Question or tone 2? How language experience and linguistic function guide pitch processing

Bettina Braun[1] and Elizabeth K. Johnson[2]

[1] University of Konstanz, Germany
[2] University of Toronto, Canada

Short title: Linguistic function guides pitch processing

Address for correspondence:

Bettina Braun

Department of Linguistics

University of Konstanz

Universitätsstr. 10

Fach 186

78467 Konstanz

bettina.braun@uni-konstanz.de

Tel: 0049-7531-882386

Abstract

How does language experience shape pitch processing? Do speakers of tone languages, which use pitch to signal lexical contrasts (e.g., Mandarin Chinese) attend to pitch movements more closely than speakers of intonation languages (e.g., English)? Contradictory findings have been reported in the literature. In the current study, we hypothesize that listeners should be particularly attentive to any pitch information that signals meaningful information in the native language. This includes pitch movements signaling lexical contrasts (present in tone languages only) as well as postlexical contrasts (present in all languages). Both Mandarin and Dutch listeners performed speeded ABX match to sample tasks on the same sets of nonsense words. As predicted, the same pitch movements were attended to differentially by the two language populations depending on the role that information played in the native language. Mandarin speakers were more attentive than Dutch listeners to pitch movements as these signaled potential lexical contrasts in Mandarin (but not Dutch). Importantly, Dutch listeners were more attentive to pitch movements signaling postlexical information than to pitch movements signaling no meaningful information. These findings underscore the importance of postlexical information in online speech processing, and explain apparent contradictions in the literature.

Keywords: Pitch, lexical tone, intonation, postlexical, cross-linguistic, Dutch, Chinese

Introduction

The speech signal is rich in information, only some of which encodes linguistic meaning. The status of this information varies across languages, and it is the task of the listener to sort out those acoustic patterns that carry linguistic meaning from those that do not. Thus, efficient decoding of spoken utterances requires that listeners deploy their attention in a strategic fashion. Some listening tactics are universal, e.g., when identifying words listeners rely more heavily on consonants than vowels (e.g., Cutler, Sebastián-Gallés, Soler-Vilageliu, & Van Ooijen, 2000) and more heavily on segments than on pitch (Cutler & Chen, 1997; Ye & Connine, 1999). Other strategies are language-specific, e.g., Russian listeners attend closely to the contrast between /d/ and /t/ because it is linguistically meaningful (phonemic) in Russian. Korean listeners, on the other hand, do not attend to this same contrast as it is contextually determined (allophonic) in Korean (Kazanina, Philipps, & Idsardi, 2006). This type of attunement to language-specific listening strategies is achieved early in infancy, and appears to be the foundation upon which subsequent language development is built (e.g.,Werker & Tees, 1984). The speed with which infants learn to ignore non-native contrasts is highly correlated with later language achievement. Infants who are slow to learn to ignore non-native contrasts display poorer language skills in early childhood than those who are fast (Kuhl, et al., 2008).

Evidence from the segmental processing literature clearly suggests that linguistically meaningful contrasts are attended to more closely than contrasts that are not contrastive in a given language (e.g., Cutler, Weber, & Otake, 2006; Flege, Bohn, & Jan, 1997; McAllister, Flege, & Piske, 2002). It should be noted that for segmental contrasts, there is only a binary distinction regarding their semantic contribution: either a certain contrast is lexically distinctive in a language (so that minimal pairs can be formed that only differ in the contrast in question) or it is not.[1] As we shall see below, prior work investigating the processing of *pitch* information has implicitly assumed a similar binary distinction for the semantic contribution of pitch movements (either they are lexically distinctive, as in tone languages, or they are not, as in non-tonal languages). However, for

---

[1] The fact that there is only a binary distinction regarding the semantic contribution of segments does not mean that speakers are unable to perceive and encode non-native segmental contrasts (cf., Best, 1994, p. 191).

