DE GRUYTER MOUTON Laboratory Phonology 2014; 5(4): 489 – 522

Vance Schaefer and Isabelle Darcy*

Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones

Abstract: Determining the factors involved in the non-native perception of the pitch patterns of tones is complicated by the fact that all languages use pitch to various extents, whether linguistic (e.g., intonation) or non-linguistic (e.g., singing). Moreover, many languages use pitch to distinguish lexical items with varying degrees of functional load and differences in inventory of such pitch patterns. The current study attempts to understand what factors determine accurate naïve (= non-learner) perception of non-native tones, in order to establish the baseline for acquisition of a tonal L2. We examine the perception of Thai tones (i.e., three level tones, two contour tones) by speakers of languages on a spectrum of lexically contrastive pitch usage: Mandarin (lexical tone), Japanese (lexical pitch accent), English (lexical stress), and Korean (no lexically contrastive pitch). Results suggest that the importance of lexically contrastive pitch in the L1 influences non-native tone perception so that not all non-tonal language speakers possess the same level of tonal sensitivity, resulting in a hierarchy of perceptual accuracy. Referencing the Feature Hypothesis (McAllister et al. 2002), we propose the Functional Pitch Hypothesis to model our findings: the degree to which linguistic pitch differentiates lexical items in the L1 shapes the naïve perception of a non-native lexically contrastive pitch system, e.g., tones.

Keywords: non-native tone perception, lexically contrastive pitch, prominence, pitch accent, word stress

DOI 10.1515/lp-2014-0016

1 Introduction

Many languages, including Thai and Mandarin Chinese, use tone to distinguish words. Tones consist of variations of the fundamental frequency (F_0), amplitude,

---

***Corresponding author: Isabelle Darcy:** Indiana University. E-mail: idarcy@indiana.edu
Vance Schaefer: Indiana University. E-mail: vkschaef@indiana.edu

and/or voice quality attached to each syllable at the suprasegmental level. Because tones serve to signal lexical contrast, they are phonemic in the same way as vowels and consonants (e.g., Lehiste 1970: 92; Van Lancker and Fromkin 1973; Abramson 1975; Wang et al. 2001), and are processed accordingly by speakers of Thai or Mandarin. They also constrain lexical access, as shown for Mandarin speakers by Lee (2007). Cross-linguistically, listeners whose native language (L1) is not a tonal language have been shown to process tones in a nonlinguistic manner, or at least not in terms of linguistic categories (Gandour and Harshman 1978; Hallé et al. 2004 for French listeners; Wang et al. 2004, among others). Wang, Jongman, and Sereno (2001) show that American English listeners process tones mainly in the right hemisphere, whereas native speakers use their left hemisphere more, supporting the linguistic–nonlinguistic processing difference. Hallé et al. (2004) report that while Mandarin listeners process native tones in a categorical manner, French listeners do not, even though they are able to perceive acoustic differences between them. French is a non-stress language which does not exploit F_0 variations at the lexical level. As a result, French listeners display a low sensitivity to syllable-level prosodic variations involving F_0 contours (Dupoux et al. 1997; Hallé et al. 2004; Dupoux et al. 2008). In other words, they are less well equipped to perceive F_0 distinctions at the word level, and may experience difficulties in discriminating, identifying, and acquiring Mandarin or Thai tonal distinctions. In contrast, previous linguistic experience with tones appears to facilitate perception of unfamiliar tones in a different language (Wayland and Guion 2003).

One question of interest examined in this study is whether speakers of non-tonal languages *in general* would experience the same difficulties as French speakers in perceiving tones. By and large, studies examining tone perception by speakers of various L1 backgrounds show that speakers of non-tonal languages perform less accurately than speakers of tonal languages on tone discrimination or identification tasks (Gandour 1983; Burnham et al. 1996; Wayland and Guion 2004; Wang et al. 2006; Francis et al. 2008; but see Hao 2012 for diverging findings). Similarly, Wayland and Guion (2003) show that perceptual discrimination of Thai lexical tone contrasts is challenging for non-tonal (American English) speakers, and is resistant to training, whereas Mandarin Chinese speakers improved after training. In addition, Wayland and Guion (2004) show that American English learners of Thai outperform naïve listeners, even though their performance remains significantly lower than native Thai listeners.

Variations in F_0 appear to be the main acoustic cue to identify tone in languages such as Thai (Gandour 1983). However, languages use F_0 variations to different degrees and with dissimilar functions. Word stress or pitch-accent languages, for instance, also use variations in F_0 (Pike 1948; Fry 1958) to distinguish words, even though they contrast lexical items using overall fewer F_0 patterns,[1] with a larger domain, and where pitch patterns spread over an entire word or multiple syllables. Not all languages express information structure (e.g., focus) through prosody. Spanish is an example, where focus or new/given distinctions are encoded through word order, and pitch accents are not dependent on such distinctions. The picture that emerges from studies of tone perception suggests that not all non-tonal speakers are equally ill-prepared to categorize or identify unfamiliar tones, perhaps as a result of the use of linguistic pitch in their respective L1s. The present study examines this hypothesis by testing the perception of Thai tones by four groups of listeners whose L1 differs in terms of the lexical use of linguistic pitch: Mandarin (lexical tone), Japanese (lexical pitch accent), English (lexical stress), and Korean (no lexically contrastive pitch). We now review in more detail how pitch is used in each of the three nontonal languages relevant for this study: Japanese, English, and Korean.

A pitch-accent language like Japanese uses pitch to distinguish words, but not to the same extent and manner as tone languages do. In Japanese, one mora of each lexically accented word receives a high pitch (i.e., the ‘word accent’) before a fall, which in part determines the pitch on all other moras of the word (Kubozono 1999). As such, the pitch pattern of a word is predictable if the position of the word accent is known (Tsujimura 2006: 74) and therefore is considered lexical information needed to learn a word. One example is /ka.ki.ga/ which can mean ‘oyster’ (HLL or initial-accented), ‘fence’ (LHL or final-accented), or ‘persimmon’ (LHH[2] or unaccented), where *ga* is a particle indicating the nominative case. Thus, pitch-accent languages resemble tone languages as they employ pitch to contrast words, but with less diversity in pitch movements. In addition, there are overall fewer minimal pairs in pitch-accent languages compared to tone languages (According to Pierrehumbert and Beckman [1988], approximately 20% of word pairs contrast for pitch accent in Japanese; see also Wu et al. [2012].)

In a stress language such as English, pitch can be used to distinguish meaning or mark which syllable receives stress. Next to a F_0 turning point (high or low),

---

[1] We are not implying that the number of possible F_0 patterns in a language like English is inferior to the number of tones found in Mandarin or Thai. We are referring to the number of lexical contrasts that are possible to realize in English using F_0, for instance. Most lexical contrasts in English are minimal pairs (e.g., ˈINsert vs. inˈSERT), and not minimal triplets or more, as would be the case for a tonal language (such as minimal sextuplets in Cantonese, for instance).

[2] Some words are unaccented and do not have a final fall. The use of notation like H and L is not meant to represent an actual binary assignment of H or L pitch values to each individual syllable. These are just shorthand for representing F_0 contours in a compact manner.

the stressed syllable generally features greater intensity, lengthened vowels, and unreduced vowel quality (Fry 1958). A typical contrast of stress in English would be the noun *INsert* and the verb *inSERT*, where capitalization indicates the stressed syllable. English and Japanese differ in one major point, however. In English, unlike Japanese, F_0 variations are rarely the only cue to lexical contrast (Cutler 1986), and vowel reduction in unstressed syllables has been shown to be used by listeners to identify the location of stress among native speakers (Kochanski et al. 2005) and to recognize words as much as, if not more than, F_0 variations (Zhang and Francis 2010). Japanese, on the other hand, differs from English mainly because variations in pitch are not accompanied by segmental changes. For speakers of a pitch-accent variety of Japanese, F_0 differences clearly constrain lexical access (Otake and Cutler 1999), indicating that listeners are able to use this F_0 information at the lexical level during on-line spoken word recognition.

However, speakers of English or Japanese may be able to exploit the use – albeit limited – their language makes of F_0 contours for lexical contrasts to discriminate tonal contrasts. Burnham et al. (1996) report that speakers of Swedish, another pitch-accent language, exhibit accuracy rates and reaction times comparable to those of Cantonese speakers in a tone discrimination task involving Thai tones. So and Best (2010) show that Japanese listeners perform similarly to Cantonese listeners in most tonal comparisons, and English listeners identified 3 out of 4 tones better than chance after a brief training. Importantly, however, the exact type of tonal contrasts examined can influence performance. As pointed out by Braun and Johnson (2011), it is important to examine specific tonal comparisons with regard to the function played by pitch in the respective L1 (see also Hao 2012). Cooper, Cutler, and Wales (2002) demonstrate that English native listeners are also able – to a limited extent – to exploit lexical stress information for lexical recognition, but do so more effectively in two-syllable words than in monosyllabic words. Indeed, not all F_0 differences are processed successfully by English listeners. In the case of tone processing, Lee and Nusbaum (1993) report that English speakers process suprasegmental information in a way similar to Mandarin speakers, but only for dynamic (contour) F_0 variations, and not for constant level pitches (see also Wood 1974; Repp and Lin 1990). The authors hypothesize that this difference is due to the fact that in English dynamic F_0 variations (e.g., final rise for questions) also convey linguistic and paralinguistic information, whereas level pitch variations do not. Braun and Johnson (2011) add empirical support to this hypothesis, finding that the processing of pitch information is guided by linguistic function: in an ABX non-word classification task with disyllabic stimuli, Dutch listeners attend to F_0 information when these correspond to contours having linguistic meaning in Dutch (e.g., question intonation).

