

Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese–English bilinguals perceive “Rose?” vs. “Rose”*

MARTA ORTEGA-LLEBARIA

Department of Linguistics, University of Pittsburgh, Learning Research & Development Center, University of Pittsburgh

MARITZA NEMOGÁ

Department of Linguistics, University of Pittsburgh

NORA PRESSON

Learning Research & Development Center, University of Pittsburgh

(Received: February 16, 2015; final revision received: September 21, 2015; accepted: October 01, 2015; first published online 28 October 2015)

Long-term experience with a tonal language shapes pitch perception in specific ways, and consequently Chinese speakers may not process pitch in English words – e.g., “Rose?” spoken as a question versus “Rose” spoken as a statement – in the same way as native speakers of non-tonal languages do. If so, what are those pitch processing differences and how do they affect Chinese recognition of English words? We investigated these questions by administering a primed lexical-decision task in English to proficient Chinese–English bilinguals and two control groups, namely, Spanish–English and native English speakers. Prime-target pairs differed in one sound and/or in pitch. Results showed specific cross-language differences in pitch processing between the Chinese speakers and the control groups, confirming that experience with a tonal language shaped the perception of English words’ intonation. Moreover, such experience helps to incorporate pitch into models of word-recognition for bilinguals of tonal and non-tonal languages.

Keywords: bilingual speakers, word recognition, pitch, F0, tone, intonation, language experience, Chinese, English, Spanish

1. Introduction

There is robust evidence that long-term experience with a tonal language shapes the perception of tone differently than experience with a non-tonal language does. Behavioral and neurophysiological studies have shown that there are differences in the way that Chinese and English speakers identify vowels and tones (e.g., Gottfried & Suiter, 1997), in their perception of tones in speech versus musical stimuli (e.g., Burnham & Francis, 1997), in the strength and scope of tone and segment integration (e.g., Repp & Lin, 1990; Lee & Nusbaum, 1993), in their perception of contour versus level tones (e.g., Gandour, 1983; Kaan, Wayland, Bao & Barkley, 2008) and in how they discriminate tones from another language such as Thai (e.g., Wayland & Guion, 2004). Moreover, dichotic listening experiments show that Chinese and English speakers also differ in their pitch lateralization (e.g.,

Wang, Jongman & Sereno, 2001); functional magnetic resonance imaging (fMRI) studies show differences in the neuronal encoding of pitch (e.g., Gandour, Dzemidzic, Wong, Lowe, Tong, Hsieh, Sathamnuwong & Lurito, 2003, Gandour, Tong, Wong, Talavage, Dzemidzic, Xu, Li & Lowe, 2004; Gandour, 2007), and event-related potential (ERP) studies show differences in tone processing (Chandrasekaran, Krishnan & Gandour, 2007b, 2009; Chandrasekaran, Gandour & Krishnan, 2007a). Altogether, these studies provide consistent evidence in favor of the idea that experience with a tonal language enhances the perception of tone. However, there is little research on intonation and whether and how language experience shapes its perception. Given that pitch variations in the signal constitute the phonetic material of both tone and intonation, a comprehensive theoretical account of the effect of language experience on pitch perception must account for intonation as well as tone.

1.1. The form-meaning mapping problem: F0, tone and intonation

While tone and intonation share F0, or rate of vocal fold vibration, as their acoustic material, they differ in the domains they apply to and meanings they convey. Tones

* We would like to thank the participants for their time and the undergraduate students who helped with data collection, in particular Erika Latham and Toni Cusimano. Special thanks to Chuck Perfetti for his sharp insights and the time he so generously gave. Thank you to the Language and Reading Group and to the Perfetti Lab for their feedback and support, especially Natasha Tokowicz, Wendy Li-Yun Chang, Xiaoping Fang and Joseph Stafura.

Address for correspondence:

Marta Ortega-Llebaria, Department of Linguistics, Cathedral of Learning, Office 2830, 4200 Fifth Avenue, Pittsburgh PA 15260, USA
mao61@pitt.edu

refer to the F0 variations that take place within words and change their meanings, and intonation includes the F0 variations that take place over an intonation phrase and convey sentence-level meanings. For example, the four tones of Mandarin work together with segments to convey lexical meaning, so that Mandarin speakers cannot know whether the word “ma” means “scold” or “hemp” until they process the F0 variation. “Ma” with tone 4, which has a falling F0, means “scold,” whereas it means “hemp” in combination with the rising F0 contour of tone 2. In contrast, in non-tonal languages like English, the F0 variation present in words does not change lexical meaning. The word “Rose” spoken with a falling F0 and “Rose” with a rising F0 refer to the same person. What these F0 shapes change is sentence-level meanings. “Rose” with a falling F0 can be an answer to the question “What’s her name?” whereas “Rose” with a rising F0 can be part of the question “Rose? Is it you?” Consequently, the different meanings that tone and intonation confer on an F0 pattern may lead to a form-meaning mapping problem – namely, the same F0 pattern, for instance a rising F0 over a word, can be interpreted either as tone 2 or as a question.

This mapping problem acquires further complexity when we consider the specific ways in which languages exploit tone and intonation, resulting in a continuum of tone-intonation interactions. Tonal languages like Mandarin and Cantonese can be placed at one end of this continuum and non-tonal, stress-accent languages like English and Spanish are at the opposite extreme. In Chinese, preserving tonal shapes is so important to word comprehension that the use of F0 to express sentence intonation becomes very limited (Chao, 1968; Yuan, 2004; Yuan & Shih, 2004; Wang & Xu, 2011). For example, Wang and Xu (2011) showed that the encoding of focus and topic in Mandarin is based exclusively on the manipulation of pitch range which preserves the tone shapes of words. However, other expressions of sentence intonation are in direct conflict with tone such as a tone 2–3 word in Cantonese placed at the end of questions (Kung, Chwilla & Schriefers, 2014; Ren, Tang, Li & Sui, 2013). In this context, the rising intonation of questions blurs the contrast between a low tone like 2–3 and a high tone like 2–5. Kung et al. (2014) showed that accuracy scores reflected speakers’ perception difficulties in this context because they were lower in questions than in statements, especially for 2–3 tones. Nevertheless, Cantonese speakers tried to recover the tonal contrast in questions, as shown by the P600 and N400 ERP components. The authors interpreted the P600 differences as evidence that speakers access competing representations of tone. N400 showed that tones were more difficult to access in questions than in statements because in questions tone shapes were distorted relative to a canonical shape. Thus, ERP evidence showed that

even in contexts where tonal contrasts were neutralized because of intonation, speakers tried to recover the tonal information, corroborating the theory that, for Chinese speakers, interpreting F0 variation in relation to words takes precedence over interpreting sentence intonation meanings.

In contrast to Chinese, this conflict between tone and intonation does not exist in languages like English or Spanish. In these languages, the F0 variation present in the speech signal is recruited exclusively for the expression of sentence intonation. It will be recalled that when processing “Rose” with a falling or a rising F0, English speakers relate this F0 variation to sentence-level meanings but not to word meanings. Consequently, it is assumed that the word recognition component in stress-accent languages like English or Spanish excludes F0 and current word recognition models reflect this assumption by including only segmental information (e.g., Marslen-Wilson & Warren, 1994; McClelland & Elman, 1986; Grosjean, 1998, 2008). Any effects of F0 on word recognition are thought to come from the interaction of the lexical component with sentence prosody (Cutler, 2012; Alves, 2010; Christophe, Peperkamp, Pallier, Block & Mehler, 2004; Michelas & D’Imperio, 2010; Sekiguchi, 2006) and that it is in this interaction where English and Spanish speakers give a specific F0 shape to the word. The details of this interaction, however, are still under debate.

Together, the form-meaning mapping problem and the language-specific interaction of tone and intonation trigger the question of whether experience with a tonal language shapes the perception of intonation in specific ways. For example, given that tone preservation in Chinese constrains the use of F0 to express intonation meanings whereas in English any F0 variation in the speech signal is recruited to express intonation, do Chinese ESL speakers perceive the above-mentioned contrast of “Rose” with a falling F0 versus “Rose?” with a rising F0 differently from English and Spanish ESL speakers? This question, which is at the core of our investigation, has implications for word recognition models and pitch perception, the areas addressed in Sections 1.2 and 1.3.

1.2. F0 and word recognition in tonal and non-tonal languages

There is consensus that F0 variations related to tone are processed at the lexical level whereas F0 variations related to intonation are processed at the post-lexical level. However, traditional models of word recognition do not consider F0 as a source of information and base the matching process between the incoming signal and the speaker’s representations only on segmental data. As noted by many researchers (Kung et al., 2014; Lee, 2007; Malins & Joanisse, 2010, 2012; Schrimmer et al., 2005; Ye & Connine, 1999; Yip, 2001), the reason behind this

exclusion is that models are based on non-tonal Indo-European languages like English where suprasegmental information, i.e., pitch, duration and loudness, has little or no role in word recognition (Cutler, 1986; Cooper, Cutler & Wales, 2002; Cutler, 1986, 2012; Cutler, Dahan & van Donselaar, 1997; Cutler, Wales, Cooper & Janssen, 2007). Even in tonal languages like Chinese, pioneering behavioral studies assumed that tone did not constrain word recognition to the same extent as segments did because tone was accessed later than segments (e.g., Cutler & Chen, 1997). However, recent eye tracking and ERP studies showed that in on-line processing there was not such difference in timing (Malins & Joanisse, 2010; Schirmer, Tang, Penney, Gunter & Chen, 2005; Malins & Joanisse, 2012). Tone and segments were accessed in the same time window, and they constrained lexical access to the same extent. Given these recent findings, some authors have highlighted the need to give a more relevant role to F0 in word recognition models for tonal languages (e.g., Malins & Joanisse, 2012).

