Research Article

Tone Attrition in Mandarin Speakers of Varying English Proficiency

Carolyn Quam^a and Sarah C. Creel^b

Purpose: The purpose of this study was to determine whether the degree of dominance of Mandarin–English bilinguals’ languages affects 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 2 languages and ages of acquisition of English. Participants viewed 2 photographs at a time while hearing a familiar Mandarin word referring to 1 photograph. The names of the 2 photographs diverged in tone, vowels, or both. Word recognition was evaluated using clicking accuracy, reaction times, and an online recognition measure (gaze) and was 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’ 2 phonetic systems interact. It suggests that bilinguals may not process pitch information language-specifically and that processing strategies from the dominant language may affect phonetic processing in the nondominant language—even when the latter was learned natively.

How does becoming bilingual affect 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 and ask whether experience with English affects 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 possible tonal patterns: high flat (tone 1; T1), rising (tone 2; T2), low dipping (tone 3; T3), and falling (tone 4; 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 such as 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, whereas 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; Quam & Swingley, 2010; Singh, Hui, Chan, & Golinkoff, 2014).

Monolingual learners of tone languages versus intonation languages appear to tune their pitch-processing strategies to their native language by an early age. However, many people speak both a tone language and a nontone language or multiple tone languages with different tonal

^aCenter for Research in Language, University of California, San Diego

^bDepartment of Cognitive Science, University of California, San Diego

Carolyn Quam is now at the Department of Speech and Hearing Sciences, Portland State University

Correspondence to Carolyn Quam: cquam@pdx.edu

Editor: Jody Kreiman

Associate Editor: Ewa Jacewicz

Received July 16, 2015

Revision received January 14, 2016

Accepted March 7, 2016

DOI: 10.1044/2016_JSLHR-S-15-0248

Disclosure: The authors have declared that no competing interests existed at the time of publication.

contours. Interpreting lexical-pitch contours language-specifically would confer a strong benefit to tone- and nontone-language bilinguals' word-recognition efficiency. Very little work has examined differential recognition of pitch cues. What happens to tone processing when individuals learn both a tone language and a nontone language is an open question. Do listeners maintain language-specific tone processing despite learning and using a language without tone distinctions? If so, experience with English should not affect 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 on the basis of evidence that bilingual listeners can process consonants differently in each of their languages. When enough information is provided to indicate the language context (for an overview, see Gonzales & Lotto, 2013), 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 such as /b/ versus /p/ in English versus 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 does English.

Evidence of language-specific consonant processing might predict that bilinguals would also process tones language-specifically. However, there are several reasons why experience with a nontone 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 (de Bot, 1992; MacWhinney, 2005; Sebastián-Gallés, Echeverría, & Bosch, 2005). 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 (L1; 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 surprising, Chang (2011) found that adult native-English speakers' English vowels began to shift midway through an intensive 6-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 effect of these factors on tone processing.

The second reason why we might find trade-offs between bilinguals' lexical-pitch processing in a tone versus 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) to contrast consonants but divided it at a different point on the voice-onset-time continuum. For the bilinguals tested here, pitch contours such as rises versus 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 not clear 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 present investigation of particular interest.

The third reason to predict limits on bilinguals' ability to match their pitch and 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 (for a review, see Burnham & Mattock, 2007). Weak status in word representations theoretically could make tone more prone to attrition. Tone appears to be less useful for lexical access and word recognition in Mandarin than segments (Cutler & Chen, 1997; Lee, 2007; 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 vowels (particularly steady-state vowels). Listeners may exploit tone as soon as it is perceptually available, which is simply later than for 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 Burnham, Luksaneeyanawin, Kantamphan, & Reid, 2013; McBride-Chang et al., 2012). 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 investigated Mandarin–English bilinguals' tone processing in Mandarin. For the three reasons mentioned previously, we predicted that experience with a nontonal 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 Joannis (2010), who were interested in Mandarin speakers' relative processing speed for tones versus 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 effect of bilingual language dominance on use of tones versus 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* [“mother”] 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 on the basis of 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 (RT) to click on the target picture.

