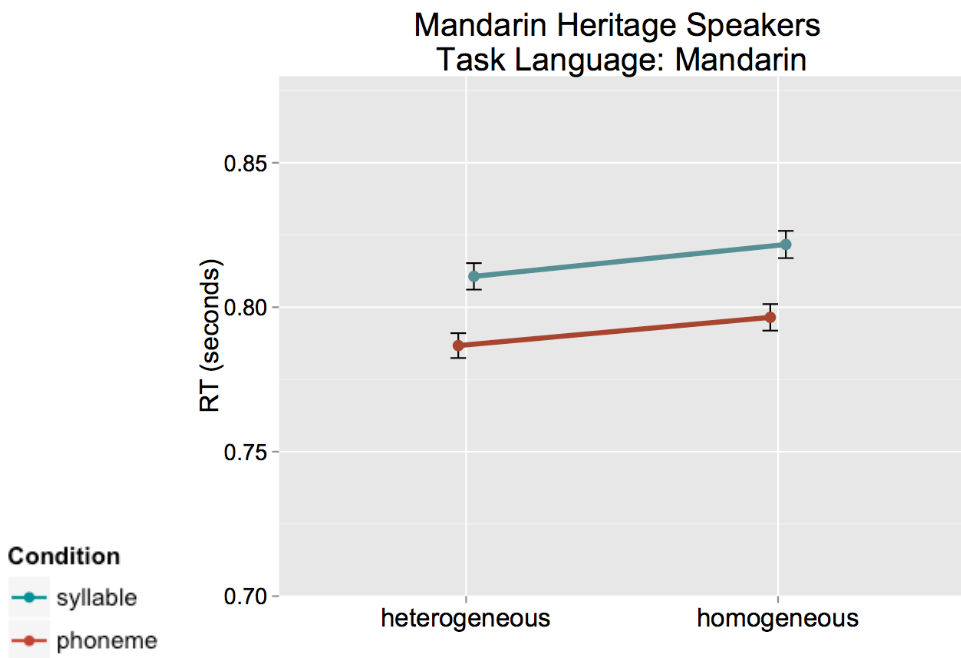

Figure 10: Response times in the English task for the Mandarin heritage speakers in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

3.3.5.1 Overall Results

In the full model for this group, the three-way interaction of condition, context, and language was marginally significant (β=.0321, t=1.8, p =.0641), indicating that the pattern of interaction between condition and context differed in the two task languages. Significant main effects of language (β=-.0234, t=-2.83, p<.01), and condition (β=.0181, t=2.28, p<.05), were also present, with slower responses in Mandarin (likely due to lower proficiency in Mandarin than in English) and in the syllable condition. A significant context by language interaction (β=.4120, t=4.36, p<.001), reflects that shared phonetic material resulted in faster response times in English, but slower response times in Mandarin. These results are pictured in Figures 9 and 10.

3.3.5.2 Results in English

In the English task, this group showed a significant main effect of context (β=.0309, t=3.04, p <.01), indicating faster response times for homogeneous sets than heterogeneous sets. This agrees with previous results for English speakers and with the pattern of results observed in this study for the English monolingual group. These results suggest that this group follows similar speech planning strategies to other English-dominant speaker groups, despite acquiring Mandarin before, or simultaneously with, English. A significant interaction of condition and context (β=.0294, t=2.11, p<.05), suggests that the benefit to response time is greater for a shared syllable than a single shared initial phoneme. This is consistent with a phoneme-based planning strategy because a shared syllable involves multiple shared phonemes. The English monolingual group did not have this interaction, instead showing a similar benefit to response time from a shared initial phoneme or syllable. Despite this difference, both groups show evidence of a phoneme-based planning strategy in English. The main effect of condition was not significant in this analysis.

3.3.5.3 Results in Mandarin

In the Mandarin task, this group showed a marginally significant main effect of condition (β=.0277, t=2.03, p <.05), with response times slower for the syllable condition than the phoneme condition. Similar to the L1 Mandarin L2 English group, the heritage speakers also showed a marginally significant main effect of context (β=-.0121, t=-1.95, p = .0511), with slower responses in the homogeneous context. Possible reasons for this pattern of results will be discussed in section 4. The interaction of condition and context was not significant in this analysis.

3.3.5.4 Comparison with other groups

As indicated by these individual language models, a shared initial syllable or onset led to faster responses in English, but this shared phonetic material led to slower responses in Mandarin. Notably, the pattern of results in English was similar to that of the English monolingual group and the pattern of results in Mandarin was similar to that of the L1 Mandarin group. This result suggests that the heritage speakers showed L1-like production strategies in both of their languages, which supports the variable strategy hypothesis.