In contrast with tonal languages, word recognition models in non-tonal languages still exclude F0 as a source of information. However, earlier studies on lexical stress showed that in contrast to English, where suprasegmental cues to stress played a minor or non-existent role in word recognition (e.g., Cutler, 1986; van Donselaar, Koster & Cutler, 2005), in languages like Spanish, German and Dutch suprasegmental cues to stress played a stronger role by reducing the number of possible word candidates (Soto-Faraco, Sebastián-Gallés & Cutler, 2001; Cutler & Pasveer, 2006; Cutler, 2012). These results were taken as supporting evidence for the interaction of the word recognition component with a higher, sentence-level prosodic component. How early in the word recognition process this higher-level prosody component is accessed is still a matter of debate (Cutler, 2012). Recently, studies in French, Portuguese and Japanese (Alves, 2010; Christophe et al., 2004; Michelas & D’Imperio, 2010; Sekiguchi, 2006) have found that pauses in sentences, which include F0 variation as one of the cues, affect word recognition, suggesting that sentence prosody, in particular pauses, has a more direct effect on word recognition than does simple bias. For example, Christophe et al. (2004) found that in French “chat grin” as in “un *chat grincheux*” (a grumpy cat) is not a good match for the word “chagrin” (grief), concluding that the prosodic analysis of sentences, which includes F0, is processed in parallel with lexical activation and recognition.

As in L1 models, in bilingual models of lexical access, such as the Bilingual Interactive Activation model (BIA, van Heuven, Dijkstra & Grainger, 1998, BIA+, Dijkstra & van Heuven, 2002), the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS, Shook & Marian, 2013), the Bilingual

Interactive Model of Lexical Access (BIMOLA, Léwy & Grosjean, 1997; Grosjean, 1998), and the Unified Model (2012), suprasegmental information is not addressed. For example, feature and phoneme nodes in BIMOLA refer exclusively to segmental information. Moreover, to our knowledge, bilingual models have yet to be tested on bilingual speakers of tonal and non-tonal languages such as Chinese–English bilinguals in whom F0 processing is especially interesting. Testing bilinguals of tonal and non-tonal languages becomes even more compelling when we take into consideration observations made in contexts of bilingualism and language contact. For example, Chinese speakers living in Spain produced words in statements with an F0 peak on the stressed syllables even though native Spanish speakers produce a post-tonic F0 peak in this context, showing that the prosodic inventories of Chinese–Spanish bilinguals lack the Spanish post-tonic pitch accent (Chen, 2007). Likewise, varieties of English with substrate tone languages like Cantonese and Nigerian English are observed to have a very reduced inventory of pitch accents, and speakers fail to deaccentuate words when the intonation of a sentence requires it (Gussenhoven, forthcoming; Gussenhoven & Udofot, 2010; Yiu, 2010). Altogether, these production patterns suggest that a tonal L1 affects the intonation of a non-tonal language, reinforcing the idea that bilingual speakers of tonal and non-tonal languages may process intonation in a stress-accent language such as English and Spanish differently from native non-tonal speakers.

In summary, F0 variation has not yet been thoroughly studied in word recognition models. Although F0 information is not included in word recognition models for non-tonal languages because it is assumed that any effect of F0 in word recognition comes from the interaction of the lexical component with a prosodic component, recent research on tonal languages has suggested that F0 information should be part of word recognition models. Bilingual models of word recognition still exclude F0 and have not yet been tested with bilingual speakers of tonal and non-tonal languages like Chinese–English bilinguals. However, this bilingual scenario offers a particularly interesting testing ground for studying F0 processing in word recognition by, for example, examining how speakers of tonal languages perceive the F0 variation of English words, which expresses intonation regardless of the fact it takes place in the word domain.

1.3. The effects of language experience on the perception of intonation

In addition to models of word-recognition, our core question – namely, whether long-term experience with a tonal language shapes the perception of intonation in English words in specific ways – has implications for pitch perception. As pointed out in Section 1,

studies on the perception of intonation are still scarce in comparison with tone studies. Many of these intonation studies used non-sense stimuli and showed that these stimuli were perceived differently by tonal and non-tonal speakers (e.g., Braun & Johnson, 2011; Braun, Galts & Kabak, 2014; Gandour et al., 2003, 2004; Gandour, 2007). For example, Braun and Johnson (2011) used the nonsense word pairs ‘mova/noba’ and ‘denu/zenu’ with two F0 contrasts, one emulating a tone and another an intonation contrast. Results from an AXB task showed that for Chinese speakers incongruent trials were more difficult to process if the F0 contrast resembled a tone, whereas for Dutch speakers the intonation trials were the most difficult, showing that Chinese and Dutch speakers perceive the same F0 stimuli differently, probably because of their language background. Moreover, Gandour and colleagues’ fMRI research (2003, 2004, 2007a) showed that these differences in the interpretation of pitch variation as tone or intonation in nonsense utterances had a neural basis. By comparing Chinese and English speakers’ pitch perception in hums and three and one-syllable nonsense utterances, they showed that regardless of the stimuli length, speech stimuli emulating a tone contrast triggered left hemisphere (LH) lateralization only in Chinese speakers. Other stimuli were processed mainly on the right hemisphere by speakers of both languages, suggesting that in speech tonal versus intonation contrasts rather than stimuli length triggered lateralization in speakers of tonal languages (for a complete overview of other factors as well as tone which trigger LH lateralization see Zatorre & Gandour, 2008).

Altogether, these studies provide strong evidence supporting the idea that language experience shapes F0 perception. The same F0 contour was perceived differently by speakers of tonal and non-tonal languages if this contour was identified as linguistically relevant by one of the language groups. Evidence was particularly compelling with the stimuli emulating tone contrasts because they triggered LH lateralization only in speakers of tonal languages. However, these results should be interpreted with caution because of the characteristics of the stimuli. Whereas using nonsense stimuli allowed a direct comparison between speakers of tonal languages and speakers of non-tonal languages, the interpretation of these stimuli in relation to either a tone meaning or an intonation meaning was based solely on the resemblance between the stimuli and the speakers’ L1. Yet, most de-contextualized F0 forms are ambiguous, and they can be related to both a tone and an intonation interpretation. Consequently, there is only indirect evidence that participants perceiving F0 variation in nonsense stimuli were indeed relating a particular F0 shape with the intended meaning. As a result, these studies demonstrate that speakers’ L1 had an effect on pitch perception but cannot provide direct evidence of whether

experience with a tonal language shaped the perception of intonation.

To our knowledge, only one study has used meaningful utterances to explore the perception of intonation by speakers of tonal and non-tonal languages. Liang and van Heuven (2007) examined the perception of tone and intonation in the Chinese dialect of Beijing by native speakers of other Chinese dialects, i.e., Nantong and Changsha, and by speakers of Uyghur, a non-tonal language unrelated to Chinese. Participants were asked to perform a tone identification task and an intonation task. Results showed that speakers of tonal dialects perceived tone more accurately and faster than intonation, whereas the non-tonal speakers perceived the boundary tone changes in the intonation task more accurately and faster than tones. This trade-off between tone and intonation showed that, despite the clear link in the stimuli between an F0 form and either a tone or an intonation meaning, speakers’ interpretations were still biased by what was linguistically relevant in their L1.

1.4. Rationale, research questions and predictions

In our study, we further explored the details of this trade-off between tone and intonation by administering a lexical decision task in English with words that contrasted in segment and F0 shape to Chinese ESL speakers and two control groups, i.e., Spanish ESL speakers and native English speakers. To ensure that results from this task gave information about the effects of a tonal language on the perception of intonation, four environmental factors were created to favor a clear mapping between the F0 shape of a given stimulus and an intonation interpretation. First, using English as the target language instead of a tonal language had the advantage of exposing participants to an unambiguous mapping between F0 shape and intonation meaning. English has no tone, and therefore any F0 shape is unambiguously related to intonation. Second, lexical decision tasks have an emphasis on meaning because speakers have to discern real words from non-words. In our task, this emphasis was amplified by making non-words different from real words in only one segment, i.e., ‘kice’ as opposed to ‘mice’ and ‘rice.’ Third, L1 interference was minimized by creating an environment that promoted the use of English. For example, the instructions and the task were in English, and the experiment was run in Pittsburgh, PA, where participants were living at the time of data collection. This pro-English environment enhanced the monolingual mode in ESL speakers, making speakers’ L1 remain as inactive as possible (Grosjean, 1998, 2008). Finally, participants were advanced ESL speakers and they used English regularly in their everyday lives because at the time of the experiment they were studying for a degree in a U.S. university. Therefore, these four environmental factors worked together to

enhance a clear mapping between F0 and intonation, which in turn ensured that the lexical decision task used in this experiment went beyond relating speakers’ L1 background to pitch perception in general and explored the effects of tone in the perception of intonation.

