*Erschienen in: Second Language Research ; 30 (2014), 3. - S. 323-350*
*https://dx.doi.org/10.1177/0267658313510926*

second
language
research

*Article*

Second Language Research
2014, Vol. 30(3) 323–350
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0267658313510926
slr.sagepub.com
SAGE

# Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation

**Bettina Braun and Tobias Galts**
University of Konstanz, Germany

**Barış Kabak**
University of Würzburg, Germany

## Abstract

Native language prosodic structure is known to modulate the processing of non-native suprasegmental information. It has been shown that native speakers of French, a language without lexical stress, have difficulties storing non-native stress contrasts. We investigated whether the ability to store lexical tone (as in Mandarin Chinese) also depends on the first language (L1) prosodic structure and, if so, how. We tested participants from a stress language (German), a language without word stress (French), a language with restricted lexical tonal contrasts (Japanese), and Mandarin Chinese controls. Furthermore, German has a rich intonational structure, while French and Japanese dispose of fewer utterance-level pitch contrasts. The participants learnt associations between disyllabic non-words (4 tonal contrasts) and objects and indicated whether picture–word pairs matched with what they had learnt (complete match, segmental or tonal mismatch conditions). In the tonal mismatch condition, the Mandarin Chinese controls had the highest sensitivity, followed by the German participants. The French and Japanese participants showed no sensitivity towards these tonal contrasts. Utterance-level prosody is hence better able to predict success in second language (L2) tone learning than word prosody.

## Keywords

encoding, French, German, intonation, Japanese, L2 tone, Mandarin Chinese, stress, word learning

**Corresponding author:**
Barış Kabak, Englische Sprachwissenschaft, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany.
Email: baris.kabak@uni-wuerzburg.de

*Konstanzer Online-Publikations-System (KOPS)*
*URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-288169*

# 1 Introduction

Acquiring the sound system of a second language (L2) necessitates the accurate perception of not only novel segmental properties of the target language, but also suprasegmental ones (intensity, duration, pitch; see Lehiste, 1970). While there is a large body of work on how non-native sound distinctions are perceived (Best, 1995; Flege, 1995; for a review, see Altmann and Kabak, 2011; Strange and Shafer, 2008), little is known about their phonological encoding in the mental lexicon. Furthermore, since studies concerning non-native segmental distinctions have been the impetus behind well-known theoretical models in the field of L2 phonetics and phonology (e.g. Best, 1995; Flege, 1995), even much less is known about the nature and organization of suprasegmentals in L2 grammars, and how suprasegmental distinctions in a learners' first language (L1) interacts with the acquisition and processing of novel suprasegmental contrasts. Below, we will first provide a brief overview of the L2 processing of suprasegmentals before postulating our research objectives and hypotheses.

In the domain of L2 stress contrasts, L1-induced variation among learners has been widely reported in both perception and production (e.g. Altmann, 2006; Archibald, 1997; Kawagoe, 2003; Ou, 2007). Recent experimental research has shown that differences in the way stress is employed in the listeners' L1 may determine their ability to store stress information in an L2 (for non-native stress perception typologies, see Altmann, 2006; Peperkamp and Dupoux, 2002). For instance, French listeners, whose L1 has no lexically contrastive stress, were shown to be impaired in encoding stress contrasts (Dupoux et al., 1997, 2001, 2008). Strikingly, language-specific modulation of stress sensitivity seems to take effect very early in human development. For example, French but not Spanish nine-month-old infants fail to discriminate stress contrasts (Skoruppa et al., 2009). This impairment, which Dupoux and colleagues termed 'stress deafness', is robustly revealed only in tasks that tap phonological representations (i.e. storage) of stress. More specifically, French speakers cannot reliably encode contrastive stress and therefore fail when the task requires long-term memory representations of stress while they can distinguish pairs contrasting in stress on the basis of acoustic information.[1] 'Stress deafness' is therefore argued to stem from processing stress at an abstract phonological level rather than at a psycho-acoustic level (see Domahs et al., 2012; Schwab and Llisterri, 2011). This suggests that at lower levels of processing, the phonetic differences may be discriminated by listeners, yet they may have a hard time encoding this information as part of the phonological representation of novel words.

Tones create yet another set of suprasegmental lexical contrasts in the languages of the world. As in the case of L2 stress, perception of L2 tones is also modulated by the usage of suprasegmentals in the L1 (e.g. Chiao et al., 2011; Gandour, 1983; Lee et al., 1996; Leung, 2008; Qin and Mok, 2011a, 2011b; So, 2005; Wang et al., 2006). For instance, examining the perception of Mandarin Chinese tones by Mandarin Chinese, Cantonese, Taiwanese, Thai, and English native speakers, Gandour (1983) found that native speakers of English were more sensitive to pitch height than to pitch movements (directionality), while native speakers of Thai, another tone language, were more sensitive to pitch movements than to pitch height.[2] Hallé et al. (2004) investigated Mandarin Chinese tone perception by native speakers of Taiwan Mandarin and French L2 speakers of Mandarin Chinese. They found that French speakers were able to discriminate Taiwan Mandarin tones, albeit clearly not as categorically as Taiwan Mandarin native speakers, suggesting that French speakers perceive tones as non-linguistic variation. Furthermore, the stronger the linguistic function of pitch is in the L1, the more sensitive to it listeners are. For instance, Braun and Johnson (2011) showed that Dutch listeners consistently matched non-words with linguistically irrelevant rising or falling pitch patterns to segmentally identical stimuli that matched or mismatched in pitch pattern. However, they were slower in doing so, if the pitch movements signalled postlexical information (question or statement) in accordance with the way Dutch employs pitch. Mandarin Chinese listeners were even more torn between pitch and segmental information as a categorization criterion since, unlike in Dutch, pitch constitutes a potential lexical contrast in their L1. Given that the function of pitch movements in the native language results in different sensitivities to pitch contrasts, it is conceivable that utterance-level prosodic contrasts can also be redeployed in the L2 processing of pitch variation, a prediction that we will revisit below.

The ability to perceive pitch differences, however, does not entail successful encoding of pitch information in establishing lexical contrasts, and this is the starting point for the current study. More specifically, at the acoustic level, the differences in pitch (fundamental frequency) are also available to listeners from non-tonal languages (Hallé et al., 2004), yet they may not be able to encode this information when building novel lexical representations due to lack of experience with storing long-term memory representations of pitch in their L1. At the segmental level, it has been widely shown that pairs of words that differ minimally with respect to a novel contrast may be encoded as homophones in the learner lexicon, even in highly fluent second language learners (e.g. Cutler et al., 2006; Pallier et al., 2001; Weber and Cutler, 2004; for further discussion, see Hayes-Harb and Masuda, 2008). Possibly, only additional metalinguistic knowledge, such as orthographic representations, may help to make learners aware of certain contrasts that are phonetically difficult to distinguish (Escudero et al., 2008). At the suprasegmental level, Wong and Perrachione (2007) showed that metalinguistic awareness also influences the learning of lexical pitch contrasts. In their study, English speakers could learn to use pitch patterns for word identification, albeit with a large amount of individual variation in learning outcomes, which correlated with their ability to identify pitch patterns in non-lexical contexts (i.e. the ability to match the pitch contour of nonce syllables to arrows representing level, rising, or falling tones on the computer screen) as well as their musical ability (for a recent study on the interaction between individual differences in perception and the design of the training paradigm, see also Perrachione et al., 2011). Finally, Hayes-Harb and Masuda (2008) tested whether L2 learners whose L1 does not have consonantal length contrasts are able to lexically store these contrasts in an L2. In their study, participants had to memorize Japanese-like non-words differing in consonantal length. The results showed that native speakers of American English were able to store distinct lexical representations of singletons and geminates (e.g. /pete/ 'dress' vs. /petːe/ 'stove') after one year of exposure. However, the same learners showed problems in producing the learned contrasts, suggesting that the consonantal length was not encoded accurately. The authors argued that the novel phoneme was initially encoded as an unfamiliar version of its closest counterpart in the L1 (e.g. target /tː/ stored as /t*/).[3]

In summary, while research on cross-linguistic speech perception has captured a great deal of variation in the discovery of phonological contrasts, very little is known about how novel suprasegmental contrasts are encoded in learners' lexical representations, and the extent to which this may be modulated by the L1 of the learner. Here we approach these issues by focusing on the acquisition of L2 tone contrasts by learners from non-tonal L1 backgrounds and test whether the function of lexical suprasegmental information in the L1, as shown for stress by Dupoux and colleagues, also has an impact on the lexical encoding of tones. Dupoux et al. (1997: 419) allude to this by making an explicit prediction that their findings 'will extend to other dimensions (tone, pitch accent, etc.)'. On a par with this conjecture, we expect L1-induced variability in the acquisition of L2 tone contrasts since non-tonal languages also use fundamental frequency ($f_0$), albeit to signal different functions, and hence may bring about different levels of sensitivity to suprasegmental contrasts. In particular, we see two levels at which languages differ, on the word-level and the utterance level, which we will unpack below.

From the perspective of word-prosodic typology, languages differ in the degree to which stressed syllables are marked in the lexicon. On the one hand, we observe so-called 'free stress' or 'lexical stress' languages, where the position of stress is defined for each word individually (and hence needs to be lexically stored, e.g. Russian or German; see below for further details). On the other hand, there are 'fixed stress' languages, where stress is typically anchored to a predictable position such as word edges (e.g. Finnish and Czech: word-initial; Turkish: word-final; Polish: penultimate), or with no lexical stress (e.g. French, where stress is assigned above the word level). Somewhere intermediate, we find languages, such as Japanese, in which not all words bear a lexically defined pitch accent, and for those that do its position in the word is not predictable. For these words, pitch accents need to be lexically stored.

From the perspective of utterance-prosodic typology (Jun, 2005), languages may be characterized by employing a relatively rich intonational system (e.g. English and German). In these languages, utterance-level prominence can be expressed by a range of different accent types and boundary tones, which signal information structure, sentence mood, etc. (Pierrehumbert and Hirschberg, 1990). At the other extreme, there are languages, such as French and Japanese, with less variability in $f_0$ at the utterance level, typically characterized by simple pitch rises or falls at the edges of prosodic domains.