Finally, some languages do not use F_0 variations to distinguish words. We will refer to these languages as *non-lexical pitch* languages, highlighting the fact that even though they use pitch and the associated features of duration and intensity mostly to distinguish discourse meaning, pitch variations are not used contrastively at the level of words to distinguish lexical meaning. They are used at the level of the phrase. Such languages include standard Korean (Kim-Renaud 2009: 22) or Parisian French. As discussed above, speakers of French do not perceive tonal contrasts categorically. The fact that F_0 is never used to carry lexical information in these languages might prevent speakers of these languages from reliably distinguishing monosyllables that differ in F_0 only. In sum, languages differ in the degree to which linguistic pitch differentiates lexical items.

In addition to considerations of the functionality of F_0 in L1, research shows that perception of tonal contrasts might also be influenced by the specific tone comparisons examined. For example, Gandour demonstrates that listeners from different L1s may vary in that respect. Using multidimensional scaling, Gandour (1983) shows that the two main tonal features used in tone identification are pitch height and pitch direction: listeners from four tonal and one non-tonal language backgrounds (Cantonese, Mandarin, Taiwanese, Thai, and English) weighted pitch height more than pitch direction to judge tone dissimilarity. Only one group (Thai) used both to almost the same extent. However, the prevalence of pitch height over direction Gandour observed may be related to the use of monosyllabic stimuli – at least in the case of English listeners. Indeed, Braun and Johnson (2011) suggest that contours that broadly mirror intonational patterns in the L1 might be perceived with more success when the stimulus is two or more syllables long. This might also allow speakers of non-lexical pitch languages to make tonal distinctions for specific dynamic (contour) F_0 patterns (see also Lee and Nusbaum 1993). However, Braun and Johnson's (2011) results do not settle the issue, since Dutch is also a stress language, and has little vowel reduction. Speakers of languages that use F_0 for lexical contrast (even in a limited number of words) appear more sensitive to F_0 variations (at least at the level of words). This may be why Dutch listeners were more sensitive to pitch variations (see also Cooper et al. [2002] for supporting evidence) than speakers of a non-lexical pitch language would have been.

In sum, it appears that a combination of L1 influence (the functionality of pitch variations to signal lexical contrast in the L1) and task variables (specific tonal comparisons made) determine tone discrimination or identification performance across languages. The goal of the present study is to contribute to a more consistent understanding of tone perception by comparing speakers of languages that differ with respect to the extent to which they exploit F_0 variations to make lexical distinctions. In this study, we tested speakers of Mandarin (a tone language), Japanese[3] (a pitch-accent language), English (a word-stress language), and Korean[4] (a non-lexical pitch language) on their perception of Thai tones.

2 The current study

We examine whether the use of pitch in the L1 to signal lexical contrast aids in the perception of non-native Thai tone, resulting in a linguistic hierarchy of perceptual accuracy. We further compare perceptual accuracy across tone pairs, to potentially elucidate whether listeners are able to use the phonetic dimensions of pitch height, direction, or both. We tested speakers from various L1s on their perception of Thai tones in an AXB categorization task, using monosyllabic stimuli. Thai was chosen over Mandarin because it is described (Abramson 1975) as having three level and two contour tones, although phonetically the 'high level' tone is not level but raises in parallel to the raising tone (Morén and Zsiga 2006; see Figure 1).

The L1s of the targeted participants utilize lexically contrastive pitch to varying degrees – Mandarin Chinese (tone), standard Japanese (pitch accent), English (word stress), and standard Korean (no lexically contrastive pitch).

2.1 Predictions for tonal categorization

As discussed above, tone-language speakers (Mandarin) are expected to successfully discriminate syllable-level F_0 variations, as a result of the use of F_0 to contrast lexical items in their L1. This is especially noteworthy as while modern Mandarin is considered a polysyllabic language (Li and Thompson 1981; DeFrancis 1984), it still possesses a large number of monosyllabic words. We expect Mandarin speakers' sensitivity to F_0 differences to be the highest.

Similar hypotheses based on the functional use of pitch in different languages are found in Van Lancker (1980), in which she relates functional use of pitch contrasts to the size of linguistic domains. As shown in Figure 2, both Thai and Mandarin Chinese use pitch contrasts most systematically, and contrasts are

---

[3] We refer to standard Japanese and the other dialects that also feature pitch accent and not to the pitch-accentless dialects such as the Fukushima or Kumamoto dialects.

[4] We refer to standard Korean and not to the Kyungsang dialect, which is a pitch-accent variety Also, we do not take into consideration the emergence of lexically contrastive pitch in the Seoul dialect (Silva 2006).

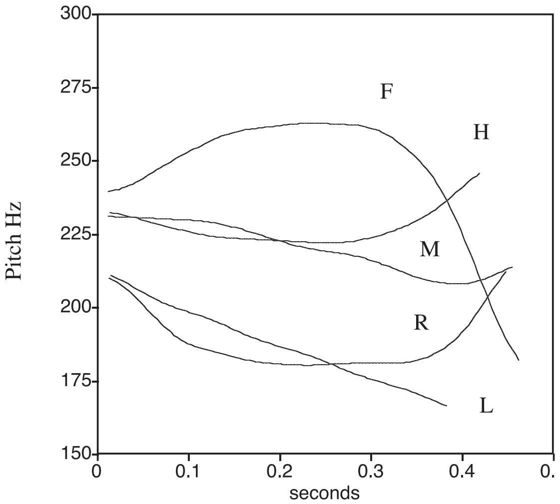

Fig. 1: Contour shapes of Thai tones in citation form, for falling (F), raising (R), low (L), mid (M), and high (H) tones. Representative examples from one speaker. From Zsiga and Nitisaroj (2007: 347).

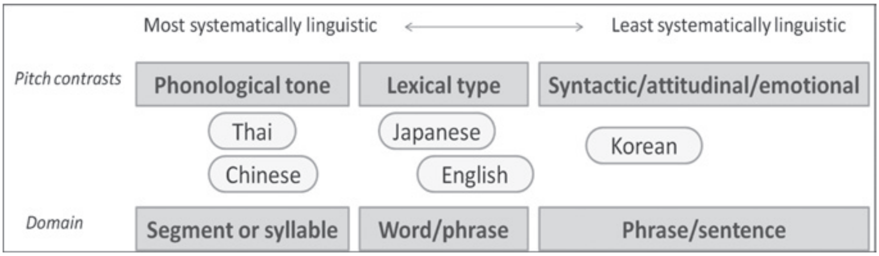

Fig. 2: Functional scale of pitch contrasts (adapted from Van Lancker 1980: 210).

made on the smallest domain, i.e., the syllable. Speakers of these languages are expected to be highly accurate in discriminating syllable-level tonal contrasts.

Following this scale, Japanese and English speakers are expected to display some sensitivity to F_0 variations. This sensitivity may be attenuated by the fact that F_0 is used to make lexical contrasts, but not at the single syllable level. On the other hand, it is possible that they may still display this sensitivity for monosyllabic stimuli. Korean speakers are expected to have the lowest sensitivity to syllable-level F_0 variations because their L1 does not use contrastive pitch for lexical distinctions.

Table 1: Respective pitch functionality according to language type and domain.

Language	Domain	Functionality
Tone (Mandarin)	Lexical, syllable/word	Maximal
Pitch accent (Japanese)	Lexical, word	High-intermediate (pitch is exclusive)
Word stress (English)	Lexical, word	Low-intermediate (pitch is non-exclusive)
Intonation only (Korean)	Non-lexical	Low

To summarize, if the lexical use of F_0 in the L1 determines heightened sensitivity for syllable-level F_0 variations, the following hierarchy of performance is predicted in terms of both accuracy and reaction times (highest/fastest to lowest/slowest): L1 Mandarin > L1 Japanese = L1 English > L1 Korean. Alternatively, English and Japanese differ on one crucial aspect. In Japanese, we consider pitch contrasts to have a slightly higher functionality than in English. If the fact that lexical contrast can be *exclusively* signaled by F_0 is the determining factor for performance, then English listeners should perform with an accuracy falling in between Japanese and Korean, since in English, F_0 is rarely used alone to distinguish words. This prediction follows from findings that stress does not strongly constrain lexical access in English (Cooper et al. 2002). As such, it is also possible that English speakers' accuracy will resemble that of Korean speakers. The respective functionality of pitch we assigned to each language type is summarized in Table 1.

2.2 Predictions for specific tonal comparisons

For the five Thai tones, there are 10 possible comparisons of tonal pairs: three subconditions compare tones described as 'level': Low-Mid, Low-High, Mid-High; one subcondition compares the Falling-Rising tones; and six subconditions for the remaining comparisons: Low-Rising, Low-Falling, Mid-Rising, Mid-Falling, High-Rising, and High-Falling. While it is difficult to generate exact predictions for each of the 10 comparisons, we expect that, overall, comparisons of pitch height might be perceptually less salient than comparisons where both height and direction vary.

Because non-native listeners cannot be expected to use phonological characteristics, they may focus more on phonetic patterns. In particular, if the 'high' tone is perceived as a raising tone, we expect a difference between Low-Mid (which we expect to be more difficult) on the one hand, and Low-High and Mid-High on the other hand. The last two are expected to be less difficult to distinguish because they differ in two phonetic dimensions: height and contour. Conversely, the High-Raising pair might pose difficulty across the board because of the phonetic similarity and parallel shape of these two tones: the only difference is in the pitch height (see Figure 1). For Mandarin, the only group for whom phonological categories might interfere with the tonal perception, we expect that contour differences might be more salient than height differences (but see Gandour [1983] for a different prediction).