3.3.6 L1 English L2 Mandarin

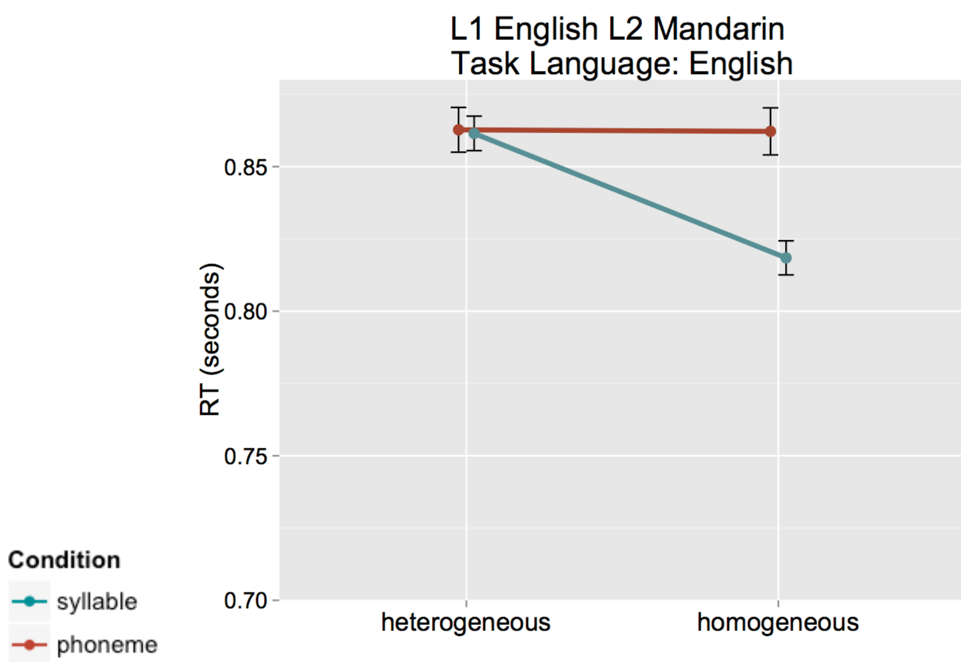

Figure 11: Response times in the English task for the L1 English L2 Mandarin group in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

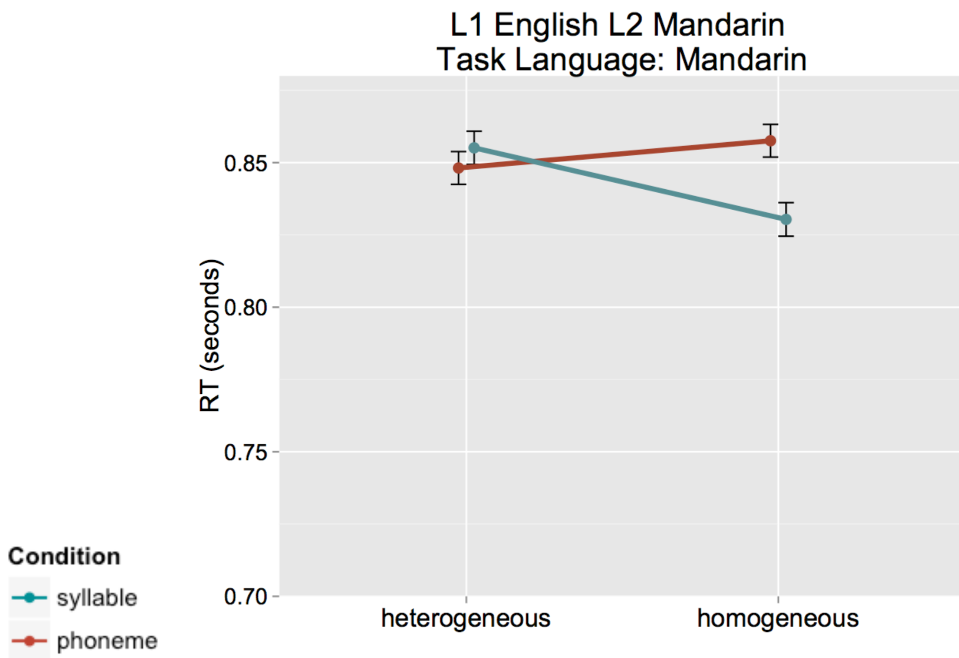

Figure 12: Response times in the Mandarin task for the L1 English L2 Mandarin group in homogeneous and heterogeneous contexts in the syllable and phoneme conditions

3.3.6.1 Overall Results

This analysis found a significant interaction of condition and context (β=.03766, t=3.253, p <.01), and a marginally significant main effect of context (β=.0129, t=1.774, p = .0762). This indicates a tendency for faster responses in the homogeneous context, but only in the syllable condition. Additionally, there was a significant main effect of language, with slower responses in L2 Mandarin than in L1 English (β=.0222, t=-2.068, p<.05). These results are pictured in figures 11 and 12.

The three-way interaction of condition, context, and language did not reach significance in this analysis, suggesting a similar pattern of responses (faster reaction times for shared initial syllables, but not phonemes) in the two languages. Possible reasons for this group's differences from English monolinguals will be discussed in section 3.4.1.

3.4 Discussion

3.4.1 Support for Variable Strategy Hypothesis

Both the L1 Mandarin L2 English group and the Mandarin heritage speakers differed in their patterns of results across Mandarin and English. This supports the variable strategy hypothesis. In English, the L1 Mandarin L2 English speakers responded more quickly when producing sets with a shared initial syllable, as compared to heterogeneous sets. However in Mandarin, shared initial phonetic material resulted, if anything, in slower responses. The heritage speakers also responded more slowly for homogeneous sets in Mandarin. In English, they responded more quickly for homogeneous sets, in both the syllable and phoneme conditions. This is similar to the results of the English monolingual controls. These results suggest that heritage speakers are able to adopt a production strategy characteristic of L1 Mandarin speakers in Mandarin and a production strategy characteristic of L1 English speakers in English.

L1 English L2 Mandarin speakers did not appear adopt different strategies in their two languages. This may be due to a lower degree of L2 proficiency than the other two bilingual groups. This group also responded more slowly in general than the other groups and unexpectedly did not show evidence of a phoneme-based strategy in L1 English. Several possible factors may have contributed to these results. One possibility is that this group experienced a transfer of L2 strategies to L1. This group participated in the Mandarin task first, which could have influenced performance on the English task. We might expect this to be more likely in the heritage speakers, who had more Mandarin experience and also completed the Mandarin task first. However, it is possible the L1 Mandarin L2 English group is less experienced at suppressing interactions between English and Mandarin. In a future experiment, testing L1 English learners of Mandarin in the English task first (or the English task only) could determine whether this is the case. It is also possible that the smaller size of this group (due to limited availability of Mandarin learners) made phoneme effects difficult to observe.