Our first question was whether relative proficiency in Mandarin versus 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 in Mandarin and scoring higher on the Bilingual Dominance Scale (BDS; Dunn & Fox Tree, 2009), which measures language background and current language use. 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 tonal and segmental contrasts (in which words diverged at their vowels). Although lexical tone is contrastive in Mandarin but not in English, most Mandarin vowels—including the ones we used—could likely be assimilated to English vowel categories (Best, 1994; Best & McRoberts, 2003). As a result, experience with English might not affect processing of these Mandarin vowel contrasts to the same degree that it affects processing of Mandarin tones. Thus, we may find selective attrition for lexical tone (rather than vowels) with increasing English proficiency (and decreasing Mandarin proficiency).

Although 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 also 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 or phonetic dimension such as pitch, then this more holistic process of attrition should affect 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/he4* and *shi1/shi3* or the fricatives in *xin1/xin4–xian1/xian4* and *qiu1/qiu2–quan1/quan2*) or because of general Mandarin subphonemic and 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 affect processing of tones selectively or instead affect both tone and vowel processing. If we find selective effects of language dominance on lexical-tone processing, however, this will provide evidence that processing of that dimension in the native language (Mandarin) is impaired by exposure to the other language (English) when a bilingual’s two languages conflict in how they treat a particular acoustic dimension in their lexical phonology. 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 et al., 2015).

Method

Participants

Participants were 72 Mandarin–English bilingual adults (49 women and 23 men, mean age = 20 years, *SD* = 2, range = 16–26 years; parental permission was obtained when necessary; age data are missing for one participant who was approximately 19 years old) who had participated in one of two novel word-learning experiments immediately before this study (Quam & Creel, 2012, in press). There were no effects of (or interactions with) the prior experiment in any of the analyses presented here, except where noted (see Footnote 1), so we have included analyses without that factor for simplicity of presentation. Thirteen additional participants were excluded from the analysis for having as much, or more, exposure to other tone languages or dialects as to Mandarin during childhood (seven); responding incorrectly in one third of tone contrast trials, indicating low proficiency (one); and experimenter error (five). Two of the 72 participants were excluded only 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 L2 learners of English rather than simultaneous bilinguals. However, 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.

Table 1. Measures of bilingual language experience.

Measure	<i>M (SD)</i>	Range
Multilingual Naming Test		
Mandarin	51.47 (10.56)	20 to 64
English	57.81 (5.87)	40 to 68
Composite	−6.33 (14.29)	−38 to 18
Bilingual Dominance Scale	−1.47 (10.51)	−23 to 16
Age of arrival (years)	8.32 (6.01)	0 to 20

The Multilingual Naming Test (MINT; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 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 and went 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 that a participant was strongly Mandarin dominant, whereas a large negative number meant that a participant was strongly English dominant. English scores were higher than Mandarin scores (see Table 1; $t(72) = 3.65, p < .001$) and less variable. Mandarin MINT scores and English MINT scores were negatively correlated (see Figure 1; $r = -.46, p < .001$). MINT scores in each language also correlated with the composite (Mandarin minus English) difference score (see 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 BDS (Dunn & Fox Tree, 2009) comprises 12 questions evaluating bilinguals' lifelong experience with each language (e.g., "At what age did you first learn or 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 "Which language do you use in your college dorm or house?" and "Which language do you use with your family?" The BDS assigns points to each language depending on participants' responses (for scoring, see Dunn & Fox Tree, 2009). 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 (see 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 (see Table 2; see Table 1 for descriptive statistics).

Stimuli

We designed a set of seven quadruplets of monosyllabic familiar Mandarin words (see Table 3) that was loosely based on stimuli created by Malins and Joannis (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. T2 and T3, 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.

When composing the quadruplets, words were selected (a) so that each word had both a tone- and a segment-disambiguated competitor and (b) to maximize imageability as much as possible so that the words 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 corpus of 193,504,018 total written characters (Da, 2004, 2005). We chose to use written character frequencies for two reasons. First, we used written rather than spoken frequencies because the spoken corpora currently available (e.g., Canavan & Zipperlen, 1996) are small. Second, we used character frequencies rather than sound-based frequencies (e.g., syllable or morpheme)—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.

Figure 1. Multilingual Naming Test vocabulary scores in Mandarin versus English.

Table 2. Correlations between language measures.

	MINT Mandarin	MINT English	MINT composite	BDS	AOA	PCA component
MINT Mandarin		-.47*	.93*	.66*	.62*	.80*
MINT English			-.76*	-.68*	-.78*	-.80*
MINT composite				.77*	.77*	.92*
BDS					.75*	.91*
AOA						.92*

Note. Multilingual Naming Test (MINT), Bilingual Dominance Scale (BDS), and age of arrival in an English-speaking country (AOA) were all highly correlated. PCA = principal-components analysis.