3 Experiment 1: Method

3.1 Participants

For this experiment, 47 participants were recruited from five language groups: Mandarin (*N* = 10; females = 6), Japanese (*N* = 12, females = 11), English (*N* = 13; females = 10), Korean (*N* = 10; females = 7) and Thai (*N* = 2, males = 2). The Thai speakers were recruited as a control to ensure that the stimuli and AXB task itself were valid for native speakers. The participants were primarily graduate students or former graduate students who were involved in language studies (i.e., language education, linguistics, applied linguistics) with the exception of 11 participants who were undergraduate students (*N* = 3) or not involved in language studies (*N* = 8) (i.e., Korean = 3, Japanese = 3, Mandarin = 3, English = 1, Thai = 1). Average ages ranged as follows: Japanese (25–50, *M* = 35.4 years); English (25–45, *M* = 31 years), Mandarin (24–31, *M* = 27.1 years), and Korean (27–47, *M* = 32.2 years). The two Thai listeners were 25 and 32 years old. The average time in an English-speaking country was 6.6 years for the Japanese, 4.5 years for the Koreans, 3.5 years for the Mandarin speakers, and 2 years for the Thai speakers. The English speakers had spent an average of 1.7 years abroad in a non-English-speaking environment.

Based on information collected in a background questionnaire, the participants were also comparable with respect to being L2 learners. Six Mandarin speakers also had various degrees of exposure to Taiwanese, another tonal language. Most had also been exposed to another Chinese dialect even if they did not consider themselves a fluent speaker of that dialect. The speakers of Japanese, a pitch- accent language, were recruited on the basis of speaking a dialect of Japanese which uses pitch accent, but not necessarily standard Japanese. Two speakers were from the Tochigi and Ibaraki prefectures, which are close to the Fukushima prefecture, known for its accentless dialect. The English speakers were native speakers of American English who had no proficiency in Thai, Mandarin, Japanese, Korean, or any tone or pitch-accent language, but who were not monolinguals. The Korean speakers were mainly from the Seoul area, but three were from the Kyungsang region where a pitch-accent dialect of Korean is spoken, and one speaker was from Cholla, an area abutting Kyungsang but with a dialect not featuring pitch accent (although these dialects do not correspond to their respective administrative borders; Lee and Ramsey [2000]). None of the non-native participants knew Thai.

Data from four individuals were excluded from the final analysis because they had significant exposure to one of the other target languages in the study or differed in background from the target group. Three English speakers had exposure to Japanese or a tone language (i.e., Mandarin or Vietnamese). One female Japanese student was an ESL student with lower exposure and proficiency in English as compared to the graduate student participants. As a result, only 11 Japanese-speaking participants and 10 English-speaking participants' data were analyzed, for a total of 43 participants.

3.2 Stimuli

The test stimuli consisted of 16 open CVV syllables with a long vowel (VV). Open syllables are considered more difficult for pitch discrimination than closed syllable words (Wayland and Guion 2003). Each syllable was recorded with each of the five different Thai tones (see Table 3 below and Figure 1 above), resulting in 80 items (41 items being real words and 39 nonwords; Thai does not feature a complete set of real words for each of the five-tone paradigms for these 16 syllables). Control stimuli were CVV and CVC syllables (all were real Thai words) composed of vowels or consonants similar to those used for the test items. The control condition also included more difficult vowels such as [ɯ], [ɛ], and [ə]. Control stimuli are stimuli that all participants should generally be able to perceive with high accuracy regardless of their L1 and thus serve as a means of comparison with the test condition (i.e., tone stimuli) to verify that participants do not have listening issues or lack an understanding of the task. Because our stimuli contain words, Thai listeners were able to approach the task using lexical knowledge for some trials. Therefore, most analyses were run excluding these listeners.

All stimuli were recorded by two native Thai speakers, one female and one male, who did not participate in the experiment. Both spoke the Central Thai dialect (i.e., standard Thai). Syllables were recorded in isolation without a carrier phrase. Each speaker recorded the 16 different test syllables three times with each tone ($16 \times 3 \times 5 = 240$). That is, the speaker was instructed to record the entire set of 16 syllables three times with the low tone, then again with the mid tone, etc., until the five tones were completed. In general, the second token of the three recordings for each syllable was selected for the task, unless the researchers felt that another token was a clearer example of the intended tone, i.e., did not contain extraneous noise, and that the tone was realized completely as decided based upon auditory and visual analysis of the sound files. Control items were only repeated twice each. The second one was chosen for the task.

The syllables were then arranged in triplets for the AXB design. The female voice was used for the A and B items while the male voice was used for the X item, which can be either of the A or of the B category. In an AXB design, four trials are needed for each comparison: AAB, ABB, BAA, BBA. For example, the two tones Low (L) and Mid (M) would be paired as LLM, LMM, MLL, and MML. If the syllable carrying such a comparison were [biː], a trial would look like the following: [biː]^{L} – [biː]^{L} – [biː]^{M} (see below).

3.3 Conditions

The experiment contains two conditions, test and control, with 48 trials each. In the test condition, the syllables within one triplet only differ by tone; the segmental make-up of the syllables remains the same. In the control condition, all syllables in the triplet have the same tone but vary in either one consonant or one vowel.

Furthermore, within the test condition, we included the 10 possible comparisons of tone pairs: Low-Mid, Mid-High, Low-High; Falling-Rising; and Low-Rising, Low-Falling, Mid-Rising, Mid-Falling, High-Rising, and High-Falling. Clearly, the same syllable carrying all 10 comparisons (e.g., [biː]) would make the task extremely monotonous in terms of segmental content, as this would result in 40 trials (4 trials × 10 comparisons) containing the syllable [biː] being spoken with various tones. In order to reduce the monotony of the experiment and to keep it to a reasonable duration, we chose to vary the segmental embedding for the tonal contrasts across trials. For example, for the Low-Rising condition, each of the four trials (AAB, ABB, BAA, BBA) uses a different segmental frame. Every test syllable appeared three times during the experiment. In order to account for this additional variance, Item is declared as a random factor in the model, which includes intercepts.

Table 2 presents an overview of the conditions used in the study. The three ‘level’ tone comparisons require 12 trials. The six mixed ‘level’ and ‘direction’ comparisons require 24 trials. The Falling-Raising condition normally would have required only 4 trials. However, in order to obtain a sufficient number of

Table 2: Overview of the tonal comparison(s) and number of trials used for each condition

Test conditions (*N* = 48)			**Control condition (*N* = 48)**
(*N* = 12)	**(*N* = 12)**	**(*N* = 24)**	
low-mid (LM)		low-rising (LR) low-falling (LF)	
low-high (LH)	falling-rising (FR)	mid-rising (MR) mid-falling (MF)	consonant vowel
mid-high (MH)		high-rising (HR) high-falling (HF)	

data points for extracting potential effects of phonological tone type (Direction), and to balance the number of trials across control and test, we created 8 additional trials for this condition (12 in total). One disadvantage of this choice is that it does not allow us to equate the number of times a given tone is heard during the experiment. As a result, the Rising and Falling tones are heard in total 24 times, whereas the other tones are heard 16 times each during the experiment. We return to this issue in the discussion.

In total, 96 trials were created (48 test and 48 control trials). All trials were randomized and put into three blocks of 32 items. A warm-up task consisted of 16 trials with feedback indicating participants' accuracy and reaction time (RT). None of these were used in the following test phase. Between each stimulus (A, X, and B), the inter-stimulus interval (ISI) was 500 ms. The experiment was timed so that after the presentation of each trial, participants had 3,000 milliseconds to make their answer, before the next trial was initiated. Reaction times were measured from the onset of the X stimulus.

3.4 Procedure

Participants were tested individually in a quiet room. On each trial, participants heard and chose whether the middle sound (i.e., X) was more similar to the first sound (i.e., A) or the third sound heard (i.e., B), by pressing two clearly identified keys on the computer keyboard. The task required 15–20 minutes in total, and was followed by a debriefing session. All procedures were performed in compliance with relevant laws and institutional guidelines, and the appropriate institutional committee(s) approved them.

4 Results

Reaction times shorter or longer than two standard deviations from the RT mean of each participant, as well as reaction times lower than 300 ms, were not analyzed (5.4% of total RTs). Data for three items in the height pairs (one each for L vs. M, L vs. H, and M vs. H) and for one item in the direction pairs (R vs. F) were excluded from analysis as one Thai participant felt that the tones were not ideal models of the targeted tone. Accuracy means for individual participants and items were screened for outliers. No item or participant was excluded.

4.1 Overall performance

Because native speakers of Thai were able to approach the task with lexical knowledge, they are excluded from further analyses, but the results are shown for comparison with the non-native participants. A linear mixed-effects model was conducted in SPSS 22 on the binary accuracy and continuous reaction times data. Language (Thai, Mandarin, Japanese, English, Korean) and Condition (test, control) were declared as fixed effects.[5] Subjects and Items were entered as random effects in the model. The significance threshold was set at $p = 0.05$ for this and all following analyses. The parameter estimates are presented in Tables 3 and 4.

When looking at the Type III tests of fixed effects, the *F*-tests showed no main effect of Condition ($F(1, 97.6) = 0.11$, $p > 0.1$), a significant effect of Language ($F(4, 38.1) = 5.36$, $p < 0.002$), and a significant interaction between the two factors ($F(4, 3809.1) = 17.71$, $p < 0.001$). The same analysis without the Thai group shows the same results.

On the test condition, we observed that Mandarin participants outperformed other groups (87% accuracy), followed by Japanese participants (77%), and by English and Korean (both at 67% accuracy). Whereas accuracy of all groups was comparable on the control condition (no effect of Language on the control condition: $F(3, 60.2) = 1.9$, $p > 0.1$), there was a significant effect of Language on the test condition, as suggested by the difference in accuracy rates ($F(3, 65.2) = 13.8$, $p < 0.001$). Mandarin listeners discriminated tonal contrasts with higher accuracy

---

[5] Another set of analyses have been conducted similarly, adding the fixed effect of Order (X = A, X = B) to the model. Order did not have any effect on accuracy rates, nor did it interact significantly with Condition and Language. For RT, Order had a significant effect ($F(1, 293.2) = 31.7$, $p < 0.001$); RT was faster when X was the same as B ($M = 1158$ ms) than when it was the same as A ($M = 1253$ ms), but this factor did not significantly interact with Condition and Language.