## II Test languages and hypotheses

Here we focus on (1) the lexical status of suprasegmental features (word-level prosody) and (2) the complexity of the pitch system (number of utterance-level contrasts) in the L1s of our learner groups. We rely on recent autosegmental-metrical approaches and compare the number of basic pitch accents, phase accents and boundary tones.

### *1 German*

Each lexical word in German has a particular syllable that attracts primary stress; longer words also have syllables with secondary stress (Wiese, 2006). The syllable with primary stress is the anchor for utterance-level pitch accents (realized by pitch modulation, increased duration and intensity). The location of primary stress cannot be determined solely on the basis of structural aspects (e.g. with respect to edges), and any syllable within a polysyllabic word can theoretically bear primary stress although it mostly occurs on one of the last three syllables (see Vennemann, 1990). Therefore, the German lexicon contains a number of words and compounds forming stress minimal pairs (e.g. [ˈgeːnʊs] ‘gender’ vs. [gəˈnʊs] ‘enjoyment’; [ˈbluːtˌʔaʀm] ‘anaemic’ vs. [ˌbluːtˈʔaʀm] ‘very poor’). Word stress also distinguishes word classes; e.g. [ˈʔaktif] ‘active (noun)’ vs. [ʔakˈtiːf][4] ‘active (adjective)’. As a weight sensitive language, it is possible to formulate rules that correctly assign lexical stress in German (Wagner, 2003). Nevertheless, native speakers of German may not be aware of these rules, and stress assignment in non-words results in considerable variation across participants (Figure 1; see also Wagner, 2003).

Regarding the use of utterance-level prosody, we will refer to the recent approach of Grice et al. (2005), which assumes two hierarchical levels of phrasing below the utterance level: the intonation phrase (IP) and the subordinate intermediate phrase (ip) (for other descriptions, see Féry, 1993; Grabe, 1998; Grice et al., 2005). Each intermediate phrase contains at least one pitch accent (associated with a metrically strong syllable), followed by a phrase accent (determining the pitch at the right edge of the ip). At the IP level, there are initial and final boundary tones (%L, %H, L%, H%). Pitch accents can have six different forms, characterized by the alignment of tonal targets (H and L) with the stressed syllable: L*, H*, L*+H, L+H*, H+L*, H+!H*.[5] As in English, pitch accent types are related to differences in information status, information structure, sentence mood, and attitudes (Baumann and Grice, 2006; Braun, 2006; Féry and Kügler, 2008; Grice et al., 2005).

## 2 *French*

At the word-level, French is often characterized as having no lexical stress or tone (e.g. Di Cristo, 1988; Rossi, 1980; Vaissière, 1991). As a consequence, French listeners do not store suprasegmental information as part of lexical representations, the primary explanation of various studies on French listeners’ reported difficulties in storing stress (see Section I).

At the utterance level, the intonational phrase (IP) is structured into accentual phrases (APs). Depending on speech rate, an AP contains one or two content words plus an optional number of function words (Jun and Fougeron, 2000, 2002). Each non-final AP within an IP is usually marked with a rising final edge tone, e.g. LH* (see Delais-Roussarie, 2000; Jun and Fougeron, 2000; Roca, 1999; Welby, 2006). An IP-final AP is however marked with a pitch fall. Depending on structural and pragmatic factors, speakers may realize an AP-initial, early rise LHi (Jun and Fougeron, 2000; Turco et al., 2012). Phonetically, this default accent can be undershot, resulting in six surface realizations (Jun and Fougeron, 2000). Overall, the ability to mark information structural contrasts via prosodic means has been shown to be rather limited in French (e.g. Turco et al., 2012).

## 3 *Japanese*

At the word-level, more than 50% of the Japanese words have one mora associated with a falling H*L pitch accent (e.g. Gussenhoven, 2004; Haraguchi, 1999; Vance, 1987).

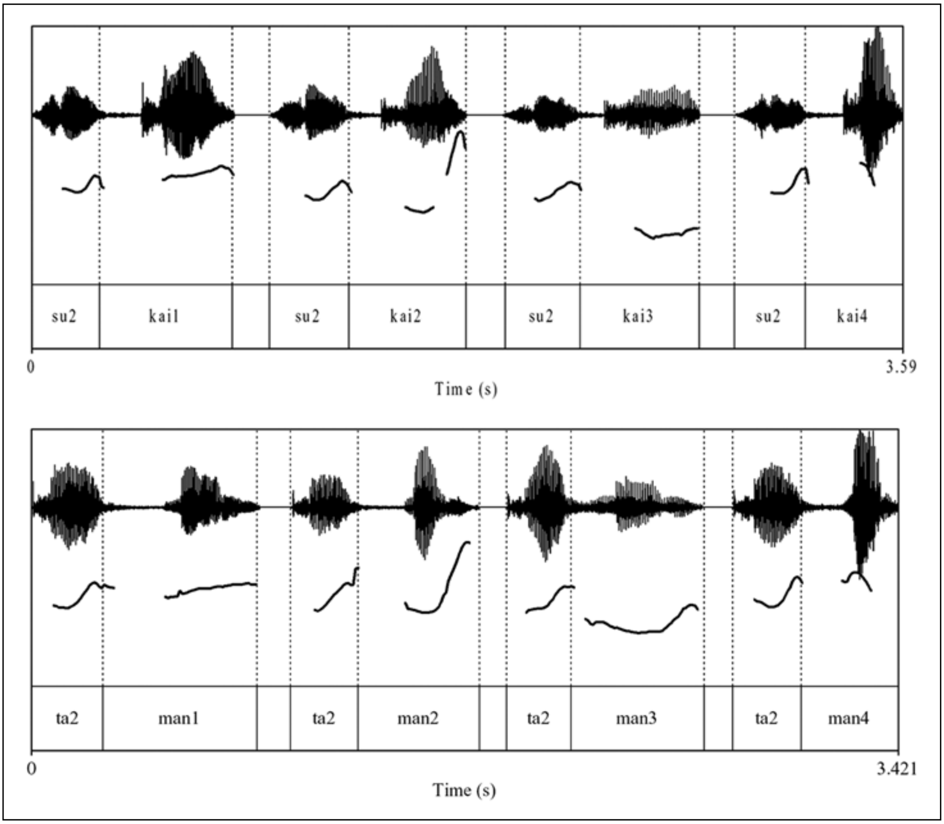


**Figure 1.** Pitch tracks of the four realizations of *sukai* (upper figure) and *taman* (lower figure).
*Notes*. The first syllable was always Tone 2, the second syllable (from left to right) Tone 1, Tone 2, Tone 3, Tone 4, respectively. Pitch range was 100–300 Hz.

Since not all morae are marked for tone, and not all words have a pitch accent specification, Japanese is regarded as a pitch-accent language (e.g. Gussenhoven, 2004; Yip, 2002), and sometimes as a very restricted tone language (Hyman, 2007).[6] In contrast to Asian tone languages like Mandarin Chinese (in which almost every syllable has a specification for tone), in Japanese there is at most one pitch accent in a word. Furthermore, the tonal contrast in Japanese is characterized by the presence vs. absence of a pitch accent on a mora, while in Mandarin Chinese a syllable can theoretically bear four different tones. Since most Japanese words (Tokyo Japanese) contain a lexical pitch accent (Gussenhoven, 2004), suprasegmental information needs to be stored for at least half of the items (e.g. [háʃi] 'chopsticks' vs. [haʃí] 'bridge').[7]

At the utterance level, the Japanese IP is made up of accentual phrases, similar to French. Each accentual phrase can contain no more than one pitch-accented word (Gussenhoven, 2004), so they are generally rather short. Other than that, the formation of Accentual Phrases is determined by style and speech rate, similar to French. The left

**Table 1.** The nature of word-level stress in Japanese, German, and French.

| | Japanese | German | French |
|---|---|---|---|
| Lexical stress | no | yes | no |
| Lexical tone | restricted tone | no | no |

**Table 2.** The number of pitch contrasts at the utterance level.

| | Japanese (Japanese Tone and Break Indices) | German (German Tone and Break Indices) | French (Jun and Fougeron, 2000) |
|---|---|---|---|
| Pitch accent types | 1 (H*L) | 6 (H*, L*, L*+H, L+H*, H+L*, H+!H*) | 2 (LH*, L*) |
| Initial phrase accents | 1 (H-) | 0 | 1 (LHi) |
| Final phrase accents | 1 (L%) | 2 (L-, H-) | 0 |
| Initial boundary tones | 1 (%L) | 2 (%H, %L) | 1 (%L) |
| Final boundary tones | 2 (L%, HL%) | 3 (L%, H%, ^H%) | 2 (L%, H%) |

edge of an IP-medial accentual phrase is marked with a high phrase accent in Japanese Tone and Break Indices (JToBI; H-; see Pierrehumbert and Beckman, 1988).[8] The right edge is marked by a falling boundary tone (L%). The IP starts with a low boundary tone (%L) and ends with either a high or falling boundary tone (H% or HL%). The main differences in pitch realization of an accentual phrase concern the timing of the rising and falling movement, which is dependent on the number of morae available and the presence or absence of a lexical pitch accent. More specifically, a word-final pitch fall is only realized when another word follows. Similar to French, Japanese speakers' ability to mark information structural contrasts is limited (Asano and Braun, 2011).

In sum, the German intonational phonology consists of more pitch accent types, more final phrase accents, and more initial and final boundary tones than Japanese and French. The nature of the accent at the word-level and the number of pitch contrasts at the utterance level in the languages studied are summarized in Tables 1 and 2, respectively.

## *4 Hypotheses*

The present study tested the learning of the association between visual objects and aurally presented disyllabic non-words (henceforth picture–word pairs). We investigated potential benefits in tone acquisition resulting from the three kinds of L1 prosodic systems and tested the following hypotheses:

- Hypothesis 1: Lexical storage of suprasegmentals
- Hypothesis 1a: Lexical storage of stress position / pitch-accent position: Listeners whose L1 uses stress or pitch accent contrastively at the word level (German, Japanese) – and therefore encode suprasegmental information as part of the lexical

representation of words – are better able to encode tone contrasts than listeners whose L1 does not use stress or pitch contrastively at the word level (French).

