Title
Tone Attrition in Mandarin Speakers of Varying English Proficiency

Permalink
https://escholarship.org/uc/item/20b7g9m9

Journal
JOURNAL OF SPEECH LANGUAGE AND HEARING RESEARCH, 60(2)

ISSN
1092-4388

Authors
Quam, C
Creel, SC

Publication Date
2017-02-01

DOI
10.1044/2016_JSLHR-S-15-0248

Peer reviewed
Tone attrition in Mandarin speakers of varying English proficiency

Carolyn Quam¹

Center for Research in Language, University of California, San Diego; 9500 Gilman Drive, La Jolla, CA 92093-0515

and

Sarah C. Creel
creel@cogsci.ucsd.edu; (858) 534-7308

Department of Cognitive Science, University of California, San Diego; 9500 Gilman Drive, La Jolla, CA 92093-0515

Keywords: bilingualism; speech perception; phonology; adults

¹ Corresponding author. Current affiliation: Departments of Speech, Language, and Hearing Sciences and Psychology, University of Arizona, 1503 E University Blvd., P.O. Box 210068, Tucson, AZ 85721. (520) 626-5720. cmquam@email.arizona.edu.
Structured Abstract

Purpose: To determine whether degree of dominance of Mandarin-English bilinguals’ languages impacts phonetic processing of tone content in their native language, Mandarin.

Method: We tested 72 Mandarin-English bilingual college students with a range of language-dominance profiles in the two languages and ages of acquisition of English. Participants viewed two photographs at a time while hearing a familiar Mandarin word referring to one photograph. The names of the two photographs diverged in either tone, vowels, or both. Word recognition was evaluated using clicking accuracy, reaction times, and an on-line recognition measure (gaze), and compared in the 3 conditions.

Results: Relative proficiency in English was correlated with reduced word recognition success in tone-disambiguated trials, but not in vowel-disambiguated trials, across all 3 dependent measures. This selective attrition for tone content emerged even though all bilinguals had learned Mandarin from birth. Lengthy experience with English thus weakened tone use.

Conclusions: This finding has implications for the question of the extent to which bilinguals’ two phonetic systems interact. It suggests that bilinguals may not process pitch information language specifically, but rather that processing strategies from the dominant language may impact phonetic processing in the non-dominant language—even when the latter was learned natively.

(200 words in abstract)
How does becoming bilingual impact sound processing in the native language? Compared with monolinguals, speech-sound processing is more complicated for speakers of two languages, who must process two different sets of sound categories. Here we consider processing of lexical tones by adult Mandarin-English bilinguals, asking whether experience with English impacts processing of tones in Mandarin. English and Mandarin Chinese use similar pitch patterns for very different functions. Mandarin is a tone language, in which pitch contours, referred to as *lexical tones*, differentiate words. Mandarin syllables carry four different possible tonal patterns: high flat (tone 1; T1), rising (T2), low dipping (T3), and falling (T4). The meaning of a syllable or word depends partly on the tonal pattern. Because pitch is important for distinguishing Mandarin words (in addition to serving other linguistic and paralinguistic functions), Mandarin listeners’ word identification is improved by attending to tone. English, on the other hand, is an intonation language, in which pitch serves functions like indicating yes/no questions or phrase boundaries, but does not contrast words. Lexical pitch is not relevant at the word level in English, so English listeners’ word identification would presumably be hindered by attending to it. Optimal lexical-pitch-processing strategies for the two languages are thus in opposition.

For monolinguals, specialization for pitch processing appears to begin early in life. Tone-language learners’ tone discrimination is maintained within the first year of life, while intonation-language learners’ tone discrimination worsens (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Yeung, Chen, & Werker, 2013; see also Harrison, 2000). At the lexical level, language-specific pitch processing seems to develop by about 11 months for recognition of word forms (Singh & Foong, 2012) and by about 17 months for word learning (Hay, Graf Estes, Wang, & Saffran, 2014; Singh, Hui, Chan, & Golinkoff, 2014; Quam &
Monolingual learners of tone vs. intonation languages appear to tune their pitch-processing strategies to their native language by an early age. But many people speak both a tone language and a non-tone language, or multiple tone languages with different tonal contours. Interpreting lexical pitch contours language-specifically would confer a strong benefit to tone+non-tone-language bilinguals’ word recognition efficiency. Very little work has examined differential recognition of pitch cues. It is an open question what happens to tone processing when individuals learn both a tone language and a non-tone language. Do listeners maintain language-specific tone processing, despite learning and using a language without tone distinctions? If so, experience with English should not impact the speed and accuracy of tone processing for native Mandarin speakers who are still using Mandarin regularly. Or, does experience with the nontone language lead to attrition in tone processing within the tone language, even when bilinguals are using both languages regularly? The present investigation addressed these questions.

One might predict that bilinguals could process lexical pitch language-specifically, based on evidence that bilingual listeners can process consonants differently in each of their languages. When enough information is provided to indicate the language context (see Gonzales & Lotto, 2013, for an overview), bilinguals can shift their segment boundaries to match the language they are hearing. For instance, bilinguals break up the voice-onset-time continuum differently when processing consonants in English vs. Spanish (Elman, Diehl, & Buchwald, 1977; Gonzales & Lotto, 2013), Dutch (Flege & Eefting, 1987), or French (Hazan & Boulakia, 1993), three languages that divide up the voice-onset-time continuum differently than English does.
Evidence of language-specific consonant processing might predict that bilinguals would process tones language-specifically. However, there are several potential reasons that experience with a non-tone language might compete or interfere with optimal lexical-pitch processing in a native tone language. First, in many studies of bilingual listeners, bilinguals have not responded as if they were two monolinguals in a single brain, but instead have shown interactions or interference between language representations (Sebastián-Gallés, Echeverria, & Bosch, 2005; MacWhinney, 2005; de Bot, 1992). If the dominant language dictates pitch processing, perhaps even relatively late acquirers of English could learn English so thoroughly, and use it so often, that their tone processing would worsen (Major, 1992). The second language (L2) can affect the sound categories of the first language (Flege, 1987, 1995). Gildersleeve-Neumann, Peña, Davis, and Kester (2009) found that 3- to 4-year-old children’s native Spanish vowel productions were affected by contact with English (their L2), at least temporarily. Even more surprisingly, Chang (2011) found that adult native-English speakers’ English vowels began to shift midway through an intensive, six-week Korean-language course. Evidence of L2 effects on L1 phonetic processing led us to assess participants’ language backgrounds and current language dominance in order to investigate the impact of these factors on tone processing.