Table 3: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB accuracy.

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
Intercept	.89	.06	91.62	14.52	.00	.76	1.01
[Lg = E]	−.13	.06	61.73	−2.10	.04	−.25	−.01
[Lg = J]	−.08	.06	61.70	−1.42	.16	−.20	.03
[Lg = K]	−.05	.06	61.67	−.84	.41	−.17	.07
[Lg = M]	−.11	.06	61.69	−1.92	.06	−.23	.00
[Condition = Test]	.08	.07	570.64	1.22	.22	−.05	.21
[Lg = E] * [Condition = Test]	−.15	.06	3809.07	−2.61	.01	−.27	−.04
[Lg = J] * [Condition = Test]	−.10	.06	3809.03	−1.72	.08	−.21	.01
[Lg = K] * [Condition = Test]	−.24	.06	3809.03	−4.19	.00	−.36	−.13
[Lg = M] * [Condition = Test]	.02	.06	3809.03	.41	.69	−.09	.14
Covariance parameters	**Estimate**	**Std. error**					
Residual	.129	.003					
Subject	.003	.001					
Item	.036	.006					

Note: Thai is the reference language; Control is the reference condition.

than the other groups, significantly outperforming both Korean and English participants ($p < 0.001$) but only marginally more accurate than the Japanese group ($p = 0.087$). Notably, Korean and English participants were not significantly different from each other ($p = 1$), as is visible in Figure 3.

Analysis of RTs was performed similarly; the data were normally distributed. A linear mixed model declaring the factors Condition (test, control) and Language (Thai, Mandarin, Japanese, English, Korean) as fixed effects, and the factors Subject and Item as random factors, was performed.

When looking at the Type III tests of fixed effects, the F-tests showed a main effect of Condition ($F(1, 96.5) = 13.7$, $p < 0.001$), a significant effect of Language ($F(4, 37.9) = 4.1$, $p < 0.008$), and a significant interaction between the two factors ($F(4, 2876.5) = 5.3$, $p < 0.001$). The same analysis without the Thai group shows that the main effect of Language disappears but the significant interaction remains ($F(3, 2718.5) = 6.6$, $p < 0.001$). The picture that emerges from the analysis of reaction times (see Figure 4) shows a similar pattern as the accuracy data. Mandarin listeners are faster than all the other non-native groups. Overall, latencies in the test condition are about 120 ms slower than on the control condition

Table 4: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB RTs.

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
Intercept	756.6	133.1	42.2	5.683	.000	487.9	1025.2
[Lg = E]	399.2	144.5	40.7	2.762	.009	107.3	691.2
[Lg = J]	466.5	143.4	40.6	3.254	.002	176.9	756.1
[Lg = K]	428.9	144.4	40.6	2.970	.005	137.2	720.7
[Lg = M]	310.5	144.5	40.7	2.149	.038	18.6	602.4
[Condition = Test]	73.6	59.4	1071.0	1.2	.216	−42.9	190.1
[Lg = E] * [Condition = Test]	135.5	59.3	2872.6	2.3	.023	19.1	251.8
[Lg = J] * [Condition = Test]	.96	58.3	2872.7	.016	.987	−113.4	115.3
[Lg = K] * [Condition = Test]	73.8	59.2	2871.6	1.248	.212	−42.2	189.9
[Lg = M] * [Condition = Test]	−9.18	58.6	2869.7	−.157	.875	−124.0	105.6
Covariance Parameters	**Estimate**	**Std. error**					
Residual	113792	3007					
Subject	31982	7719					
Item	16216	3067					

Note: Thai is the reference language; Control is the reference condition.

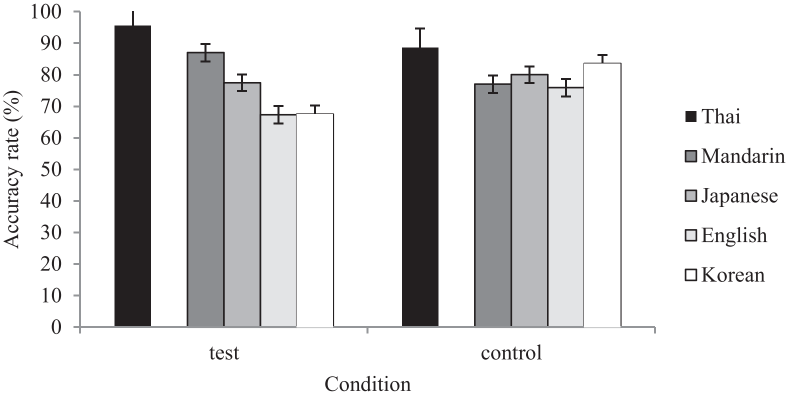

Fig. 3: Accuracy rate (%) for each language group in the test vs. control condition. Error bars enclose +/−1 SE. (Thai listeners are displayed in black for comparison purposes.)

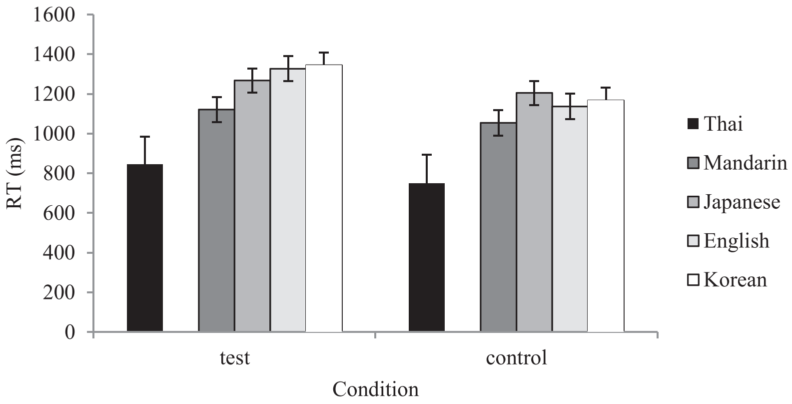

Fig. 4: Reaction time (ms) for each language group in the test vs. control condition. Error bars enclose +/−1 SE. (Thai listeners are displayed in black for comparison purposes.)

(1265 ms vs. 1141 ms). In the same way as for accuracy, the mean reaction times of all groups was comparable on the control condition (no effect of Language on the control condition: $F(3, 40.2) = 1.3$, $p > 0.1$); there was a significant effect of Language on the test condition ($F(3, 41.5) = 3.0$, $p < 0.05$).

The overall accuracy pattern that emerged from these accuracy and RT data confirms in large part the predicted hierarchy. The functionality of pitch in the L1 appears to determine accuracy in a phonological discrimination task. Against our prediction, however, the data also revealed that English and Korean participants pattern identically, perhaps suggesting that F_0 information is less readily accessible for phonological discrimination in these two groups. This finding is consistent with data from Cooper, Cutler, and Wales (2002) indicating that word stress information can be used by English listeners but less effectively in monosyllables. We will return to possible reasons for the lack of difference between Koreans and English listeners in Experiment 2, where we tested monolingual Korean speakers.

4.2 Performance in specific conditions

Turning now to performance in specific conditions, we analyzed accuracy and reaction times as a function of the 10 tone conditions (within subjects) for each non-native group (between subjects: Mandarin, Japanese, English, and Korean).

A linear mixed-effects model was conducted in SPSS 22 on the binary accuracy and continuous reaction times, excluding the Thai participants. Language

Table 5: Mean accuracy and RT across non-native groups for each subcondition.

Subcondition	Mean accuracy		Mean RT (ms)	
	Mean	Std. error	Mean	Std. error
LH	.76	.068	1244	65.3
LM	.55	.068	1342	71.1
MH	.70	.068	1315	67.7
HF	.83	.059	1273	58.5
HR	.59	.059	1466	62.6
MF	.69	.059	1284	60.7
MR	.85	.059	1277	58.0
LF	.73	.059	1365	60.0
LR	.85	.059	1215	58.2
FR	.82	.037	1218	43.9

(Mandarin, Japanese, English, Korean) and Subcondition (LH = Low-High, LM = Low-Mid, MH = Mid-High; HF = High-Falling, HR = High-Rising, MF = Mid-Falling, MR = Mid-Rising, LF = Low-Falling, LR = Low-Rising, FR = Falling-Rising) were declared as fixed effects. Subjects and Items were entered as random effects in the model. When looking at the Type III tests of fixed effects on accuracy, the *F*-tests showed a main effect of Subcondition ($F(9, 34) = 3.5$, $p < 0.01$), and a significant effect of Language ($F(3, 40.9) = 7.5$, $p < 0.001$), but no significant interaction between the two factors ($F(27, 1690.1) = 1.4$, $p = 0.071$). The same analysis on the RT showed a main effect of Subcondition ($F(9, 31.3) = 2.5$, $p < 0.05$), no main effect of Language ($F(3, 38.1) = 1.9$, $p > 0.1$) and no interaction ($F(27, 1213.9) = 1.4$, $p = 0.076$). This suggests that specific conditions differ in overall difficulty, as shown in Table 5.

Furthermore, the lack of significant interactions indicates that the difference between groups does not vary as a function of subcondition. Figures 5a and 5b display non-native groups' accuracy and reaction times on each subcondition. As is visible, the overall hierarchy among groups holds in most subconditions. The parameter estimates for accuracy and RT are presented in Appendix B.

