

BUCLD 36 Proceedings
To be published in 2012 by Cascadilla Press
Rights forms signed by all authors

What's in a Rise? Effects of Language Experience on Interpretation of Lexical Tone

Carolyn Quam and Sarah C. Creel

Languages use acoustic dimensions differently to contrast words, so children must learn which acoustic dimensions are relevant at different levels of structure in their particular language. For example, vowel duration is an important cue to vowel identity in Dutch. In English, though, vowel duration varies with suprasegmental factors like speaking rate, but does not contrast words. Children learning these languages treat vowel length accordingly by 18 months (Dietrich, Swingley, & Werker, 2007). Here we consider acquisition and interpretation of pitch categories in adult learners. English and Mandarin Chinese use similar pitch patterns for very different functions. For example, in English, a rising pattern can indicate a yes/no question, while a fall can convey a statement. In Mandarin, these pitch contours actually differentiate words. Quam and Swingley (2010) demonstrated that adult speakers and 2.5-year-old learners of English know that pitch is not relevant at the lexical level in English. Despite being taught a word with consistent pitch pronunciation, which suggested that pitch might be a relevant feature, listeners' experience with English governed their interpretation of a clearly audible pitch change. Importantly, this experiment was designed so that English listeners would not be pushed into treating pitch as relevant. This enabled investigation of adults' and children's *interpretations* of pitch, not whether they *could* exploit it in word learning.¹

The fact that the dimensions indicating lexical differences differ between languages means, clearly, that learners must discover the relevant dimensions from their linguistic input. The learning problem becomes yet more complicated for learners acquiring *multiple* languages, who must presumably keep track of two different sets of phonetic/phonological categories. In the present work, we focus primarily on adult (undergraduate) native speakers of Mandarin, whose age of acquisition of English varied between birth and adulthood. Because very

* Carolyn Quam, University of Arizona, cmquam@email.arizona.edu. Sarah C. Creel, University of California, San Diego, creel@cogsci.ucsd.edu. Thanks to Heather Pelton, Sally Chou, Roger Levy, Tamar Gollan, Jing Shen, Allen Hsu, Christina Lin, Abraham Chen, and Vivian Yu. Funding was provided by NIH 5T32DC000041-18 (via UCSD Center for Research in Language) and NIH 1F32HD065382-01A1 to CQ; and a Hellman Foundation fellowship to SCC.

¹ Contrast this with approaches like Wong and Perrachione's (2007), where English speakers were required to learn minimal tone pairs in order to succeed in the task.

similar pitch information is used at different levels of linguistic structure in the two languages, listeners' dominant-language categories might interfere with processing words in the non-dominant language. A person dominant in Mandarin might hear tones in English words. This would impair recognition of English words (e.g., "shoe"), since they occur with both rising patterns ("Do you see the shoe?") and falling patterns ("Yes, it's a shoe.") Conversely, a person dominant in English, who is used to processing pitch at the intonation level, might be less efficient at interpreting words as having tones.

The present eye-tracked word-recognition study investigated how Mandarin-English bilinguals and English monolinguals learn and recognize words with pitch content that is potentially tonal. We addressed three questions. First, how quickly is tone information exploited in on-line word recognition? Second, does use of tone information differ for newly learned words (Experiment 1) versus familiar Mandarin words (Experiment 2)? Third, does it differ depending on language experience? To examine these questions, we used visual fixations to pictured referents to evaluate the speed of exploiting tones relative to segmental information as a function of Mandarin experience (no experience; bilingual but more English dominant; bilingual but more Mandarin dominant). A previous study by Malins and Joanisse (2010) suggests that tone is exploited as rapidly as segments in familiar words, and we attempt to confirm those findings in Experiment 2.

1. Experiment 1

We are motivated by two main questions. Our first question is whether tone is exploited the same way in newly learned words and familiar words. There are two advantages to investigating people's representations of tone in new words. First, using novel words controls for people's experience with the words. This is important because if we find that less-proficient speakers of Mandarin use tone less than proficient speakers, the proficient speakers might simply be benefiting from greater familiarity. Novel words allow us to rule out the possibility that less tone use is due to reduced familiarity with the words. Second, novel words provide a window into how people *learn* words with tones—not just how they process tone once they know the words extremely well.