- Hypothesis 1b:Lexical storage of pitch information: Listeners who have to encode pitch information in their L1 lexicon (Japanese) are better at encoding tone contrasts than listeners who do not have to encode pitch information lexically in their L1 (German, French).
- Hypothesis 2: Complexity of the utterance-level pitch inventory: Listeners whose L1 has a rich $f_0$-inventory (German) are better able to encode lexical tones than those whose L1 has a simpler $f_0$-inventory (French, Japanese).

A better encoding of lexical tones is operationalized by a higher sensitivity to tonal contrasts (specifically: higher d' scores; see Macmillan and Creelman, 2005). We further recorded reaction times (RT) as a measure of task difficulty. If learners are sensitive to tonal information, we expect longer RTs due to competition between the newly learned word forms with different tones (Shatzman and McQueen, 2006). Lack of sensitivity to tones should yield shorter reaction times since there is no competition. Since we focus on the very early stages of vocabulary acquisition in a tone language, we make no predictions with regard to the end-state of tone acquisition, which may ultimately be influenced by learner internal and external variables (see Moyer, 1999; Obler, 1989; Orie, 2006).

## III Experiment

The present study provides a cross-linguistic comparison of the lexical encoding of Chinese tones in a word learning paradigm. All of our language groups have previously been shown to be sensitive to pitch differences in an auditory same–different task (Japanese: So and Best, 2010; German: Chiao et al., 2011; French: Hallé et al., 2004). Our study targets the next step in L2 acquisition, the lexical encoding of tone.

### 1 Methods

*a Participants.* We tested 8 German (4 male, 4 female, aged between 24 and 33, mean age 26.3), 8 French (3 male, 5 female, aged between 20 and 34, mean age 25.1) and 8 Japanese (3 male, 5 female, aged between 23 and 35, mean age 29.4) participants. Eight native speakers of Mandarin Chinese (4 male, 4 female, aged between 26 and 39, mean age 31.9) served as controls. All of the participants were recruited at the University of Konstanz, Germany, and they participated in the study voluntarily. None of the participants reported any hearing or seeing impairments. None of the German, French, and Japanese participants had any prior knowledge of Mandarin Chinese or any other tone language. The number of participants for each group was determined by means of a power test (Faul et al., 2009).[9]

*b Materials.* As the target language, we used Mandarin Chinese, which has a 4-way tonal contrast: Tone 1 with a high-level pitch, Tone 2 a high-rising pitch, Tone 3 a low-dipping pitch and Tone 4 a high-falling pitch (e.g. Chao, 1948, 1968; Ho, 1976; Howie, 1976; Lin, 1988; Wang, 2006).[10]

Eight disyllabic Mandarin Chinese non-words belonging to two segmental sets ([tʰaːman] and [suːkʰai]) were chosen to serve as auditory stimuli. The segments and syllable structures occur frequently in all the language groups tested. The tone on the first syllable was always Tone 2 while the tone on the second syllable was produced with all 4 possible tones (for the pitch tracks of the 4 chosen *sukai* and *taman* realizations, see Figure 1). Tone 2 was used as default tone for the first syllable of all stimuli in order to avoid tone sandhi effects, i.e. the alternation of a particular tone in certain tonal contexts (see Chen, 2000; Chen and Yuan, 2007; Shih, 1986; Zhang, 2007). Furthermore, the fact that [tʰa] is not attested with Tone 2 and that [suː] with Tone 2 only occurs in a few infrequent words should prevent lexical interference in the Chinese group. We used disyllabic stimuli since this structure is frequent in the Chinese lexicon (see He and Li, 1987). Furthermore, storing and recalling disyllabic words is more challenging than monosyllabic words, which may prevent ceiling effects in certain non-native language groups. We tested a total of 8 contrasts (4 tonal and 2 segmental ones), the segmental contrasts serving as a control condition.

All 8 non-words were produced by a male native speaker of Mandarin Chinese from Hebei (province surrounding Beijing). They were recorded digitally (44,100 kHz, 16Bit) in a sound attenuated booth at the University of Konstanz. The speaker was asked to produce several tokens of each non-word. After recording, stimuli were selected on the basis of acoustic similarity of the tonal realizations of the first syllable in the two segmental sets. The average $f_0$ excursion in the first syllable of the *taman* items was 4.4 semitones (SD 0.46) with an average duration of 268.8 ms (SD = 11.1 ms). The average $f_0$ excursion in the first syllable of the *sukai* items was 3 semitones (SD 0.32) with an average duration of 300 ms (SD 17.4 ms); see Appendix 1: Table 7. A non-parametric Fligner–Killeen test of homogeneity of variances (Conover, et al., 1981) showed no differences across the two sets of items in the variance of the $f_0$ range ($\chi^2 = 1.76$, $df = 1$, $p > 0.15$) and duration ($\chi^2 = 1.46$, $df = 1$, $p > 0.2$). Each auditory stimulus was randomly paired with a picture (1600 × 900 pixels) depicting a colourful object (Table 3).

*c Procedure.* Each participant was tested individually in a quiet room. They were first asked to fill out a questionnaire about their language background and language use. The experiment lasted on average 40 minutes and consisted of three phases (learning phase, practice task, and main task) administered in a consecutive order.

In the learning phase, the participants were instructed to learn the picture–word pairs by means of simultaneous auditory and visual presentation. They were told that the words were from a foreign language, but no further details were given (i.e. they were not instructed to focus on tone). The pictures were presented on the laptop screen using Psyscope X B51 (Cohen et al., 1993), and the corresponding auditory stimuli were played in stereo through Sennheiser HD 215 headphones. Each trial started with a '+' sign, which was shown in the centre of the screen for 500 ms. Then each picture was presented for 2 seconds, together with the auditory presentation (0 ms stimulus onset asynchrony). The inter-stimulus interval (ISI) was 500 ms. Each auditory word–picture pairing was presented 5 times in a randomized order, yielding 40 learning trials altogether.

**Table 3.** List of auditory and visual stimulus pairs.

| | [tʰaːmɑn] | [suːkʰaɪ] |
|---|---|---|
| Tone 2 – Tone 1 |  |  |
| Tone 2 – Tone 2 |  |  |
| Tone 2 – Tone 3 |  |  |
| Tone 2 – Tone 4 |  |  |

Following the learning phase, a practice task was administered to ensure successful learning of at least the segmental contrasts. The participants saw combinations of pictures and sounds, which matched or did not match the association they had learned before. Their task was to indicate, by using their dominant hand, as quickly and correctly as possible whether the auditory word matched the picture or not. The timing of trials was identical to the learning phase, and participants had a maximum of 2 seconds to respond. The practice task contained 32 picture–word pairs. Each picture was presented 4 times: twice with the matching auditory word (Complete Match Condition, e.g. ta2man1–ta2man1) and twice with a segmentally mismatching auditory word but with identical tones (segmental Mismatch Condition, e.g. ta2man1–su2kai1). No tonal minimal pairs were presented at this stage. At the end of the practice task, the participants received feedback about the number of errors but not about their nature. When they scored less than 75% correct in the practice task, they were instructed to repeat the learning phase and the practice task until they achieved at least 75% correctness in the practice task (but not more than 3 times).

The set-up of the main task was identical to that of the practice task but also included tonal minimal pairs (e.g. ta2man1 vs. ta2man2). Each picture was presented 6 times with the correctly assigned auditory stimulus (Complete Match), 4 times with each of the remaining 3 mismatching auditory stimuli that were segmentally homophonous to the item displayed in the picture but differed from it in tone (Tonal Mismatch), and three times with another mismatching auditory stimulus only differing in segments but with identical tones (Segmental Mismatch). No stimuli differed in both segments and tone from the learned associations. Due to the large number of stimuli in the Tonal Mismatch conditions (24 mismatches), the pairs in this condition were separated into two different, counterbalanced lists (see Table 4), and the participants were assigned randomly to one of these lists. The total of 120 picture–word pairs (48 in the Complete Match condition + 48 in the Tonal Mismatch condition + 24 in the Segmental Mismatch condition; see Table 5) was randomized anew for each participant. The reaction times (RT) relative

**Table 4.** Counterbalanced lists for the Tonal Mismatch condition with the specification of the tone (T) of the second syllable.

| List A (picture word) | List B (picture word) |
|---|---|
| T1–T2 | T2–T1 |
| T1–T3 | T3–T1 |
| T1–T4 | T3–T2 |
| T2–T3 | T4–T1 |
| T2–T4 | T4–T2 |
| T3–T4 | T4–T3 |

**Table 5.** Stimulus conditions.

| | Complete match (tone ID, segment ID) | Tonal Mismatch (tone ¬ID, segment ID) | | Segmental Mismatch (tone ID, segment ¬ID) |
|---|---|---|---|---|
| | | List A | List B | |
| Pairs | 8 | 12 | 12 | 8 |
| Repetitions | 6 | 4 | 4 | 3 |
| Total | 48 | 48 | 48 | 24 |

*Note*. Half of the pairs contain *sukai*, the other half *taman*.

to the offset of the auditory stimuli and the accuracy of every answer were recorded. Responses that were made later than 2 seconds after the offset of the auditory stimuli were excluded due to timeout.

## 2 Results

*a Practice task.* We first report the number of learning phases necessary to achieve the criterion level (75% correct) before we analyse the accuracy in the Tonal Match and Segmental Mismatch conditions. While all the German and French participants needed only two trials to achieve the criterion level in the practice task, some of the Japanese and Mandarin Chinese participants needed three trials. Particularly striking was the pattern of the Mandarin Chinese controls, none of whom reached the criterion in the first run (while more than a third of the participants in the other language groups did); see Table 6. Furthermore, more than 60% of the Chinese participants needed the third run.