3.4.2 Agreement with previous work

English results for monolinguals and Mandarin heritage speakers agreed with previous work. As in results for English speakers in O'Seaghdha et al. (2010), a shared initial phoneme was sufficient to induce faster response times for these groups. This is consistent with an account of phonemes as the proximate unit in English.

3.4.3 Differences with previous work

Especially for the Mandarin data, not all of our results followed expected patterns from previous work. Chen et al., (2002), O'Seaghdha et al. (2010), and Chen & Chen (2013) all reported faster response times in Mandarin for sets of words with shared initial syllables (but not phonemes alone) in a series of implicit priming tasks. However both the L1 Mandarin L2 English group and the Mandarin heritage speakers in the present study showed no trace of this syllable preparation benefit in Mandarin. In fact, these groups tended to respond more slowly in conditions with shared initial phonetic material, both in the syllable and phoneme conditions. This is particularly surprising given that we used the same set of Mandarin words as Chen & Chen (2013). The effect was marginally significant, but the fact that it was present in both of these groups suggests that it might not be entirely spurious. This result is clearly different from the pattern of results for these two groups in English, but does not clearly determine whether phonemes or syllables might be the proximate units for these groups in Mandarin.

One important difference between this study and previous implicit priming experiments in Mandarin is that participants in this study were bilingual. Bilingualism can affect speech production in several ways. For example, bilinguals name pictures more slowly than monolinguals (e.g. Ivanova & Costa, 2008) and lexical representations in both languages are thought to be active regardless of which language a bilingual is speaking (Kroll, Bobb, Misra, & Guo, 2008). Gollan and Goldrick (2012) found that bilinguals (including Mandarin-English bilinguals) are more adversely affected by tongue twisters than monolinguals. Although this particular study's tongue twisters were composed of materials with onsets alternating in an ABBA or ABAB pattern (e.g. *moat nap mop nut* or *moss knife noose muff*), sequences of words with shared initial phonetic material can also interfere with speech planning and lead to slower productions (Sevald & Dell, 1994). It is possible that homogeneous sets in the implicit priming paradigm create adverse speaking conditions that affect bilinguals more strongly than monolinguals. A direct comparison of Mandarin monolinguals and Mandarin-English bilinguals would allow this possibility to be tested.

3.4.4 Concluding Remarks

Overall, the results of this study support the variable strategy hypothesis. The differences in patterns of results across English and Mandarin for the L1 Mandarin L2 English group and the Mandarin Heritage speakers suggest that speech planning strategies may differ across languages for bilinguals. Degree of L1 and L2 experience also affected speakers' production strategies. The L1 English L2 Mandarin group (least L2 experience) did not vary strategies across languages. The L1 Mandarin L2 English group (intermediate L2 experience) varied strategies across language, but did not adopt an L1-like syllable-based strategy in English. The Mandarin heritage speakers (most L2 experience) adopted L1-like strategies in both English and Mandarin. However, Mandarin results for the Mandarin heritage speakers and L1 Mandarin L2 English group did not show the expected pattern for either phonemes or syllables as the proximate unit. The reason for this result is an open question to be answered with future research. It is possible that the bilingual populations in this study perform the implicit priming task differently than Mandarin monolinguals. A comparison between bilinguals and monolinguals in the implicit paradigm would help to determine whether this is the case.

4 Conclusions

4.1 Comparison Across Experiments

Many of the bilingual subjects in experiments 1 and 2 participated in both experiments. For these subjects, it is possible to compare results in perception and production. This allows us to ask several questions about the relationship between perception and production that would otherwise not be possible to answer. First, we ask whether a domain-general factor, such as language proficiency or processing speed, affects response times in perception and production. Second, we ask whether a common mechanism underlies the selection of a speaker's perception and production strategies.

In the following sections, we compare response times within the same language and the same subjects, but across modalities. We examine correlations between results from the Mandarin conditions of Experiments 1 (perception) and 2 (production). We also compare results from the English bimorphemic nasal condition of Experiment 1 (because its stimuli were the most comparable to the Mandarin stimuli) and the English condition of Experiment 2. These potential correlations are examined in all three of our bilingual groups: L1 Mandarin L2 English speakers, Mandarin heritage speakers, and L1 English L2 Mandarin speakers. Because of the post-hoc nature of these comparisons and the small group sizes, the results of these analyses should be considered preliminary, but they may point toward possible directions for future work.

4.1.1 Relationship Between Reaction Times in Perception and Production

Perception and production are quite different skills, but there are several reasons to believe that ability in these two areas might be related. In everyday conversation, both perception and production skills are practiced, so more experience in conversation would lead to more experience in both. Additionally, speakers are generally able to hear their own productions, so perception experience is obtained any time that production experience is. We might also expect that some individuals might be faster (or slower) across tasks because of individual variation in a more general domain (e.g. processing speed).

On the other hand, practice of perception without any accompanying production practice is possible and may be common for L2-learners in a classroom environment or heritage speakers who may frequently hear the heritage language at home, but speak it less often. Additionally, the measures we compare here are average response times in two tasks not specifically designed to measure proficiency or degree of language experience. It could be the case that more language experience does not lead to faster responses in these specific tasks. Table 13 summarizes the correlation between average reaction times for subjects in the English and Mandarin tasks. Reported r-values were obtained from a Pearson's product moment correlation and reported p-values are one-tailed, with the prediction of a positive direction of correlation.