The second question we investigate is whether people's language experience affects their success at exploiting tone. We investigate language experience in three ways. (1) We include both bilingual speakers of Mandarin and English and speakers of just English. This enables us to ask whether experience with Mandarin improves efficiency/accuracy for encoding tone information in new words and using it in word recognition. (2) We consider bilinguals' language-dominance profiles in the two languages. Do the dominant language and degree of dominance (highly dominant vs. just mildly dominant) affect sensitivity to tones in novel words? Since English uses pitch much differently than Mandarin, does hearing and speaking English more often than Mandarin interfere with efficient processing of Mandarin tones? (3) Does the

language context of the experiment affect people's attention to tones? Are listeners capable of learning in an English vs. a Mandarin processing mode, as Elman, Diehl, and Buchwald (1977) demonstrated for consonant voice-onset time boundaries? Will this affect their likelihood of attending to tones?

1.1. Method

1.1.1. Participants

English speakers. Twenty-four English speakers from the UCSD community, verified to have had no (or limited) exposure to Mandarin, were included in the English-context version (12 females, mean age = 21, $SD = 2$, range = 16-26; age data were not collected for 2 participants who were approximately 21 and 22). Six were Spanish-English bilinguals, and one was a Hindi-English bilingual, but none had had significant experience with a tone language. The seven bilingual participants' responses were not meaningfully different from monolinguals'. Ten more participants were excluded from the analysis: five for not reaching the training criterion within two hours, one for excessive eye-tracking data loss (18% of the analysis window), two for exposure to tone languages, one for experimenter error, and one for withdrawing from the experiment.

Mandarin-English bilinguals. Forty-eight speakers of both Mandarin and English were included: 24 in the Mandarin-context version (17 females, mean age = 20, $SD = 2$, range = 18-26; age data were not collected for 1 participant who was approximately 19), and 24 in the English-context version (15 females, mean age = 20, $SD = 2$, range = 16-23). Sixteen more were excluded from the analysis: four for not reaching the training criterion within two hours, one for eye-tracking data loss (32% of the analysis window in Experiment 2), five for having as much or more exposure to other tone languages or dialects during childhood as to Mandarin, one for responding incorrectly in 1/3 of familiar-word tone-contrast trials (in Experiment 2), and five for experimenter error. All 48 bilingual participants self-reported as at least "fairly proficient" in both English and Mandarin, and to have been exposed more to Mandarin during childhood than to any other Chinese dialects. These criteria encompassed a wide range of language backgrounds and language-dominance profiles (see **Language-Dominance Assessments** for details). Most of our participants acquired Mandarin before English, so by a traditional definition most of our participants 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.

1.1.2. Stimuli

We designed a set of 16 novel words (see **Table 1**). The words' phonemes and phonotactics were designed to be compatible with both English and Mandarin, but they contained Mandarin tones 1–4 on their first syllables (and

the neutral Mandarin tone on their second syllables). The first syllable of each word contained a consonant and vowel/diphthong that both occur in Mandarin, but never together; this was to minimize effects of syllable frequency on processing. The sixteen words formed four quadruplets, each word of which (e.g., “biu3fu”) was contrasted with two similar-sounding words in the quadruplet that differed in first-syllable *tone* (“biu4fa”) or *segments* (“bou3fa”) as well as in second-syllable vowel. The tone vs. segmental contrast in the 1st syllable was designed to compare the speed with which participants could exploit tone vs. segmental information. The vowel contrast in the 2nd syllable ensured that listeners could identify the words without exploiting tone. Tone 2 and Tone 3, the pair most difficult to distinguish (Shen & Lin, 1991), were never contrasted directly.