In addition to examining pitch perception, the lexical decision task also examined segment processing by comparing prime-target pairs that differed in one segment. These segment contrasts took place in onset and in coda position, i.e., rice-mice, plate-plane. Because in previous bilingual word recognition research segmental processing has been thoroughly studied whereas F0 information has not yet been addressed, the results of this lexical decision task on segments allow a comparison with previous research, providing a baseline that F0 results cannot provide because of its novelty. Previous research on segment perception found consistent evidence in favor of the idea that sub-syllabic components affected segment processing in languages like English and Spanish (Alloppenna, Magnuson & Tanenhaus, 1998; Desroches, Joannisse & Robertson, 2006; Desroches, Newman & Joannisse, 2009). However, results were not as clear in Chinese. Some authors proposed a holistic processing of the Chinese syllable based on evidence showing that, in Mandarin, syllabic temporal effects were stronger and more pervasive than those of their components (Zhao, Guo, Zhou & Shu, 2011). In contrast, results from recent studies failed to find supporting evidence for this holistic approach and showed that onsets and rhymes in Mandarin behaved similarly to the way they do in English (Malins & Joannisse, 2012).

Thus, the lexical decision task in this experiment examines both F0 and segment processing. Whereas results on segmental processing are expected to support those of previous research, F0 results constitute a new contribution to the pitch perception literature in the sense that, to our knowledge, no study has explored the perception of F0 variation in English words by bilingual speakers of tonal and non-tonal languages. In particular, we answer the research questions below.

1. In a lexical decision task where English words contrast in one segment and/or F0 shape, what are the processing similarities between speakers of tonal languages like Chinese and speakers of non-tonal languages like Spanish and English?
2. Are there any cross-language processing differences? In particular, are there any differences in F0 processing between Chinese ESLs on the one hand and Spanish ESL and English native speakers on the other, suggesting that long-term experience with a tonal language shapes the perception of intonation in English words in specific ways?
3. If so, what are the implications of the above F0 processing differences for word recognition models,

especially in models for bilingual speakers of tonal and non-tonal languages?

2. Methodology

2.1. Participants

The experimental group consisted of 40 Chinese native speakers (31 females, 9 males) with ages ranging from 18 to 45 years ($M = 27.5$ $SD = 7.05$). English native speakers ($N = 53$, 33 females and 20 males) and native Spanish speakers ($N = 41$, 19 females and 22 males) constituted the non-tonal control groups. English speakers’ ages ranged from 18 to 33 ($M = 21.33$ $SD = 2.7$) and Spanish speakers’ ages ranged from 21 and 44 ($M = 32.5$ $SD = 5.29$). At the time of the experiment, participants were undergraduate or graduate students at the University of Pittsburgh. The Chinese and Spanish speakers rated their English proficiency in speaking, listening, reading and writing as either intermediate-high or high. They used English every day, both at work and in social contexts. Out of the 81 L2 English speakers, 77 had lived in the US between two and four years. The remaining four participants had lived 10 years or more in the US. Participants reported having normal speech and hearing.

2.2. Materials and recordings

A list of 128 monosyllabic words with a CVC, CCVC or CVCC syllabic structure was created to generate materials for the lexical decision task. These 128 words were grouped into 32 sets of 4. Each quartet consisted of a target word, e.g., mice, a non-word target, e.g., kice, a related prime, e.g., rice, and an unrelated prime e.g., gold. Sixteen quartets contained prime-target pairs that differed in one segment in onset, e.g., rice-mice. Another 16 quartets contained pairs that contrasted in one segment in coda, e.g., plate-plane. The unrelated primes shared no segment with the target, i.e., gold-mice. Words in each quartet had similar frequencies according to the Spoken Frequency Database (Pastizzo & Carbone, 2007). In general, words’ relative frequency ranged from 0.00 to 0.04.

Two native speakers of English were asked to produce each of the 128 words within two carrying sentences, i.e., “She said _____,” which yielded words with a falling F0, and the incredulity question “She said _____?” which gave words with a rising F0 (see Figure 1). The recording was created with Praat at a sampling rate of 44100 Hz (8 bits) with a high quality condenser microphone, Sennheiser Evolution e845, in a quiet room. Two trained phoneticians, two native English speakers, and two native Chinese speakers listened to these recordings and selected the speaker with the clearest diction and his most clearly spoken words. These selected words were used for the

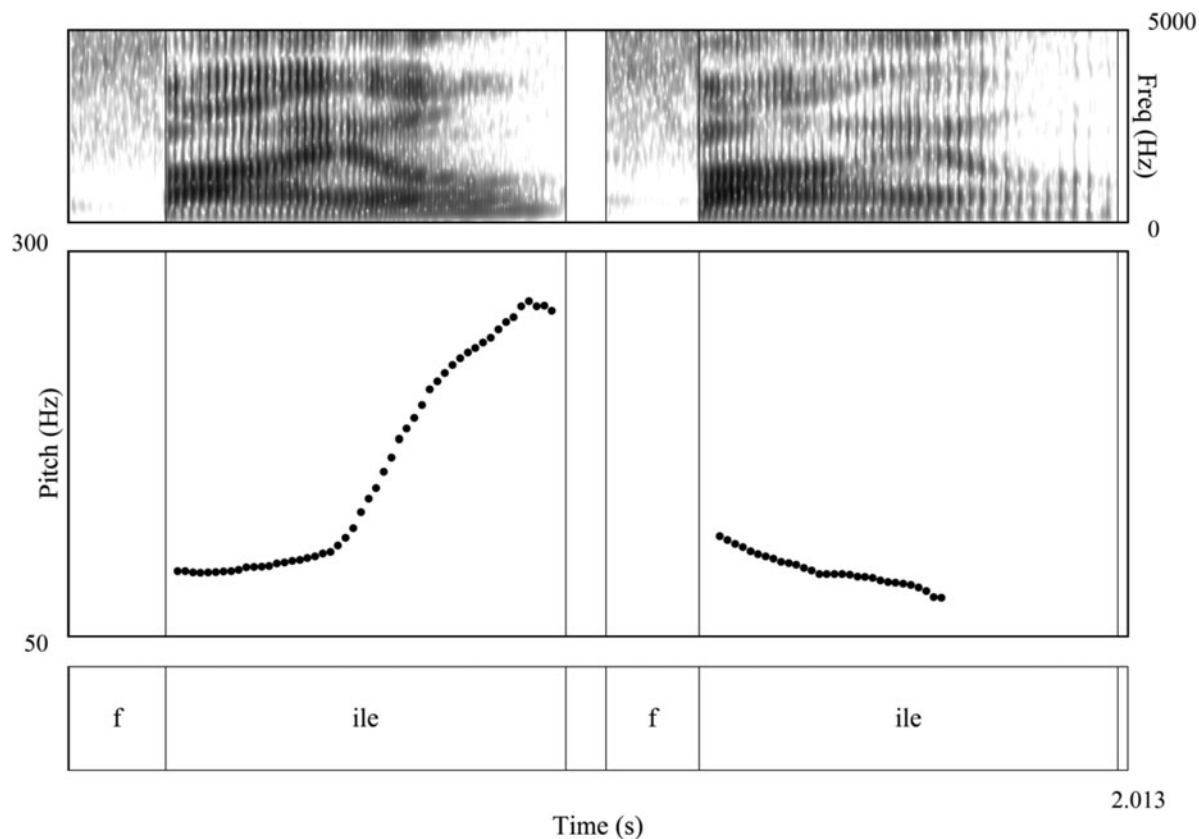


Figure 1. Spectrograms and F0 tracks of the word file spoken as a question (left) or a statement (right)

lexical decision task (80 words * 2 F0 contours = 160 words; see Appendix 1 for a complete word list).

To further examine the differences between falling and rising F0 contours of the selected 160 words, we measured F0 at three different points in the pitch contour, i.e., within the initial 20 ms, at the middle and within the last 20 ms (means and standard deviations are summarized in Appendix 2). The mean differences between falling and rising contours were around 20 Hz at the beginning of the contour, increasing to over 100 Hz towards the end. Since the just noticeable difference (jnd) remains constant at 3 Hz in pure tones ranging from 100 Hz to 1000 Hz (Shower & Biddulph, 1931; Wever & Wedell, 1941), the pitch differences between falling and rising contours in the test words remained clearly audible all along the word.