Table 13: Correlation between reaction times across English and Mandarin tasks

Group	**English perception (bimorphemic nasal condition) vs. English production**			**Mandarin perception vs. Mandarin production**			**Perception vs. Production (pooled across languages)**		
L1 Mandarin L2 English	n= 13	r = -.247	p= .803	**n= 9**	**r = .617**	**p= .029**	n= 17	r=.027	p=.459
Mandarin heritage speakers	**n= 13**	**r= .817**	**p= .0003**	n= 13	r= .331	p= .135	**n= 18**	**r= .605**	**p= .0039**
L1 English L2 Mandarin	n= 6	r= -.061	p= .552	n= 8	r= -.373	p= .838	n= 10	r= -.174	p= .685
All groups	n= 32	r= .213	p=.121	n= 30	r= .105	p= .290	n= 45	r=.097	p= .263

For the L1 Mandarin L2 English group and the Mandarin heritage speakers, we see significant correlations between reaction time in perception and production for their dominant languages (Mandarin and English, respectively), but not for their non-dominant languages. These correlations are pictured in Figures 13 and 14. One possible explanation for this pattern of results is that the dominant language is frequently used in conversations, which gives language users the opportunity to practice production and perception in similar amounts. Having a very high degree of experience in both perception and production may also lead to reaction times that are limited primarily by an individual's processing speed, rather than linguistic knowledge or experience. For a language learned primarily in a classroom or a heritage language, perception experience and ability may be substantially higher than production experience and ability, leading to less correlation in performance across the two modalities. The correlation between perception and production times was also significant for the Mandarin heritage speakers when average response times were pooled across languages. This may have occurred because this group had relatively extensive experience in both Mandarin and English. This correlation is pictured in Figure 15. L1 English L2 Mandarin speakers did not show any significant correlation across modalities, which could be due to a smaller group size.

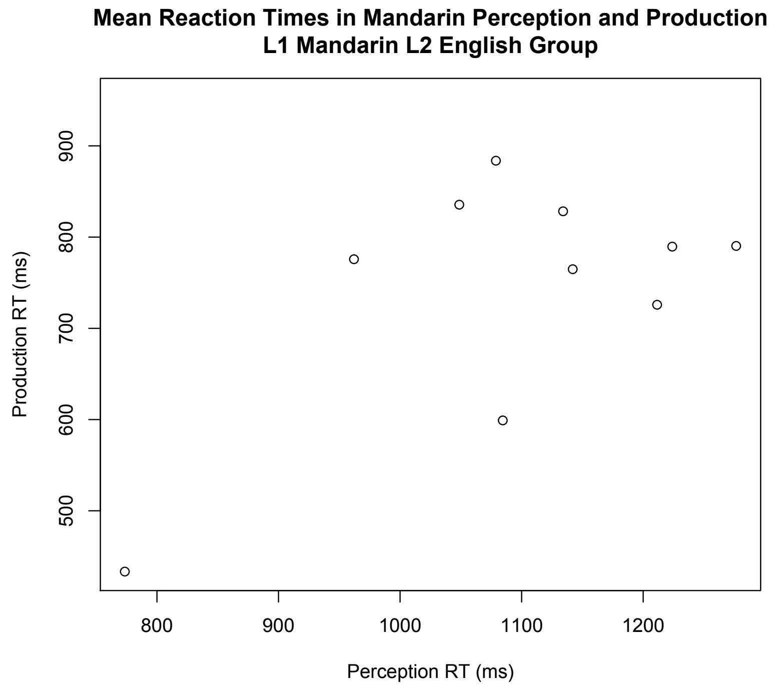

Figure 13: Reaction times in Mandarin perception and production experiments for the L1 Mandarin L2 English group

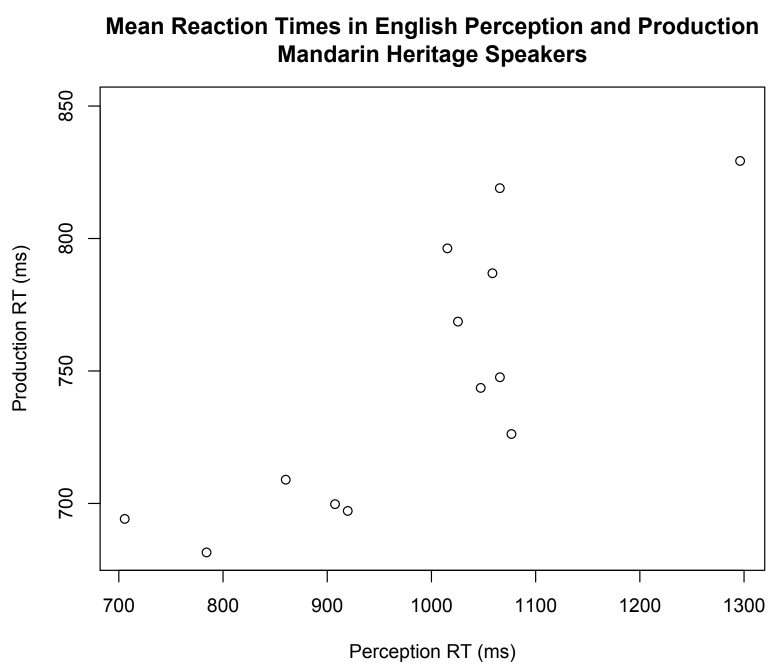

Figure 14: Reaction times in English perception and production experiments for Mandarin heritage speakers