* $p < .001$.

All included characters were ranked 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 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: .38:1; vowel disambiguated: .37:1; tone&vowel-disambiguated: .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 dB in Praat (Boersma & Weenink, 2009). Sentences were recorded by a native Mandarin speaker who was born in Taiwan and learned English at age 7 years. The speaker was slightly English dominant (with a Mandarin dominance score of -3 on the BDS and -4 on the MINT). We used a relatively balanced bilingual speaker rather than a more strongly Mandarin-dominant

speaker for two reasons: (a) to better match the average language-dominance profile of our participants and (b) 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 (because she was slightly English dominant), we presented a new set of Mandarin-English bilingual listeners with sentences recorded by the same speaker intermixed with sentences recorded by 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 by clicking on a graphically depicted continuum with labeled endpoints, *sounds like they are from the United States* to *sounds like they are from China or Taiwan*. Participants rated our speaker as much more likely to be from China or Taiwan than from

Table 3. Seven familiar-word quadruplets.

No.	Quadruplet					
	Mandarin word	English meaning	Character frequency	Mandarin word	English meaning	Character frequency
1	<i>cha1</i>	“fork”	5,299	<i>cha2</i>	“tea”	19,404
	<i>chuang1</i>	“window”	26,978	<i>chuang2</i>	“bed”	27,283
2	<i>mu3</i>	“mother”	71,462	<i>mu4</i>	“tree”/“wood”	54,433
	<i>mi3</i>	“rice”	69,668	<i>mi4</i>	“honey”	6,919
3	<i>shi1</i>	“lion”	5,456	<i>shi3</i>	“feces”	1,257
	<i>shu1</i>	“book”	151,357	<i>shu3</i>	“mouse”/“rat”	10,476
4	<i>xian1</i>	“first place”	210,628	<i>xian4</i>	“thread”	98,261
	<i>xin1</i>	“heart”	392,228	<i>xin4</i>	“envelope”	230,248
5	<i>hua1</i>	“flower”	103,367	<i>hua4</i>	“painting”	38,263
	<i>he1</i>	“to drink”	32,112	<i>he4</i>	“crane”	3,133
6	<i>qiu1</i>	“autumn”	24,030	<i>qiu2</i>	“ball”	63,913
	<i>quan1</i>	“circle”	17,115	<i>quan2</i>	“fountain”	11,237
7	<i>ta3</i>	“pagoda”	29,454	<i>ta4</i>	“to tread on”/“to step on”	9,656
	<i>tu3</i>	“dirt”	79,809	<i>tu4</i>	“rabbit”	4,418

Note. Spellings are in Pinyin romanization. Moving horizontally within each quadruplet creates a tone contrast (e.g., *hua1* vs. *hua4*); moving vertically within each quadruplet creates a vowel contrast (e.g., *hua1* vs. *he1*). Character frequencies were calculated using a written corpus containing 193,504,018 total character tokens (Da, 2004, 2005).

Figure 2. Example photographs used as visual stimuli from one of the seven quadruplets. The four pictures correspond to the following words: *cha1* (“fork”), *cha2* (“tea”), *chuang1* (“window”), and *chuang2* (“bed”). Photographs appeared in pairs. Each picture appeared with all three competitors across trial types.

the United States (paired $t(13) = 5.37, p < .001$); on a scale of 0 to 1, bilinguals rated her Chinese-accent compatibility .78 and her American-accent compatibility only .25. Our speaker's ratings did not differ significantly from those of the strongly Mandarin-dominant speaker from China (Chinese-accent compatibility: .74; American-accent compatibility: .22; $t(13) = -0.07, p = .94$), but did differ significantly from those of all four English-speaking talkers (average Chinese-accent compatibility: .37; American-accent compatibility: .60; all $t(13) > 5.00$, all $p < .001$).

Visual stimuli consisted of photographs (taken either from Microsoft Office's ClipArt online repository or from Flickr Creative Commons-licensed photos) that were 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 press) and 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 English-like novel words in an English language context. Within the first group, of 48 participants, 24 (1/2) learned words in a Mandarin language context (i.e., the research assistant

spoke in Mandarin, and words were presented in Mandarin carrier phrases), and 24 (1/2) learned words in an English language context (i.e., the research assistant spoke in English, and words were presented in English carrier phrases). Because this latter distinction did not affect 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 analysis reported in Footnote 1.