pitch movements there is a four-way distinction in terms of the *type of information* that is conveyed (lexical, postlexical, paralinguistic, non-linguistic[2]). First, pitch information may be lexically contrastive. In tone languages, such as Thai or Mandarin Chinese pitch serves a lexical function and is assumed to be part of a words' mental representation (e.g., Cutler & Chen, 1997). Consequently, the meaning of segments (e.g., ma) is dependent on the pitch contour associated with them. In Mandarin Chinese, for instance, there are four lexical tones and a neutral one: Tone 1 is a high level tone produced at the top range of a speakers' register, tone 2 is a rising tone where the rise occurs late in the syllable. Tone 3 is low falling to the bottom of a speakers' register and often accompanied by laryngalization; if produced in isolation it is considerably longer than the other tones and ends slightly rising. Tone 4, finally, is a high falling tone. Stylized versions of the tonal movements are shown in Figure 1. The neutral tone has been argued to also have a tonal specification but it is much weaker than that of the other tones, which leads to more variation in pitch realization. Furthermore, syllables with a neutral tone are much shorter than syllables with a full tone (Chen & Xu, 2006).

Insert Figure 1 about here

Second, in addition to carrying lexical information, pitch movements may also signal postlexical information, i.e., information that is linguistically meaningful but does not contribute to the lexical meaning of a word. More specifically, pitch movements may provide cues to syntactic constituency (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Snedeker & Trueswell, 2002; Speer, Kjelgaard, & Dobroth, 1996), convey the domain of focus (e.g., Birch & Clifton, 2002; Welby, 2003), mark contextually old and new information (e.g., Baumann, Grice, & Steindamm, 2006; Bock & Mazzella, 1983; Braun, 2006; Cutler & Foss, 1977; Dahan, Tanenhaus, & Chambers, 2002; Gussenhoven, 1984; Kohler, 1991; Terken & Nooteboom, 1987), and differentiate between sentence types such as statements and questions (e.g., Van Heuven & Haan, 2002). Developmental studies have suggested that listeners are closely tuned to

---

[2] Note that the term 'non-linguistic' does not refer to non-speech pitch (such as musical tones) but to pitch movements that do not signal an obvious linguistic function

postlexical information from the very early stages of development (e.g., Johnson & Seidl, 2008).

A third type of information signaled by pitch movements is paralinguistic information, such as attitudes and emotions (e.g., Liberman, 1975; Liberman & Sag, 1974; Scherer, Ladd, & Silverman, 1984), see also the 'Nine ways of saying yes' in Crystal (1995). This type of information can be highly distracting for those who have not yet learned to distinguish it from linguistically meaningful pitch movements (Singh, Morgan, & White, 2004). Finally, some pitch information may simply be non-attested in a language and therefore sound odd or have no apparent linguistic function. Listeners appear to map unfamiliar contours onto existing intonation contours (Braun, Kochanski, Grabe, & Rosner, 2006; Pierrehumbert & Steele, 1989) and are even able to interpret sentences with an unfamiliar intonation contour (Braun, Dainora, & Ernestus, 2011). Hence, truly non-linguistic pitch movements may be rare.

In this study we investigate how the processing and short term retention of pitch information is modulated by the exact *kind* of information conveyed by it. Do listeners attend to information more closely and encode it in short term memory more readily when that information is *lexically* meaningful as compared to information that signals 'only' postlexical information? Previous studies have not explicitly manipulated postlexical or paralinguistic contributions of pitch but have instead operationalized the binary distinction between non-lexical and lexical pitch information when investigating the influence of language experience (and hence linguistic function) on pitch processing, with different outcomes.

There is a growing body of experimental evidence suggesting that the processing of pitch information is indeed modulated by language experience, just like the processing of segmental information is (for an overview see Burnham, 1986). Behavioral and neurophysiological evidence suggests that there are differences in the way Chinese and English listeners identify Chinese tones and vowels (Gottfried & Suiter, 1997), how they process level tones vs. contour tones (e.g., Gandour, 1983) and in how they discriminate tones from another tone language such as Thai (Wayland & Guion, 2004). Furthermore there are differences in lateralization of pitch processing (e.g., Wang, Jongman, & Sereno, 2001), differences in the neuronal encoding of pitch (e.g., Krishnan, Xu,

Gandour, & P., 2005), and differences in the processing of lexical tone (e.g., Chandrasekaran, Krishnan, & Gandour, 2007 using mismatch negativity). Wang et al. (2001), for instance, had American English listeners (who had received a 30 minute training session) and Mandarin Chinese listeners identify Chinese tones presented to their right or left ear. While Chinese listeners showed lateralization (more errors when stimuli where presented to the left ear), American English listeners did not. This was interpreted as showing linguistic processing for tone in Chinese listeners but not in American English listeners. Chandrasekaran et al. (2007) used the passive oddball paradigm presenting one syllable with a certain tone (e.g., ma3) as standard and the same syllable with a different tone as deviant (e.g., ma1, or vice versa). Mismatch negativities (MMNs) following the detection of a change in stimulus differed across participant groups, showing higher amplitudes and later MMNs for native Chinese participants compared to English participants.