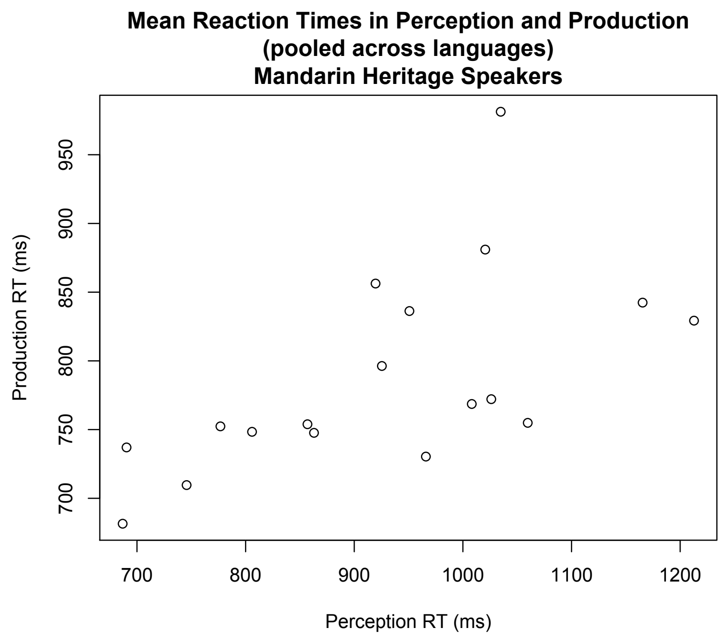

Figure 15: Reaction times (pooled across language) in perception and production experiments for Mandarin heritage speakers

4.1.2 Relationship Between Effect Sizes in Perception and Production

In this section we ask whether a common mechanism underlies the selection of a speaker's perception and production strategies. We address this question by examining whether subjects who show English-like or Mandarin-like perception patterns also show patterns characteristic of the same language in production. Below, we describe the pattern shown by native speakers in each experiment and what measure we used to index how "English-like" or "Mandarin-like" each subject's strategies were.

In perception, L1 Mandarin listeners responded much more quickly when spotting CV sequences (e.g. xi-) in words beginning with a CV syllable (e.g. xi1nan2), as opposed to spotting CVC (e.g. xin-) sequences in the same words. This indicates evidence of a syllable-based segmentation strategy. English monolinguals and L1 English speakers did not show this effect, either in English or in Mandarin. We use the difference in reaction time between CV sequences and CVC sequences in the English and Mandarin CV conditions as our measure of Mandarin-likeness in perception. Individuals with a bigger difference between these two conditions (in the direction of longer response times for CVC condition) show a strategy more similar to the syllable-based segmentation strategy of L1 Mandarin listeners while individuals with smaller differences are more like L1 English listeners.

In production, English monolinguals responded more quickly when producing sets of words beginning with a common initial phoneme (homogeneous phoneme condition), compared to when producing sets of words that did not share initial phonetic material (heterogeneous phoneme condition). As described in Chapter 3, we take this as evidence of a phoneme-based production strategy. Mandarin monolinguals in previous studies and L1 Mandarin L2 English speakers in our study do not show this evidence of phoneme-based speech planning. If anything, L1 Mandarin L2 English speakers in our study responded more slowly in the homogeneous phoneme condition than the heterogeneous phoneme condition in Mandarin. We use the difference in reaction time between the homogeneous and heterogeneous phoneme conditions as our measure of Mandarin-likeness in production. Subjects who responded more quickly in the homogeneous phoneme condition show a more English-like phoneme-based strategy while subjects who do not speed up, or even slow down in the homogeneous condition show a more Mandarin-like strategy. Table 14 summarizes the correlation of these effect sizes across perception and production in English and Mandarin for each group. Reported r-values were obtained from a Pearson's product moment correlation and reported p-values are one-tailed, with the prediction of a positive correlation in English and a negative correlation in Mandarin (due to the negative effect of a shared phoneme for L1 Mandarin speakers the production experiment).

Table 14: Correlation between effect sizes across English and Mandarin tasks

Group	**English perception (bimorphemic nasal condition) vs. English production**			**Mandarin perception vs. Mandarin production**		
L1 Mandarin L2 English	n= 13	r= -.373	p= .905	n= 9	r= .060	p= .565
Mandarin heritage speakers	n= 12	r= -.155	p= .693	**n= 12**	**r= -.387**	**p= .096**
L1 English L2 Mandarin	n= 6	r= -.007	p= .493	n= 9	r= .145	p= .355

In the Mandarin data, we see a marginally significant correlation between perception and production strategies for the Mandarin heritage speakers. Mandarin heritage speakers with a larger syllable effect in Mandarin perception tended toward a more negative phoneme effect in Mandarin production. This relationship is pictured in Figure 16. This correlation may indicate that speakers who have successfully adopted a perception strategy from the non-dominant language are also likely to have successfully adopted that language's production strategy. These correlations may not appear in the dominant language because speakers have already fully adopted the characteristic perception and production strategies of that language. The L1 English L2 Mandarin group also showed no significant correlations, which may be due to a combination of smaller n and lower L2 proficiency. No significant correlation between perception and production strategies was found the in English data. This may be because speakers from all three groups followed a similar response pattern, with average difference in response times between the CV and CVC targets near zero.

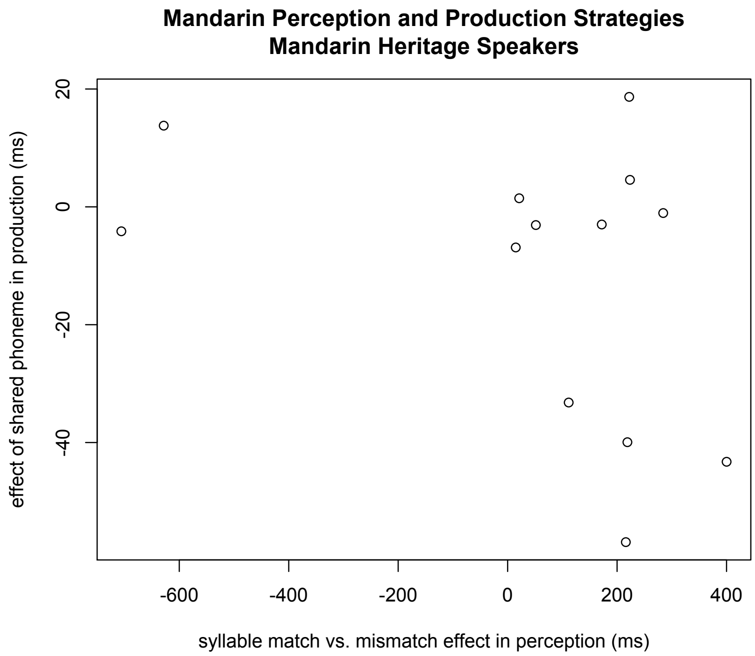

Figure 16: Correlation between Mandarin perception and production strategies in Mandarin heritage speakers