Table 1: Four novel-word quadruplets. Moving horizontally within each quadruplet creates a tone contrast in the first syllable (e.g., “biu3fu” vs. “biu4fa”); moving vertically, a segmental contrast (e.g., “biu3fu” vs. “bou3fa”).

biu3fu	biu4fa	fi2pi	fi4pu	sei1tu	sei2ti	fai1di	fai3da
bou3fa	bou4fu	fao2pu	fao4pi	sua1ti	sua2tu	fiao1da	fiao3di

Auditory stimuli were recorded in a sound-attenuated chamber and normalized to a mean amplitude of 70 decibels (using a Praat script; Boersma & Weenink, 2009). Sentences were recorded by a bilingual speaker, who learned Mandarin from birth in Taiwan, and English at age seven. She is now slightly English dominant according to both the MINT test (she named 61 pictures correctly in English and 57 in Mandarin) and the BDS (she scored a -4: slightly more proficient in English than Mandarin; see below for test details). Each target word was recorded once in an English carrier (“Choose the [bou4fa].”) and twice in a Mandarin carrier (“Qing3 xuan3 [bou4fa].”). The best token from the Mandarin carrier was spliced into English and Mandarin carriers, to ensure identical acoustics for the target word across both contexts. Two Mandarin-speaking research assistants verified tone pronunciations. In response to their observation that Tones 2 and 3 sounded very similar, we resynthesized the pitch contours of Tone 2 stimuli using Praat sound-editing software (Boersma & Weenink, 2009) to rise more linearly (they had been slightly dipping).

In the experiment, each word was matched with an unfamiliar black-and-white shape, used in several prior experiments (e.g., Creel, Aslin, & Tanenhaus, 2006, 2008; Creel, Tanenhaus, & Aslin, 2006; Creel & Tumlin, 2011). There were four assignments of shapes to words across trial orders, but words from the same quadruplet (which were similar-sounding) never had the same base shape.

1.1.4. Language-Dominance Assessments

After the eye-tracking, participants completed the Multilingual Naming Test (MINT; Gollan et al., 2011) and then a questionnaire. The MINT evaluates

bilinguals' ability to name pictures in their two languages; English and Mandarin are among the languages for which it is designed.² Participants viewed a series of 68 pictures of estimated increasing difficulty, naming them separately in both English and Mandarin (language order was counterbalanced across participants). They then completed a questionnaire asking about their medical history, impressions of the experiment, and language background. For bilinguals, the language-background questions included the twelve questions from the Bilingual Dominance Scale (BDS; Dunn & Fox Tree, 2009), which evaluate 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?"

Both the MINT and the BDS provided a score for each bilingual participant in Mandarin and English. The MINT score was simply the number of pictures named in each language. The BDS assigns points to each language depending on participants' responses (see Dunn & Fox Tree, 2009, for scoring). For each test, we then subtracted the English score from the Mandarin score to compute the degree of Mandarin dominance: a large positive number means someone is strongly Mandarin dominant, while a large negative number means strongly English dominant. **Figure 1** plots the dominance scores for each bilingual participant on each measure. There was a strong correlation between the two tests ($r = .74, p < .001$), but MINT scores appeared to have slightly more predictive power, so throughout the paper we use MINT dominance scores to describe participants' language dominance.

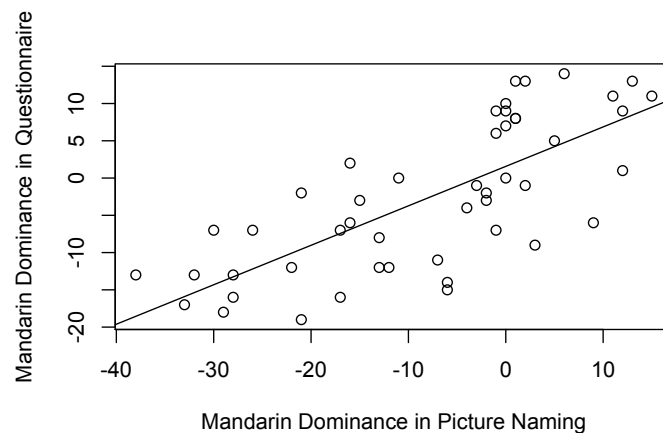


Figure 1: Two Language-Dominance Measures for Each Bilingual Participant.