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 each picture once, one at a time, and heard a sentence containing the picture's name in order to become familiarized with the label for each image (for a similar procedure, see Creel, Aslin, & Tanenhaus, 2008). 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 in tone, vowel, or both tone and vowel. The vowel-disambiguated pairs diverged in the first vowel (*mi* vs. *mu*) to make the temporal unfolding of vowels as comparable as possible to the temporal unfolding of tones, which are carried on the vowel. Each word served as the target three times—once each with its vowel; tone; and tone and vowel competitors (84 trials total). In each trial, two pictures (200 × 200 pixels) appeared on the left and right sides of the screen centered at 25% and 75% of the screen width, respectively, and 50% of the height. After 500 ms, 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 RT.

An Eyelink Remote eye tracker (SR Research, Mississauga, Ontario, Canada) sampled gaze position every 4 ms. Participants wore a small targetlike sticker on their foreheads, which the eye tracker used to identify their head position. We ran Eyelink software on a personal computer (PC) tower in DOS mode. A Mac Mini computer (10.4.1; Apple, Cupertino, CA) presented experimental stimuli using custom MATLAB (7.6.0, R2008a; MathWorks, Natick, MA) 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 and 39.06% of the horizontal span of the screen). Looks to target or nontarget 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 are displayed in the gaze time-course plot and the correlation plots. In analyses of variance (ANOVA; by subject and by item), we considered the effect of trial type (tone&vowel-disambiguated, 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 ($F_1(2, 142) = 8.55$, $p < .001$; $F_2(2, 54) = 7.97$, $p < .001$). Planned comparisons indicated that clicking accuracy in tone&vowel-disambiguated trials was significantly higher than in tone-disambiguated trials—paired $t_1(71) = 3.57$, $p < .001$; paired $t_2(27) = 3.82$, $p < .001$ —but not significantly higher than in vowel-disambiguated trials. Clicking accuracy was also significantly higher in vowel-disambiguated trials than in tone-disambiguated trials (paired $t_1(71) = 2.67$, $p < .01$; paired $t_2(27) = 2.41$, $p < .05$).

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 on the basis of 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 because all three are highly positively correlated (see Table 2). In an analysis of covariance (ANCOVA) with trial type as a within-subject categorical predictor of clicking accuracy and the Mandarin-dominance component as a between-subjects covariate, we found a significant main effect of trial type ($F(2, 140) = 9.27$, $p < .001$) and 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 ($r = .32$, $p < .01$; see Figure 3, middle). This suggests that dominance was particularly affecting use of tone—not vowel—information.

RTs

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 SD s 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 an ANOVA (by subject and by item). Trial type was a significant predictor of RT ($F_1(2, 142) = 29.1$, $p < .001$; $F_2(2, 54) = 11.67$, $p < .001$). Planned comparisons confirmed that RTs were significantly slower in tone-disambiguated trials ($M = 830$ ms, $SD = 180$ ms) than in either vowel-disambiguated trials ($M = 750$ ms, $SD = 140$ ms; paired $t_1(71) = -5.29$, $p < .001$; paired $t_2(27) = -3.63$, $p < .005$) or tone&vowel-disambiguated trials ($M = 730$ ms, $SD = 130$ ms; paired $t_1(71) = -6.63$, $p < .001$; paired $t_2(27) = -5.05$, $p < .001$). Vowel-disambiguated trials and 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-subject 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 as Mandarin dominance increased in tone-disambiguated trials ($r = -.29$, $p = .01$; see Figure 4, middle) but not in vowel-disambiguated 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.

The carrier phrase we used—*Qing3 xuan3 [cha2]* (“Please choose [cha2]”)—ended in T3, the dipping tone, and was recorded naturalistically for each stimulus. A tone sandhi rule in Mandarin causes a T3 preceding another T3 to essentially change to T2 (the rising tone; Wang & Li, 1967). Thus, trials in which either the target or the competitor word contained T3 (*mu3*, *mi3*, *shi3*, *shu3*, *ta3*, or *tu3*) might have been disambiguated more quickly than other trials, particularly for more proficient bilinguals. To investigate whether the ability to exploit tone sandhi cues might have contributed to our effects of Mandarin dominance on tone processing, we conducted an ANCOVA on RT in tone-disambiguated trials, with tone type as a within-subject factor (trials either disambiguated by T3 or not) and the Mandarin dominance PCA component as a between-subjects covariate. The interaction of tone type and Mandarin dominance did not approach significance, $F(1, 70) = 1.80$, $p = .18$.