4.1.3 Discussion

In this section, we found some support for possibility that domain-general processing speed may affect response times in both perception and production. This relationship was primarily evident in individuals' dominant language, suggesting that it may only surface when differences in linguistic knowledge and experience are not a major source of variability. We found limited evidence of a common mechanism underlying the adoption of perception and production strategies in individuals' non-dominant language. Further study with additional measures of individuals' perception and production strategies and larger groups could help to determine more clearly the nature of this relationship.

4.2 Future directions

In the process of answering the question of what happens when a bilingual's languages promote different perception and production strategies, we found results that raise several new questions. These questions may be pursued in future work that builds on this study's findings. These questions are explained in the following sections, accompanied by descriptions of possible future research directions.

4.2.1 Perception of tone by speakers of lexical-stress languages

In Experiment 1 (perception), English speakers with a few years of Mandarin experience readily adopted a syllable-based segmentation strategy for Mandarin stimuli. We raised the possibility that English speakers may be interpreting all or most of the Mandarin syllables as stressed. Given that English speakers tend to segment speech at stressed syllable boundaries (Cutler & Norris, 1988), this could help to facilitate the adoption of a syllable-based segmentation strategy in Mandarin. The question of whether English listeners perceive stress on Mandarin syllables raises more generally the question of whether, and how, speakers of languages with contrastive lexical stress (but not tone) perceive stress in input from tonal languages. We can also ask the reverse question of whether speakers of lexical tone languages perceive tone in input from (non-tonal) lexical stress languages.

Loanwords provide some evidence that listeners may make mappings between tone and stress systems in their two languages. For example, English loanwords in Mandarin (Wu, 2006) and Cantonese (Silverman, 1992) generally have a tone containing a high pitch on syllables corresponding to English stressed syllables. Disyllabic Mandarin loanwords in English often have stress on both syllables (e.g. feng shui, mahjong, Beijing). However, mappings between stress and tone systems are not straightforward for all pairs of languages. Different pairs of languages adapt loan words using a variety of strategies, some of which reflect source language prosody and some of which do not (see Davis, Tu, & Tsujimura, 2012 and Kang, 2010 for reviews). For example, in Lhasa Tibetan (a tonal language) tones assigned to loanwords from English and Mandarin are not influenced by stress and tone in the source language, but instead assigned based on phonemic content of the first syllable, with voiced initial consonants resulting in a low tone on the initial syllable and voiceless onsets resulting in a high tone (Hsieh & Kenstowicz, 2008).

In order to clarify how mappings between stress and tone systems across languages are formed, future research could investigate online perception of non-native lexical prosodic contrasts and their mappings to native prosodic categories. For example, in an artificial language learning paradigm, a listener might learn a minimal pair of words that contrast only in a native prosodic category. In a later phase, participants might be presented with a word with the same phonemic content, but non-native prosody, and asked to judge which member of the minimal pair it is more similar to.

4.2.2 Neighborhood effects in Mandarin

Another finding of Experiment 1 was that L1 Mandarin L2 English listeners detected phoneme sequences matching the syllable structure of Mandarin stimulus words more quickly than those that did not match, but only if that word began with a CV syllable. For words beginning with a CVC syllable, these listeners detected CVC sequences no more quickly than CV sequences. We suggested as one possible explanation that these CV sequences all correspond to Mandarin morphemes, often many morphemes. This could lead to a high degree of activation for the sequences, causing them to be recognized quickly.

With only about 1200 distinct syllables (or around 400 if tone is not taken into account) and mostly monosyllabic morphemes, large numbers of homophones and even larger lexical neighborhoods (groups of words differing by a single phoneme) are possible. Lexical neighbors sometimes facilitate and sometimes inhibit production and perception of words (see Chen & Mirman 2012 for a review). With the large neighborhoods possible in Mandarin, avoiding potential inhibitory effects might be more important. Additionally, the role of syllables in Mandarin speech planning might affect the role of lexical neighbors in lexical selection. This leads to several possible questions for future research. Do neighbors differing from a word by a single phoneme exert a similar influence on production as in languages with phonemes as the basic planning unit? What would be the effect of "syllable neighbors" that differ from a word by one syllable? The high degree of homophony and dense lexical neighborhoods found in Mandarin and their consequences for speech production and perception are areas deserving of further study in future work.

4.2.3 Alternative paradigms for determining production strategies

In Experiment 2, we observed the surprising result that highly proficient Mandarin speakers (both the L1 Mandarin L2 English group and the Mandarin heritage speakers) tended to produce responses more slowly for sets of Mandarin words sharing an initial phoneme. This condition was intended to, if anything, facilitate faster production, but led to slower responses for this group. The result was a somewhat ambiguous finding regarding the status of phonemes as planning units in Mandarin. Shared phonemes clearly did not have the facilitative effect that they do for L1 English speakers. On the other hand, they did influence speakers' reaction times in Mandarin, so they clearly play some role in Mandarin speech planning. The potential for simultaneous competition and facilitation from the shared initial phonemes complicates the interpretation of results from implicit priming tasks. Future work investigating cross-language variability in proximate units could avoid this problem by using different experimental paradigms. These might include more traditional priming paradigms that involve the visual presentation of a prime (e.g. a picture of an object beginning with the same phoneme as the target), rather than producing sets of words with shared phonetic material. This may be especially important for investigating proximate units in bilinguals, who may be more vulnerable to interference from overlapping phonetic material (Gollan & Goldrick, 2012).