5 Discussion

This first experiment examined the effects of L1 experience on the processing of tonal contrasts by Mandarin, Japanese, English, and Korean listeners. In particular, we examined (a) to what extent the degree of pitch functionality to signal

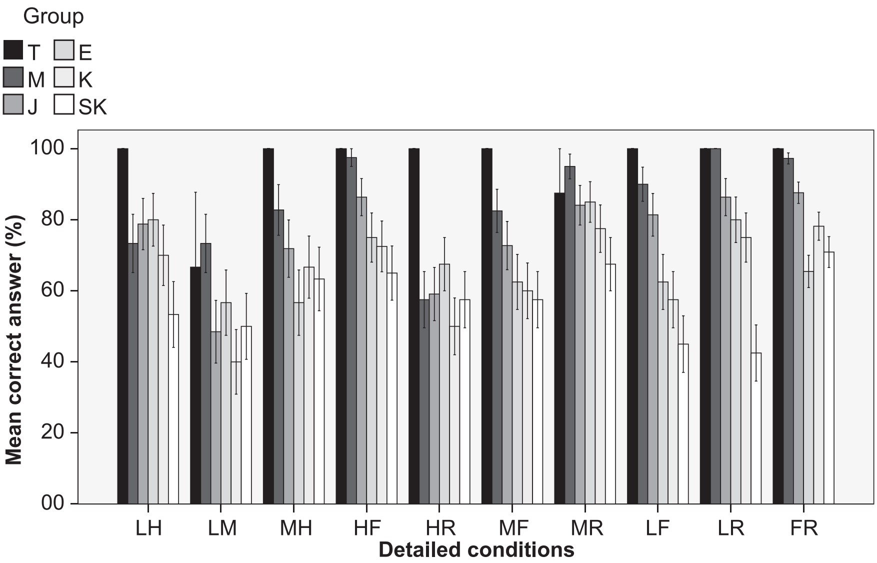

Fig. 5a: Accuracy rate on each condition for each Language group. Error bars enclose +/−1 SE. LH = Low-High, LM = Low-Mid, MH = Mid-High (for the Height condition); HF = High-Falling, HR = High-Rising, MF = Mid-Falling, MR = Mid-Rising, LF = Low-Falling, LR = Low-Rising (for the Mixed condition), FR = Falling-Rising (for the Direction condition). T = Thai, M = Mandarin, J = Japanese, E = English, K = Korean, SK = Seoul Korean (Experiment 2).

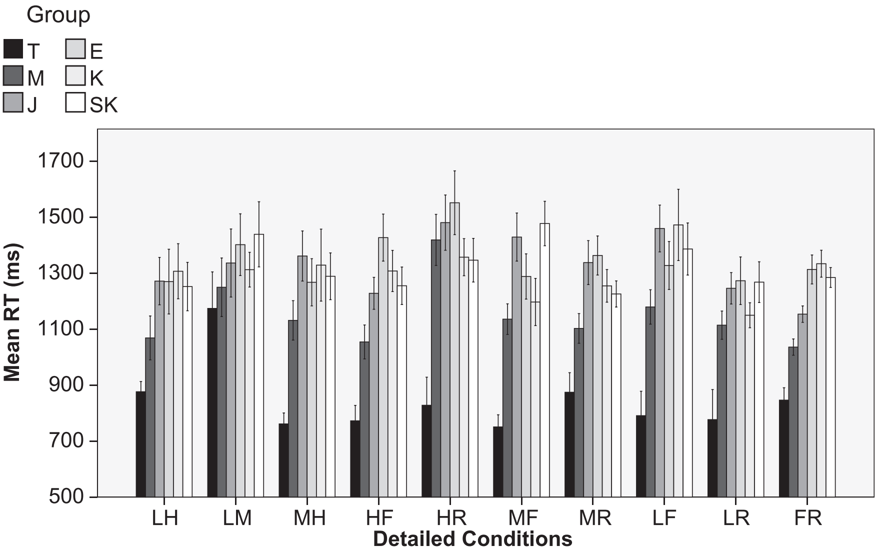

Fig. 5b: Mean RT (ms) on each condition for each Language group. Error bars enclose +/−1 SE. Label abbreviations are as in Figure 5a.

lexical contrast in the L1 aids in the non-native perception of Thai tone, and (b) whether certain tonal comparisons are easier or more difficult to differentiate for listeners.

We observed an effect of the L1 on non-native tone perception, suggesting that the functionality of linguistic pitch to signal lexical contrast in the L1 shapes the non-native perception of pitch in a gradient fashion. In an AXB task comparing tonal contrasts and segmental contrasts, we observed no effect of L1 in segmental perception (control condition), but a hierarchy in accuracy on tonal contrasts with the following pattern: L1 Mandarin > L1 Japanese > L1 English = L1 Korean. Mandarin listeners outperformed other groups (significantly more accurate than both English and Korean, but not Japanese), likely because of their use of lexically contrastive pitch at the smallest domain, therefore having the highest functionality among the four groups. Japanese listeners' accuracy was intermediate, suggesting that the use of pitch-accent contrasts in the L1 aids in non-native tone perception. Interestingly, English listeners' global performance in accuracy and in reaction times was strikingly similar to Korean listeners' performance, and was significantly different from Mandarin listeners' performance. The fact that lexical contrast can be signaled by F_0 *exclusively* hence appears to play an important role in determining performance in tone discrimination, as shown by the different patterns obtained by Japanese and English listeners. The fact that F_0 is rarely used alone to distinguish words in English appears to yield the same performance as if F_0 were not used at all to signal lexical contrast (English = Korean). This explanation would be consistent with findings showing that stress constrains lexical access only to a limited extent in English (Cooper et al. 2002). However, there are other possible explanations to this effect.

Indeed, this experiment does not allow us to distinguish whether performance of the English listeners is lower than expected or the performance of the Koreans is higher than expected. It is possible that Korean listeners were equally as accurate as English because they have learned English as a second language. All our participants had learned at least one additional (non-tonal) language, including the English-speaking participants. It is possible that if Korean listeners have acquired English word-stress patterns (see Darcy et al. [2011] for supporting evidence), they might use it in the present task. Another possibility is that some of our Korean speakers have been exposed to a pitch-accent dialect in Southern Korea (Kyungsang and possibly Cholla). To start answering this question, in Experiment 2, we tested a group of Korean listeners who do not know English and whose exposure to a pitch-accent dialect is strictly controlled (i.e., Seoul dialect speakers).

Regarding the second question examining the 10 possible tonal comparisons, we observe that the overall hierarchical pattern statistically holds across most tonal comparisons as well, as shown by the absence of a significant interaction between Language and Subcondition. However, it seems the case that some comparisons are more difficult than others (as shown by the main effect of Subcondition on accuracy and RTs). In particular, we note (see Figure 5a) that the Low-Mid and High-Rising comparisons trigger lower accuracy rates for the non-native listeners. On the Low-Mid comparison, all groups, except for the Mandarin speakers, performed at chance. This performance is in line with previous findings (Wayland and Guion 2004), and may reflect the difficulty of gauging the difference in pitch height of the two tones. This effect is perhaps further compounded by the different genders of the two speakers in our study (although see Lee [2007] for contrary evidence in the case of Mandarin L1 perception). Similarly, the confusion between the high and rising tones may be attributed to the phonetic similarity of the two pitch patterns (see Figure 1) despite their difference in phonological descriptions (i.e., 'level' vs. 'contour'). As well, the fact that our stimuli were spoken by speakers of different genders is expected to make this pitch height comparison more difficult. Hence, the difference in pitch height only between two tones with similar pitch contours appears to cause perceptual confusion among listeners. While this also suggests that the tonal comparisons involving direction seem to be easier to discriminate, this result should be interpreted with caution because the number of comparisons involving direction is lower (only Falling-Rising and perhaps High-Falling) compared to the mixed or height comparisons. As discussed in Bohn (1995), fewer levels of contrasts can yield higher accuracy, requiring this issue to be reexamined.

6 Experiment 2

A reanalysis of the Korean data from Experiment 1 (Schaefer and Darcy 2014) examined performance as a function of dialectal exposure. The data suggest that D1 (first dialect) Korean speakers of the Kyungsang dialect (which uses pitch accent; Kim and de Jong 2007; Kim 2011) perform at comparable levels to L1 speakers of Japanese. We concluded that this effect most likely was due to the presence of lexically contrastive pitch accent in the D1/L1. Hence, the inclusion of three Kyungsang speakers and one Cholla speaker in Experiment 1 appears to have skewed the data of the L1 Koreans speakers as a group, and boosted the group's performance to a level comparable to the L1 English speaker group. After removing the Kyungsang Korean speakers, the remaining Seoul Korean speakers performed at the accuracy levels we had originally predicted, that is, less accurately than the L1 speakers of English. To bolster the claim that the lack of lexically contrastive pitch accent in standard (Seoul) Korean reduces accuracy in our Thai tonal discrimination task, the second experiment tested a new group of Korean speakers who speak the Seoul dialect as their D1.

6.1 Experiment 2: Method

6.1.1 Participants

Ten monolingual participants (females = 6) were recruited and tested in Seoul, Korea, to ensure that Kyungsang Korean speakers were not present among the pool of Korean speakers and to minimize the effect of English exposure. All participants self-reported that they were native speakers of the Seoul dialect. Average ages for this group of Koreans ranged from 26–41 ($M = 30.4$ years). The participants were not involved in language studies or linguistics, but a variety of occupations.

6.1.2 Stimuli

The stimuli remain the same as reported in Experiment 1.

6.1.3 Conditions

The conditions remain the same as reported in Experiment 1.

6.1.4 Procedure

The procedure remained the same as reported in Experiment 1. The participants, however, were tested in Seoul, Korea, in similar conditions.