2.3. Task

The selected 160 words were paired into prime-target sets according to five experimental conditions, i.e., Full Match (FM), Full Mismatch (FMM), Mismatch in Segment (MMS), Mismatch in F0 (MMF0) and Mismatch in Segment and F0 (MMSF0). In Full Matches, prime and target words had the same segments and F0 contour, e.g., *miceR- miceR* in which *R* refers to rising F0. In the Full Mismatch condition, primes had no segments in

common with the target and had opposite F0 contours, e.g., *goldF- miceR* in which *F* refers to falling F0. In the Mismatch in Segment condition, the mismatched segment could take place in onset or in coda, e.g., *riceR- miceR*, *plateF- planeF*. In the Mismatch F0 condition, primes and targets differed in the pitch contour, e.g., *miceF- miceR*. As explained in Section 2.3, F0 differences between falling and rising contours were audible all along the contour from the beginning of the word. Finally, prime-target pairs in the Mismatch in Segment and F0 condition differed in one segment and in the speech contour, e.g., *riceR- miceF*. Each condition was created with word and non-word targets.

All the stimuli combined yielded a total of 400 pairs. To ensure that the task did not last more than 30 minutes, only 120 of the 200 target non-word pairs were used, reducing the number of pairs to 320 (200 with word targets and 120 with non-word targets). With these 320 word pairs, we built three lists, one with 50 ms between prime and target, a second with 250 ms and a third with 750 ms in order to examine whether processing differences between L1 groups increased with larger ISIs which promote a more linguistic and less acoustic mode of perception (e.g., Lee, 2007). These lists were listened to by a similar number of participants within each language group (Chinese: 15, 15, 10; English: 21, 16, 16; Spanish: 15, 15, 11).

2.4. Procedure

The 134 participants listened to the priming lexical decision task via Sennheiser HD555 headphones connected to the computers in the Phonetics Laboratory at the University of Pittsburgh. The participants’ task consisted of listening to word pairs and deciding if the second word in a pair was a real English word. They were instructed to press the red button in the response pad (Cedrus RB840) if they heard a real English word and the blue button if they heard a non-word. After reading the instructions in the computer screen, participants had a practice session that consisted of 10 randomly selected pairs (6 non-word targets, 4 word targets) not included in the task. Once participants had completed the session, they were encouraged to ask questions before starting the task. The task started with a warm-up of 20 pairs (10 word targets, 10 non-word targets) to give participants time to get acquainted with the task and the response pad. These 20 pairs were not included in the results. The task proper consisted of the next 320 pairs (200 word targets, 120 non-word targets). Participants were given a one-minute break after every 40 stimuli, for a total of seven breaks. The practice, warm-up and experimental task took approximately 30 minutes.

2.5. Measurements and statistics

Speakers are faster at processing their native language than an L2. For example, the native English speakers in our study were generally faster ($N = 10263$, $M = 502.76$ ms, $SD = 359$) than the L2 groups (Chinese: $N = 7643$, $M = 611.14$ ms, $SD = 439$; Spanish: $N = 7797$, $M = 717.19$ ms, $SD = 448$). However, our goal was to capture cross-language differences related to the experience with tonal (L1 Chinese) versus non-tonal languages (L1 English, L1 Spanish) rather than speed differences related to being a native versus a non-native English speaker. To ensure the capture of the cross-language differences related to a tonal versus a non-tonal background, we ran our models on two score types, i.e., facilitation scores and log-transformed RTs. Facilitation scores were calculated as the difference between the Full Mismatch condition (FMM) and each of the other four conditions, i.e., $FMM - FM$, $FMM - MMS$, $FMM - MMF0$, and $FMM - MMSF0$. For example, the RTs obtained by participant 1 in the FM pair *miceF-miceF*, in the MMS pair *riseF-miceF*, etc were each subtracted from the RT he obtained in the FMM word pair *goldR-miceF*. Larger scores represented stronger facilitation effects (see Figure 2). Thus, facilitation scores were less sensitive than RTs to speed differences because of the participants’ L1-L2 difference as they were normed against a control condition. Thus, we were interested in factors that became significant in the facilitation scores. Yet we also

considered factors that were significant in both measures to ensure that significant factors from the facilitation scores were robust enough to remain significant even when there was L1-L2 variation in the data and were not a mere artifact of the subtraction in the control condition.

The facilitation scores and log-transformed RTs were computed on correctly identified English words (Chinese speakers correctly identified 90.4% of real English words, English speakers 91.8% and Spanish speakers 90%) and were analyzed by means of hierarchical linear modeling (package **lmer** in R; Bates, Maechler, Bolker & Walker, 2014) with random intercepts for participant and stimulus pairs. There was no evidence of variability in random slopes for these terms. Because there was no consistent evidence of differences between the three values of ISI for both facilitation scores and log-transformed RT data, data were collapsed across ISI for the analyses reported here. Two models were analyzed for each outcome based on a priori prediction: main effects of L1 (Chinese, English, Spanish), Condition (FM, MMF0, MMS, MMSF0) and Intonation (target with rising F0, target with falling F0) and their interactions in section 3.1., and the main effects of L1, Condition and Syllable Position (Onset, Coda) and their interactions in section 3.2.. For each model, *p*-values, least square means and differences of least square means for fixed effects were estimated with the *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2014).

3. Results

3.1. L1. Intonation and Condition

A hierarchical linear model with the fixed effects of L1 (Chinese, English, Spanish), Intonation (falling F0, rising F0) and Condition (FM, MMS, MMF0, MMSF0, FMM) and the random effects of subject and item were run separately on facilitation scores and on log-transformed RTs on correctly identified words. Results are displayed in Table 1 below.

As shown in Table 1, the three-way interaction, the two-way interactions of L1 with Intonation and Condition and the main factor of Condition were significant at $p < 0.05$ in both datasets, ratifying that these effects were robust. Whereas the significant main factor of Condition shows that there are some cross-language resemblances, the significant interactions with L1 show that there are cross-language differences in the modulating effect of the participants’ native language.

Cross-language similarities showed up in the significant main factor of Condition (see Table 1) which denotes the way in which participants processed the degree of similarity between primes and targets. Regardless of the speakers’ language background, participants obtained the largest facilitation scores in Full Matches ($N = 4927$, $M = 153.6$, $SD = 499$). Facilitation

Table 1. Effects of L1, condition and intonation on facilitation scores and log-transformed RTs.

	Facilitation Scores				Log-Transformed Scores			
	NmDF	DnDF	F. value	Pr(>F)	NmDF	DnDF	F. value	Pr(>F)
L1	2	129	0.52	0.596	2	131	25	<0.0001***
Condition	3	19646	155	<0.0001***	4	24144	181.2	<0.0001***
Intonation	1	38	2.281	0.01392	1	24144	0.1	0.72419
L1:Condition	6	19646	2.128	0.04691*	8	24144	3.8	0.00020***
L1:Intonation	2	19672	3.962	0.01904*	2	24144	21.4	<0.0001***
Cnd:Intn	3	19645	1.378	0.24743	4	24144	5.5	0.00022***
L1:Cnd:Intn	6	19645	2.443	0.02310*	8	24144	3.1	0.00149**