² Note that the MINT is still being normed for use in Mandarin.

1.1.5. Apparatus and Procedure

In the Mandarin context, pre-experiment instructions were in Mandarin, and stimulus sentences were the Mandarin equivalent of “Please choose [bou4fa].” In the English context, all instructions and stimulus sentences (“Choose the [bou4fa].”) were in English. In all trials, two pictures appeared on the left and right sides of the screen, each 200 X 200 pixels, centered at 25% and 75% of screen width, respectively, and 50% of height. After 500 milliseconds, a sentence from headphones labeled one of the pictures (e.g., “Choose the bou4fa.”). Participants clicked the computer mouse on the picture they thought matched the last word in the sentence, guessing if necessary.

During the training phase, the names for the two pictures on the screen came from distinct quadruplets, and contained different tones whenever possible, to make learning easier and reduce the likelihood that participants would detect the experimental manipulation. Each picture was the target eight times per block, appearing equally often in each screen position, both as the target and as the competitor. During training, participants received feedback: after one picture was clicked, the correct picture stayed on the screen while the incorrect picture disappeared. Training trials were presented in 128-trial blocks, and the Matlab script tabulated accuracy at the end of each block. Once participants reached 90% accuracy within a training block, they proceeded to the test phase. Training took 2.44 training blocks on average (bilinguals: 2.44 blocks, $SD = 1.00$; English-speakers: 2.46 blocks, $SD = 0.93$).

In the test phase, no response feedback was given. There were three different trial types. Half were *baseline* trials, comparable to training trials in that the two pictures had very distinct names (though the exact word-pairs from training were not repeated). In the other half of test trials, word pairs were taken from the same quadruplet; half of these differed in their *segments* (e.g., “biu4fa” vs. “bou4fu”), and half differed in their *tone* (e.g., “biu4fa” vs. “biu3fu”).

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 (10.4.11) presented experimental stimuli using custom Matlab (7.6.0, R2008a) scripts that relied on PsychToolBox 3 (Brainard, 1997; Pelli, 1997) and the embedded 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 square region extending 100 pixels above, below, left and right of each location were counted as looks to that location. Looks to target or non-target pictures were then averaged within trial type for each participant, and binned into 50-ms chunks.

1.2. Results and Discussion

Participants' **eye-gaze responses** revealed effects of language background. **Clicking accuracy** is not described here due to space limitations, but showed similar patterns. Eye movements revealed that participants were fastest to recognize words on tone+vowel trials, next fastest in vowel-differentiated trials, and slowest in tone-differentiated trials. **Figure 2** depicts target – competitor fixations (which range from roughly 0, chance looking, to 1, looking only at the target) in each trial type. To statistically compare these responses, we averaged this measure across the time window 200–2000 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). Trial lengths were variable, ending when participants clicked on a picture, so we extended the final fixation of each trial out to 2000 ms. We selected 2000 ms as the end of the analysis window to best reflect the asymptote in **Figure 2**.

The between-subjects manipulation of hearing the words in a Mandarin vs. an English context exerted no effect on responses, so the two groups were collapsed for subsequent analyses. As **Figure 2** displays, bilinguals fixated the target (vs. the competitor) most for highly dissimilar (“baseline”) words (M , 0.57; SD , 0.10), next highest for segment-differentiated words (M , 0.46; SD , 0.11), and least for tone-diff. words (M , 0.30; SD , 0.13). All three trial types were significantly different ($t(47) > 4.7$; $p < .001$). But was the disadvantage for tone driven by lower proficiency Mandarin speakers? We separated bilingual participants into two groups using a median split on MINT language-dominance scores. We then compared target fixation in the three trial types for these two groups as well as English speakers. Trial type ($F(2,207) = 92.39$, $p < .001$) and language group ($F(2,207) = 9.55$, $p < .001$) were significant predictors of target advantage scores, and their interaction trended toward significance ($F(4,207) = 1.99$, $p = .10$).