Visual Fixations

Analyzing eye movements provides additional information about participants' fine-grained temporal processing

Figure 3. Clicking accuracy plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score on the Multilingual Naming Test) in the three trial types. Lines are linear-regression models. The y-axis range is restricted to allow visualization of correlations at high accuracy, and data points are jittered to improve their visibility.

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 2,000 ms so that all trials contributed equally across the full time course.

To statistically compare responses, we computed the e-logit transform on target fixations and competitor fixations (Barr, 2008) and analyzed the target minus competitor difference score, or target advantage, for these transformed values. We then averaged e-logit-transformed target advantage across the time window 200 to 800 ms after the onset of the noun. The start of this window (200 ms) 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 best reflect the asymptotes in

Figure 5 while keeping the percentage of the analyzed time window that had been extended from the final fixation (see the previous paragraph) as low as possible (10.03%). Two of the 72 participants were excluded from the gaze analyses (leaving 70 participants) because their target-plus-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 ANOVA (by subject and by item). Trial type was a significant predictor of target advantage ($F_1(2, 138) = 48.77, p < .001$; $F_2(2, 54) = 14.41, p < .001$). Planned comparisons confirmed that target advantage was higher in tone&vowel-disambiguated trials ($M = 0.58, SD = 0.11$) than in both vowel-disambiguated trials— $M = 0.54, SD = 0.11$; paired $t_1(69) = 3.13, p < .005$; paired $t_2(27) = 1.98, p = .058$ —and tone-disambiguated trials ($M = 0.43, SD = 0.14$; paired $t_1(69) = 9.01, p < .001$; paired $t_2(27) = 5.56, p < .001$). Target advantage was also higher in vowel-disambiguated trials than in tone-disambiguated

Figure 4. Reaction times plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score on the Multilingual Naming 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. Error bars are standard errors.

trials (paired $t_1(69) = 6.70, p < .001$; paired $t_2(27) = 2.99, p < .010$).

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-subject categorical predictor and the Mandarin dominance component as a between-subjects covariate.¹ We found significant main effects of trial type ($F(2, 136) = 50.26, p < .001$) and 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 in both tone-disambiguated trials ($r = .31, p = .01$; see Figure 6, middle) and tone&vowel-disambiguated trials ($r = .26, p < .05$; see Figure 6, right) but not in vowel-disambiguated trials.

¹The gaze ANCOVAs 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 ($F(1, 66) = 3.27, p = .026$), and 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-disambiguated and tone&vowel-disambiguated trials were driven by participants who learned Mandarin-compatible novel words in the previous experiment (tone disambiguated: $r = .46, p = .001$; tone&vowel disambiguated: $r = .44, p = .002$) 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). Experiment did not interact with the critical Language Dominance $\times$ Trial Type interaction.

As with RT, we investigated whether more Mandarin-dominant bilinguals were better able to exploit tone sandhi cues, by comparing gaze patterns for trials containing T3-bearing words with trials containing words bearing other tones. We conducted an ANCOVA on tone-disambiguated trials with factors tone type (disambiguated by T3 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 T3-disambiguated trials (.58) than in other trials (.32). This suggests that listeners overall were using the tone sandhi information to execute eye movements more quickly on these trials. As in the previous analysis, 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 T3 ($r = .45, p < .001$), not in trials with other tones.

Along with the accuracy and RT data, the gaze results provide converging evidence that Mandarin dominance particularly affected use of tone information but not vowel information. However, we wanted to rule out a potential confound. 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 eight were born in the United States), 26 had mainland Chinese dialect exposure (21 had lived in mainland China, four were born in the United States, and one was born in Singapore), six were born in the United States and did not report dialect information, and four had other (three) or unknown (one) 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 participants 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 unlikely to explain the relationship we found between Mandarin dominance and ease of tone processing.

Discussion

We asked how listeners with varying levels of experience with a tone language versus a nontone language process lexical tones versus 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

Figure 6. Target advantage (gaze) plotted against Mandarin dominance (displayed here as Mandarin vocabulary score minus English vocabulary score on the Multilingual Naming Test) in the three trial types. Lines are linear regression models, and data points are jittered to improve their visibility.