Signif. codes: p<.0001 **** 0.001 *** 0.01 ** 0.05 * † 0.1 † † 1

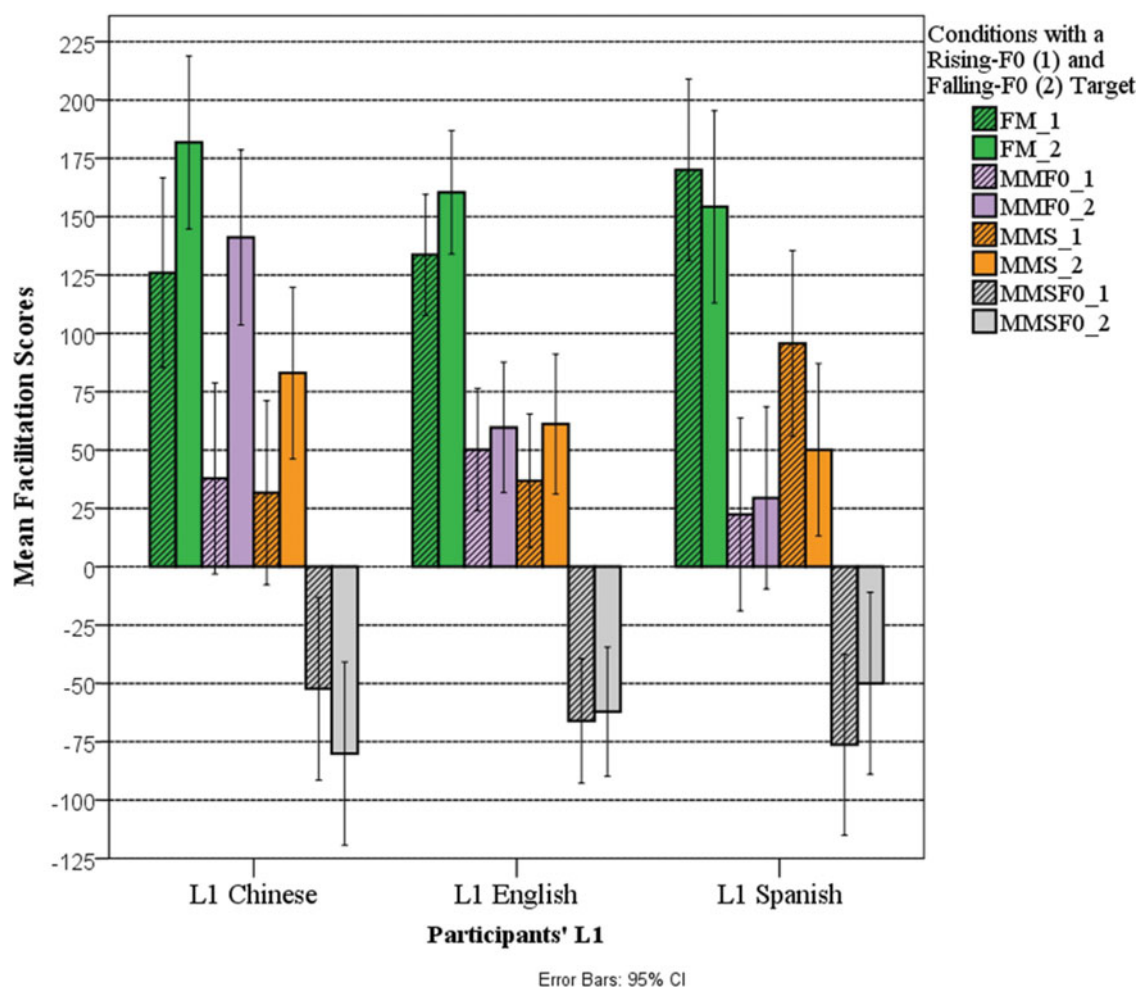


Figure 2. (Colour online) Participants' Facilitation Scores by L1, Condition and Intonation. Participants' L1 is on the x-axis. Facilitation scores in ms. are along the y-axis and represent the difference of each condition with Full Mismatches, the control condition.

scores became progressively smaller in conditions where the prime and the target differed by one dimension, i.e., Mismatched F0 (N = 4963, M = 56.5, SD = 509), Mismatched Segment (N = 4989, M = 58.7, SD = 509) and finally by two dimensions, i.e., Mismatches

F0 and Segment (N = 4981, M = -64.4, SD = 503). This progression suggested that more similar pairs triggered larger facilitation effects than less similar pairs. As depicted in Figure 2, these similarity effects took place within each language group and in both targets.

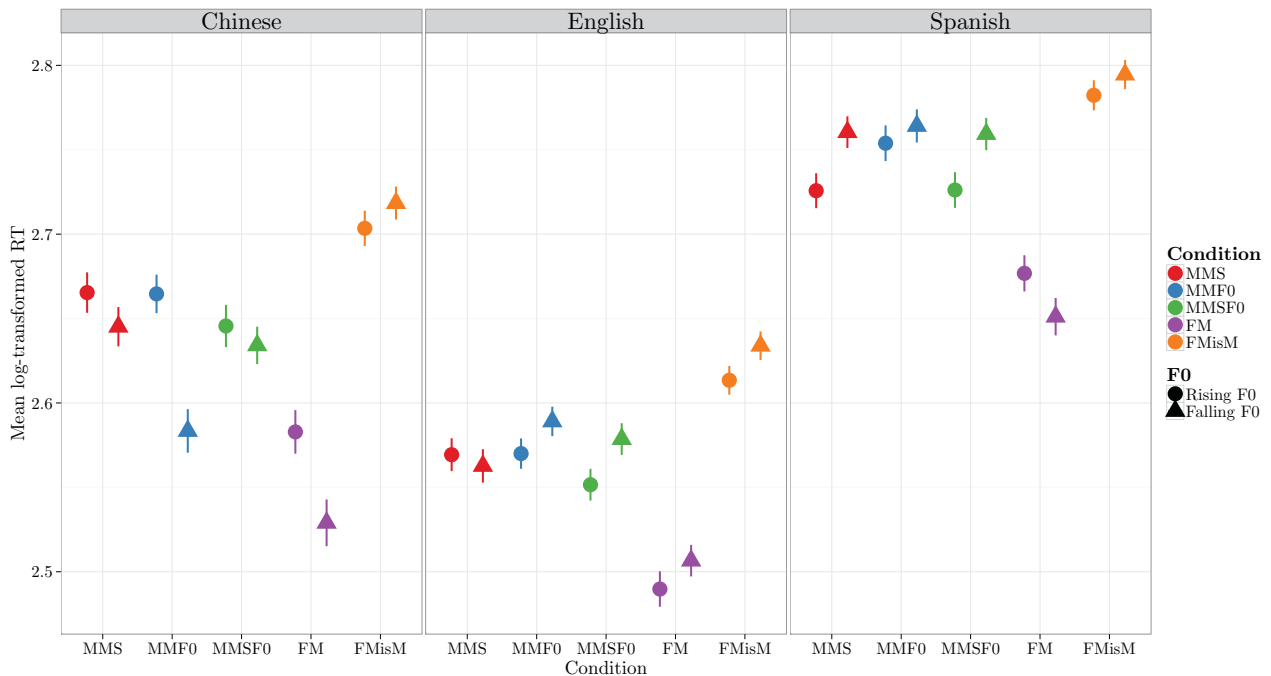


Figure 3. (Colour online) Participants’ log-transformed RTs by L1, Condition, and Intonation.

Within each language group, the blue bars for the Full Matches obtained larger facilitation scores than the green and red bars for the Mismatched F0 and Mismatched Segment conditions respectively, and these were larger than the orange bars for the Mismatched F0 and Segment conditions.

Regarding the cross-language differences, the three-way interaction showed that, for Chinese speakers, facilitation scores were larger in targets with a falling F0 (bars in darker shades in Figure 2) than in targets with a rising F0 (bars in lighter shades), and that this difference became especially large in the Mismatched F0 condition. The two-way interactions corroborate these results. The L1 * Intonation interaction confirms that, for Chinese speakers, a target with a falling F0 (N = 2948, M = 81.2, SD = 529) triggered more facilitation than one with a rising F0 (N = 2937, M = 35.6, SD = 556), $t(139.6) = -4.08, p = 0.0001$, and that in contrast with Chinese, differences between falling and rising F0 targets did not reach significance in English, $t(94.1) = -0.98, p = 0.3295$, or in Spanish, $t(135.8) = -0.13, p = 0.8953$. The L1 * Condition interaction ratifies that for Chinese speakers, a mismatch in F0 produced significantly larger facilitation than in English (M diff = 35), $t(633.9) = 2.91, p = 0.0038$, and Spanish (M diff = 63), $t(652.1) = 3.95, p = 0.0001$. Thus, Chinese speakers were the only group showing a bias for targets with a falling F0, which became especially noticeable in F0 mismatches.

In order to examine the three-way interaction further, hierarchical linear models with the fixed effects of

Intonation and Condition and the random effects of subject and item on facilitation scores were run separately for each language group. As summarized in Table 2, results showed that only Chinese speakers obtained significant differences for Intonation and Intonation*Condition ratifying that Chinese speakers’ performance was different from that of non-tonal language speakers. The significant Intonation factor confirmed that only Chinese speakers showed a bias towards targets with a falling F0, and the significant Intonation * Condition interaction corroborated that this bias was especially large in F0 mismatches. In addition, results showed that Condition was significant in each language group, confirming that computations of similarity between primes and targets were similar across language groups, i.e., full matches were more similar than mismatches in one dimension and those were more similar than mismatches in two dimensions.

3.2. Effects of onset and coda

As explained in Section 2, syllabic position was relevant only for conditions containing mismatched segments. Therefore, a second model with the random factors of subject and item and the fixed factors of L1, Condition and Syllable Position (Onset Coda) was applied to the facilitation scores of Mismatched Segment and Mismatched Segment and F0 conditions. The model was run on log-transformed RTs. Recall that facilitation scores and log transformed RTs were based on participants’

Table 2. Effects of condition and intonation on facilitation scores within each language group.

		Facilitation Scores			
		NmDF	DnDF	F. Value	Pr (>F)
Chinese	Condition	3	5728	43.96	<0.0001***
	Intonation	1	37.8	5.7772	0.0212*
	Cnd:Intn	3	5727	3.79	0.0099**
English	Condition	3	5564	44.54	<0.0001***
	Intonation	1	36.4	0.204	0.6541
	Cnd:Intn	3	5564	1.251	0.2896
Spanish	Condition	3	7930	76.5	<0.0001***
	Intonation	1	38	1.77	0.1913
	Cnd:Intn	3	7929	0.32	0.811

Signif. codes: p < .0001 **** 0.001 *** 0.01 ** 0.05 * † 0.1 †† 1

Table 3. Effects of L1, condition and syllable position on facilitation scores and log transformed RTs.