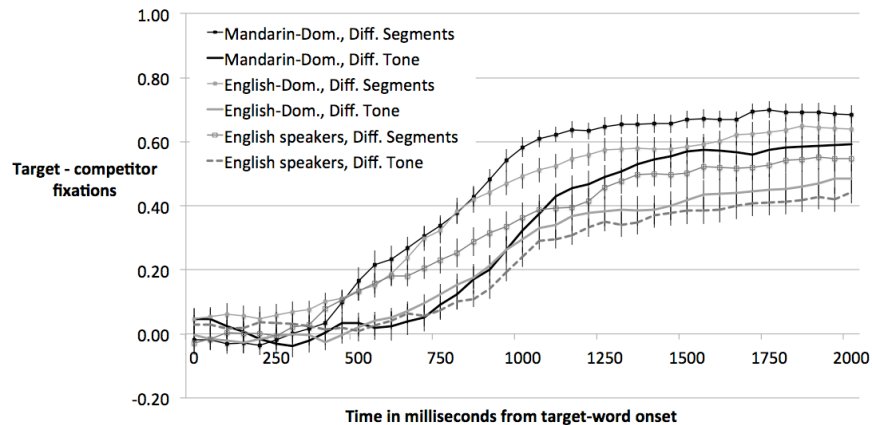


Figure 2: Target Advantage for Novel Words By Language Group.

Because we were specifically interested in how dominance impacts processing of different types of information (e.g., tones vs. segments), we conducted planned comparisons to see which trial types differed between the three groups, and which groups processed segments better than tones. All three groups showed significantly higher target advantages for segments than tones (more Mandarin-dominant: mean difference, 0.15; more English-dominant: mean diff., 0.17; English speakers: mean diff., 0.11; $t(23) > 4.5, p < .001$). The two groups of bilinguals did not significantly differ in any individual trial-type, though there were trends for the English-dominant group to be better in baseline (diff., -0.05, $t(43.12) = -1.64, p = 0.11$) and worse in tone trials (diff., 0.06, $t(43.71) = 1.66, p = 0.10$) trials. To consider language dominance as a continuous predictor of tone processing, we computed the Pearson's correlation between target advantage and language-dominance scores (see **Figure 3**, left). As Mandarin dominance increased, target advantage increased in tone-diff. trials ($r = 0.3502, p < .05$), but not segment-diff. trials. Still, 13/14 bilinguals who scored as more proficient in Mandarin than English identified the target better on segment-diff. trials than on tone-diff. trials (binom. $p < .005$).

How did English speakers compare to bilinguals? Compared with the more English-dominant bilinguals, English speakers were significantly worse in vowel trials (diff., 0.09, $t(45.86) = 2.25, p < .03$) but not in tone trials. Compared with the more Mandarin-dominant bilinguals, they were significantly worse with both vowels (diff., 0.13, $t(41.28) = 3.77, p < .001$) and tones (diff., 0.09, $t(45.98) = 2.89, p < .01$). English speakers therefore showed lower target advantages than bilinguals with *both* types of similar-sounding words. We suspect that this stems from two separate effects: (1) worse processing of tone for people with no Mandarin experience, which was expected, and (2) worse processing of segments for these people as well, because our speaker had a Chinese accent. Our novel words were designed to be equally compatible with the sound categories and regularities of English and Mandarin, and our speaker was slightly English-dominant. Still, having to produce Mandarin tones may have caused her to pronounce the words with more Mandarin-like vowel quality. We are exploring this possibility in ongoing work.

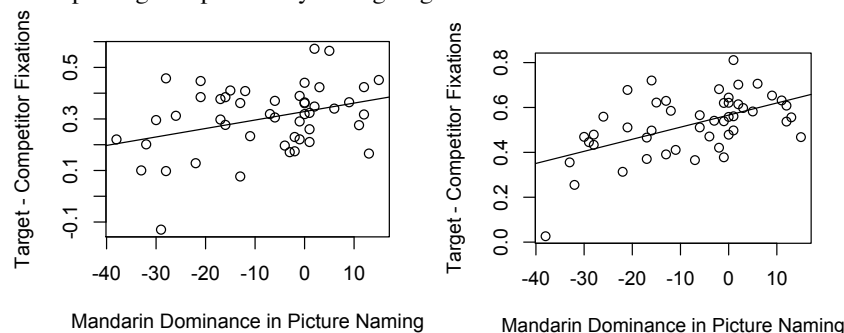


Figure 3: Correlation Between Target Advantage and Mandarin Dominance in

Novel-Word (Left) and Familiar-Word (Right) Tone Trials.