4.2.4 Bilingual flexibility in other linguistic domains

Another potential area for future work is determining more generally what factors promote bilingual flexibility and what factors make the learning of certain aspects of L2 difficult or impossible. This could be accomplished by examining bilingual flexibility in other linguistic domains. Existing findings in the domains of phoneme category learning and syntax raise the question of when similarities between L1 and L2 promote bilingual flexibility and when they may hinder it.

A common theme in several models of L2 phoneme learning (e.g. the Perceptual Assimilation Model (Best, 1994), the Speech Learning Model (Flege, 1987, 1995), and the Native Language Magnet model (Khul et al., 2008)) is that phoneme categories in the L1 influence those of the L2. Under these models, learning new sounds in a second language is more difficult when the new sounds are very similar to an existing sound in the learner's first language. This is because the L2 sounds may be perceived as belonging to an L1 phoneme category. For example, adult English speakers show difficulty distinguishing dental from retroflex stops, such as /t̪/ vs. /ʈ/, which both tend to be perceived as the English consonant /t/. English-acquiring infants are able to quickly learn this contrast at the age of 6-8 months, but lose this ability by the age of 1 year as English phoneme contrasts become established (Werker & Tees, 1984). New contrasts involving sounds that are sufficiently different from familiar sounds are learned more easily because speakers have no existing category to (erroneously) map them on to. For example English speakers are able to discriminate minimal pairs of click consonants from Zulu with a high degree of accuracy with little to no training (Best, McRoberts, & Sithole, 1989). This in some ways parallels our finding that English-dominant listeners are able to adopt a syllable-based segmentation strategy in Mandarin (highly dissimilar from English) when previous work has found that they did not do so in French (more similar to English).

In contrast to the case of phoneme categories, grammatical gender appears to be learned more easily in an L2 if a gender system is also present in the L1. Sabourin, Stowe, & De Haan (2006) compared the performance of L1 English speakers and L1 Romance language speakers on grammaticality judgments of sentences requiring gender-agreement between nouns and relative pronouns in L2 Dutch. L1 speakers of English, which does not have grammatical gender, performed more poorly than speakers of Romance languages, which do have grammatical gender. This was true even though Romance and Dutch gender systems differ substantially, including different available gender categories and rules for constructions requiring gender agreement. These findings suggest that rather than interfering with the application of gender rules, the presence of rules for gender in the L1 grammar facilitates learning to correctly apply gender agreement in the L2. This difference between findings in the domains of syntax and phoneme contrast learning in L2 acquisition suggests that future work might attempt to generalize when linguistic similarities support native-like performance in the L2 and when they hinder it.

4.3 Conclusions

This study set out to determine what happens when a bilingual's two languages promote different perception and production strategies. Below, we summarize our findings for each specific question we addressed.

4.3.1 Can bilinguals adopt perception and production strategies from multiple languages?

Our results strongly support the conclusion that bilinguals can adjust perception and production strategies across task languages. In Experiment 1, we found that all three of our bilingual groups showed some evidence of syllable-based segmentation in Mandarin while no groups showed evidence of syllable-based segmentation in English. In Experiment 2, we found that L1 Mandarin L2 English speakers and Mandarin heritage speakers differed in their production strategies across languages. In particular, Mandarin heritage speakers showed L1-like response patterns in both Mandarin and English. The L1 English L2 Mandarin group did not show evidence of differing strategies across language, possibly due to lower L2 proficiency than the other bilingual groups.

4.3.2 How does degree of experience in L1 and L2 affects a bilingual's choice of strategy?

In Experiment 1, all bilingual groups were able to change perception strategies across languages, suggesting that even a modest amount of experience can promote the adoption of new perception strategies. However, English monolinguals did not adopt a syllable-based segmentation strategy for Mandarin stimuli, suggesting that at least some L2 experience is necessary for learning a new perception strategy. In Experiment 2, the L1 English L2 Mandarin group, which had the least L2 experience of our groups, did not change production strategies across Mandarin and English. This suggests that more L2 experience may be required to learn a new production strategy, as compared to a new perception strategy, or that this group of bilinguals had more experience in Mandarin perception than Mandarin production.

4.3.3 Are bilinguals more likely to apply L1 strategies in L2 environments that more closely resemble L1?

We found no evidence that bilinguals are more likely to apply L1 strategies in L2 environments that are more similar to L1. In Experiment 1, we compared segmentation strategies of L1 Mandarin speakers in English environments that were more Mandarin-like (words with nasal syllable codas) and less Mandarin-like (words with fricative and liquid codas, which are not allowed in Mandarin). L1 Mandarin speakers did not apply a syllable-based segmentation strategy in either of these environments.

4.3.4 Is there a relationship between perception and production strategies within individuals?

In the first section of this chapter, we found that Mandarin heritage speakers who have more fully adopted the perception strategy of Mandarin tend to have also more fully adopted the production strategy of Mandarin. This preliminary finding hints at the possibility of a deeper relationship between perception and production strategies. If syllables generally play a more prominent role in Mandarin than in English, language learners may detect this difference and learn to adjust their perception and production systems together toward or away from syllable-based strategies. Future work comparing production and perception strategies in larger groups of individuals is needed to explore this possibility more fully.