The second reason that we might find trade-offs between bilinguals’ lexical-pitch processing in a tone vs. an intonation language is that pitch is used dramatically differently in the two languages. In prior work on consonant processing, bilinguals’ two languages both relied on the same acoustic dimension (voice-onset time; VOT) to contrast consonants, but divided it at a different point on the VOT continuum. For the bilinguals tested here, pitch contours like rises vs. falls are used at the lexical level (as well as other levels) in Mandarin, but not in English. It would seem that optimal phonetic processing would entail attending to lexical pitch for word
differentiation in Mandarin and disregarding it in English. It is an open question whether bilinguals can adapt their phonetic processing to the language when the two languages differ so strikingly in their use of a dimension. This makes the current investigation of particular interest.

The third reason to predict limits on bilinguals’ ability to match their pitch/tone processing to the language context—as has been demonstrated for consonants—is that lexical tone is potentially more weakly represented in the lexicon than consonants and vowels (see Burnham & Mattock, 2007, for a review). Weak status in word representations could theoretically make tone more prone to attrition. Tone appears to be less useful for lexical access and word recognition in Mandarin than segments (Lee, 2007; Cutler & Chen, 1997; see also Patel, Xu, & Wang, 2010). Differences between the acoustical properties of tones and segments might account for some tone disadvantages in spoken-word recognition (Ye & Connine, 1999). Tones are often characterized by dynamic contours, which unfold more slowly in the acoustic signal than consonants and (particularly steady-state) vowels. Listeners may exploit tone as soon as it is perceptually available, which is simply later than vowels (Ye & Connine, 1999). However, similar tone disadvantages in lexical access were recently demonstrated in orthographic tasks, providing stronger evidence for differences within lexical representations themselves (Davis, Schoknecht, Kim, & Burnham, 2015). Awareness of tones also appears to be weaker than awareness of other segments even when both tones and phones are represented orthographically in the language (Burnham et al., 2011; see also McBride-Chang et al., 2012; Burnham, Luksaneeyanawin, Kantamphan, & Reid, 2013). In short, it may be the case that tones are represented more weakly and thus are more susceptible to attrition processes than are segments.
The Present Study

The present study investigated Mandarin-English bilinguals’ tone processing in Mandarin. For the three reasons mentioned above, we predicted that experience with a non-tonal language—English—would weaken listeners’ ability to exploit tones in their native language, Mandarin. While being eye-tracked, native speakers of Mandarin who learned English at varying ages were asked to select the picture to which a spoken Mandarin word referred. Our materials were loosely inspired by an earlier study by Malins and Joanisse (2010), who were interested in Mandarin speakers’ relative processing speed for tones vs. vowels. However, the focus of our study is very different: we tested Mandarin-English bilinguals with a range of language-experience profiles in order to investigate the impact of bilingual language dominance on use of tones vs. vowels in word recognition.

On some trials, the two pictured words differed by a vowel (e.g., mu3—mother—vs. mi3—rice; numbers indicate Mandarin tones 1-4). On others, the two pictured words differed only by a tone (e.g., mu3 vs. mu4—tree/wood). On the third trial type, pictured words differed by both vowel and tone information (e.g., mu3 vs. mi4—honey). To succeed in tone-disambiguated trials, listeners had to distinguish words based on a single tone difference. Even if accuracy were high, however, there might be differences in the rate at which listeners processed tone information, which could be picked up by the rapidity of visual fixation increases to the target item on tone trials, as well as by reaction time to click on the target picture.

Our first question was whether relative proficiency in Mandarin vs. English would affect bilinguals’ ability to exploit tone information. In other words, would we see attrition of tone processing with increasing English exposure? Bilingual participants were all native speakers of
Mandarin. If learning Mandarin from birth confers uniformly good tone processing, then all listeners should exploit tone equally well to disambiguate familiar words. On the other hand, many of our participants had become dominant in English, exhibiting higher vocabulary scores in English than Mandarin and scoring higher on the Bilingual Dominance Scale, which measures language background and current language use (Dunn & Fox Tree, 2009). Thus, participants with greater English experience might have undergone attrition of their Mandarin phonological skills, and might show weaker use of tone, even for familiar words.

Our second question was, if we did find attrition of Mandarin phonetic processing, would this attrition be specific to lexical tones? To test this question, we included both tone and segmental contrasts (in which words diverged at their vowels). While lexical tone is contrastive in Mandarin but not in English, most Mandarin vowels, including the ones we used (see Table 3), could likely be assimilated to English vowel categories (Best, 1994; Best & McRoberts, 2003). As a result, experience with English might not impact processing of these Mandarin vowel contrasts to the same degree that it will impact processing of Mandarin tones. Thus, we may find selective attrition for lexical tone (rather than vowels) with increasing English proficiency (and decreasing Mandarin proficiency).

While an assimilation-based model (e.g., see Reid et al., 2015, for a recent extension of the Perceptual Assimilation Model to lexical tones) would likely predict attrition for tone content but not vowel content in our stimuli, there are other reasons to think we might find attrition for vowel content. First, if attrition operates at the level of familiarity (or “rustiness”) with particular words rather than at the level of an acoustic/phonetic dimension like pitch, then this more holistic process of attrition should impact processing of words regardless of the dimension on which they contrast. Second, it could be that even if vowels can be roughly assimilated into English
categories—at least to the degree that they are discriminable by English speakers—proficiency with Mandarin might still lead to greater ease of processing segmental contrasts. This could be because of the inclusion of some sounds that are fairly un-English-like (e.g., the vowels in “he1/4” and “shi1/3,” or the fricatives in “xin1/4”-“xian1/4” and “qiu1/2”-“quan1/2”; see Table 3), or because of general Mandarin subphonemic/accentual content.

It is therefore an open question whether exposure to English will lead to attrition of Mandarin phonetic processing, and, if so, whether it will selectively impact processing of tones or instead impact both tone and vowel processing. If we find selective effects of language dominance on lexical-tone processing, however, this will provide evidence that when a bilingual’s two languages conflict in how they treat a particular acoustic dimension in their lexical phonology, processing of that dimension in the native language (Mandarin) is impaired by exposure to the other language (English). This would further suggest that bilinguals are not maintaining two distinct processing “modes” for pitch content, corresponding to each language, but instead that there are trade-offs between a bilingual’s two languages. Note, however, that weaker status of tone in word representations could also make tones more prone to attrition than consonants and vowels (e.g., Davis, Schoknecht, Kim, & Burnham, 2015).

Method

Participants

Participants were 72 Mandarin-English bilingual adults (49 women, mean age = 20, $SD = 2$, range = 16–26; parental permission obtained when necessary; age data missing for 1 participant who was approximately 19), who had, immediately prior, participated in one of two novel-word-learning experiments (Quam & Creel, 2012, in preparation). There were no effects
of (or interactions with) prior experiment in any of the analyses presented here except where noted (see Footnote 2), so we have included analyses without that factor for simplicity of presentation. Thirteen additional participants were excluded from the analysis for as much or more exposure to other tone languages or dialects during childhood as to Mandarin (7); responding incorrectly in 1/3 of tone-contrast trials, indicating low proficiency (1); and experimenter error (5). Two of the 72 included participants were excluded just from gaze analyses because of poor eye-tracking quality (see Results).