The accuracy data were analysed using binomial logistic regression models[11] with Language Group and Condition as fixed factors (Agresti, 2002; Baayen et al., 2008; Jaeger, 2008) and Items and Participants as crossed random factors (adjusting for both the intercept and the slope of the within-group factor; see Barr et al., 2013; Cunnings, 2012). Results showed a significant interaction between Language Group and Condition (see Figure 2). Further analyses showed that in the Complete Match condition there was no effect of Language Group ($p > 0.35$ for all pairwise comparisons), indicating that

**Table 6.** Number of participants of each language group that needed one, two, or three learning sessions before reaching the criterion level in the practice task.

| Number of participants in each language group | Number of sessions needed to proceed to the main task | | |
|---|---|---|---|
| | One learning session | Two learning sessions | Three learning sessions |
| Chinese | 0 | 3 | 5 |
| French | 5 | 3 | 0 |
| Japanese | 3 | 3 | 2 |
| German | 3 | 5 | 0 |

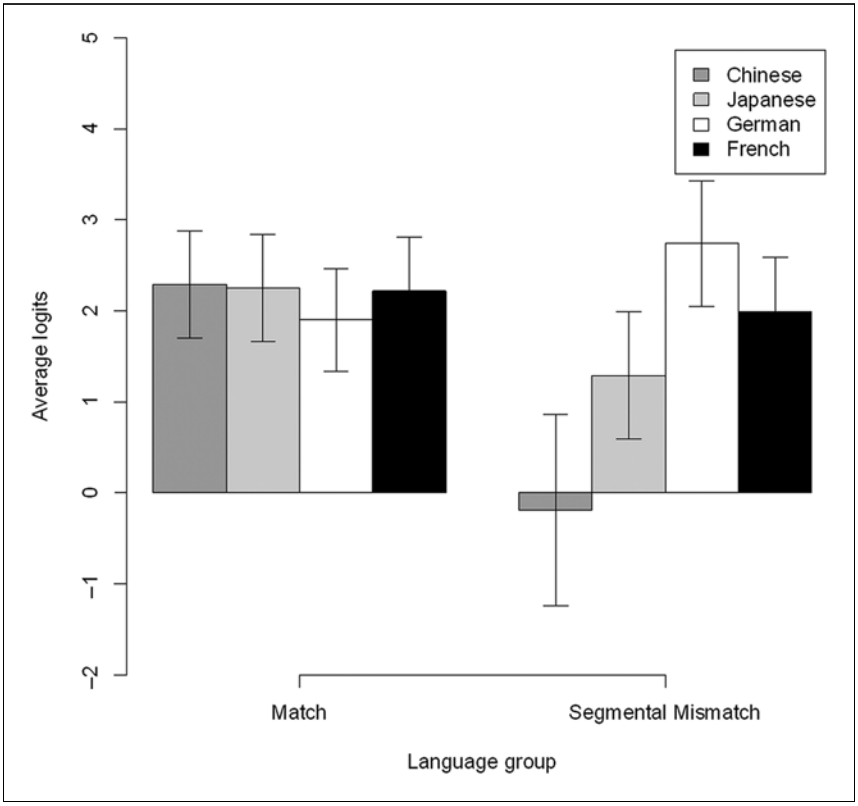


**Figure 2.** Average logits of the three language groups in each participant's last practice test, split by condition.

*Notes*. A logit of 0 corresponds to 50% correct, a logit of 1 to 73% correct and a logit of 2 to 88% correct. In this and the subsequent figures, whiskers indicate standard errors, as calculated by the statistical model.

all language groups learned the picture–word associations. In the Segmental Mismatch condition, the Mandarin Chinese participants performed worse than the French ($\beta = 2.37$, *SE* = 1.0, $p < 0.05$) and the German participants ($\beta = 3.1$, *SE* = 1.1, $p < 0.005$), but they did not differ from the Japanese participants ($p > 0.15$). The latter group only differed from the German participants ($\beta = 1.6$, *SE* = 0.7, $p < 0.05$), but not from the French participants ($p > 0.2$). Average accuracy scores across language groups and conditions are listed in Appendix 1: Table 8.

The low accuracy of the Mandarin Chinese group in the segmental Mismatch condition is puzzling. But note that the standard error of the Chinese group is approximately 1.5 larger than the standard error of the other groups. A closer inspection of the Mandarin Chinese participants' background questionnaires revealed two sub-groups regarding their proficiency in a non-tonal L2, English or German. More specifically, there were 3 Mandarin Chinese participants who indicated frequent usage of English or German in their daily activities (henceforth, 'L2-dominant'), and 5 who predominantly used Mandarin Chinese (henceforth, 'L1-dominant'). Notably, the participants in the L2-dominant group only needed two learning phases to proceed to the main task, while the L1-dominant Mandarin Chinese participants all needed three. We recalculated the logistic regression model with five participant groups, separating the Mandarin Chinese into two subgroups (L1-dominant and L2-dominant). As shown in Figure 3, in the Complete Match condition the L1- and L2-dominant participants did not differ from one another and not from the other language groups ($p > 0.4$ for all pairwise comparisons). On the other hand, in the Segmental Mismatch condition only the L1-dominant Mandarin Chinese participants performed significantly worse than the other groups (in comparison to the L2-dominant Mandarin Chinese: $\beta = 5$, *SE* = 0.84, $p < 0.0001$, in comparison to Japanese: $\beta = 3.3$, *SE* = 0.81, $p < 0.0001$, in comparison to German: $\beta = 4.9$, *SE* = 0.73, $p < 0.0001$, in comparison to French: $\beta = 4.2$, *SE* = 0.64, $p < 0.0001$). Furthermore, the L2-dominant Mandarin Chinese group even outperformed the Japanese participants ($\beta = 1.7$, *SE* = 0.83, $p < 0.05$). The comparisons between the other language groups are the same as detailed above.

*b Main task.* In total, 179 of the 3,840 data points (32 participants × 120 trials) had to be discarded due to timeout (4.7%).[12] The analysis of the last round of the practice task ensured that all participants performed equally well in the Match condition. The aim of the main task was to compare participants' sensitivity in the Tonal Mismatch condition to their sensitivity in the Segmental Mismatch condition. We analysed participants' sensitivity to the two kinds of cross-modal contrasts (Segmental Mismatch and Tonal Mismatch) using the premises of Signal Detection Theory (Green and Swets, 1966; Macmillan and Creelman, 2005), which has been widely used in previous studies (e.g. Altmann et al., 2011; Dupoux et al., 2008). To that end, we calculated each participant's hit rate by dividing the number of correct responses in the Tonal Mismatch condition by the total number of trials in that condition. The same was done for the Segmental Mismatch condition. The false alarm rate of each participant was calculated by dividing the number of incorrect responses in the Complete Match condition by the total number of Match trials. We used the R-function *AnotA* (Brockhoff and Christensen, 2010) to calculate each participant's d' score from his or her hit and false alarm rates.[13] These d' scores were then subjected to a general linear regression model, with language as fixed factor. To be on the safe side, we added the number of attempts in the learning phase as a factor,

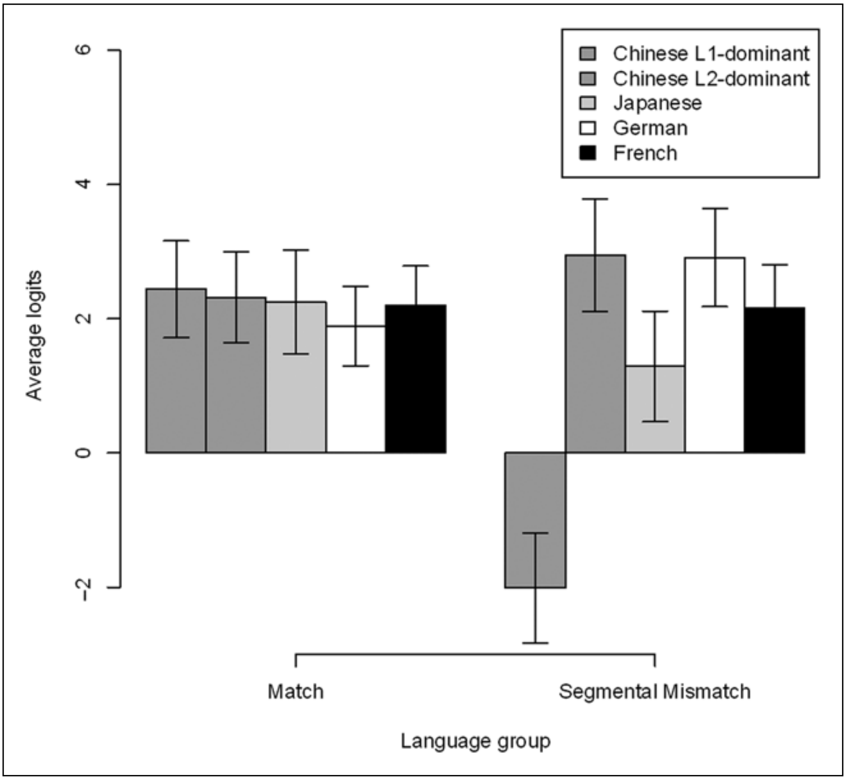


**Figure 3.** Average logits in the three language groups in each participant's last practice test block, split by condition.
*Note*. The Mandarin Chinese control group is separated into L1- and L2-dominant.

as participants with more attempts had more exposure to the tonal contrasts than those with fewer attempts. However, there was neither a main effect of number of attempts in the learning phase, nor an interaction with language (all *p*-values > 0.2), which is why this factor is not discussed anymore. To ensure the validity of the model, data points with residuals beyond 2.5 standard deviations of the mean were removed, and the model was refitted (Baayen, 2008). As shown in Figure 4, results showed an interaction between Condition and Language Group.

In the Segmental Mismatch condition, the Mandarin Chinese participants had d' scores that were lower than those of all other groups (Chinese vs. Japanese: $\beta = 2.34$, $SE = 0.26$, $p < 0.0001$, Chinese vs. German: $\beta = 1.30$, $SE = 0.27$, $p < 0.0001$, Chinese vs. French: $\beta = 1.79$, $SE = 0.26$, $p < 0.0001$). Furthermore, the German group had a lower sensitivity than the Japanese group ($\beta = 1.04$, $SE = 0.26$, $p < 0.0005$). The difference between the German and French participants approached significance ($\beta = 0.49$, $SE = 0.26$, $p = 0.07$). The French participants in turn had significantly lower d' scores than the Japanese participants ($\beta = 0.54$, $SE = 0.25$, $p < 0.05$).