4.3.5 Overall Conclusions

Overall, our findings suggest flexibility in bilingual perception and production systems. Even bilinguals with much less L2 than L1 experience adjusted strategies according to task language. Our perception results contrast with previous findings that suggested limitations on the flexibility of bilingual speech segmentation strategies. Cutler et al. (1988) found that English-dominant listeners did not adopt a syllable-based segmentation strategy in French, even though French promoted such a strategy, and even with lifelong experience speaking French. However, we found that English speakers adopt a syllable-based segmentation strategy in Mandarin after only a few years of Mandarin experience. This suggests that the differences between English and Mandarin, such as differences in syllable structure and syllable-morpheme correspondence, create an environment that promotes the adoption of different segmentation strategies and allows bilingual flexibility to surface. These findings underscore the importance of studying bilingual speakers of typologically distant languages.

Our findings are also consistent with the view that a bilingual's two languages are always active to some degree (e.g. Marian & Spivy, 2003). In our production experiment, we find evidence that, even in the dominant language, bilingual performance is not necessarily the same as monolingual performance. Shared phonetic material in Mandarin that led to faster response times for monolingual speakers in Chen & Chen, 2013 did not for the L1 Mandarin L2 English group and the Mandarin heritage speakers. If anything, this shared material slowed these speakers down, suggesting that they may have experienced competition from the shared phonemes.

Under the view that bilinguals never completely deactivate a language that is not in use, they must constantly suppress interference between their two languages. This interference could come from any part of the language, including the grammar, the lexicon, or the phonetic inventory. For bilinguals whose two languages differ in perception and production strategies, these L1 and L2 strategies are yet another aspect of language that they must learn to selectively activate and suppress. Our results suggest that bilinguals are indeed able to learn multiple perception and production strategies and switch between them effectively depending on the task language.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: the Transition From Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human perception and performance*, *14*(3), 345.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, *19*(2), 290.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240-250.

Bock, K. (1991). A sketchbook of production problems. *Journal of Psycholinguistic Research*, *20*(3), 141-160.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.3.53, retrieved 21 February 2015 from http://www.praat.org/

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249-268.

Chen, J. Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient: Special Issue: Cognitive Processing of the Japanese and Chinese Languages II*, 43, 15–26.

Chen, J. Y., Chen, T. M., and Dell, G. (2002). Word-form encoding in mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46(4), 751–781.

Chen, J. Y., Lin, W. C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology*, *45*(1), 107-120.

Chen, T. M., & Chen, J. Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review*, *20*(1), 154-162.

Chen, T. M., Dell, G. S., & Chen, J. Y. (2007). A cross-linguistic study of phonological units: Syllables emerge from the statistics of Mandarin Chinese, but not from the statistics of English. *Chinese Journal of Psychology*, 49(2), 137-144.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech–Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, *63*(1), 47-75.

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, *16*(5-6), 609-636.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400.
Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3), 133-142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113.
Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilingualism. *Nature*, 340(6230), 229-230.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The Monolingual Nature of Speech segmentation by bilinguals. *Cognitive Psychology*, 24(3), 381-410.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*(6), 824-844.

Davis, S., Tu, J. Y., & Tsujimura, N. (2012). Toward a taxonomy of loanword prosody. In *Catalan Journal of Linguistics* (Vol. 11, pp. 0013-39).

Dell, G. S. (1995). Speaking and miss-speaking. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science. Language* (Vol. 1, pp. 183–208). Cambridge, MA: MIT Press.
Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress 'deafness' in simultaneous French–Spanish bilinguals. *Cognition*, 114(2), 266-275.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47-65.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 229–273). Timonium, MD: York Press.

Gollan, T. H., & Acenas, L. A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 246.

Gollan, T. H., & Goldrick, M. (2012). Does bilingualism twist your tongue? *Cognition*, *125*(3), 491-497.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220-1234.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*(01), 63-83.

Hanulikova, A., Mitterer, H., & McQueen, J. M. (2011). Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition*, 14(04), 506-521.

Hsieh, F. F., & Kenstowicz, M. J. (2008). Phonetic knowledge in tonal adaptation: Mandarin and English loanwords in Lhasa Tibetan. *Journal of East Asian Linguistics*, *17*(4), 279-297.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277-288.

Kang, Y. (2010). Tutorial overview: Suprasegmental adaptation in loanwords. *Lingua*, *120*(9), 2295-2310.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and speech*, *51*(4), 343-359.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69-89.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, *6*(02), 97-115.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 298-305.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. Mora rhythm. *Language and Speech*, *50*(1), 77-99.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*(2), 258-278.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), B9-B17.

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(35), 14265–70.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*,*39*(3), 484-507.

Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language*, *68*(1), 300-305.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition,* 53(2), 91-97.

Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology*, *9*(02), 289-328.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281-284.

Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 531-547.

Tseng, S. C. (2005). Syllable contractions in a Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics*, 10(1), 63-83.

Verdonschot, R.G., Nakayama, M., Zhang, Q., Tamaoka, K., & Schiller, N.O. (2013). The Proximate Unit of Chinese-English bilinguals: Proficiency Matters. *PLOS One*, 8(4) e61454.

Verdonschot, R.G., Lai, J., Chen, F., Tamaoka, K., & Schiller, N.O. (2015). Construction initial phonology in Mandarin Chinese: Syllabic or subsyllabic? A masked priming investigation. *Japanese Psychological Research*, 57(1), 61-68.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597-607.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866-1878.

Wu, H. H. I. (2006). Stress to tone: A study of tone loans in Mandarin Chinese. *MIT Working Papers in Linguistics*, *52*, 227-253.

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PloS One*, *7*(10), e46595.