6.2 Results

For the analysis of the group of Seoul Korean participants, a linear mixed-effects model was conducted in SPSS 22 on the binary accuracy and continuous reaction times data. Condition (test, control) was declared as a fixed effect. Subjects and Items were entered as random effects in the model. There was a main effect of Condition on accuracy ($F(1, 90) = 16.5$, $p < 0.001$), indicating that performance on the test condition ($M = 59.9\%$ correct) was less accurate than on the control condition ($M = 77.1\%$ correct). Similarly, there was a marginal effect of Condition on RTs ($F(1, 535.2) = 3.98$, $p = 0.05$): RT on the test condition tended to be slower ($M = 1294$ ms) than on the control condition ($M = 1247$ ms). The parameter estimates are presented in Tables 6 and 7.

To compare the results of the Seoul Korean participants to the other four non-native groups, a linear mixed-effects model was conducted in SPSS 22 on the binary accuracy and continuous reaction times data. Condition (test, control) and

Table 6: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB accuracy.

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
Intercept	.77	.037	30.8	20.7	.000	.69	.85
[Condition = Test]	−.17	.043	90	−4.1	.000	−.26	−.09
Covariance parameters	**Estimate**	**Std. error**					
Residual	.179	.009					
Subject	.005	.003					
Item	.025	.006					

Note: Control is the reference condition.

Table 7: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB reaction times.

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
Intercept	1247	57.3	9.6	21.8	.000	1119	1375
[TestCondition = 1]	47.0	23.9	535.2	2.0	.050	.073	94.04
Covariance Parameters	**Estimate**	**Std. error**					
Residual	74999	4590					
Subject	30314	14995					
Item	*	.0					

*Note: Control is the reference condition. *This parameter is set to zero because it is redundant.*

Language (Mandarin, Japanese, English, Korean, Seoul Korean) were declared as fixed effects. Subjects and Items were entered as random effects in the model. When looking at the Type III tests of fixed effects for accuracy, the *F*-tests showed no main effect of Condition ($F(1, 90.0) = 2.7$, $p > 0.1$), a significant effect of Language ($F(4, 46.1) = 5.82$, $p < 0.002$), and a significant interaction between the two factors ($F(4, 4537.1) = 21.0$, $p < 0.001$). In other words, the exact same pattern of results was obtained with this new group added to the analysis. When looking at the Type III tests of fixed effects for the RT analysis, again, the exact same pattern of results as in the previous analysis emerges: the *F*-tests showed a main effect of Condition ($F(1, 85.9) = 16.3$, $p < 0.001$), no significant effect of Language ($F(4, 46.0) = 1.6$, $p > 0.1$), and a significant interaction between the two factors ($F(4, 3256.7) = 6.4$, $p < 0.001$).

Scores for this group on the 10 tonal comparisons are presented in Figures 5a and 5b but are not discussed further here.

To more clearly show the effect of exposure to pitch accent in the two groups of Korean participants, we reassigned the participants of the first Korean group according to whether they had exposure to pitch accent in their native dialect. The six who did not were grouped with the new Seoul Korean participants ($N = 16$), and the four pitch-accent Korean speakers were in a separate group. Figure 6 displays the respective accuracy scores for these two Korean groups (white and dotted bars), in comparison to the other non-native groups. It becomes clear that the non-pitch-accent Korean participants in this reanalysis now corroborate our originally predicted hierarchy of performance. Interestingly, the four

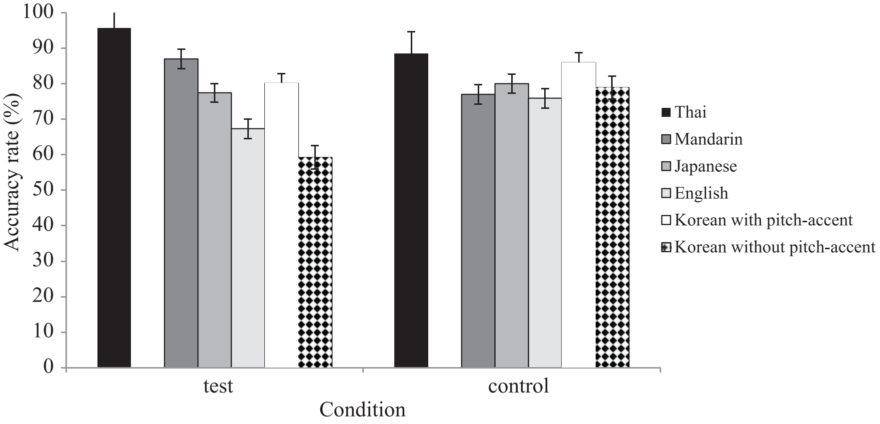

Fig. 6: Accuracy rate (%) for each language group in the test vs. control condition. Error bars enclose +/−1 SE. (Thai listeners are displayed in black for comparison purposes.)

Korean speakers with exposure to pitch accent perform at the same level as the Japanese speakers. A linear mixed-effects model on the accuracy scores declared Condition (test, control) and Language (Thai, Mandarin, Japanese, English, Kyungsang Korean/pitch-accent, Seoul Korean/no pitch-accent) as fixed effects. Subjects and Items were entered as random effects in the model. The Type III tests of fixed effects for accuracy were identical to the above analysis: there was no main effect of Condition ($F(1, 98.7) < 1$), a significant effect of Language ($F(5, 47.0) = 9.8$, $p < 0.001$), and a significant interaction between the two factors ($F(5, 4718.1) = 20.6$, $p < 0.001$). Post-hoc pairwise comparisons also indicate that the Seoul Korean group was significantly less accurate than the English group on the test condition ($p = 0.041$). This group was also significantly less accurate than all other groups (all $p < 0.001$) on the test condition – but not on the control condition.

In sum, Experiment 2 clearly demonstrates that the reason for the equal performance of the Korean and English groups in Experiment 1 was not due to the English speakers performing less accurately than expected, but indeed to the Koreans performing more accurately. We determined that the presence of speakers of the Kyungsang pitch-accent Korean dialect in our original Korean group was the determining factor in the equal performance between the L1 Korean and L1 English group. Therefore, our originally predicted hierarchy of performance based on pitch functionality in L1 is further supported.

7 General discussion

The current study derived specific predictions for the naïve perception of tone based on the functional salience of pitch in the L1 (cf. Feature Hypothesis; McAllister et al. 2002). Taken together, the results of Experiments 1 and 2 suggest that the functionality of linguistic pitch to signal lexical contrast in the L1 shapes the perception of non-native pitch in a gradient fashion. Our study established a baseline for tone perception focusing on the functional use of linguistic pitch in four language types. We propose the *Functional Pitch Hypothesis* to model our data. Accordingly, the degree to which pitch differentiates lexical items in the L1 shapes the naïve (= non-learner) perception of a non-native lexically contrastive pitch system, as in this case, of a non-native tone system. Within this framework, the definition of pitch functionality in the L1 takes into consideration the prosodic domain of pitch contrasts. More specifically, functionality is not determined only by whether or not a language makes lexical use of pitch, in an all or none fashion. It is rather constrained by the functional prosodic domain (e.g., syllable, foot, prosodic word, phonological phrase) where pitch variations are realized.

Indeed, our findings suggest that the specific prosodic domain in which pitch differentiates lexical items also constrains performance: sensitivity to pitch variations is highest where the functional domains overlap (Mandarin and Thai), and is greatly reduced where functional domains do not align (Japanese and Korean). For instance, Mandarin uses pitch to signal lexical contrast at the syllable/word domain, which is also the case in Thai, whereas in Japanese, the prosodic domain of pitch contrasts is rather the prosodic word. When domains do not overlap, it appears more difficult to map L1-pitch usage to the non-native pitch contrasts. In the case of English, pitch variation is an unreliable correlate of word-level stress and, as such, appears not to aid in tone perception. On the other hand, pitch variation plays an important role in the phrase-level intonational system of English, raising the question of whether intonational categories which typically require a phrasal domain can be mapped or associated with tonal contrasts in a smaller prosodic domain such as the syllable. Our findings suggest that, regardless of the fact that all languages use F_0 variations in intonation at larger size domains, applying this sensitivity to F_0 variations from a larger domain to a smaller, syllable-sized domain appears not to be easy. This claim makes interesting testable predictions for second language acquisition of tonal contrasts.

Three additional properties should also be taken into account when defining pitch functionality: (1) *exclusivity* to signal lexical contrast, (2) *functional load*, and (3) *inventory of pitch patterns*. *Exclusivity* refers to whether lexically-contrastive pitch is used by itself and not in combination with other phonetic parameters to differentiate words in the L1. For example, in Japanese this appears to be the case, while word stress in English includes the other correlates of vowel length, spectral quality, and intensity in addition to pitch. *Functional load* refers to the extent and/or number of minimal pairs differentiated in the L1. That is, tone languages such as Mandarin require pitch to distinguish a far larger number of lexical items in comparison to pitch-accent and word stress languages. Finally, *inventory of pitch patterns* refers to the number and type of patterns possible. For example, Mandarin has four tones, one level and three contour tones, while Thai has five tones, three level and two contour tones, implying a possible bias toward pitch height and/or direction, which may be constrained by the phonetic characteristics of specific tones. Similarly, differences exist within phonological tone types, i.e., the high tones in Thai and Mandarin vary phonetically (Chang et al. 2008).

In addition to these points of consideration, we also define L1 in the narrow sense where one's dialect (e.g., pitch-accented Kyungsang Korean dialect, non-pitch-accented Fukushima Japanese dialect) also impacts pitch functionality in a listener's L1 phonological system (cf. Weinreich 1953; Pallier et al. 1997; Otake and Cutler 1999).