In summary, we found gradient effects of experience with Mandarin on overall accuracy. In **Figure 2**, the more English-dominant bilinguals appear to pattern in between the more Mandarin-dominant group and the English speakers in their target advantage for similar-sounding words. Statistically, the two groups of bilinguals did not differ on any particular trial-type, but target advantage for tones increased significantly with increasing language dominance. Compared to English speakers, English-dominant bilinguals showed better performance with segments, but not tones.

2. Experiment 2

Bilingual participants identified familiar Mandarin words in a procedure analogous to the Experiment 1 test phase. This allowed us to examine whether tone processing is similar for newly learned words and familiar Mandarin words. We also evaluated whether bilinguals' language-dominance profiles influenced their ability to exploit tone in familiar words, as they did for novel words.

Malins and Joanisse (2010; M&J) found that Mandarin speakers exploited tone information as rapidly as segmental information when recognizing familiar Mandarin words. M&J compared participants' speed in identifying the target (e.g., “mi4,” honey) in the presence of a tone-disambiguated competitor (e.g., “mi3,” rice) vs. a segmentally disambiguated competitor (e.g., “mian4,” noodles). In most cases, the late-segments–disambiguated competitor diverged from the target after the first vowel (the /i/ for “mi”/“mian”); this divergence point was intended to be comparable to the divergence point for tone contrasts. Word frequency in Mandarin was roughly balanced across word-types (target words vs. different competitor types), but word frequency in the experiment, due to the design, was such that the target words were labeled over four times more often than competitor words. This appears to have reduced competition effects from the competitor words, which may in turn have obscured differences in the timing of tone vs. late-segmental information. Here, we ask whether M&J's findings hold when a balanced design controls both global and within-experiment word frequencies.

2.1. Method

2.1.1. Stimuli

We designed a set of 7 quadruplets of monosyllabic, familiar Mandarin words, analogous to the quadruplets in Experiment 1 (see **Table 2**), and based on M&J's (2010) stimuli. Within each quadruplet, each word (e.g., “chuang2”) had competitors differing in tone (“chuang1”), segments (“cha2”), and both tone and segments (“cha1”). Crucially, each word occurs equally often in each role. For example, “chuang2” was the tone-disambiguated competitor for “chuang1” (and vice-versa), the segment-disambiguated competitor for “cha2” (and vice-

versa), and the tone+segment-disambiguated competitor for “cha1” (and vice-versa). Each word was therefore its own frequency control across trial-types (balancing global word frequencies), and all words were labeled equally often. The 28 target words were embedded in the same Mandarin sentence as in Experiment 1. Visual stimuli consisted of photographs (taken either from ClipArt online or from Flickr creative-commons licensed photos), edited and presented on the left and right sides of the computer screen.

Table 2: Seven familiar-word quadruplets.

cha1 (fork)	cha2 (tea)	hua1 (flower)	hua4 (painting)
chuang1 (window)	chuang2 (bed)	he1 (to drink)	he4 (crane)
mu3 (mother)	mu4 (tree/wood)	qiu1 (autumn)	qiu2 (ball)
mi3 (rice)	mi4 (honey)	quan1 (circle)	quan2 (fountain)
shi1 (lion)	shi3 (poop)	ta3 (pagoda)	ta4 (to tread/step on)
shu1 (book)	shu3 (mouse/rat)	tu3 (dirt)	tu4 (rabbit)
xian1 (first place)	xian4 (thread)		
xin1 (heart)	xin4 (envelope)		