Language-Dominance Assessments

All 72 bilingual participants self-reported as at least “fairly proficient” in both languages, but they exhibited a wide variety of language backgrounds. Most had learned Mandarin before English and would be considered second-language learners of English rather than simultaneous bilinguals. But by college, many had had much more exposure to English than to Mandarin, so that they had become dominant in English. Bilingual participants’ language experience was quantified via age of arrival in an English-speaking country plus two test measures, as summarized in Table 1.

—INSERT TABLE 1 HERE—

The Multilingual Naming Test (MINT; Gollan et al., 2012) evaluates bilinguals’ ability to name pictures in their two languages; it is designed for English and Mandarin, among other languages. (Note that at the time we used the MINT, it was still being normed for use in Mandarin.) Participants named a series of 68 pictures of estimated increasing difficulty, going through the set twice, once in English and once in Mandarin (language order was
counterbalanced across participants). The MINT score for each language was the number of pictures named in that language. We then subtracted the English score from the Mandarin score to compute Mandarin-dominance scores: a large positive number meant someone was strongly Mandarin dominant, while a large negative number meant strongly English dominant. English scores were higher and less variable than Mandarin scores (Table 1; t(72) = 3.65, p < .001). Mandarin MINT scores and English MINT scores were negatively correlated (Figure 1; r = -0.46, p < .001). MINT scores in each language also correlated with the composite (Mandarin – English) difference score (Table 2). This difference score was used as a measure of “Mandarin dominance,” where highly positive scores indicate strong Mandarin dominance, and highly negative scores indicate strong English dominance.

The Bilingual Dominance Scale (BDS; Dunn & Fox Tree, 2009) is composed of 12 questions evaluating bilinguals’ life-long experience with each language (e.g., “At what age did you first learn / feel comfortable speaking each language?”) and current language use (e.g., “When doing math in your head, which language do you use?”). We modified some questions to tailor the assessment to our population. For example, we split the question “Which language do you use at home?” into “…in your college dorm or house?” vs. “…with your family?” The BDS assigns points to each language depending on participants’ responses (see Dunn & Fox Tree, 2009, for scoring). As with the MINT, we subtracted the English score from the Mandarin score to compute Mandarin-dominance scores. Scores on both the MINT and the BDS were slightly skewed toward English dominance (Table 1). The BDS and MINT Mandarin-dominance scores
correlated strongly with each other and with age of arrival in the United States or another English-speaking country (Table 2; see Table 1 for descriptive statistics).

—INSERT TABLE 2 HERE—

Stimuli

We designed a set of 7 quadruplets of monosyllabic, familiar Mandarin words (see Table 3), loosely based on stimuli created by Malins and Joanisse (2010). Within each quadruplet, each word (e.g., “chuang2”) had competitors differing in tone (“chuang1”), vowels (“cha2”), and both tone and vowels (“cha1”). Vowel-disambiguated trials also differed in their following consonants. However, the first divergence point was always at the vowel, so we use the term *vowel-disambiguated* throughout. Tone 2 and Tone 3, the pair most difficult to distinguish (Shen & Lin, 1991), were never contrasted directly. Each word occurred equally often as a target and as a competitor. For example, “chuang2” was the tone-disambiguated competitor for “chuang1” (and vice-versa), the vowel-disambiguated competitor for “cha2,” and the tone&vowel-disambiguated competitor for “cha1.” Each word was therefore its own frequency control across trial types (thus balancing within-experiment word frequency), and all words were labeled equally often.

—INSERT TABLE 3 HERE—

Words were selected to comprise the quadruplets, in which each word had both a tone- and a segment-disambiguated competitor, but also to maximize imageability as much as possible,
so that they could be used to label pictures in the eye-tracking paradigm (see Figure 2). Word frequency was also maximized as much as possible overall (see Table 3). Word frequencies were calculated from the Modern Chinese Character Frequency List, which was compiled using a large corpus of 193,504,018 total written characters (http://lingua.mtsu.edu/chinese-computing/statistics/char/list.php?Which=MO; Da, 2004). We chose to use written character frequencies for two reasons. First, we used written rather than spoken frequencies because spoken corpora currently available (e.g., Canavan & Zipperlen, 1996) are small. Second, we used character frequencies rather than sound-based (e.g., syllable or morpheme) frequencies—even though particular characters can have multiple pronunciations—because of the prevalence of homophones in Mandarin. We judged that analyzing character frequencies rather than syllable frequencies was a better proxy for evaluating the frequency of words’ sounds when they indicate a particular meaning, which was important given that our task required matching spoken words to their visual referents.

All included characters were ranked within roughly the top third (within the top 33.6%) of characters in the corpus. Ratios of word-pair frequencies (e.g., the frequency ratio for the characters corresponding to “chuang2,” bed, vs. “chuang1,” window) were also balanced as much as possible across trial-types. The average frequency ratios (less-frequent:more-frequent) for word pairs in the three trial types were very comparable (tone-disambiguated: 0.38:1, vowel-disambiguated: 0.37:1, and tone&vowel-disambiguated: 0.38:1).
The 28 target words were embedded in the Mandarin carrier phrase “Qing3 xuan3 cha2]” (Please choose [cha2].), recorded anew for each stimulus. Auditory stimuli were recorded in a sound-attenuated chamber and normalized to a mean amplitude of 70 decibels in Praat (Boersma & Weenink, 2009). Sentences were recorded by a native Mandarin speaker who was born in Taiwan, and learned English at age seven. The speaker was slightly English dominant (with a Mandarin-dominance score of -3 on the BDS and of -4 on the MINT). We used a relatively balanced bilingual speaker, rather than a more strongly Mandarin-dominant speaker, for two reasons: (1) to better match the average language dominance profile of our participants; and (2) because the same speaker recorded stimuli for the novel-word experiment that preceded the present, familiar-word experiment, and the novel-word experiment required recording both Mandarin and English carrier phrases.

To assess whether the speaker who recorded our stimuli was identifiable as a speaker of Mandarin (since she was slightly English-dominant), we presented a new set of Mandarin-English bilingual listeners with sentences recorded by the same speaker, intermixed with one strongly-Mandarin-dominant speaker from mainland China (with a Mandarin-dominance score on the BDS of 13) and four American-English speakers with no prior instruction in Mandarin. Bilinguals (N = 14) were asked to rate speakers on a continuum from “sounds like they are from the US” to “sounds like they are from China or Taiwan.” Participants rated our speaker to be much more likely to be from China or Taiwan than from the US (paired t(13) = 5.37; p < .001); on a scale from 0 to 1, bilinguals rated her compatibility with Chinese-accent 0.78, and her compatibility with American-accent only 0.25. Our speaker’s ratings did not differ significantly from those of the strongly-Mandarin-dominant speaker from China (Chinese-accent compatibility: 0.74; American-accent compatibility: 0.22; t(13) = -0.07, p = 0.94), but did differ
significantly from those of all four English-speaking talkers (average Chinese-accent compatibility: 0.37; American-accent compatibility: 0.60; all t(13) > 5; all p < .001).