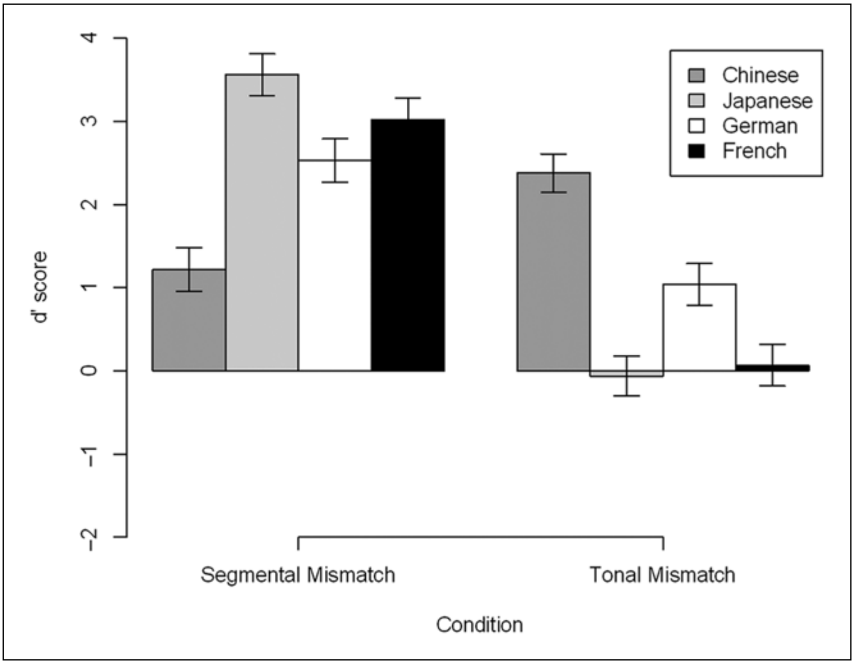


**Figure 4.** Average d' scores in the Segmental and Tonal Mismatch conditions.

In the Tonal Mismatch condition, the Mandarin Chinese group showed the highest d' scores (Chinese vs. Japanese: $ß = 2.44$, $SE = 0.29$, $p < 0.0001$, Chinese vs. German: $ß = 1.35$, $SE$ 0.32, $p < 0.0005$, Chinese vs. French: $ß = 2.32$, $SE = 0.29$, $p < 0.0001$). Furthermore, the German participants' d' scores were significantly higher than those of Japanese and French participants (German vs. Japanese: $ß = 1.09$, $SE = 0.32$, $p < 0.005$, German vs. French: $ß = 0.97$, $SE = 0.33$, $p < 0.01$). The latter two groups did not differ from each other ($p > 0.65$).[14] A one-sample, one-tailed $t$-test showed that French and Japanese d' scores did not differ from zero (French: $t(7) = 0.5$, $p > 0.25$, Japanese: $t(7) = –1.1$, $p > 0.8$), while the German participants 'd' scores were significantly larger than zero ($t(7) = 2.2$, $p < 0.05$). Average accuracy scores are provided in Appendix 1: Table 10.

In addition, we analysed RTs to evaluate whether the low sensitivity to tonal contrasts in Japanese and French speakers is caused more by an inability to store the contrast (long reaction times) or by strong lexical competition with a false decision (fast reaction times). As introduced before, we assumed that the successful detection of form–meaning mismatches requires the prior establishment of novel representations that include tonal information. Since more detailed representations would increase the number and strength of competitors in the task, we expected longer overall RTs for language groups with a higher sensitivity towards the tonal contrast. Conversely, if the lexical representation of the newly acquired lexical items is impoverished or incomplete with respect to tonal specification (as is expected for the French and the Japanese participants), the effect and

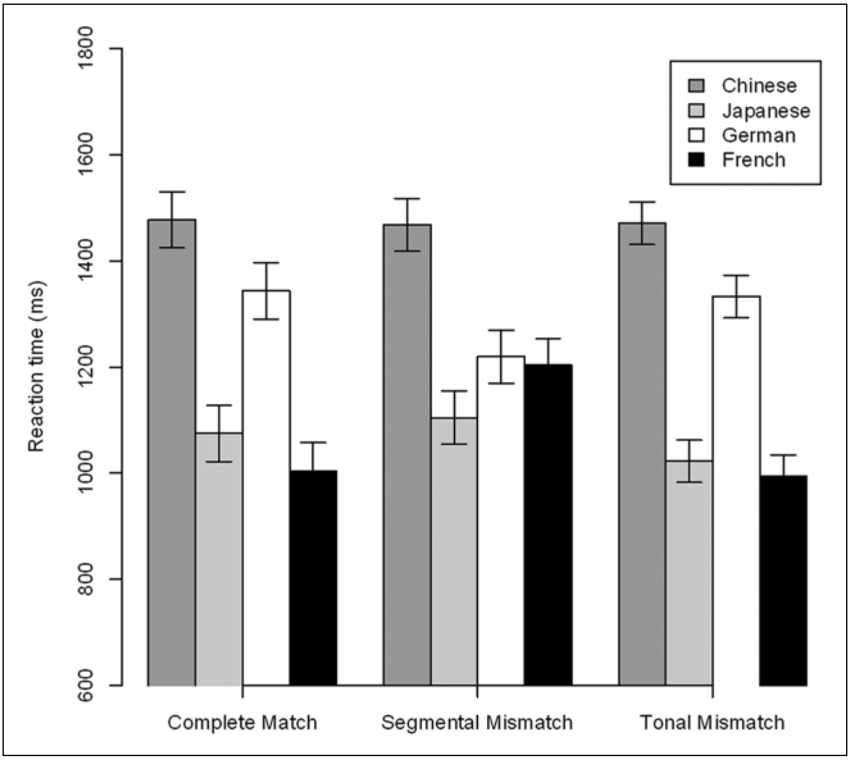


**Figure 5.** Reaction times in the main task, split by condition and language group.

number of competitors would go down (due to possible homophony), reducing the overall RTs across conditions. Accordingly, the following order of gross average RTs across language groups would corroborate the d' scores reported above: Mandarin Chinese > German > French = Japanese.

RTs for correct and incorrect trials were analysed using linear mixed effects regression models with Language Group, Condition, and Repetition as fixed factors. The initial models always included all factors and interactions and a maximal random effects structure (Barr et al., 2013; Cunnings, 2012). Model selection was performed using backwards elimination: Factors that did not have a significant main effect ($p$ >; 0.1) were removed if they did not occur in significant interactions and if this did not deteriorate the fit of the model (as estimated by model comparisons based on the Akaike Information Criterion; see Akaike, 1974). Subsequently, the model was refitted. Finally, we excluded data points with residual errors beyond 2.5 standard deviations from the mean and refitted the model to test its validity. For the linear mixed effects regression models, $p$-values cannot be estimated; so we report $t$-values instead. A $t$-value larger than 2 (or smaller than −2) indicates a significant difference at $\alpha < 0.05$.

The results showed a main effect of Repetition and Language Group as well as an interaction between Language Group and Condition (see Figure 5). There was an overall learning effect: participants became faster with each encounter of the stimulus ($ß$ = −12.9, $SE$ = 3.1, $t$ = −4.1). The effect of language group was as follows: The Mandarin Chinese participants had the longest RTs overall, which did not differ significantly from the German participants ($t$ = 1.70). The RTs of the fastest language groups, Japanese and French, did not differ from each other ($t$ = 0.14); they were significantly faster than the German group (Japanese vs. German: $ß$ = 237.3, $SE$ = 90.0, $t$ = 2.6, French vs. German: $ß$ = 247.9, $SE$ = 111.8, $t$ = 2.2) and the Mandarin Chinese group (Japanese vs. Chinese: $ß$ = 411.1, $SE$ = 61.8, $t$ = 6.6, French vs. Chinese: $ß$ = 421.7, $SE$ = 90.7, $t$ = 24.6). This pattern of overall results leaves two groups of languages, Chinese and German on the one hand, and French and Japanese on the other.

The analysis of RTs in the three conditions separately revealed two patterns: In the Segmental Mismatch condition, RTs of Japanese, German and French participants did not differ from one another (Japanese vs. German: $ß$ = 98.4, $SE$ = 79.5, $t$ = 1.2, French vs. German: $ß$ = 39.4, $SE$ = 107.8, $t$ = 0.37, Japanese vs. French: $ß$ = 137.8, $SE$ = 89.8, $t$ = 1.5). All of these three groups had faster RTs than the Mandarin Chinese control group (Japanese vs. Chinese: $ß$ = 368.2, $SE$ = 8.3.8, $t$ = 4.4, German vs. Chinese: $ß$ = 269.8, $SE$ = 97.3, $t$ = 2.8, French vs. Chinese: $ß$ = 230.4, $SE$ = 130.8, $t$ = 2.0). This corroborates the finding that the Segmental Mismatch condition was difficult only for the Mandarin Chinese participants. In the Complete Match and the Tonal Mismatch condition, RTs were equally fast for the Japanese and the French participants ($t$ = 0.6). Their RTs were significantly faster than those of the German group (Japanese vs. German: $ß$ = 270.1, $SE$ = 102.4, $t$ = 2.6, French vs. German: $ß$ = 316.3, $SE$ = 119.1, $t$ = 2.7) and of the Mandarin Chinese group (Japanese vs. Chinese: $ß$ = 422.0, $SE$ = 63.8, $t$ = 6.6, French vs. Chinese: $ß$ = 468.1, $SE$ = 88.3, $t$ = 5.3). RTs for German participants did not differ from those of Mandarin Chinese participants ($t$ = 1.4). These results are largely in line with the predicted order of reaction times (Mandarin Chinese > German > French = Japanese). The fact that the Mandarin Chinese group had longer RTs in the Tonal Mismatch condition than the German group suggests that the competition was stronger in the native language control group compared to the German group.