2.1.2. Procedure

The constraints of matching segmentally and tone-disambiguated words meant that some of our words were not especially imageable (e.g., ta4, *to tread*). So that participants would know which picture each word referred to, they first saw each picture in turn, and heard a sentence referring to the picture. Next, they saw two pictures at a time and heard a sentence labeling one of them. The objects’ names contrasted either in tone (e.g., target “mi4” and competitor “mi3”), segments (competitor “mu4”), or tone and segments (competitor “mu3”). The segment-disambiguated pairs diverged in the first vowel (“mi” vs. “mu”). Since tone information is carried on the vowel, this made the time-course of disambiguation comparable.³ Each word served as the target three times: once each with its segmental, tone, and tone+segments competitors (making 84 total trials). The eye-tracking procedure was analogous to the novel-word test phase. The data were collected immediately after the novel-word test.

2.2 Results and Discussion

Participants clicked on the referents of familiar Mandarin words with high accuracy in all trial types (tone+vowel, $M = 99.85\%$, $SD = 0.72\%$; vowel, $M = 99.48\%$, $SD = 1.47\%$; tone, $M = 98.59\%$, $SD = 2.83\%$). Even in tone trials, the lowest score for any participant was 89.29%, suggesting that all our bilingual

³ Note that Malins & Joanisse’s (2010) segment-disambiguated competitors diverged later in the word, e.g., “chuang” versus “chuan,” or “mi” versus “mian.”

participants were fluent comprehenders of Mandarin. Since accuracy was at ceiling, we analyzed only eye movements, which allowed us to investigate the speed of word recognition and degree of uncertainty in people’s responses to the different phonological-competitor types.

Eye movements revealed that participants were fastest to recognize words on tone+vowel trials, next fastest in vowel-differentiated trials, and slowest in tone-differentiated trials. **Figure 4** depicts target – competitor fixations in each trial type. To statistically compare these responses, we averaged this measure across the time window 200–1000 milliseconds (ms) after the onset of the noun. As in Experiment 1, we extended the final fixation of each trial out to 2000 ms. We selected 1000 ms as the end of the analysis window to best reflect the asymptote in **Figure 4**. Over this window, target – competitor fixation was highest in tone+vowel trials ($M = 0.65$; $SD, 0.09$), lower in vowel-differentiated trials ($M = 0.60$, $SD, 0.10$), and lowest in tone-differentiated trials ($M = 0.52$, $SD, 0.14$). Trial type was a significant predictor of target fixation proportions in an analysis of variance (ANOVA; $F(2,144) = 10.6$, $p < .001$). Planned comparisons confirmed that target fixation was significantly higher in segment+tone trials than in segment-differentiated (paired $t(47) = 3.06$, $p < .005$) and tone-differentiated trials (paired $t(47) = 7.21$, $p < .001$). Target fixation was also significantly higher in segment trials than in tone trials (paired $t(47) = 4.14$, $p < .001$), and this pattern held for 38/48 (79%) of the participants (binomial $p < .001$).

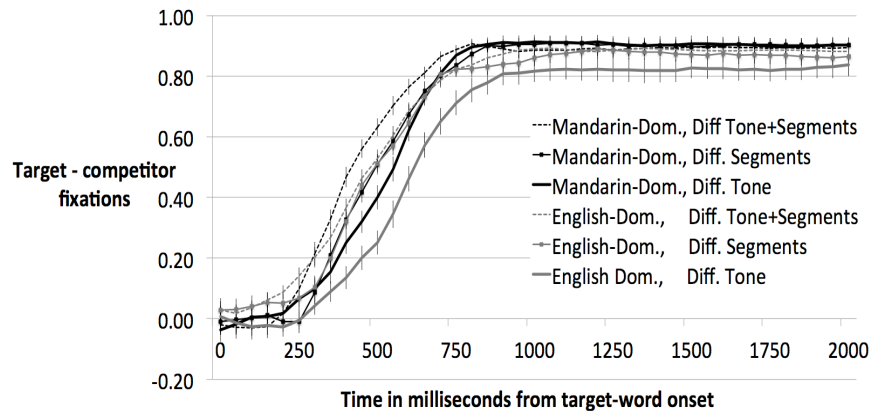


Figure 4: Eye-Gaze to Familiar Words, Split By Language Dominance.