	Facilitation Scores				Log-Transformed Scores			
	NmDF	DnDF	F. value	Pr(>F)	NmDF	DnDF	F. value	Pr(>F)
L1	2	9959	0.537	0.58	2	131	27.1	<0.0001***
Condition	1	9959	152.38	<0.00001***	2	14394	91	<0.0001***
SyllPos	1	9959	0.329	0.56	1	18	4	0.0597†
L1:Condition	2	9959	10.454	0.63	4	14394	1.5	0.2049
L1:SyllPos	2	9959	0.296	0.74	2	14395	5	0.0065**
SyllPos:Cond	1	9959	125.5	<0.00001***	2	14394	94.6	<0.0001***
L1:SyllP:Cnd	2	9959	0.834	0.43†	4	14394	1.9	0.1088

Signif. codes: p < .0001 **** 0.001 *** 0.01 ** 0.05 * † 0.1 †† 1

reaction times to correctly identified words. As in the previous models, results were based on significant factors from both datasets. Results are summarized in Table 3.

The effect of Syllable Position was significant in both datasets only in their interaction with Condition, showing that speakers detected mismatched segments faster in onset position than in coda. In the log-transformed RT dataset, L1 was significant as a main effect, showing that native English speakers were faster overall ($N = 4087$, $M = 470.45$, $SD = 352.146$) than L2 English speakers (Chinese: $N = 3054$, $M = 577.85$, $SD = 432.796$, Spanish: $N = 3120$, $M = 665.06$, $SD = 437.606$). Moreover, in log-transformed RTs, Syllable Position was marginally significant as a main effect and significant as an interaction with L1. As a main effect, it showed that mismatches in onset were detected 28 ms faster than in coda. In the interaction with L1, English speakers detected mismatches in onsets 16 ms faster than in codas, Chinese speakers 40 ms faster and Spanish speakers 33 ms faster. Overall, these results showed that speakers from the three

language groups detected mismatches in onsets faster than in codas.

3.3. Summary

Participants from the three language backgrounds exhibited similar patterns regarding the syllable position of the mismatched segment and the number of mismatched dimensions in the prime. Pairs with mismatched segments in onset, e.g., *riceF-miceF*, triggered more facilitation than pairs with mismatched segments in the coda position, e.g., *birdF-birthF*. Regarding the number of mismatched dimensions, fully matched pairs obtained higher facilitation scores than pairs mismatched in the segmental dimension e.g., *riceF-miceF*, or the F0 dimension, e.g., *miceF-miceR*. In turn, the latter pairs obtained higher facilitation scores than pairs mismatched in the two dimensions, e.g., *riceF- miceR*. Thus, more similar pairs triggered greater facilitation effects than less similar pairs. Altogether, the above results

indicate that cross-language similarities were related to segment processing and general computations of similarity involving F0.

In contrast, cross-language differences were related to F0 processing and singled out Chinese speakers from the other two language groups. First, Chinese speakers obtained higher facilitation scores when the target word was produced with a falling F0 and these differences occurred within each experimental condition. In contrast, F0 variation in the target word had no significant effects for English and Spanish speakers. Second, Chinese speakers obtained higher facilitation scores in F0 mismatches than English and Spanish speakers, especially in F0 mismatches with a falling F0 target, *miceR-miceF*. Thus, these cross-language differences set Chinese speakers apart from the other language groups by showing a Chinese bias towards English words with a falling F0, which triggered more facilitation, especially in the MMF0 condition.

4. Discussion

4.1. Cross-language similarities

Cross-language similarities were related to both segment and F0 processing. Similarities in segment processing showed that for speakers from the three L1 backgrounds mismatches in onsets triggered temporally different effects from mismatches in codas. Speakers obtained higher facilitation scores in onsets than in codas, suggesting that they recovered faster from a mismatch that occurred earlier in the word. Detecting a segment mismatch in onset in a CVC word allowed facilitation to start from the vowel. However, detecting a segment mismatch in coda delayed the whole recovery process. Previous research in English and Spanish L1 obtained similar results (Alloppenna et al., 1998; Desroches et al., 2006, 2009). However, previous research in Chinese obtained controversial results. On the one hand, some studies showed that in Mandarin the syllable as a whole had temporal effects that go above and beyond those of their parts, suggesting a more holistic processing of the syllable (Zhao et al., 2011). On the other hand, results from recent studies failed to find supporting evidence for this holistic approach while showing that onsets and rhymes in Mandarin behaved similarly in English (Malins & Joannis, 2012). Our results showed that speakers from the three language backgrounds processed segment mismatches in onsets and codas similarly, suggesting that the three languages, Chinese, English and Spanish, have similar onset-coda temporal effects. If so, this would support the findings that onsets and codas in Chinese are processed similarly in English. However, if Chinese syllables are in fact processed holistically and onsets and codas behave differently from English, then our results would indicate

that our Chinese speakers did learn the English syllabic structure and processed it in a native-like manner.

The second cross-language similarity referred to F0 processing. Speakers from the three language groups found full matches more similar than mismatches in one dimension, i.e., a mismatch in segment or in F0, and those more similar than mismatches in two dimensions, i.e., a mismatch in both segment and F0. That means that speakers from the three language groups found prime-target pairs with the same segments more similar when they did not differ in F0, i.e., the pair *riceF-miceF* was more similar than *riceR-miceF* and the pair *miceF-miceF* was more similar than *miceR-miceF*. Thus, regardless of whether they had a tonal or a non-tonal L1, speakers could not ignore F0 variation in English words when making a lexical decision. This result was unexpected, especially for native English speakers and Spanish ESLs, because, as reflected in current models of word recognition for non-tonal languages, F0 information is not part of the computations of lexical access and recognition. Yet our results indicate that speakers from the three language backgrounds used F0 in the English lexical decision task.

One possible explanation is that, as described in Section 1, speakers from tonal and non-tonal languages accessed F0 by means of different mechanisms. On the one hand, Chinese speakers could have accessed F0 during the word recognition process by transferring to English those mechanisms that computed tone. On the other hand, in English and Spanish, F0 information was computed at the post-lexical level in the sentence prosody component, and F0 could have been accessed in word recognition through the interaction with this component. Yet it seems implausible that such different mechanisms – one operating at the lexical level and another at the post-lexical – yield similar outcomes.

A common mechanism based on the integral processing of segments and F0 information provides an alternative explanation. There is solid evidence that speakers of both tonal and non-tonal languages cannot ignore F0 variation in the speech signal when processing segmental information. Traditional Garner-type studies (e.g., Lee & Nusbaum, 1993; Miller, 1978; Repp & Lin, 1990) showed that both speakers of Chinese and speakers of English could not avoid processing F0 variation when attending to segmental contrasts. Cross-language differences resided in the degree and scope of this integration, showing more integration for speakers of tonal languages. For example, Lee and Nusbaum (1993) showed that Chinese speakers perceived segments and tones integrally in both a tone task and a flat pitch task whereas English speakers perceived them integrally only in the tone task. Likewise, identification and discrimination studies using word-level stimuli provide further evidence for different levels of integration (e.g., Braun & Johnson, 2011; Braun et al., 2014). In an AXB

task, Braun and Johnson (2011) found that speakers of non-tonal languages could not ignore F0 in incongruent trials, albeit speakers of tonal languages used the F0 dimension more often – i.e., 7.1% by Dutch speakers versus 43% by Mandarin speakers. Altogether, these studies showed that despite cross-language differences in degree, F0 and segment information were processed integrally by speakers of tonal and non-tonal languages, and, as a result, speakers from different language backgrounds cannot ignore F0 variation even in a lexical decision task in English where F0 variation becomes irrelevant for word meaning. Since integral perception of segmental and F0 information takes place at lower levels of processing, this cross-language mechanism operates with acoustic information and without relating F0 to lexical meaning. In this respect, it resembles the processing of F0 to identify the gender of the speaker. A mechanism of this sort would suffice to account for the cross-language pattern whereby prime-target pairs with identical segments are less similar if they differ in F0.

In summary, speakers from the three language backgrounds processed English words similarly regarding syllable structure and the acoustic integration of the segmental and F0 information in the speech signal. Regardless of the nuances of syllable structure in the speakers' L1, i.e., a more holistic approach in Chinese than in English and Spanish, speakers were sensitive to the linearity of the speech signal in that they recovered sooner from mismatches in onsets than in codas. It was clear as well that speakers from the three language groups integrated segmental and F0 information at lower levels of processing and that this acoustic perception accounted for their initial computations of similarity between prime-target pairs where pairs with identical segments were less similar if they differed in F0.

4.2. *Cross-language differences*

Whereas speakers from the three language groups were similar in that they could not ignore F0 while processing segmental information, Chinese speakers differed from English and Spanish speakers in another two aspects of F0 processing. These differences are carefully examined in this section in order to evaluate whether they support the idea that experience with a tonal language shapes the perception of intonation in specific ways, answering in this way our second research question.