## IV Discussion

The present study investigated how the lexical encoding of non-native (L2) tonal contrasts is modulated by the way a learner's L1 employs pitch information. The participants had to learn the form–meaning association between 8 auditorily presented disyllabic non-words (4 tonal × 2 segmental contrasts) and visually presented objects. Because of the 8-way semantic contrast employed, the task may seem fairly challenging. However, the results from both the practice and the main task show that all participant groups had an average accuracy of at least 88% correct in the Complete Match condition. High accuracy scores were also achieved in the Segmental Mismatch condition, albeit with a strikingly low performance by the Mandarin Chinese controls. Further analyses revealed that it was mainly the Chinese-dominant participants who faced difficulties in this condition. The Mandarin Chinese participants who had indicated frequent use of English or German, on the other hand, did not differ from the other language groups in the Segmental Mismatch condition. Due to the small sample size, however, no strong conclusions about language dominance in bilingual participants can be drawn. Altogether, we take the relatively high scores in the Complete Match as well as the Segmental Mismatch condition to suggest that the form–meaning pairings were learned during the learning phase by all participant groups.

In the Tonal Mismatch condition, the focus of our article, we found interesting group differences. In particular, significantly lower d' scores were obtained from the Japanese, German, and French participants in comparison to the Mandarin Chinese controls. This overall L2 effect is not surprising given the test groups' lack of experience with storing 4-way tonal contrasts. Interestingly, we observed that the differences among the non-tonal language groups were modulated by the prosodic structure of the participants' L1. More specifically, the French and Japanese participants had remarkably low average d' scores, which did not differ from zero. This indicates an almost complete lack of sensitivity towards tonal minimal pairs in the auditory stimuli. Instead, they consistently reported a match when the tones in the auditory stimuli were different from the one they had been presented with in the learning phase. The near lack of sensitivity to tonal mismatches in these two groups suggests that they were not able to encode tonal contrasts in the long-term memory representations of novel words. This claim is corroborated by the faster RTs of the Japanese and French participants than the German and Mandarin Chinese participants. Apart from having longer RTs than the Japanese and French participants (in the Match and Tonal Mismatch condition), the German participants additionally had significantly higher d' scores. These two measures indicate that the German participants showed a high degree of sensitivity to the lexically contrastive use of pitch.

Our findings are not readily compatible with Hypothesis 1. In particular, Hypothesis 1a, which predicted that the presence of suprasegmental lexical contrasts in the L1 (word stress in German and lexical accent in Japanese) is beneficial for lexicalizing tone, must be rejected because of the Japanese participants' surprisingly low sensitivity to tonal contrasts (low d' scores and very fast RTs). In a similar vein, Hypothesis 1b, which predicted an overall advantage for the Japanese participants due to the lexical use of $f_0$ in their L1, must also be rejected. Instead, the outcome of the study supports Hypothesis 2, which predicted that a richer system of pitch contrasts at the utterance level (in terms of the number of boundary tones and pitch accents) is beneficial for building long-term representations of lexical tone. As discussed in Section II, German has a larger number of pitch accents and boundary tones than French and Japanese, which are used for signalling pragmatic, post-lexical information. French is characterized by a typical rising intonation pattern within each non-IP-final accentual phrase (for a detailed discussion, see Post, 2000). Apart from this and the utterance-final declarative (falling) and interrogative (rising) pitch patterns, the language arguably has very few pitch variations (e.g. Turco et al., 2012). The same is true for Japanese, which has been shown to rely little on $f_0$-variation to signal pragmatic, postlexical contrasts (see Abe, 1998; Asano and Braun, 2011). Therefore, we argue that the number of L1 utterance-level pitch contrasts, rather than the availability of word-level suprasegmental contrasts in the lexicon, offers a more plausible explanation for the cross-language influence we have found in our study.

It should be noted that our results cannot be explained by a psycho-acoustic account either. Such an account would attribute the cross-linguistic differences to a general inability to perceive tonal contrasts (as a consequence of 'tone deafness' on the part of the

French and Japanese participants, analogous to 'stress deafness'). While we do not have a direct measure of this in our study, independent evidence in the literature rules out this explanation. In particular, So and Best (2010), in a 4-alternative forced-choice tone identification task after a brief exposure to Mandarin Chinese tones, found that Japanese listeners performed equally well as Cantonese listeners whose L1 employs lexical tones. Furthermore, French listeners had also previously been reported to discriminate Mandarin Taiwan tones accurately in tasks that require phonetic, but not phonological, processing (Hallé et al., 2004), which is on a par with Dupoux et al.'s (1997, 2001, 2008) observation that L1 effects in the processing of word stress emerge when long-term phonological, rather than acoustic, representation of stress is called for. Furthermore, both the French and the Japanese participants in our study performed similarly to the Mandarin Chinese controls in the Complete Match condition, which we take to suggest that the inability to learn tonal contrasts should not be directly linked to a perceptual failure.

The low sensitivity to tonal contrasts for the Japanese participants is particularly striking as pitch accent information has been shown to play a significant role in word recognition (Cutler and Otake, 1999). On the other hand, the Japanese varieties differ in the realization of the lexical pitch accent and some varieties have no pitch accents at all. We assume that the low density in the Japanese lexicon may in part be responsible for the low sensitivity towards newly acquired rich tonal contrasts. Possibly, the reason why both the Japanese and the French participants failed to lexicalize the tonal contrasts in our study is a combination of the low functional load of pitch contrasts at the word-level (or their complete absence in French) as well as the poverty of the sentence-level pitch events in their L1.

Before we turn to the conclusions, we will briefly discuss the low accuracy rates of the Mandarin Chinese (and in particular the L1-dominant Mandarin Chinese) participants in the Segmental Mismatch conditions in both the practice and the main task. These may seem unexpected at first glance, given the information carried by segments in all the languages of the world. One possible explanation is the use of real objects for the non-word labels in the current study, which might have led to lexical interference in the Mandarin Chinese group. However, these objects were well known to the other language groups as well (with different lexical representations). Therefore, it seems unlikely that the use of real objects would hinder the Mandarin Chinese group in particular. Instead, we turn to cue-weighting as a potential explanation. Previous research on the role of segmental and tonal information in lexical activation in Mandarin Chinese (e.g. Lee, 2007; Ye and Connine, 1999) as well as available experimental evidence on Mandarin Chinese speakers' attention to these cues in other perceptual experiments (e.g. Lee and Nusbaum, 1993; Miller, 1978; Repp and Lin, 1990) suggest that, although both are relevant, pitch may outweigh segmental information in Mandarin Chinese. In a recent study, similar results have been found in an ABX task in which Mandarin Chinese participants were much more likely than Dutch participants to match a target stimulus X according to tone, rather than segments (Braun and Johnson, 2011). On the basis of these results, it can be suggested that the Mandarin Chinese group experienced difficulties with allocating selective attention to segments and hence suppress tonal information when judging segmental mismatches in the task. Furthermore, our finding is on a par with Dupoux et al. (1997), who showed that Spanish listeners experienced problems with ignoring suprasegmental information in discriminating segmentally mismatching pairs with the same stress pattern. They described stress for the Spanish as ‘a nondetachable aspect of phonological information, whereas for the French, this information is not represented [at the same level]’ (Dupoux et al., 1997: 418). A similar explanation could be provided for the L1-dominant Chinese participants in our study, who could not detach tonal information from segmental information. Interestingly, the frequent use of a non-tonal L2 (English or German) reduced the strong influence of tone in this group.

## V Conclusions

This study investigated the possible redeployment of L1 suprasegmental information in the early, untutored L2 acquisition of lexical tones. The considerably better results of German participants in comparison to French and Japanese participants suggest that the necessity to store suprasegmental lexical contrasts in the L1 has a beneficial effect on the lexical encoding of Mandarin Chinese tones. However, the poor results of the Japanese participants casts doubts on the relationship between the lexical storage of meaningful pitch variations in the L1 and the ability to lexically store tone in an L2 since the pitch accent is stored lexically in Japanese. This suggests that predicting learner outcomes in the acquisition of a suprasegmental feature on the basis of its lexical status in the L1 is limited in scope. At best, it yields an ad hoc assumption that lexical encoding of stress in German must be different from that of lexical pitch accents in Japanese. This naturally raises the question as to what the actual status and functional load of pitch accents in Japanese are, and whether the language should be classified as a restricted type of tone language (see Hyman, 2007).[15]

Given the similar results of the French and Japanese participants, it appears that their rather limited pitch inventory at the utterance level has a negative effect on the ability to store L2 lexical tones. Accordingly, it is plausible to postulate that linguistic experience with different pitch patterns helps the learner in storing tonal patterns. However, since pitch variations are primarily used to signal or decode pragmatic, post-lexical information in German, an additional mechanism is necessary to map L2 lexical tone contrasts onto already existing L1 pitch patterns. Therefore, the learner’s overall experience with pitch variation across different levels of linguistic organization (i.e. lexical and post-lexical) rather than concrete strategies to map novel contrasts to already existing L1 categories is at stake here. Thus, we conclude that learners’ combined experience with pitch variations both at the lexical and the utterance level provides a head start for the L2 learner in the early stages of tone acquisition. Since we only investigated the initial stages of language acquisition, we cannot predict whether the lexical encoding of tone may improve with training, or whether that may also be subject to L1 influence. Furthermore, it is still an open question as to whether the perception–production dichotomy that was attested in Hayes-Harb and Masuda’s (2008) study also emerges in the acquisition of tonal contrasts. Future studies should investigate the long-term effects of cross-language influence on the perception and production of L2 tones.