Bilinguals processed tones more slowly than segments, but was this driven by the less proficient listeners? We divided participants into two groups, above and below the median score for language dominance on the MINT. We then looked at both groups’ gaze patterns (see **Figure 4**) and added language-dominance group to the ANOVA. Both trial type ($F(2,138) = 16.52$, $p < .001$) and dominance ($F(1,138) = 13.44$, $p < .001$) were significant predictors of target fixation proportions, and their interaction trended toward significance ($F(2,138)$

= 2.53, $p = .08$). Because we were specifically interested in how dominance impacts processing of different types of information (e.g., tones vs. segments), we conducted planned comparisons to see which trial types differed between the two groups. The more Mandarin-dominant group showed significantly higher target-fixation proportions in tone-differentiated trials (diff., 0.12, $t(39.62) = 3.33$, $p < .005$) and tone+segments trials (diff., 0.05, $t(45.00) = 2.05$, $p < .05$). The two groups did not differ in segments-differentiated trials (diff., 0.02). Only the English-dominant group was significantly worse with tones than segments (diff., 0.12, paired $t(23) = 4.55$, $p < .001$), though 16/24 participants in the more Mandarin-dominant group looked at the target more in segment-diff. trials (binomial $p = .15$). Target advantage in tone-differentiated trials (**Figure 3**, right), but not segment-differentiated trials, increased significantly with increasing Mandarin dominance ($r = 0.54$, $p < .001$). Of participants who were more dominant in Mandarin than in English, however, 9/14 still showed a numerical disadvantage for tone relative to segments.

3. General Discussion

We asked how listeners with varying levels of tone-language experience (from no experience up to Mandarin-dominant Mandarin-English bilinguals) encoded and processed lexical tones in novel and familiar words. Our studies suggest slower or less-reliable use of tone than segmental information unless word familiarity and Mandarin proficiency are high. For familiar Mandarin words (Experiment 2), only more Mandarin-dominant bilinguals exploited tones as rapidly as segments. For novel words (Experiment 1), all bilinguals (as well as non-tone-language listeners) showed more accurate encoding of segments than tones. Nonetheless, language dominance still influenced novel word learning, as it was significantly correlated with target advantage in tone trials.

Our work provides a more nuanced view than previous studies exploring the impact of language background on tone interpretation (Malins & Joanisse, 2009; Wong & Perrachione, 2007). Malins and Joanisse, also using a visual-world eye-tracking paradigm, reported that Mandarin listeners used tones as rapidly as segments in recognizing familiar words. However, we found that tone use was somewhat weaker than segments and varied with degree of language dominance. Wong and Perrachione previously taught English speakers novel words with tones with some success. However, their participants were forced to attend to tones to attain high accuracy, because the word-set consisted of sets of three words differing only in tone. Their paradigm also did not compare tone-word-discrimination accuracy to use of some other (segmental) distinction. In contrast, our paradigm provided tone information, but listeners did not have to use it to distinguish words. In these circumstances, non-tone-language speakers clearly use segmentally-cued distinctions more readily than tone distinctions. Models of sound-categorization and word learning must address how second-language learners apply existing categories to learn a new language. They should also

address how dominance patterns, as well as phonological and environmental contexts, shape attention to phonetic dimensions in established bilinguals.

4. References

- Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.07) [Computer program]. Retrieved May 12, 2009, from <http://www.praat.org/>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.
- 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., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language (JML), 65*, 264-285.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *JML, 54*, 1-19.
- 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.
- Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *JEP: Learning, Memory, and Cognition, 32*, 15-32.
- Dietrich, C., Swingle, D., & Werker, J.F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *PNAS, 104*, 16027-16031.
- 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. *J. Acoust. Soc. Am., 62*, 971-974.
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2011). 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*, Available on CJO 2011 doi:10.1017/S1366728911000332
- Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *JML, 62*, 407-420.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.
- Quam, C. & Swingle, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *JML, 62*, 135-150.
- Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*, 145-156.
- Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholing., 28*, 565-585.