The present results can be useful for investigations of tone acquisition in a second language. In a way similar to theoretical models developed for segmental contrasts such as PAM (Best 1995) and SLM (Flege 1995), which predict acquisition difficulties based on initial perceptual confusions between segmental categories, our data allow generating predictions using similar mechanisms. For example, we posit that within the *Functional Pitch Hypothesis*, tone-to-tone mapping may be applied where there is domain overlap, and where pitch categories are robust, as seems to be the case for L1 tone-language speakers. In this case, a model like PAM-L2 (Best and Tyler 2007) might be applied to make straightforward predictions for cross-linguistic tone-to-tone perception, although such an approach needs to be more clearly elaborated upon and subsequently tested. Current approaches are examining this possibility (Hao 2012; or So and Best 2010). However, for non-tonal language speakers, the question remains as to the extent to which lexically contrastive pitch *categories* can be defined (e.g., Stagray and Downs 1993). A case for the existence of lexically contrastive pitch categories could be made for pitch-accent languages such as Swedish or Japanese, as such languages are defined as a subclass of tone languages (Yip 2002: 2) and therefore allow the application of a tone-to-tone mapping approach as well.

To conclude, this study was conducted to further our understanding of cross-linguistic perception of tonal contrasts and expand current models of L2 phonology by advancing the *Functional Pitch Hypothesis* as a first step in defining naïve perception of lexically contrastive pitch and a baseline for L2 tone acquisition.

Acknowledgments: We would like to thank our participants who made this research possible. We are grateful to Dr. Jeffrey Holliday for his help in testing participants in Seoul, Korea. We also thank Professors Kenneth de Jong, Rex Sprouse, Charles Lin, Öner Özçelik, and the Department of Second Language Studies at Indiana University for their guidance and support, and Stephanie Dickinson (Senior Consultant & Managing Director, Indiana Statistical Consulting Center) for help and advice on statistical analyses. We further thank Jennifer Cole, Mirjam Ernestus, and two anonymous reviewers for their excellent suggestions, and audiences at the 13th Conference on Laboratory Phonology in Stuttgart, at the 2012 Second Language Research Forum in Pittsburgh, as well as at the 2013 New Sounds conference in Montréal for insightful feedback.

References

Abramson, Arthur S. 1975. The tones of central Thai: some perceptual experiments. In Jimmy G. Harris & James R. Chamberlain (eds.), *Studies in Tai linguistics. In honor of William J. Gedney*. Bangkok: Central Institute of English Language.

Best, Catherine. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research,* 171–204. Baltimore, MD: York Press.

Best, Catherine T., & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Second-language speech learning: The role of language experience in speech perception and production. A festschrift in honor of James E. Flege,* 13–34. Amsterdam: John Benjamins.

Braun, Bettina, & Elizabeth Johnson. 2011. Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics* 39. 585–594.

Burnham, Denis, Elizabeth Francis, Di Webster, Sudaporn Luksaneeyanawin, Francisco Lacerda, & Chayada Attapaiboon. 1996. Facilitation or attenuation in the development of speech mode processing? Tone perception over linguistic contexts. In Paul McCormack & Alison Russell (eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology,* 587–592. Canberra: Australian Speech Science and Technology Association.

Chang, Yueh-Chin, Pierre Hallé, Catherine T. Best, & Arthur Abramson. 2008. Do non-native language listeners perceive Mandarin tone continua categorically? Paper presented to the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers. Peking, China.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Cooper, Nicole, Anne Cutler, & Roger Wales. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45. 207–228.

Cutler, Anne. 1986. Forbear is a homophone: lexical prosody does not constrain lexical access. *Language and Speech* 29. 201–220.

Darcy, Isabelle, Hanyong Park, & Chung-Lin Yang. 2011. Underlying sources of individual differences: Cognitive abilities in L2 phonological development. Paper presented to the Second Language Research Forum, Ames, IA.

DeFrancis, John. 1984. *The Chinese language: Fact and fantasy.* Honolulu: University of Hawaii Press.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastián-Gallés, & Jacques Mehler. 1997. A destressing “deafness” in French? *Journal of Memory and Language* 36. 406–421.

Dupoux, Emmanuel, Nuria Sebastián-Gallés, Eduardo Navarrete, & Sharon Peperkamp. 2008. Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition* 106. 682–706.

Flege, James E. 1995. Second language speech learning. Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience. Issues in cross-language research,* 233–277. Timonium, MD: York Press.

Francis, Alexander L., Valter Ciocca, Lian Ma, & Kimberly Fenn. 2008. Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers. *Journal of Phonetics* 36. 268–294.

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1. 126–152.

Gandour, Jackson. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11. 149–175.

Gandour, Jackson, & Richard Harshman. 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Language and Speech* 21. 1–33.

Hallé, Pierre A., Yueh-Chin Chang, & Catherine T. Best. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32. 395–421.

Hao, Yen-Chen. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40. 269–279.

Kim, Jungsun. 2011. Perception of lexical pitch accent by Kyungsang and Cholla Korean listeners. In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences 2011 [ICPhS XVII],* 1070–1073. Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Kim, Jungsun, & Kenneth de Jong. 2007. Perception and production in the pitch accent system of Korean. In J. Trouvain & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences 2007 [ICPhS XVI]*, 1273–1277. Dudweiler: Pirrot.

Kim-Renaud, Young-Key. 2009. *Korean: An essential grammar*. New York: Routledge.

Kochanski, Greg, Esther Grabe, John Coleman, & Burton Rosner. 2005. Loudness predicts prominence: fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118. 1038–1054.

Kubozono, Haruo. 1999. Mora and syllable. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 31–61. Malden, MA: Blackwell Publishing.

Lee, Chao-Yang. 2007. Does horse activate mother? Processing lexical tone in form priming. *Language and Speech* 50. 101–123.

Lee, Iksop, & S. Robert Ramsey. 2000. *The Korean language*. Albany, New York: State University of New York Press.

Lee, Lisa, & Howard C. Nusbaum. 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Attention, Perception, and Psychophysics* 53. 157–165.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Li, Charles N., & Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar.* University of California Press.

McAllister, Robert, James E. Flege, & Thorsten Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

Morén, Bruce, & Elizabeth Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language and Linguistic Theory* 24(1). 113–178.

Otake, Takashi, & Anne Cutler. 1999. Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics* 27. 229–253.

Pallier, Christophe, Laura Bosch, & Nuria Sebastian-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64. B9–B17.

Pierrehumbert, Jane, & Mary Beckman. 1988. *Japanese tone structure. Linguistic inquiry monograph 15*. Cambridge: MIT Press.

Pike, Kenneth. 1948. *Tone languages*. Ann Arbor: University of Michigan Press.

Repp, Bruno H., & Hwei-Bing Lin. 1990. Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics* 18. 481–495.

Schaefer, Vance, & Isabelle Darcy. 2014. Pitch prominence matters: Perception of Thai tones by Seoul Korean and Kyungsang Korean speakers. *Proceedings of the International Symposium on the Acquisition of Second Language Speech: Concordia Working Papers in Applied Linguistics 5*, 597–611.

Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23. 287–308.

So, Connie K., & Catherine T. Best. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53. 273–293.

Stagray, James R., & David Downs. 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics* 21. 143–163.

Tsujimura, Natsuko. 2006. *An introduction to Japanese linguistics*. Malden, MA: Blackwell Publishing.

Van Lancker, Diana. 1980. Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication* 13. 201–277.

Van Lancker, Diana, & Victoria A. Fromkin. 1973. Hemispheric specialization for pitch and 'tone': Evidence from Thai. *Journal of Phonetics* 1. 101–109.

Wang, Yue, Allard Jongman, & Joan A. Sereno. 2001. Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language* 78. 332–348.

Wang, Yue, Allard Jongman, & Joan A. Sereno. 2006. L2 acquisition and processing of Mandarin Chinese tones. In Ping Li, Elizabeth Bates, Li Hai Tan, & Ovid J. L. Tseng (eds.), *The Handbook of East Asian Psycholinguistics*, 250–256. Cambridge: Cambridge University Press.

Wang, Yue, Dawn M. Behne, Allard Jongman, & Joan A. Sereno. 2004. The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics* 25. 449–466.

Wayland, Ratree, & Susan Guion. 2003. Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics* 24. 113–129.

Wayland, Ratree, & Susan Guion. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54. 681–712.

Weinreich, Uriel. 1953. *Languages in contact*. The Hague: Mouton Publishers.

Wood, Charles. 1974. Parallel processing of auditory and phonetic information in speech discrimination. *Attention, Perception, and Psychophysics* 15. 501–508.

Wu, Xianghua, Jung-Yueh Tu, & Yue Wang. 2012. Native and nonnative processing of Japanese pitch accent. *Applied Psycholinguistics* 33. 623–641.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Zhang, Yanhong, & Alexander Francis. 2010. The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics* 38. 260–271.

Zsiga, Elizabeth, & Rattima Nitisaroj. 2007. Tone features, tone perception, and peak alignment in Thai. *Language and Speech* 50. 343–383.

Appendix A: Syllables used for the test comparisons

Thai target syllables for the five tones, including gloss for real words. Gray shading indicates that the combination produces a nonword.