Visual stimuli consisted of photographs (taken either from Microsoft Office’s ClipArt online repository—no URL is available because the Microsoft Office ClipArt repository has since been replaced with the Bing image search engine—or from Flickr creative-commons licensed photos), edited and presented on the left and right sides of the computer screen (see Figure 2).

**Apparatus and Procedure**

Participants first completed one of two eye-tracked novel-word-learning experiments (reported in Quam & Creel, 2012, in preparation), then the eye-tracked familiar-word test reported here. They then completed the MINT and BDS language-dominance tests. The language mode of the preceding novel-word experiment was intentionally manipulated across participants. Of the 72 participants, 48 (2/3) learned phonologically and phonotactically Mandarin-compatible novel words, while 24 (1/3) learned strongly English-like novel words. Within the first group of 48 participants, half (24) learned Mandarin-compatible words in a Mandarin language context, in which the research assistant spoke in Mandarin and words were presented in Mandarin carrier phrases. The other half (24)—and the second group of 24 participants who learned English-like words—learned words in an English language context, in which the research assistant spoke in English and words were presented in English carrier phrases. As this latter distinction did not impact tone processing in the novel-word experiment, we included only the former distinction (N=48 who learned Mandarin-compatible novel words vs. N=24 who learned English-compatible novel words) in the “Experiment” factor included in the analysis reported in Footnote 2.
The familiar-word test had two phases: prefamiliarization and test. The constraints of matching vowel- and tone-disambiguated Mandarin words within quadruplets (see Table 3) meant that some of our words were not especially imageable (e.g., ta4, *to tread*). Therefore, in a brief prefamiliarization phase, participants saw pictures one at a time, once each, and heard a sentence containing the picture’s name, so that participants were familiarized with the label for each image (see Creel, Aslin, & Tanenhaus, 2008, for a similar procedure). Then, in the eye-tracked test, they saw two pictures at a time and heard a sentence labeling one of them. The objects’ names contrasted either in tone, vowel, or tone and vowel. The vowel-disambiguated pairs diverged in the first vowel (“mi” vs. “mu”), to make the temporal onset of vowel differences as comparable as possible to the temporal onset of tone differences, which are carried on the vowel. Each word served as the target three times: once each with its vowel, tone, and tone&vowel competitors (making 84 total trials). In each trial, two pictures appeared on the left and right sides of the screen, each 200 x 200 pixels and centered at 25% and 75% of screen width, respectively, and 50% of height. After 500 milliseconds, a sentence, played over headphones, labeled one of the pictures. Participants clicked the computer mouse on the picture they thought matched the last word in the sentence, guessing if necessary. We measured visual fixations to each picture as well as clicking accuracy and reaction time.

An Eyelink Remote eye-tracker (SR Research, Mississauga, Ontario; www.sr-research.com) sampled gaze position every four milliseconds. Participants wore a small target-like sticker on their foreheads, which the EyeLink used to identify their head position. Eyelink software ran on a PC tower in DOS mode. A Mac Mini computer (10.4.1) presented experimental stimuli using custom Matlab (7.6.0, R2008a) scripts that relied on PsychToolBox 3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002).
The Mac sent messages to the PC marking trial onset, sound onset, and response selection; the PC then interpolated these messages with time stamps into the eye tracking data stream. After the experiment, we condensed the data by variables of interest. Looks within a rectangular region centered on each picture location, and extending 100 pixels beyond the picture on each side, were counted as looks to that location (each analyzed region thus subsumed 52.09% of the vertical span of the screen, and 39.06% of the horizontal span of the screen). Looks to target or non-target pictures were then averaged within trial type for each participant, and binned into 50-ms chunks.

Results

Accuracy. Participants clicked on the referents of familiar Mandarin words with high accuracy in all trial types (tone&vowel-disambiguated $M = 99.8\%$, $SD = 0.9\%$; vowel-disambiguated $M = 99.5\%$, $SD = 1.4\%$; tone-disambiguated $M = 98.5\%$, $SD = 2.7\%$). Even in tone-disambiguated trials, where stimuli were minimal tone pairs, the lowest score for any participant was 89.3%, thus confirming that all our bilingual participants were fluent comprehenders of Mandarin.

To statistically analyze clicking accuracy, we first computed the empirical-logistic (e-logit) transform on accuracy scores (Barr, 2008). E-logit transformed values are used for statistical analyses of both clicking and eye-gaze measures throughout to correct for the fact that both these measures are bounded distributions and thus are not normally distributed (see Barr, 2008; Jaeger, 2008). For ease of interpretation, raw means are reported in the text, and raw data is displayed in the gaze time-course plot (Figure 5) and the correlation plots (Figures 3 and 6). In analyses of variance (ANOVAs, by-subject and by-item), we considered the effect of Trial
Type (tone\&vowel, vowel-disambiguated, and tone-disambiguated; within-subjects and within-items) on e-logit-transformed clicking accuracy. Trial Type was a significant predictor of clicking accuracy (F1(2,142) = 8.55, p < .001; F2(2,54) = 7.97, p < .001). Planned comparisons indicated that clicking accuracy in tone\&vowel trials was significantly higher than in tone-disambiguated trials (paired t1(71) = 3.57, p < .001; paired t2(27) = 3.82, p < .001) but not significantly higher than in vowel-disambiguated trials. Clicking accuracy was also significantly higher in vowel-disambiguated than tone-disambiguated trials (paired t1(71) = 2.67, p < .01; paired t2(27) = 2.41, p < .05).

To address whether tone use varied as a function of Mandarin proficiency or Mandarin dominance, we used Principal Components Analysis (PCA) to compute a component based on the shared variance of all three language-dominance measures: MINT, BDS, and Age of Arrival. PCA is an appropriate and parsimonious way to take all three measures of language dominance into account simultaneously, since all three are highly positively correlated (see Table 2). In an ANCOVA with Trial Type as a within-subjects categorical predictor of clicking accuracy and the Mandarin Dominance component as a between-subjects covariate, we found a significant main effect of Trial Type (F(2,140) = 9.27, p < .001), no significant main effect of Mandarin Dominance, but a significant interaction of the two (F(2,140) = 6.92, p < .005). In Pearson’s correlation tests investigating the interaction of Trial Type and Mandarin Dominance, clicking accuracy increased significantly only in tone-disambiguated trials as Mandarin Dominance increased (r = .32, p < .01; Figure 3, middle). This suggests that dominance was particularly affecting use of tone—not vowel—information.