The first F0 pattern that differentiated Chinese speakers from speakers of non-tonal languages is that Chinese speakers obtained significantly higher facilitation scores in F0 mismatches, suggesting that, in comparison with the control condition, Chinese speakers detected F0 variation faster than speakers of non-tonal languages. Higher facilitation scores and faster reaction times in F0 mismatches have also been obtained in lexical

decision tasks in tonal languages (e.g., Yip, 2001) and in training studies with word-object pairs (e.g., Braun et al., 2014). For example, Braun et al. (2014) showed that Chinese speakers were more accurate and faster at detecting F0 mismatches than speakers of German, Russian and French, suggesting a finer perception of F0 by speakers of tonal languages. This finer F0 perception by Chinese speakers in comparison with speakers of non-tonal languages has been replicated not only in words but in a wide range of stimuli such as level and contour tones (Gandour, 1983; Chandrasekaran et al., 2007b; Kaan et al., 2008), and speech versus musical tones (Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda & Keller, 1996) indicating that it affects perception of F0 contours in general and not only those related to lexical meaning. Moreover, neurophysiological research provides evidence for a relation between earlier and higher levels of processing, making it feasible that an earlier fine perception of tone is related to experience with a tonal language. For instance, MMN studies showed that this finer F0 perception by Chinese speakers in contrast to English speakers took place at a pre-attentive level of processing (Chandrasekaran et al., 2009). Connections from cortical areas to the brainstem (e.g., Heald & Nusbaum, 2014) provide neurological evidence for a direct relation between a pre-attentive level of processing with higher-level processing. Altogether, these results show that experience with a tonal language refines perception of F0 contours in general regardless of conveyance of tonal or intonation meanings, failing to support the idea of a special relation between experience with tonal language and perception of intonation.

However, the second F0 pattern that differentiates the current sample of Chinese speakers from English and Spanish speakers constitutes compelling evidence in support of the idea that experience with a tonal language shapes the perception of intonation in specific ways. Chinese speakers obtained higher facilitation scores when the target word had a falling F0. This difference occurred in each condition and it became especially large in F0 mismatches. For example, while a full match like *miceF- miceF* obtained higher facilitation scores than *miceR- miceR*, an F0 mismatch like *miceR- miceF* obtained notably higher facilitation scores than *miceF- miceR*, showing that, for Chinese speakers, the shapes of the F0 contours played a stronger role in their recognition of English words than they did for English and Spanish speakers. For Chinese speakers, English words with a falling F0 were easier to retrieve than those with a rising F0, suggesting that words with falling F0, the pitch contour of words in citation form, constituted a closer match to their representation of English words. Thus, Chinese ESL speakers represented English words with a rather fixed F0 contour as they did with tone in Chinese

words. In contrast, for English and Spanish speakers, both F0 contours were equally acceptable. Words with falling or rising F0 were equally easy to retrieve, suggesting that their representation of English words does not contain a fixed or preferred F0 contour. Thus, these results showed that despite the clear link in our stimuli between F0 and intonation meanings, the pro-English context promoting a monolingual English mode and the advanced proficiency of our ESL speakers, L1 was still active and Chinese speakers processed English words with a tone-like, fixed F0 shape, failing to process intonation contours in a similar way to English and Spanish speakers.

In summary, some aspects of F0 processing were common across languages whereas others singled out Chinese speakers from English and Spanish speakers. Cross-language similarities showed that speakers from tonal and non-tonal languages processed both segmental and tonal information in English words. Despite having no bearing on lexical meaning, F0 could not be ignored in the English lexical decision task by any of the language groups, presumably because the F0 and spectral acoustic information of the speech signal were processed integrally at early stages of perception. This perceived F0 variation had the effect of reducing facilitation scores in prime-target pairs that contained the same segments but had an additional mismatch in F0, i.e., *miceF-miceR* had less facilitation than *miceR-miceR*. This initial cross-linguistic perception of F0 was shaped further by experience with a tonal language, as illustrated by the two F0 patterns that singled out Chinese speakers from English and Spanish speakers. First, Chinese speakers were faster at detecting F0 mismatches. Since this finer F0 perception has been obtained in a large variety of stimuli, it reflects a more detailed acoustic perception of F0 contours regardless of their link with tone or intonation meanings. However, the second F0 pattern provided solid evidence that tonal language experience shaped the perception of intonation in specific ways. Chinese speakers represented the intonation of English words with a rather fixed F0 contour reminiscent of a tone, which in turn, favored recognition of words with a falling F0 over those with a rising F0. This tone-like representation demonstrated that Chinese speakers failed to process intonation as English and Spanish speakers did and this cross-language difference appeared despite the clear link between F0 shapes and intonation meanings in English words, the English monolingual mode and the advanced proficiency of our ESL speakers.

4.3. Implications for word recognition models

To our knowledge, models of bilingual word recognition have not yet addressed F0 processing. We fill this gap by discussing how the three F0 processing patterns described above could be implemented in current bilingual models.

The first F0 pattern showed that contrary to expectations speakers from the three language backgrounds found prime-target pairs with identical segments more similar if they did not differ in F0, showing that regardless of whether they had a tonal or non-tonal L1, speakers processed F0 information in our lexical decision task in English, a language where F0 was assumed not to be part of the lexical access and recognition process. We suggested that this unexpected cross-language behavior could be accounted for by the integral perception of segmental and F0 information at early stages of processing. This integral perception could be represented in word recognition models by feature detectors that analyze both the spectral and the F0 information of the speech signal. However, feature detectors in current bilingual models only analyze cues to segmental information. For example, in BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997), the speech signal is linked to French and English phonemes by a set of 16 segmental features inspired by the original Chomsky and Halle (1968) features. In BLINCS (Shook & Marian, 2013), the speech signal is mapped into English and Spanish phonemes by means of three-dimensional vectors that capture place, voice and manner characteristics in segments. By considering cues that capture F0 information, such as the tonal center of gravity (Barnes, Veilleux, Brugos & Shattuck-Hufnagel, 2010) in addition to cues to segmental information, these models could include F0 information in their feature detectors, allowing low-level processing of F0 information in bilingual word recognition.

Second, our results showed that Chinese ESL speakers were faster at detecting F0 variation in English words than speakers of non-tonal languages, indicating that the fine tuning to F0 used by Chinese speakers when processing Chinese words is also used when the same Chinese speakers processed English words. In contrast to Chinese ESL speakers, Spanish ESL speakers did not show these effects. This cross-language difference can be modeled by shared F0 feature detectors by Chinese and English on the one hand and by Spanish and English on the other, providing further supporting evidence for the ideas that feature detectors are shared by the two languages of a bilingual speaker, as they are in many models like BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997), BIA+ (Dijkstra & van Heuven, 2002), and BLINCS (Shook & Marian, 2013), and that they are shaped by the speakers' long-term exposure to tone (Chandrasekaran et al., 2007b, 2009). Therefore, by virtue of these cross-language, especially-sensitive F0 feature detectors, bilingual speakers of a tonal and a non-tonal language will have a more nuanced perception of F0 in their non-tonal language than native speakers of that same non-tonal language.

The third F0 pattern showed that Chinese–English bilinguals had a bias for English words with a falling

F0 so that these words were recognized faster than words with a rising F0. This bias suggested that, after detecting the acoustic F0 characteristics of English words via feature detectors, Chinese–English speakers kept on processing F0 information. In Chinese words, this latter F0 processing was a required step for accessing tone in a lexical representation (i.e., Malins & Joanisse, 2012). In English words, however, it was not so required, as shown by the absence of a similar bias in our native English speakers and Spanish–English bilinguals, who nevertheless successfully recognized English words. Therefore, this late F0 processing of English words by Chinese–English bilinguals arose from the connection between the phono-lexical representations of a tonal and a non-tonal language, which can be modeled by self-organizing maps (SOM) like those used in BLINCS (Shook & Marian, 2013) and the Unified Model (MacWhinney, 2004).

Self-organizing maps work on the Hebbian learning principle that “what fires together wires together” (Hebb, 1949). For example, in the phono-lexical SOMs in BLINCS (Shook & Marian, 2013), English and Spanish words organized themselves according to their phonological similarity, equivalent translations and word frequency so that a word like *tenedor* (“fork” in Spanish) activated a within language competitor, *tortuga* “turtle,” a cross-linguistic competitor, *tent*, and the translation equivalent, *fork*. In Chinese words, however, segmental information is wired together with F0 information. This means that in phono-lexical SOMs for Chinese–English bilinguals, words from the two languages are organized by F0 similarity in addition to segmental and word frequency criteria. Moreover, in English, words with a falling F0 (i.e., words in citation form, words at the end of statements) are more frequent than those with a rising F0. As a result of the F0 and word frequency organization criteria, Chinese–English bilinguals recognized English words with a falling F0 faster than those with a rising F0.

In contrast to models like BLINCS (Shook & Marian, 2013) and the Unified Model (MacWhinney, 2004), it would be difficult to reproduce this bias in a model like BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997) where languages are kept separate, albeit interconnected, in the phoneme and lexical representations. Interlanguage connections in BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997) are regulated by top-down effects like language mode and linguistic context, and bottom-up effects, like language-specific phonotactics. In our experiment both top-down effects and phonotactics strongly favored a unilingual English mode, whereby interlanguage activation was kept at a low level. Therefore, in this context, the late F0

processing of English words should be avoided, making it difficult to model a falling F0 bias on English words.