Segments	High tone	Mid tone	Low tone	Falling tone	Rising tone
[baː]		bar	shoulder	crazy	
[puː]		land; ground		male; person	
[tʃʰaː]	slow	tea			
[diː]		good			
[huː]	shrunken				ear
[kʰaːi]		to spit out	net; limit	camp	to sell
[maː]	horse	to come			dog
[maːi]	wood			window	to indicate
[miː]		to have	egg noodles		bear
[naː]	younger maternal uncle/aunt	rice field		face; season	thick
[pʰiː]				older brother/sister	ghost
[ruː]	to know	hole			
[suː]			to arrive at	to fight	
[tʰaː]	to dare	to smear on		if; posture	
[waː]		Thai measure of length		to say	
[jaː]			leader	bright	particle to call s.o.'s attention

Appendix B: Parameter estimates

Table B.1: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB accuracy

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
Intercept	.950	.083	133.009	11.421	.000	.785	1.115
[Lg = E]	−.100	.095	611.546	−1.051	.294	−.287	.087
[Lg = J]	−.109	.093	611.546	−1.174	.241	−.292	.073
[Lg = K]	−.175	.095	611.546	−1.840	.066	−.362	.012
[Detailedconditions = FR]	.023	.093	118.185	.245	.807	−.161	.206
[Detailedconditions = HF]	.025	.112	118.185	.223	.824	−.197	.247
[Detailedconditions = HR]	−.375	.112	118.185	−3.343	.001	−.597	−.153
[Detailedconditions = LF]	−.050	.112	118.185	−.446	.657	−.272	.172
[Detailedconditions = LH]	−.217	.121	118.185	−1.788	.076	−.457	.023
[Detailedconditions = LM]	−.217	.121	118.185	−1.788	.076	−.457	.023
[Detailedconditions = LR]	.050	.112	118.185	.446	.657	−.172	.272
[Detailedconditions = MF]	−.125	.112	118.185	−1.114	.267	−.347	.097
[Detailedconditions = MH]	−.119	.122	121.038	−.979	.330	−.361	.122
[Lg = E] * [Detailedconditions = FR]	−.218	.103	1689.950	−2.116	.034	−.420	−.016
[Lg = E] * [Detailedconditions = HF]	−.125	.125	1689.950	−1.001	.317	−.370	.120
[Lg = E] * [Detailedconditions = HR]	.200	.125	1689.950	1.602	.109	−.045	.445
[Lg = E] * [Detailedconditions = LF]	−.175	.125	1689.950	−1.402	.161	−.420	.070
[Lg = E] * [Detailedconditions = LH]	.167	.135	1689.950	1.236	.217	−.098	.431
[Lg = E] * [Detailedconditions = LM]	−.067	.135	1689.950	−.494	.621	−.331	.198
[Lg = E] * [Detailedconditions = LR]	−.100	.125	1689.950	−.801	.423	−.345	.145
[Lg = E] * [Detailedconditions = MF]	−.100	.125	1689.950	−.801	.423	−.345	.145
[Lg = E] * [Detailedconditions = MH]	−.164	.136	1690.382	−1.210	.226	−.430	.102
[Lg = J] * [Detailedconditions = FR]	.012	.101	1689.950	.123	.902	−.185	.210
[Lg = J] * [Detailedconditions = HF]	−.002	.122	1689.950	−.019	.985	−.242	.237
[Lg = J] * [Detailedconditions = HR]	.125	.122	1689.950	1.025	.306	−.114	.364

Table B.1 (cont.)

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
[Lg = J] * [Detailedconditions = LF]	.022	.122	1690.209	.176	.860	−.218	.261
[Lg = J] * [Detailedconditions = LH]	.164	.132	1689.950	1.242	.214	−.095	.422
[Lg = J] * [Detailedconditions = LM]	−.139	.132	1689.950	−1.058	.290	−.398	.119
[Lg = J] * [Detailedconditions = LR]	−.027	.122	1689.950	−.224	.823	−.267	.212
[Lg = J] * [Detailedconditions = MF]	.011	.122	1689.950	.093	.926	−.228	.251
[Lg = J] * [Detailedconditions = MH]	.003	.133	1690.941	.023	.982	−.258	.264
[Lg = K] * [Detailedconditions = FR]	−.016	.103	1689.950	−.154	.877	−.218	.186
[Lg = K] * [Detailedconditions = HF]	−.075	.125	1689.950	−.601	.548	−.320	.170
[Lg = K] * [Detailedconditions = HR]	.100	.125	1689.950	.801	.423	−.145	.345
[Lg = K] * [Detailedconditions = LF]	−.150	.125	1689.950	−1.201	.230	−.395	.095
[Lg = K] * [Detailedconditions = LH]	.142	.135	1689.950	1.051	.294	−.123	.406
[Lg = K] * [Detailedconditions = LM]	−.158	.135	1689.950	−1.174	.241	−.423	.106
[Lg = K] * [Detailedconditions = LR]	−.075	.125	1689.950	−.601	.548	−.320	.170
[Lg = K] * [Detailedconditions = MF]	−.050	.125	1689.950	−.400	.689	−.295	.195
[Lg = K] * [Detailedconditions = MH]	.011	.136	1690.382	.081	.935	−.255	.277

Covariance parameters	**Estimate**	**Std. error**
Residual	.156	.005
Subject	.006	.002
Item	.010	.003

Note: MR is the reference category; Mandarin is the reference language; Lg = language.

Table B.2: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AXB reaction times.

Fixed effects	Estimate	Std. error	*df*	*t*	Sig.	95% confidence interval	
						Lower	Upper
Intercept	1118.7	94.7	99.9	11.81	0.000	930.7	1306.7
[Lg = E]	258.7	124.4	96.6	2.08	0.040	11.7	505.8
[Lg = J]	223.9	121.2	95.4	1.85	0.068	−16.6	464.5
[Lg = K]	152.5	125.3	99.3	1.22	0.227	−96.2	401.2
[Detailedconditions = FR]	−75.3	79.1	103.7	−0.95	0.343	−232.1	81.5
[Detailedconditions = HF]	−47.1	96.1	105.2	−0.49	0.625	−237.7	143.5
[Detailedconditions = HR]	281.6	107.2	156.3	2.63	0.009	69.9	493.3
[Detailedconditions = LF]	64.8	96.6	107.2	0.67	0.504	−126.7	256.3
[Detailedconditions = LH]	−9.1	109.1	126.8	−0.08	0.933	−225.1	206.8
[Detailedconditions = LM]	115.3	110.4	132.7	1.04	0.298	−103.1	333.7
[Detailedconditions = LR]	−13.7	96.1	105.3	−0.14	0.887	−204.3	176.9
[Detailedconditions = MF]	27.9	99.4	119.0	0.28	0.780	−168.9	224.6
[Detailedconditions = MH]	49.4	108.0	122.2	0.46	0.648	−164.4	263.1
[Lg = E] * [Detailedconditions = FR]	15.2	97.7	1210.2	0.16	0.877	−176.6	206.9
[Lg = E] * [Detailedconditions = HF]	111.2	117.9	1207.4	0.94	0.346	−120.2	342.5
[Lg = E] * [Detailedconditions = HR]	−78.1	130.1	1210.2	−0.60	0.549	−333.4	177.2
[Lg = E] * [Detailedconditions = LF]	−98.9	122.2	1220.7	−0.81	0.419	−338.5	140.8
[Lg = E] * [Detailedconditions = LH]	−77.2	130.8	1206.5	−0.59	0.555	−333.7	179.4
[Lg = E] * [Detailedconditions = LM]	−97.9	145.7	1233.2	−0.67	0.502	−383.9	188.0
[Lg = E] * [Detailedconditions = LR]	−59.0	117.2	1204.4	−0.50	0.615	−289.0	170.9
[Lg = E] * [Detailedconditions = MF]	−107.6	125.1	1216.5	−0.86	0.390	−352.9	137.8
[Lg = E] * [Detailedconditions = MH]	−85.9	139.4	1219.6	−0.62	0.538	−359.4	187.6
[Lg = J] * [Detailedconditions = FR]	−115.0	92.8	1205.2	−1.24	0.215	−297.0	67.0
[Lg = J] * [Detailedconditions = HF]	−75.4	112.4	1203.7	−0.67	0.503	−295.9	145.2
[Lg = J] * [Detailedconditions = HR]	−126.3	130.8	1207.9	−0.97	0.335	−383.0	130.4
[Lg = J] * [Detailedconditions = LF]	41.3	114.7	1207.8	0.36	0.718	−183.6	266.3

Table B.2 (cont.)

Fixed effects	**Estimate**	**Std. error**	***df***	***t***	**Sig.**	**95% confidence interval**	
						Lower	**Upper**
[Lg = J] * [Detailedconditions = LH]	−61.4	127.5	1205.2	−0.48	0.630	−311.6	188.8
[Lg = J] * [Detailedconditions = LM]	−105.0	141.5	1219.8	−0.74	0.458	−382.6	172.7
[Lg = J] * [Detailedconditions = LR]	−76.8	112.5	1204.3	−0.68	0.495	−297.5	143.9
[Lg = J] * [Detailedconditions = MF]	78.9	120.2	1209.0	0.66	0.511	−156.8	314.7
[Lg = J] * [Detailedconditions = MH]	−19.3	130.1	1212.5	−0.15	0.882	−274.6	235.9
[Lg = K] * [Detailedconditions = FR]	163.5	97.6	1205.0	1.67	0.094	−28.1	355.0
[Lg = K] * [Detailedconditions = HF]	132.8	120.8	1208.1	1.10	0.272	−104.3	369.8
[Lg = K] * [Detailedconditions = HR]	−168.6	136.8	1214.5	−1.23	0.218	−437.0	99.9
[Lg = K] * [Detailedconditions = LF]	148.9	126.0	1228.5	1.18	0.238	−98.3	396.1
[Lg = K] * [Detailedconditions = LH]	40.5	134.2	1207.6	0.30	0.763	−222.7	303.8
[Lg = K] * [Detailedconditions = LM]	−1.1	153.7	1218.7	−0.01	0.994	−302.6	300.5
[Lg = K] * [Detailedconditions = LR]	−59.0	118.8	1204.9	−0.50	0.620	−292.1	174.2
[Lg = K] * [Detailedconditions = MF]	−55.4	126.4	1207.9	−0.44	0.661	−303.4	192.5
[Lg = K] * [Detailedconditions = MH]	57.7	136.7	1207.3	0.42	0.673	−210.6	325.9

Covariance parameters	**Estimate**	**Std. error**
Residual	117391.5	4789.4
Subject	43570.1	11073.0
Item	5765.1	2536.8

Note: MR is the reference condition; Mandarin is the reference language; Lg = language.