The carrier phrase we used (“Qing3 xuan3 [cha2]” (Please choose [cha2].) ended in Tone 3, the dipping tone, and was recorded naturalistically for each stimulus. A tone sandhi rule in
Mandarin causes a Tone 3 preceding another Tone 3 to essentially change to Tone 2 (the rising tone; Wang & Li, 1967). Thus, trials in which either the target or the competitor word contained Tone 3 (mu3, mi3, shi3, shu3, ta3, or tu3) might have been disambiguated more quickly than other trials, particularly for more proficient bilinguals. To investigate this possibility, we conducted an analysis of covariance (ANCOVA) on tone-disambiguated trials, with Tone Type as a within-subjects factor (trials either disambiguated by Tone 3 or not) and the Mandarin Dominance PCA component as a between-subjects covariate. The ANCOVA revealed a main effect of Tone Type (F(1,70) = 15.71, p < .001), indicating slightly lower accuracy for Tone 3—disambiguated trials (98.15%) than for other trials (98.78%). As reported in the previous paragraph. Mandarin Dominance was significant (F(1,70) = 8.10, p < .01), and it also interacted with Tone Type (F(1,70) = 7.64, p < .01). Pearson’s correlations investigating the interaction revealed that accuracy increased significantly with Mandarin Dominance only in trials disambiguated by Tone 3 (r = .43, p < .001), not in trials with other tones.

——INSERT FIGURE 3 HERE——

Reaction Times. Reaction times (RTs) provide a more continuous measure that might reveal processing differences even when participants ultimately clicked on the correct picture. We first excluded RTs that were more than 3 standard deviations above or below the mean for each participant (this excluded 2.33% of trials). For trials in which participants clicked on the correct picture, we then considered the effect of Trial Type on RT in analyses of variance (ANOVAs, by-subject and by-item). Trial Type was a significant predictor of RT (F1(2,142) = 29.1, p < .001; F2(2,54) = 11.67, p < .001). Planned comparisons confirmed that RTs were
significantly slower in tone-disambiguated trials \((M = 830 \text{ ms}, SD = 180 \text{ ms})\) than in either vowel-disambiguated trials \((M = 750 \text{ ms}, SD = 140 \text{ ms}; \text{ paired } t1(71) = -5.29, p < .001; \text{ paired } t2(27) = -3.63, p < .005)\) or tone\&vowel-disambiguated trials \((M = 730 \text{ ms}, SD = 130 \text{ ms}; \text{ paired } t1(71) = -6.63, p < .001; \text{ paired } t2(27) = -5.05, p < .001)\); vowel- vs. tone\&vowel-disambiguated trials did not differ significantly.

To address whether tone use varied as a function of Mandarin proficiency or Mandarin dominance, we conducted an ANCOVA on RT with Trial Type as a within-subjects categorical predictor and the Mandarin Dominance component as a between-subjects covariate. We found a significant main effect of Trial Type \((F(2,140) = 31.97, p < .001)\) but no significant main effect of Mandarin Dominance. However, the two factors interacted \((F(2,140) = 7.81, p < .001)\). In Pearson’s correlation tests investigating the interaction of Trial Type and Mandarin Dominance, RTs decreased significantly in tone-disambiguated trials as Mandarin dominance increased \((r = -.29, p = .01; \text{ Figure 4, middle})\), but not in vowel-disambiguated trials or tone\&vowel-disambiguated trials. The RT results thus converge with the accuracy results and again indicate that Mandarin dominance was particularly affecting use of tone (not vowel) information.

As with accuracy, we compared RT for trials containing Tone 3—bearing words vs. words bearing other tones, in an ANCOVA on tone-disambiguated trials, with factors Tone Type (disambiguated by Tone 3 or not) and the Mandarin Dominance covariate. Unlike in the ANCOVA on accuracy, the interaction of Tone Type and Mandarin Dominance did not reach significance \((F < 2)\).
Visual fixations. Analyzing eye movements provides additional information about participants’ fine-grained temporal processing, and can reveal processing dynamics even before participants have decided which picture to select. Participants’ eye-gaze responses across time are depicted in Figure 5, which displays raw target minus competitor fixations or “target advantage” in correct trials only. This number ranges from roughly 0, chance looking, to 1, looking only at the target. Trial lengths were variable, ending when participants clicked on a picture, so we extended the final fixation of each trial to 2000 ms so that all trials contributed equally across the full time course.

To statistically compare responses, we computed the empirical-logistic (e-logit) transform on target fixations and competitor fixations (Barr, 2008) and analyzed the target – competitor difference score, or “target advantage,” for these transformed values. We then averaged e-logit-transformed target advantage across the time window 200–800 milliseconds (ms) after the onset of the noun. The start of this window represents the earliest point at which adults can initiate an eye-movement response (Hallett, 1986). The end of this window, 800 ms, was selected to balance two opposing factors: best reflecting the asymptotes in Figure 5, while keeping the percentage of the analyzed time window that had been extended from the final fixation (see previous paragraph) as low as possible (10.03%). Two of the 72 participants were excluded from the gaze analyses (leaving 70 included participants) because their target + competitor fixations in correct trials were below 80% during the analyzed time window, indicating poor eye-tracking quality.
For trials in which participants clicked on the correct picture, we first considered the effect of Trial Type on target advantage in an analysis of variance (ANOVA, by-subject and by-item). Trial Type was a significant predictor of target advantage (F1(2,138) = 48.77, p < .001; F2(2,54) = 14.41, p < .001). Planned comparisons confirmed that target advantage was higher in tone\&vowel-disambiguated trials (M = 0.58; SD, 0.11) than both vowel-disambiguated (M = 0.54, SD, 0.11; paired t1(69) = 3.13, p < .005; paired t2(27) = 1.98, p = .058) and tone-disambiguated trials (M = 0.43, SD, 0.14; paired t1(69) = 9.01, p < .001; paired t2(27) = 5.56, p < .001), and higher in vowel-disambiguated than tone-disambiguated trials (paired t1(69) = 6.70, p < .001; paired t2(27) = 2.99, p < .01).

To address whether tone use varied as a function of Mandarin proficiency or Mandarin dominance, we conducted an ANCOVA on target advantage, with Trial Type as a within-subjects categorical predictor and the Mandarin Dominance component as a between-subjects covariate.\(^2\) We found significant main effects of Trial Type (F(2,136) = 50.26, p < .001) and of Mandarin Dominance (F(1,68) = 4.93, p < .05), and a significant interaction of the two (F(2,136) = 3.11, p < .05). In Pearson’s correlation tests investigating the interaction of Trial Type and Mandarin Dominance, target advantage increased significantly as Mandarin dominance increased.