In summary, the first two F0 patterns affected early processing and were amenable to implementation in a wide selection of models. They would require that word recognition models of both tonal and non-tonal languages included F0 feature detectors shared by the two languages. However, the last of the three F0 patterns, the falling F0 bias in English words, showed interlingual interference at later processing levels which was better handled by bilingual models with highly interconnected languages, i.e., BLINCS (Shook & Marian, 2013) and the Unified Model (MacWhinney, 2004). Devices like phono-lexical SOMs allowed modeling of the falling F0 bias on English words by organizing both English and Chinese words according to F0 shape in addition to segmental similarity and word frequency, showing that the interconnection between English and Chinese lexical forms is stronger than that predicted by selective models like BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997).

5. Conclusion

The results of this experiment illustrate how F0 variation in English words is processed by speakers of tonal and non-tonal languages. Cross-language similarities suggested an initial computation of acoustic similarity based on the integral processing of segmental and F0 variation in the speech signal. After this initial stage, language experience shaped F0 perception further, as suggested by the two patterns that singled out Chinese speakers from English and Spanish speakers. First, Chinese speakers detected F0 mismatches faster. Because this pattern has been found in a wide range of stimuli, it was interpreted as an enhanced acoustic representation of F0 contours owed to experience with a tonal language. Second, Chinese speakers more quickly processed words with a falling F0, suggesting a bias towards a fixed F0 representation in English words reminiscent of tone despite any F0 variation in English words being unambiguously related to intonation. In light of these results, it is suggested that models of word recognition should include F0 information at lower levels of processing not only in tonal languages but also in non-tonal languages like English. Moreover, bilingual models of word recognition for Chinese–English speakers need to address the falling F0 bias. It has been argued that models with highly interconnected lexicons such as BLINCS (Shook & Marian, 2013) could account better for this bias than more selective models like BIMOLA (Grosjean, 1998, 2008; Léwy & Grosjean, 1997).

Appendix 1. Word quartets

Related Prime	Target Word	Non-Word Target	Unrelated Prime
Rice	mice	kice	gold
free	tree	pree	month
men	ten	len	house
flue	glue	plue	cake
silk	milk	pilk	judge
line	fine	hine	group
joy	toy	doy	wise
cat	pat	nat	winds
pain	gain	zane	suit
kiss	miss	fiss	tape
bird	birth	birn	twice
grease	greed	greach	booth
fight	file	fipe	wave
lake	laid	laip	noun
plate	plane	plabe	fifth
roof	root	roop	oil
team	teach	teag	air
paid	pain	paik	staff
lunch	lung	lunt	mouse
moon	mood	moob	ear

Appendix 2. F0 values in word stimuli

Means and standard deviations of the words’ F0 values at the beginning (F0₁), the middle (F0₂), and the end of the F0 track (F0₃).

Contrast	Pitch	F0 ₁ M	SD	F0 ₂ M	SD	F0 ₃ M	SD
Onset	Rising	102.98	9.49	148.94	42.24	248.16	51.24
	Falling	125.16	20.51	126.89	15.51	97.59	11.53
Coda	Rising	98.26	7.51	144.02	45.12	250.67	49.19
	Falling	124.26	15.24	104.03	12.38	83.53	4.62

Means and standard deviations of the non-words’ F0 values at the beginning (F0₁), the middle (F0₂), and the end of the F0 track (F0₃).

Contrast	Pitch	F0 ₁ M	SD	F0 ₂ M	SD	F0 ₃ M	SD
Onset	Rising	101.67	9.64	174.23	28.78	247.95	28.78
	Falling	117.93	14.85	104.77	16.02	91.85	9.12
Coda	Rising	98.26	8.50	146.90	53.97	237.31	51.85
	Falling	124.82	16.84	102.46	10.54	84.46	2.63

References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.
- Alves, D. (2010). Phonological phrase boundary restrictions in lexical access by BP adult speakers. In *Proceedings of Speech Prosody*, Chicago, IL.
- Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2010). The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Speech Prosody*, Chicago.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. <http://CRAN.R-project.org/package=lme4> (retrieved 1/26/2015).
- Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. *Southeast Asian linguistic studies in honour of Vichin Panupong*, 29–47.
- Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research*, 30, 323–350.
- Braun, B., & Johnson, E. K. (2011). Question or tone? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39, 585–594.
- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In *Spoken Language, 1996. ICSLP 96. Proceedings of the Fourth International Conference on Spoken Language Processing*, Vol. 1, pp. 2514–2517.
- Chandrasekaran, B., Gandour, J. T., & Krishnan, A., (2007a). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25, 195–210.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain research*, 1128, 148–156.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Sensory processing of linguistic pitch as reflected by mismatch negativity. *Ear and Hearing*, 3, 552–558.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press, Berkeley, CA.
- Chen, Y. (2007). *A Comparison of Spanish Produced by Chinese L2 Learners and Native Speakers—an Acoustic Phonetics Approach*. ProQuest.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row: New York.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. *Adult data*. *Journal of Memory and Language*, 51, 523–547.
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and speech*, 45, 207–228.
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201–220.
- Cutler, A. (2012). *Native listening: Language Experience and the recognition of spoken words*. Cambridge, MA: MIT Press.
- Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165–179.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, 40(2), 141–201.
- Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In *Proceedings of the XVI international congress of phonetic sciences*, pp. 1913–1916.
- Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In *3rd International Conference on Speech Prosody*. TUD press.
- Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, 100(3), B32–B42.
- Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 21, 1893–1906.
- Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197.
- Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, 11(2), 149–175.
- Gandour, J. (2007). Neural substrates underlying the perception of linguistic prosody. *Tones and Tunes*, 2, 3–25.
- Gandour, J., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Sathamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and language*, 84(3), 318–336.
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., Li, X. & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, 23, 344–357.
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25, 207–231.
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1, 131–149.
- Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press, Oxford, UK.
- Gussenhoven, C., & Radboud, U. (forthcoming). On the intonation of tonal varieties of English. *The Oxford Handbook of World Englishes*.
- Gussenhoven, C., & Udofot, I. (2010, May). Word melodies vs. pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English. In *Proc. of Speech Prosody*.
- Heald, S. L., & Nusbaum, H. C. (2014). Talker variability in audio-visual speech perception. *Frontiers in psychology*, 5.

- Kaan, E., Wayland, R., Bao, M. & Barkley, C. M. (2008). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, 1148, 113–122.
- Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, 53, 293–309.
- Kuznetsova, A., Brockhoff, P., & Christensen, R. H. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-11. <http://CRAN.R-project.org/package=lmerTest>.
- Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, 50, 101–123.
- Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, 53, 157–165.
- Léwy, N., & Grosjean, F. (1997). A computational model of bilingual lexical access. *Manuscript in preparation*, Neuchâtel University, Neuchâtel, Switzerland.
- Liang, J., & van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. *Tones and tunes*, 2, 27–61.
- MacWhinney, B. (2004). A unified model of language acquisition. *Handbook of bilingualism*, 49. <http://repository.cmu.edu/cgi/>
- Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, 62, 407–420.
- Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, 50, 2032–2043.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological review*, 101(4), 653.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1), 1–86.
- Michelas, A., & D’Imperio, M. (2010). Accentual phrase boundaries and lexical access in French. In *Proceedings of Speech Prosody*, Chicago, Illinois. pp 4. <https://hal.archives-ouvertes.fr/hal-00463205>
- Miller, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics*, 24, 175–180.
- Pastizzo, M. J., & Carbone, R. F. (2007). Spoken word frequency counts based on 1.6 million words in American English. *Behavior Research Methods*, 39, 1025–1028.
- Ren, G. Q., Tang, Y. Y., Li, X. Q., & Sui, X. (2013). Pre-attentive processing of Mandarin tone and intonation: Evidence from event-related potentials. In *Tech*. <http://creativecommons.org/licenses/by/3.0>.
- Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *The Journal of the Acoustical Society of America*, 87, S46–S46.
- Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1–12.
- Sekiguchi, T. (2006). Effects of lexical prosody and word familiarity on lexical access of spoken Japanese words. *Journal of psycholinguistic research*, 35, 369–384.
- Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, 16, 304–324.
- Shower, E. G., & Biddulph, R. (1931). Differential pitch sensitivity of the ear. *Journal of the Acoustical Society of America* 3, 275–287.
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45(3), 412–432.
- van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A*, 58, 251–273.
- van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39, 458–483.
- Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language*, 78, 332–348.
- Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics*, 39, 595–611.
- Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681–712.
- Wever, E. G., & Wedell, C. H. (1941) Pitch discrimination at high frequencies, *Psychological Bulletin* 38, 727.
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609–630.
- Yip, M. C. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia*, 44(3), 223–229.
- Yiu, S. Y. S. (2010). Intonation of English Spoken in Hong Kong. *Hong Kong Baptist University*, 2010.
- Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling*. Doctoral dissertation, Cornell University.
- Yuan, J., & Shih, C. (2004). Confusability of Chinese intonation. In *Speech Prosody 2004, International Conference*, Nara, Japan.
- Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, 49(7), 1761–1770.