### Declaration of conflicting interest

The author declares that there is no conflict of interest.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Notes

1. French listeners, who mark phonological phrase boundaries by differences in $f_0$, duration, and segmental energy, have been shown to deploy these cues in online word segmentation (e.g. Christophe et al., 2004; Kabak et al., 2010).
2. Having tone in the native language does not necessarily facilitate second language tone perception. For instance, So (2005, 2006) found that native speakers of Cantonese have fewer problems in distinguishing the Mandarin Chinese tonal contrasts than Japanese speakers, which she attributed to the more complex tonal system of Cantonese (So, 2006: 442; see also Chiao et al., 2011).
3. Alternatively, their inability to produce the consonantal length contrast may be rooted in articulatory planning and motor programmes (e.g. Benner et al., 2007; Levelt and Wheeldon, 1994; for a recent study on the L2 learners' difficulties in producing Italian geminates, see Kabak et al., 2011).
4. In German, tense vowels are shortened in unstressed syllables (except for /aː/ and /ɛː/, for which vowel length is contrastive).
5. The asterisk indicates which of the tones is associated with the stressed syllable. Note that the contrast between H+L* and H+!H* is disputed in the literature (Rathcke and Harrington, 2010).
6. Hyman's argument for classifying Japanese as tone language is that '[a] language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes' (Hyman, 2007: 167).
7. The accent indicates the position of the lexically falling accent.
8. Gussenhoven (2004) argues that the left edge of an accentual phrase (α) should be marked by a rising phrase accent ($L_\alpha H_\alpha$).
9. The power test used an expected effect size of 0.7 for d' scores in an AX task (see Altmann et al., 2012) and a power of 0.85.
10. In certain contexts, a neutral tone, i.e. a syllable having no tonal contour at all – sometimes also referred to as Tone 5 – may occur (e.g. Chao, 1948, 1968; Norman, 1988). The neutral tone is often found on particles expressing grammatical functions such as the question marker 吗 *ma*, the perfect marker 了 *le*, or it may be the result of tonal reduction in certain cases such as in 钥匙 *yàoshi* ('key') or 谢谢 *xièxie* ('thank you'), where the second syllable carries an underlying Tone 4 but surfaces as neutral tone (Duanmu, 2007).
11. We chose logit mixed effects models for our statistical analyses since categorical outcome variables such as accuracy data lead to wrong interpretations when analysed with widely used statistical procedures such as ANOVA (see Jaeger, 2008). For a comprehensive treatment of the use of mixed effects models in second language research, see Cunnings (2012).
12. Timeouts were strongly influenced by repetition (decreasing overall from 65 instances in the first encounter of a stimulus to 38 in the third encounter) and by condition (40 in the Complete Match condition, 69 in the Segmental Mismatch condition and 70 in the Tonal Mismatch condition). The Japanese participants had only 13 timeouts, the other language groups more than 50 (Chinese: 57, German: 58, French: 51); for details, see Appendix 1: Table 9.
13. When the hit or false alarm rate was zero, one instance was added; if they reached ceiling, one instance was subtracted.
14. For the sake of completeness, we also calculated d' scores for each speaker for each tonal

mismatch (taking hit rates for e.g. Tone 1 – Tone 2 Mismatch and false alarm rates to the Tone 1 – Tone 1 and Tone 2 – Tone 2 Match condition). Results showed a main effect of language group (see main text) and of tonal contrast, but no interaction. For all language groups, it was more difficult to distinguish Tone 2 vs. Tone 3 and Tone 3 vs. Tone 4 mismatches compared to Tone 1 vs. Tone 2 or Tone 1 vs. Tone 4 mismatches ($p < 0.05$). No further significant differences were found.

15. The relatively simplex tonal system of Japanese does not pose any constraint in tone discrimination in comparison to other tone languages (see So, 2005, 2006; So and Best, 2010). Since those studies had no non-tone language as control (and our study had no other complex tonal language as control), this issue remains to be tested further.

## References

Abe I (1998) Intonation in Japanese. In Hirst D and Di Cristo A (eds) *Intonation systems: A survey of twenty languages*. Cambridge. Cambridge University Press, 360–75.

Agresti A (2002) *Categorical data analysis*. New York: John Wiley.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transaction on Automatic Control* 19: 716–23.

Altmann H (2006) The perception and production of second language stress: A cross-linguistic experimental study. Unpublished PhD dissertation, University of Delaware, Newark, DE.

Altmann H, Berger I, and Braun B (2012) (A)symmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research* 28: 387–413.

Altmann H and Kabak B (2011) Second language phonology. In: Kula N, Botma B and Nasukawa K (eds) *The continuum companion to phonology*. London/New York: Continuum, 298–319.

Archibald J (1997) Acquisition of English stress by speakers of non-accentual languages. Lexical storage versus computation of stress. *Linguistics* 35: 167–81.

Asano Y and Braun B (2011) Integrating lexical and postlexical suprasegmental information in native and non-native Japanese. *NINJAL International Conference on Phonetics and Phonology (ICPP 2011)*. Kyoto.

Baayen RH (2008) *Analyzing linguistic data: A practical introduction to linguistics*. Cambridge: Cambridge University Press.

Baayen HR, Davidson DJ, and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412.

Barr D J, Levy R, Scheepers C, and Tily H (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255–78.

Baumann S and Grice M (2006) The intonation of accessibility. *Journal of Pragmatics* 38: 1636–57.

Benner U, Flechsig I, Dogil G, and Möbius B (2007) Coarticulatory resistance in a mental syllabary. *Proceedings of the 16th International Congress of the Phonetic Sciences*. Saarbrücken, 485–88.

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

Braun B (2006) Phonetics and phonology of thematic contrast in German. *Language and Speech* 49: 451–93.

Braun B and Johnson EK (2011) Question or tone 2? How language experience and linguistic function guide attention to pitch. *Journal of Phonetics* 39: 585–94.

Brockhoff PB and Christensen RHB (2010) Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* 21: 330–38.

Chao YR (1948) *Mandarin primer*. Cambridge: Harvard University Press.

Chao YR (1968) *A grammar of spoken Chinese*. Los Angeles, CA: University of California Press.

Chen MY (2000) *Tone sandhi: Patterns across Chinese dialects*. Cambridge. Cambridge University Press.

Chen YY and Yuan JH (2007) A corpus study of the 3rd tone sandhi in Standard Chinese. In: van Hamme H and van Son R (eds) *Proceedings of 8th International Conference of the International Speech Communication Association (Interspeech 2007)*. Adelaide: Causal Productions, 2749–52.

Chiao W-H, Kabak B, and Braun B (2011) When more is less: Non-native perception of level tone contrasts. *Proceedings of the Psycholinguistic Representation of Tone Conference 2011*. Hong Kong, 42–45.

Christophe A, Peperkamp S, Pallier C, Block E, and Mehler J (2004) Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language* 51: 523–47.

Cohen JD, MacWhinney B, Flatt M, and Provost J (1993) Psyscope: A graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers* 25: 257–71.

Conover WJ, Johnson ME, and Johnson MM (1981) A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23: 351–61.

Cunnings I (2012) An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28: 369–82.

Cutler A and Otake T (1999) Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America* 105: 1877–88.

Cutler A, Weber A, and Otake T (2006) Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*34: 269–84.

Delais-Roussarie E (2000) Vers une nouvelle approche de la structure phonologique [Towards a new approach to phonological structure]. *Langue française* 126: 92–127.

Di Cristo A (1998) Intonation in French. In: Hirst D and Di Cristo A (eds) *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press, 195–218.

Domahs U, Genç S, Knaus J, Wiese R, and Kabak B (2012) Processing (un)predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes* 28: 335–54.

Duanmu S (2007) *Phonology of Standard Chinese*. 2nd edition. Oxford: Oxford University Press.

Dupoux E, Pallier C, Sebastián N, and Mehler J (1997) A destressing 'deafness' in French? *Journal of Memory and Language* 36: 406–21.

Dupoux E, Peperkamp S, and Sebastián-Gallés N (2001) A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America* 110: 1608–18.

Dupoux E, Sebastián-Gallés N, Navarrete E, and Peperkamp S (2008) Persistent stress 'deafness': The case of French learners of Spanish. *Cognition* 106: 682–706.

Escudero P, Hayes-Harb R, and Mitterer H (2008) Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60.

Faul F, Erdfelder E, Buchner A, and Lang A-G (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41: 1149–60.

Féry C (1993) *German intonational patterns*. Tübingen: Niemeyer.

Féry C and Kügler F (2008) Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36: 680–703.

Flege JE (1995) Second-language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 229–73.

Gandour JT (1983) Tone perception in Far Eastern languages. *Journal of Phonetics* 11: 149–75.

Grabe E (1998) Pitch accent realization in English and German. *Journal of Phonetics* 26: 129–43.
Green DM and Swets JA (1966) *Signal detection theory and psychophysics*. New York: Wiley.
Grice M, Baumann S, and Benzmüller R (2005) German intonation in autosegmental-metrical phonology. In: Sun-Ah J (ed.) *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, 55–83.
Gussenhoven C (2004) *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Hallé PA, Chang YC, and Best CT (2004) Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32: 395–421.
Haraguchi S (1999) Accent. In: Tsujimura N (ed.) *The handbook of Japanese linguistics*. Malden, MA: Blackwell Publishers, 1–30.
Hayes-Harb R and Masuda K (2008) Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research* 24: 5–33.
He K and Li D (1987) *现代汉语三千常用词表 Xiàndài Hànyǔ sānqián chángyòng cí biǎo [Three thousand most commonly used words in Standard Chinese]*. Beijing: 北京师范大学出版社 Běijīng Shīfàn Dàxué Chūbǎnshè.
Ho A (1976) The acoustic variation of Mandarin tones. *Phonetica* 33: 353–67.
Howie JM (1976) *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
Hyman LM (2007) How (not) to do phonological typology: The case of pitch-accent. *UC Berkeley Phonology Lab Annual Report 2007*. Berkeley, CA: UC Berkeley Phonology Lab, 654–85.
Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–46.
Jun S-A (ed.) (2005) *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.
Jun S-A and Fougeron C (2000) A phonological model of French intonation. In: Botinis A. (ed.) Intonation: *Analysis, modeling and technology*. Dordrecht: Kluwer Academic, 209–42.
Jun S-A and Fougeron C (2002) The realizations of the accentual phrase in French intonation. *Probus* 14: 142–72.
Kabak B, Maniwa K, and Kazanina N (2010) Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology* 1: 207–24.
Kabak B, Reckziegel T, and Braun B (2011) Timing of second language geminates and singletons. *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong, 994–97.
Kawagoe I (2003) Acquisition of English word stress by Japanese learners. In: Liceras JM, Zobl H and Goodluck H (eds) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA) 2002*. Somerville, MA: Cascadilla Proceedings Project, 161–67.
Lee C-Y (2007) Does horse activate mother? Processing lexical tone in form priming. *Language and Speech* 50: 101–23.
Lee L and Nusbaum HC (1993) Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics* 53: 157–65.
Lee Y, Vakoch D, and Wurm L (1996) Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research* 25: 527–44.
Lehiste I (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.
Leung AS (2008) Tonal assimilation patterns of Cantonese L2 speakers of Mandarin in the perception and production of Mandarin tones. In: *Proceedings of the Annual Conference of the*

*Canadian Linguistic Association 2008*. Retrieved from: http://homes.chass.utoronto.ca/~cla-acl/actes2008/actes2008.html (November 2013).