\(^2\) The gaze analyses of covariance also revealed an interaction of the language-dominance component with experiment (which novel-word-learning experiment participants had completed immediately prior to the present experiment). An ANCOVA including Experiment, Trial Type, Language Dominance, and their interactions as predictors revealed a main effect of Trial Type (F(2,132) = 49.46, p < .001), a main effect of Language Dominance (F(1,66) = 3.27, p = .026), no significant main effect of Experiment, but a significant interaction of Language Dominance and Experiment (F(1,66) = 4.85, p = .031) and a marginal interaction of Language Dominance and Trial Type (F(2,132) = 3.06, p = .050). Pearson’s correlation tests to investigate the interaction of Language Dominance and Experiment were conducted separately in each trial type to make them comparable to the tests reported in the main text. These correlation tests revealed that the effects of Language Dominance on target advantage in tone- and tone\&vowel-disambiguated trials were driven by participants who learned Mandarin-compatible novel words in the prior experiment (tone\&disambiguated: r = 0.46, p = .001; tone\&vowel\&disambiguated: r = 0.44, p = .002) rather than English-compatible novel words (tone\&disambiguated: r = 0.05, p = .81; tone\&vowel\&disambiguated: r = -.06, p = .80). It is possible that the preceding manipulation put the latter group in a more English-like processing mode and thus dampened effects of language dominance. The latter group was also roughly half the size (N=23 included in gaze analyses) of the former group (N=47 included in gaze analyses). Importantly, Experiment did not interact with the critical Language Dominance x Trial Type interaction.
in both tone-disambiguated trials ($r = .31, p = .01$; **Figure 6**, middle) and tone\&vowel trials ($r = .26, p < .05$; **Figure 6**, right), but not in vowel-disambiguated trials.

Again, we compared gaze patterns for trials containing Tone 3—bearing words vs. words bearing other tones, in an ANCOVA on tone-disambiguated trials, with factors Tone Type (disambiguated by Tone 3 or not) and the Mandarin Dominance covariate. There was a significant main effect of Tone Type ($F(1,68) = 83.29, p < .001$), indicating higher target advantage scores in Tone 3—disambiguated trials (0.58) than in other trials (0.32), suggesting that listeners overall were using the tone sandhi information to execute eye movements more quickly on these trials. As above, Mandarin Dominance was significant ($F(1,68) = 10.11, p < .005$). The two factors also significantly interacted ($F(1,68) = 14.31, p < .001$). Pearson’s correlations investigating the interaction revealed that target advantage increased significantly with Mandarin Dominance only in trials disambiguated by Tone 3 ($r = .45, p < .001$), not in trials with other tones.

For both gaze and accuracy, effects of Mandarin Dominance on tone processing seemed to particularly impact trials in which either the target word or competitor contained Tone 3, providing some evidence that the most highly proficient Mandarin listeners were most able to exploit tone sandhi to anticipate the target word. However, for the reaction-time measure, Mandarin Dominance did not interact significantly with Tone Type (whether the target and competitor were disambiguated by Tone 3 or not), perhaps because reaction time is assaying a later stage of processing than gaze. Additionally, listeners overall were actually slightly less accurate in trials disambiguated by Tone 3, which is the reverse of the effect in the gaze data. Thus, while we find some evidence that the ability to exploit tone sandhi in sentence processing
was contributing to effects of Mandarin dominance on tone processing, the contribution of tone sandhi to our effects is somewhat nuanced.

Along with the accuracy and reaction-time data, the gaze results provide converging evidence that Mandarin dominance was particularly affecting use of tone information, but not vowel information. However, one potential confound that we wanted to rule out was that because our participants were from a range of geographic areas (e.g., Taiwan vs. different areas of mainland China), if the more Mandarin-dominant participants tended to be from Taiwan, they would be more familiar with the speaker’s accent and this—rather than their Mandarin dominance—could account for their better tone recognition. We addressed this by calculating the average Mandarin-dominance scores (assessed via the MINT vocabulary test) for groups from different regions of origin. Of the 72 participants, 36 had Taiwanese dialect exposure (28 had lived in Taiwan, and 8 were born in the US); 26 had mainland-Chinese dialect exposure (21 had lived in mainland China; 4 were born in the US; 1 was born in Singapore); 6 were born in the US and did not report dialect information; and 4 had other (3) or unknown (1) backgrounds. The 26 participants with mainland Chinese dialect exposure tended to be, if anything, slightly more Mandarin-dominant ($M = -2.23, SD = 15.48$) than the 36 with Taiwanese dialect exposure ($M = -5.58, SD = 11.00$), so Mandarin dominance is not conflated with Taiwanese dialect exposure. Thus, familiarity with the speaker’s dialect is not likely to explain the relationship we found between Mandarin dominance and ease of tone processing.

—INSERT FIGURE 6 HERE—
Discussion

We asked how listeners with varying levels of tone- vs. non-tone-language experience process lexical tones vs. vowels in familiar words. The most interesting outcome of this experiment was the effect of language dominance on tone processing: bilinguals with more experience with English showed less use of tone information—that is, they showed tone attrition. These effects of language dominance on tone processing are notable given that all participants had learned Mandarin from birth.

All three measures (accuracy, RT, and gaze) revealed effects of Mandarin dominance in tone-disambiguated trials, and the gaze measure revealed additional effects of Mandarin dominance in tone&vowel trials. In tone&vowel trials, tone information was helpful for word disambiguation, but words also differed in their segments. Thus, these trials might have patterned intermediately between tone-only and vowel-only trials, in terms of the impact of language dominance. However, conclusions based on the gaze measure alone should be made with caution. It might be that gaze is a more sensitive measure than accuracy and RT, particularly in tone&vowel trials, where accuracy was 99.8% on average. Still, we cannot be sure why this pattern emerged only for gaze.

Interestingly, effects of Mandarin dominance on accuracy and gaze in tone-disambiguated trials appear to be at least partly driven by those trials in which tone-sandhi information disambiguated the target word from the competitor before the onset of the target word. Tonal coarticulation is pervasive in Mandarin sentences, and listeners compensate for this coarticulation (Xu, 1994). Thus, it is possible that attrition impacts the ability to recognize tones in fluent speech more than (or before) it impacts memory for individual words’ tones. However, the RT measure did not show this pattern, and accuracy for bilinguals overall was actually lower
in Tone 3 trials, so further research is needed to tease apart the relative contributions of fluent use of tone sandhi vs. word-level tone recognition to tone attrition effects.