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-disambiguated trials. In tone&vowel-disambiguated 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 effect of language dominance. However, conclusions that are 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-disambiguated trials, where accuracy was 99.8% on average. Still, we cannot be sure why this pattern emerged only for gaze.

Effects of Mandarin dominance on 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. This suggests that the most highly proficient Mandarin listeners were most able to exploit tone sandhi to anticipate the target word. Tonal coarticulation is pervasive in Mandarin sentences, and listeners compensate for this coarticulation (Xu, 1994). Thus, it is possible that attrition affects the ability to recognize tones in fluent speech more than (or earlier than) it affects memory for individual words' tones. However, the ability to exploit tone sandhi did not account for Mandarin dominance effects on RTs, perhaps because RT is assaying a later stage of processing than gaze. Thus, although we found some evidence that the ability to exploit tone sandhi in sentence processing contributed to effects of Mandarin dominance on tone processing, the contribution of tone sandhi to our effects is somewhat nuanced. Further research is needed to tease apart the relative contributions of fluent use of tone sandhi versus word-level tone recognition to tone attrition effects.

The attrition we found in the ability to process tones fluently in an L1 after experience with a nontonal language has implications for the question of the extent to which bilinguals' two phonetic systems interact. It suggests (a) that bilinguals may not process pitch information language-specifically and (b) that when English has become the dominant language, processing strategies from English affect phonetic processing in the nondominant language, Mandarin—even when Mandarin was learned natively. It is not clear whether weaker use of tone information in English-dominant bilinguals results from more exposure to English or from less exposure to Mandarin. 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 online recognition of lexical tones (Malins & Joanisse, 2010) to consider the effect 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 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 versus 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 T3—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, although all were native Mandarin speakers, may use tone information more slowly than monolingual Mandarin listeners would. Malins and Joanisse 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. On the basis of their description, their bilingual participants were likely 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) than in 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, which worsened Mandarin word recognition. Related investigations of the effects of dominance on tone processing in newly learned words (e.g., Quam & Creel, 2012, in press; Singh & Quam, 2015) 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.

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—likely could 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 with 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 on the basis of 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 (Cutler & Chen, 1997; Davis et al., 2015; Lee, 2007; Ye & Connine, 1999) and that listeners are less aware of lexical tone than of segmental content, even in languages with orthographic systems for representing tone (Burnham et al., 2013; Davis

et al., 2015). Perhaps tones are more vulnerable to attrition than vowels due to weaker representational strength. However, we found that Mandarin proficiency was related to both fluent use of tone during sentence processing and knowledge of individual words' tones. The effect 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 why familiarity with the tone patterns of particular words is unlikely to fully account for the dominance effects. First, accuracy, even for minimal tone pairs, was very high for all bilinguals: 98.5% on average and no lower than 89.0% for individual participants. Another reason to doubt word familiarity as the explanation for dominance effects is that accuracy, gaze, and RT scores did not improve with increasing word frequency (calculated from the Modern Chinese Character Frequency List; Da, 2004, 2005). Pearson's correlation tests investigating the relationship between tone processing and word frequency in fact showed a significant negative correlation for gaze ($r = -.39, p < .05$) and nonsignificant correlations for accuracy and RT (also in the opposite direction of what would be predicted on the basis of a word-familiarity explanation). Also, even when only correct trials were considered, language dominance still affected 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 affected by their experience with English (or reduced exposure to Mandarin).

Last, on a word-familiarity account of language-dominance effects on tone attrition, participants who are less Mandarin dominant should be less accurate and slower in word recognition across the board, regardless of the type of phonological competitor present on a given trial. However, we did not find dominance effects in vowel-disambiguated trials and instead found them 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 four tones in Mandarin take advantage of the full

pitch space (high level T1 vs. low dipping T3; rising T2 vs. falling T4). The most confusable contrast is between T2 and T3 (Shen & Lin, 1991); for that reason, we did not include that contrast. Still, future work should assess whether bilingual 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 versus English, tone use during recognition of familiar Mandarin words was correlated with degree of Mandarin language dominance. It is interesting to note that, for gaze, 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 affect both (a) the ability to rapidly exploit tone sandhi cues during sentence processing to anticipate the tone of the target word and (b) recognition of the word itself. This suggests that phonetic processing strategies from the dominant language affect sentence processing and word recognition in the nondominant language, even when the latter was learned natively.