Levelt WJM and Wheeldon L (1994) Do speakers have access to a mental syllabary. *Cognition* 50: 239–69.

Lin MC (1988) *普通话声调的声学特性和知觉征兆* Pǔtōnghuà shēngdiào de shéngxué tèxìng hé zhījué zhēngzhào [Tone characteristics and percepts of Mandarin Chinese]. *中国语言 Zhōngguó Yǔyán* 3, 182–193.

Macmillan NA and Creelman CD (2005) *Detection theory: A user's guide*. 2nd edition. Mahwah, NJ: Lawrence Erlbaum.

Miller JL (1978) Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics* 24: 1175–80.

Moyer A (1999) Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21: 81–108.

Norman J (1988) *Chinese*. New York: Cambridge University Press.

Obler L (1989) Exceptional second language learners. In: Gass S, Madden S, Preston D and Selinker L (eds) *Variation in second language acquisition: Volume II: Psycholinguistic issues*. Clevedon: Multilingual Matters, 141–59.

Orie ỌỌ (2006) L2 acquisition and Yoruba tones: Issues and challenges. In: Arasanyin OF and Pemberton MA (eds) *Selected proceedings of the 36th Annual Conference on African Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 121–28.

Ou SC (2007) Linguistic factors in L2 word stress acquisition: A comparison of Chinese and Vietnamese EFL learners' development. In: Trouvain J and Barry WJ (eds) *Proceedings of the 16th International Congress of Phonetic Sciences 2007*. Dudweiler: Pirrot, 1681–84.

Pallier C, Colomé A, and Sebastián-Gallés N (2001) The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries. *Psychological Science* 12: 445–49.

Peperkamp S and Dupoux E (2002) A typological study of 'stress deafness'. In: Gussenhoven C and Warner N (eds) *Papers in Laboratory Phonology 7*. Berlin: de Gruyter, 203–40.

Perrachione TK, Lee J, Ha LYY, and Wong PCM (2011) Learning novel phonological contrasts depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America* 130: 461–72.

Pierrehumbert JB and Beckman M (1988) *Japanese tone structure*. Linguistic inquiry monograph 15. Cambridge, MA: MIT Press.

Pierrehumbert JB and Hirschberg J (1990) The meaning of intonational contours in the interpretation of discourse. In: Cohen PR, Morgan J and Pollack ME (eds) *Intentions in communication*. Cambridge: MIT Press, 271–311.

Post B (2000) Pitch accents, liaison and the phonological phrase in French. *Probus* 12: 127–64.

Qin Z and Mok P (2011a) Perception of Cantonese tones by Mandarin, English and French speakers. In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*. Hong Kong, 1654–57.

Qin Z and Mok P (2011b) Discrimination of Cantonese tones by Mandarin, English and French speakers. *Proceedings of the Psycholinguistic Representation of Tone Conference (PLRT)*, Hong Kong, pp. 50–53.

Rathcke T and Harrington J (2010) The variability of early accent peaks in Standard German. In: Fougeron C, Kühnert B, D'Imperio M and Vallée N (ed.) *Papers in laboratory phonology: Volume 10*. Berlin: De Gruyter Mouton, 533–55.

Repp BH and Lin H-B (1990) Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics* 18: 481–95.

Roca IM (1999) Stress in the Romance languages. In: Van der Hulst H (ed.) *Word prosodic systems in the languages of Europe*. New York: de Gruyter, 659–812.

Rossi M (1980) Le français, langue sans accent? [French: A non-stressed language?]. In: Fónagy I and Léon P (eds) *L'accent en français contemporain* [Stress in modern French]. Paris: Didier, 13–51.

Schwab S and Llisterri J (2011) Are French speakers able to learn to perceive lexical stress contrasts? In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*. Hong Kong, 1774–77.

Shatzman KB and McQueen JM (2006) Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science* 17: 372–77.

Shih CL (1986) The prosodic domain of tone sandhi in Chinese. Unpublished PhD dissertation, University of California, San Diego, CA, USA.

Skoruppa K, Pons F, Christophe A, Bosch L, Dupoux E, and Sebastián-Gallés N (2009) Language-specific stress perception of stress by nine-month-old French and Spanish infants. *Developmental Science* 12: 914–19.

So CK (2005) The influence of L1 prosodic background on the learning of Mandarin tones: Patterns of tonal confusion by Cantonese and Japanese naïve listeners. In: Gurski C (ed) *Proceedings of the 2005 Annual Conference of the Canadian Linguistic Association*. Retrieved from: http://westernlinguistics.ca/Publications/CLA-ACL/So.pdf (November 2013).

So CK (2006) Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. In: Warren P and Watson CI (eds) *Proceedings of the 11th Australian International Conference on Speech Science and Technology* 2006, 438–43.

So CK and Best C (2010) Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53: 273–93.

Strange W and Shafer VL (2008) Speech perception in second language learners: The re-education of selective perception. In: Hansen-Edwards JG and Zampini ML (eds) *Phonology and second language acquisition*. Amsterdam: John Benjamins, 153–92.

Turco G, Dimroth C, and Braun B (2012) Intonational means to mark verum focus in German and French. *Language and Speech*. Advance online publication. DOI: 10.1177/0023830912460506.

Vaissière J (1991) Rhythm, accentuation and final lengthening in French. In: Sundberg J, Nord L and Carlson R (eds) *Music, language, speech and brain*. New York: Macmillan, 108–20. Wenner–Gren International Symposium Series, vol. 59.

Vance TJ (1987) *An introduction to Japanese phonology*. Albany: State University of New York Press.

Vennemann T (1990) Syllable structure and simplex accent in modern standard German. In: Ziolkowski M, Noske M and Deaton K (eds) *Papers from the Regional Meeting of the Chicago Linguistic* Society 26: Volume II: The parasession. Chicago, IL: Chicago Linguistic Society, 399–412.

Wagner PS (2003) Improving the prediction of German lexical stress. In: Solé MJ, Recasens D and Romero J (eds) *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Futurgraphic, 2059–62.

Wang P (2006) *北京话单音字调和元音的统计性研究* Běijīnghuà dānyīn zìdiào hé yuányīn de tǒngjìxìng yánjiū [A statistical study of the tones and vowels of Beijing Mandarin monosyllables]. Unpublished PhD dissertation, *南开大学* Nankai University, China.

Wang Y, Jongman A, and Sereno JA (2006) L2 acquisition and processing of Mandarin Chinese tones. In: Li P, Bates E, Tan LH and Tseng O (eds) *The handbook of East Asian Psycholinguistics*. Cambridge: Cambridge University Press, 250–56.

Weber A and Cutler A (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25.

Welby P (2006) French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34: 343–71.

Wiese R (2006) *The phonology of German*. 2nd edition. Oxford: Oxford University Press.
Wong PCM and Perrachione TK (2007) Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28: 565–85.
Ye Y and Connine CM (1999) Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes* 14: 609–30.
Yip M (2002) *Tone*. Cambridge: Cambridge University Pres.
Zhang J (2007) A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics* 16: 259–302.

## Appendix 1

**Table 7.** Overview of the acoustic measurements ($f_0$-minimum, $f_0$-maximum, semitone difference of the rise and duration) in the first syllable of the 8 auditory stimuli.

| | $f_0$-min of 1st syllable (Hz) | $f_0$-max of 1st syllable (Hz) | $f_0$-range of 1st syllable (st) | Duration of first syllable (ms) |
|---|---|---|---|---|
| ta2man1 | 145 | 186 | 4.3 | 282 |
| ta2man2 | 141 | 185 | 4.7 | 267 |
| ta2man3 | 144 | 180 | 3.9 | 255 |
| ta2man4 | 146 | 194 | 4.9 | 271 |
| su2kai1 | 167 | 193 | 2.5 | 279 |
| su2kai2 | 154 | 183 | 3.0 | 318 |
| su2kai3 | 152 | 183 | 3.2 | 310 |
| su2kai4 | 165 | 198 | 3.2 | 293 |

**Table 8.** Cross-tabulation of accuracy scores of the last practice task (as percentage) according to Language Group and Condition.

| Condition | Chinese total (L1/L2-dominant) | Japanese | German | French |
|---|---|---|---|---|
| Match | 88.3 (87.5/89.5) | 85.2 | 84.4 | 87.5 |
| Segmental Mismatch | 46.1 (17.5/93.8) | 69.5 | 92.2 | 87.5 |

**Table 9.** Cross-tabulation of timeouts according to Language Group and Condition.

| Condition | Chinese total (L1/L2-dominant) | Japanese | German | French | Total |
|---|---|---|---|---|---|
| Match | 10 (5/5) | 7 | 10 | 13 | 40 |
| Segmental Mismatch | 38 (31/7) | 1 | 10 | 21 | 70 |
| Tonal Mismatch | 9 (1/8) | 5 | 38 | 17 | 69 |
| Total | 57 (37/20) | 13 | 58 | 51 | 179 |

**Table 10.** Cross-tabulation of accuracy scores of the main task (as percentage) according to Language Group and Condition.

| Condition | Chinese total (L1/ L2-dominant) | Japanese | German | French |
|---|---|---|---|---|
| Match | 85.6 (83.4/89.2) | 97.3 | 86.9 | 97.3 |
| Segmental Mismatch | 54.5 (32.8/84.6) | 95.8 | 91.8 | 89.5 |
| Tonal Mismatch | 88.8 (86.6/92.6) | 1.1 | 33.5 | 2.2 |