One possibility is that when the target or competitor contained Tone 3, this provided an additional tonal cue to the distinction between the words, as two words—not just the target word, but also the last word in the carrier phrase—contrasted in their tone realizations. For example, tone realizations throughout the phrase for the target words ‘shi1’ vs. ‘shi3’ would differ in two tones: “Qing2 xuan3 shi1.” vs. “Qing2 xuan2 shi3,” because the tone sandhi rule caused the Tone 3 on the word “xuan” to change to Tone 2 before another tone 3 (in “shi3”). By contrast, tone realizations for phrases ending in ‘xian1’ vs. ‘xian4’ would only differ in the tone of the target word: “Qing2 xuan3 xian1.” vs. Qing2 xuan3 xian4.”

The attrition we found in the ability to fluently process tones in an L1 after experience with a non-tonal language has implications for the question of the extent to which bilinguals’ two phonetic systems interact. It suggests that bilinguals may not be processing pitch information language specifically, but rather that when English has become the dominant language, processing strategies from English impact phonetic processing in the non-dominant language, Mandarin—even when Mandarin was learned natively. Whether weaker use of tone information in English-dominant bilinguals results from more exposure to English, or from less exposure to Mandarin, is not clear. Nevertheless, it is interesting that tone processing appears to be affected more than vowel processing, and that processing of tones in fluent speech (i.e., the ability to exploit tone sandhi to anticipate the target word) seems to be affected, in addition to memory for tones in representations of individual words.

Further, our investigation extends previous research on on-line recognition of lexical tones (Malins & Joanisse, 2010) to consider additionally the impact of bilingual language
dominance on tone processing. In an eye-tracking study, Malins and Joanisse (2010) found that Mandarin speakers exploited tones as rapidly as segments in recognizing familiar words. Our findings seem to contradict Malins and Joanisse’s in that we found that tone information was utilized more slowly than vowel information. However, the seeming discrepancy between our findings and those of Malins and Joanisse may be driven by two factors.

First, Malins and Joanisse (2010) designed their stimuli so that word pairs diverged in the second vowel of diphthongs, with the idea that tone contrasts take longer to disambiguate than vowel contrasts. Our vowel contrasts diverged from the first vowel in the word. In a separate experiment using a gating paradigm, we examined the time-course of tone vs. vowel differentiation by an additional set of bilingual listeners. Listeners generally disambiguated vowel contrasts more quickly than tone contrasts—except when tone contrasts could be disambiguated earlier because of tone sandhi preceding tone 3—supporting an information-availability account of slower tone-based word identification.

Second, it is possible (and would be consistent with the Mandarin-dominance effect we observed) that our sample of bilinguals, while all were native Mandarin-speakers, may all use tone information more slowly than monolingual Mandarin listeners would. Malins and Joanisse (2010) reported that their Mandarin-English bilingual participants “were from Mainland China and had been living in North America for a mean of 3.9 years.” Based on their description, it is possible that their bilingual participants were more Mandarin dominant on average than ours.

Our interpretation of our results is that increased exposure to English and reduced exposure to Mandarin led bilinguals to less successfully exploit tone information. We propose that this reflects the fact that pitch information—as opposed to the segmental contrasts we used—is used in fundamentally different ways in English (an intonation language) vs. Mandarin
(a tone language). Thus, exposure to English, in which lexical pitch changes do not indicate changes in meaning, caused interference or competition with attention to tonal contrasts in Mandarin, worsening Mandarin word recognition. Related investigations of effects of dominance on tone processing in newly learned words (e.g., Quam & Creel, 2012, in preparation; Singh & Quam, under review) will help address these questions by controlling for word familiarity across individual listeners. They will also shed light on how representations of tone information are formed.

**Limitations and alternative explanations**

It is important to acknowledge here that there are at least two possible explanations for the selective attrition effect we found for tones but not vowels. The first potential explanation, and the one we favor, is that tone perception has undergone more attrition than vowel perception for our listeners because tone is not used at the lexical level in English—individual English words do not contain an underlying tonal pattern. Thus, experience with English might have weakened tone discrimination while maintaining vowel discrimination, as the Mandarin vowels we used (see Table 3)—and many Mandarin vowels in general—could likely be assimilated into English vowel categories. Future work could test this explanation by comparing attrition of tones among speakers of a tonal L1 and a non-tonal L2 to attrition of L1 vowels that fall within a single L2 category. For instance, the Mandarin vowel system contains roundedness contrasts that English does not. If these vowel contrasts display the same attrition effects as tonal contrasts, this would support an explanation based on overlap between L1 and L2 phonetic categories.

The second possible explanation for the selective tone attrition we found is that tones are qualitatively different from consonants and vowels in terms of their representations. Evidence
suggests that tones constrain lexical access less efficiently than other segments (Lee, 2007; Cutler & Chen, 1997; Ye & Connine, 1999; Davis, Schoknecht, Kim, & Burnham, 2015), and that listeners are less aware of lexical tone than segmental content even in languages with orthographic systems for representing tone (Davis, Schoknecht, Kim, & Burnham, 2015; Burnham, Luksaneeyanawin, Kantamphan, & Reid, 2013). Perhaps, due to weaker representational strength, tones are more vulnerable to attrition than vowels. However, we found that Mandarin proficiency was related to the ability to use tone (including tone sandhi cues) to rapidly disambiguate the target and competitor words. Attrition seemed to affect both fluent use of tone during sentence processing and knowledge of individual words’ tones. The impact of attrition on use of anticipatory tone sandhi cues we found in the gaze measure suggests that listeners are exploiting tone information to constrain word recognition as rapidly as possible. This argues against the notion that tones constrain word recognition more slowly than other segments, and is more consistent with the notion that tone information is used as soon as it becomes available.

Still another possible explanation of our results is that much of the tone attrition we found was driven by word-level familiarity or “rustiness” with particular words. However, there are several reasons that familiarity with the tone patterns of particular words is unlikely to fully account for the dominance effects. First, the fact that Mandarin dominance seemed to impact the use of tone sandhi, rather than attention to tone more globally, calls a word-level explanation into question. Second, accuracy even for minimal tone pairs was very high for all bilinguals: 98.5% on average, and no lower than 89% for individual participants. Another reason to doubt word familiarity as the explanation for dominance effects is that neither accuracy, gaze, nor RT scores improved with increasing word frequency (calculated from the Modern Chinese Character
Frequency List, \url{http://lingua.mtsu.edu/chinese-computing/statistics/char/list.php?Which=MO}; Da, 2004). Pearson’s correlation tests investigating the relationship between tone processing and word frequency in fact showed a significant negative correlation for gaze ($r = -0.39$, \( p < .05 \)), and non-significant correlations for accuracy and RT (also in the opposite direction of what would be predicted based on a word-familiarity explanation). Also, even when only correct trials were considered, language dominance still impacted participants’ ability to use tone to visually identify the target picture and to click rapidly on it. This indicates that even when participants knew the tone content of familiar words—which was true 98.5% of the time—their ability to efficiently map words onto referents using tone was impacted by their experience with English (or reduced exposure to Mandarin).