Acknowledgments

The first author was supported by National Institutes of Health (NIH) Institutional Training Grant T32 DC00041-12 via the Center for Research in Language, University of California, San Diego; NIH Grant F32 HD065382; and NIH Grant K99-R00 DC013795. The second author was supported by National Science Foundation (NSF) Grants BCS-1057080 and BCS-1230003. 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.

References

- Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.
- Best, C. (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* (pp. 167–224). Cambridge, MA: MIT Press.
- Best, C., & McRoberts, G. (2003). Infant perception of nonnative consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183–216.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.07) [Computer program]. Retrieved from <http://www.praat.org/>
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology*, 108, 693–712.
- Burnham, D., Luksaneeyanawin, S., Kantamphan, S., & Reid, A. (2013). Phonics vs. whole-word instruction in a tone language. *Written Language & Literacy*, 16, 60–76.
- Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 259–280). Amsterdam, the Netherlands: Benjamins.
- Canavan, A., & Zipperlen, G. (1996). *CALLHOME Mandarin Chinese speech* [DVD]. Philadelphia, PA: Linguistic Data Consortium.
- Chang, C. B. (2011). Systemic drift of L1 vowels in novice L2 learners. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the XVII International Congress of Phonetic Sciences* (pp. 428–431). Hong Kong: City University of Hong Kong.
- Cornelissen, F. W., Peters, E., & Palmer, J. (2002). The Eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments & Computers*, 34, 613–617.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106, 633–664.
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165–179.
- Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In Z. Pu, T. Xie, & J. Xu (Eds.), *The studies on the theory and methodology of the digitalized Chinese teaching to foreigners: Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese* (pp. 501–511). Beijing, China: Tsinghua University Press.
- Da, J. (2005). *Modern Chinese character frequency list*. Retrieved from <http://lingua.mtsu.edu/chinese-computing/statistics/char/list.php?Which=MO>
- Davis, C., Schoknecht, C., Kim, J., & Burnham, D. (2015). The time course for processing vowels and lexical tones: Reading aloud Thai words. *Language and Speech*. doi:10.1177/0023830915586033
- de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, 13(1), 1–24.
- Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition*, 12, 273–289.
- Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, 62, 971–974.
- Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). Timonium, MD: York Press.
- Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6, 185–202.
- Gildersleeve-Neumann, C. E., Peña, E. D., Davis, B. L., & Kester, E. S. (2009). Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact. *Bilingualism: Language and Cognition*, 12, 259–272.

- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 15, 594–615.
- Gonzales, K., & Lotto, A. J. (2013). A *bafri*, un *pafri*: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological Science*, 24, 2135–2142.
- Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (sections 10.25–10.28). New York: Wiley.
- Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, 110, 581–616.
- Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2014). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development*, 86, 10–22.
- Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, 36, 17–38.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, 50, 101–123.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford, United Kingdom: Oxford University Press.
- Major, R. (1992). Losing English as a first language. *The Modern Language Journal*, 76, 190–208.
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, 62, 407–420.
- Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10, 241–265.
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106, 1367–1381.
- McBride-Chang, C., Lin, D., Liu, P. D., Aram, D., Levin, I., Cho, J.-R., ... Zhang, Y. (2012). The ABC's of Chinese: Maternal mediation of Pinyin for Chinese children's early literacy skills. *Reading and Writing*, 25, 283–300.
- Patel, A. D., Xu, Y., & Wang, B. (2010, May 11–14). *The role of F0 variation in the intelligibility of Mandarin sentences*. Paper presented at the Proceedings of Speech Prosody 2010, Chicago, IL.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Quam, C., & Creel, S. C. (2012). What's in a rise? Effects of language experience on interpretation of lexical tone. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 487–499). Somerville, MA: Cascadia Press.
- Quam, C., & Creel, S. C. (in press). Mandarin-English bilinguals process lexical tones in accordance with the language context. *PLOS-ONE*. doi:10.1371/journal.pone.0169001
- Quam, C., & Swingle, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, 62, 135–150.
- Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perception assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, and Psychophysics*, 77, 571–591.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52, 240–255.
- Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34, 145–156.
- Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, 124, 128–142.
- Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, 17, 94–109.
- Singh, L., & Quam, C. (2015). Can bilingual children turn one language off? Evidence from perceptual switching. *Journal of Experimental Child Psychology*, 147, 111–125.
- Wang, W. S., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, 10, 629–636.
- Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95, 2240–2253.
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tonal information. *Language and Cognitive Processes*, 14, 609–630.
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68, 123–139.