Finally, on a word-familiarity account of language dominance effects on tone attrition, less-Mandarin-dominant participants should be less accurate and slower in word recognition across the board, regardless of the type of phonological competitor present on a given trial. Yet we did not find dominance effects in vowel-disambiguated trials, but instead only when tone content helped differentiate words. A potential alternative explanation for this difference is that vowel contrasts could have been more discriminable (or more rapidly disambiguated) than tones. As contour tones unfold over the course of the vowel, the possibility that tone discriminations are harder (or slower) to make than vowel discriminations could be an intrinsic feature of tone processing. Nevertheless, Mandarin has a relatively small inventory of tones compared with other contour-tone languages (e.g., Cantonese). Just as languages with a small vowel inventory tend to use the point vowels because they are maximally distinctive, the 4 tones in Mandarin take advantage of the full pitch space (high level tone 1, T1, vs. low dipping T3; rising T2 vs. falling T4). The most confusable contrast is between tones 2 and 3 (Shen & Lin, 1991), and for that
reason we did not include that pair. Still, future work should assess whether listeners could maintain highly similar vowel distinctions despite losing tone distinctions.

In conclusion, the present study found a relationship between long-term language experience and listeners’ use of lexical-tone content for recognizing words in fluent speech. For native speakers of Mandarin with varying experience with Mandarin vs. English, tone use during recognition of familiar Mandarin words was correlated with degree of Mandarin language dominance. Interestingly, for the gaze and accuracy measures, this relationship was strongest when the final word of the carrier phrase contained tone sandhi cues to the tone of the target word. Thus, degree of dominance in Mandarin seemed to impact both the ability to rapidly exploit tone sandhi cues during sentence processing to anticipate the tone of the target word and recognition of the word itself. This suggests that phonetic processing strategies from the dominant language are impacting sentence processing and word recognition in the nondominant language, even when the latter was learned natively.

Acknowledgements

This work would not have been possible without the tremendous assistance of bilingual undergraduate research assistants Sally Chou, Abraham Chen, Vivian Yu, Allen Hsu, Christina Lin, Guan Wang, Angela Wang, and Kaili Guo, as well as lab manager Heather Pelton. Tamar Gollan graciously allowed us to use her Multilingual Naming Test materials; Alexandra Dunn and Jean Fox Tree responded thoughtfully and promptly to questions about their Bilingual Dominance Scale; Jing Shen and Crane Huang consulted on the Mandarin language; and Roger Levy consulted on both Mandarin and statistics. CQ was supported by NIH institutional training
grant T32 DC00041-12 via Center for Research in Language, UCSD; NIH F32 HD065382; and NIH K99 DC013795. SCC was supported by NSF BCS-1057080 and NSF BCS-1230003.

References


Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of


Table 1: Measures of bilingual language experience.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINT Mandarin</td>
<td>51.47 (10.56)</td>
<td>20 – 64</td>
</tr>
<tr>
<td>MINT English</td>
<td>57.81 (5.87)</td>
<td>40 – 68</td>
</tr>
<tr>
<td>MINT composite</td>
<td>-6.33 (14.29)</td>
<td>-38 - 18</td>
</tr>
<tr>
<td>Bilingual Dominance Scale</td>
<td>-1.47 (10.51)</td>
<td>-23 - 16</td>
</tr>
<tr>
<td>Age of arrival (years)</td>
<td>8.32 (6.01)</td>
<td>0 - 20</td>
</tr>
</tbody>
</table>
Figure 1: MINT vocabulary scores in Mandarin vs. English.
Table 2: Correlations between language measures. Multilingual Naming Test (MINT), Bilingual Dominance Scale (BDS), and age of arrival in English-speaking country (AOA) were all highly correlated.

<table>
<thead>
<tr>
<th></th>
<th>MINT Mandarin</th>
<th>MINT English</th>
<th>MINT composite</th>
<th>BDS</th>
<th>AOA</th>
<th>PCA component</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINT Mandarin</td>
<td>-.47***</td>
<td>.93***</td>
<td>.66***</td>
<td>.62***</td>
<td>.80***</td>
<td></td>
</tr>
<tr>
<td>MINT English</td>
<td>-.76***</td>
<td>-.68***</td>
<td>-.78***</td>
<td>-.80***</td>
<td>.92***</td>
<td></td>
</tr>
<tr>
<td>MINT composite</td>
<td></td>
<td>.77***</td>
<td>.77***</td>
<td>.91***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BDS</td>
<td></td>
<td></td>
<td></td>
<td>.75***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.92***</td>
</tr>
</tbody>
</table>

***p < .001
Table 3: Seven familiar-word quadruplets. Spellings are in Pinyin Romanization. Moving horizontally within each quadruplet creates a tone contrast (e.g., “hua1” vs. “hua4”); moving vertically, a vowel contrast (e.g., “hua1” vs. “he1”). Character frequencies, in italics, were calculated using a written corpus containing 193,504,018 total character tokens (http://lingua.mtsu.edu/chinese-computing/statistics/char/list.php?Which=MO; Da, 2004).

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cha1</td>
<td>cha2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5,299</td>
<td>19,404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chuang1</td>
<td>chuang2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27,283</td>
<td></td>
<td></td>
</tr>
<tr>
<td>he1</td>
<td>he4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32,112</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3,133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mu3</td>
<td>mu4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71,462</td>
<td>54,433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mi3</td>
<td>mi4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69,668</td>
<td>6,919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qiu1</td>
<td>qiu2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24,030</td>
<td>63,913</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shu1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151,357</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tu3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79,809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xian1</td>
<td>xian4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210,628</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xin1</td>
<td>xin4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>392,228</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5,299</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27,283</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32,112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71,462</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69,668</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24,030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151,357</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79,809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210,628</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>392,228</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: **Example photographs used as visual stimuli, from one of the seven quadruplets.** The 4 pictures correspond to the following words: cha1 (fork), cha2 (tea), chuang1 (window), and chuang2 (bed). Photographs appeared in pairs. Each picture appeared with all 3 competitors across trial-types.
Figure 3: Clicking accuracy plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score, on the MINT test) in the three trial-types. Lines are linear-regression models. Y-axis range is restricted to allow visualization of correlations at high accuracy, and data points are jittered to improve their visibility.
Figure 4: Reaction times plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score, on the MINT test) in the three trial-types. Lines are linear-regression models, and data points are jittered to improve their visibility.
Figure 5: Eye-gaze target-advantage across time. Only correct trials are included. Throughout, error bars are standard errors.
Figure 6: Target advantage (gaze) plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score, on the MINT test) in the three trial-types. Lines are linear-regression models, and data points are jittered to improve their visibility.