Hence there seems to be robust experimental evidence for language-specific processing of pitch. Interestingly, while all of these studies report differences across language groups, results are not uniform across *tones*. For instance, American English listeners in the lateralization study (Wang, et al., 2001) made significantly more errors for Tone 4 than for Tone 1 or 3. In the MMN study by Chandrasekaran and colleagues, MMN amplitudes were larger for Chinese listeners only in the Tone 1/Tone 3 comparisons but not for Tone 2/Tone 3 comparisons. These differences in results across tones is not surprising if we give up the binary view that tonal movements are either meaningful (for Chinese listeners) or not (for English listeners). Although none of the tones signals lexical information for someone only exposed to an intonation language, *some* of the tones may signal what could be interpreted as postlexical or paralinguistic information. The rising Tone 2 in Mandarin Chinese, for instance, if realized on the final syllable (or on a monosyllabic stimulus) may be interpreted as questioning, while the falling Tone 4 might be interpreted as a declarative. Anecdotal evidence also supports this finding at least for Tone 4: Listeners from a non-tonal language with no experience of a tone language sometimes interpret normal Chinese conversation as aggressive or assume that there is an argument. This impression may be invoked by the sharp falling Tone 4, which signals finality in English (e.g., Steedman, 2000). Such an interference

from the L1 prosodic system was experimentally shown by Broselow (1987) who tested English listeners' perception of Chinese tones. After a short period of training, English and Chinese listeners had to identify the tonal pattern of a certain syllable in different positions (in monosyllabic stimuli or at various positions in di- or trisyllabic sequences). She showed that Tone 4 was the easiest to be identified in sentence-final position, probably owing to its similarity to a declarative pitch movement. It was, however, the hardest tone to be identified in non-final positions.

Further complicating the picture, other tasks have shown very similar results for Chinese and English listeners. Cutler and Chen (1997), for instance, engaged Cantonese and Dutch speakers without tone language experience in a same-different task. They heard pairs of monosyllabic Cantonese stimuli that differed in consonant, vowel, tone, or combinations thereof. Results showed that both Cantonese and Dutch listeners were slower and less accurate when the difference was only in tone compared to when it was in the consonant or vowel. Cutler and Chen (1997) attribute these similarities across language groups to 'simple perceptual processing rather than to linguistic knowledge' (p. 165). This ties in with findings that English and Chinese listeners do not differ in how well they can discriminate f0 in non-speech stimuli (Bent, Bradlow, & Wright, 2006), a task which only relies on acoustic comparisons. Also, in a recent tone-learning study where Chinese and English learners were trained to recognize Cantonese tones, performance during the pre-training phase of the experiment was similar across language groups (Francis, Ciocca, Ma, & Fenn, 2008). Using the speeded classification task introduced by Garner (1970) both Repp and Lin (1990) and Lee and Nusbaum (1993) compared Mandarin Chinese and English listeners with respect to their classification speed for segments and pitch. In addition to testing pitch information that was *not* considered lexically distinctive in the two languages (Repp & Lin, 1990: low level vs. rising-falling pitch; Lee & Nusbaum, 1993: mid level vs. low level-pitch condition similar to Wood, 1974), they also included two existing Chinese tones (Repp & Lin, 1990: Tone 1 vs. Tone 4; Lee & Nusbaum, 1993: Tone 3 vs. Tone 4). Overall, both studies lead to very comparable results across language groups. Lee and Nusbaum (1993), for instance, reported that Chinese listeners showed mutual integrality of pitch and segments (i.e., influence of pitch manipulation on classification times for consonants

and vice versa in the orthogonal condition) for Chinese tones (Tone 3 vs. Tone 4) as well as for constant pitch stimuli that are no Chinese tones (mid level vs. low level pitch). English speakers, on the other hand, only showed mutual integrality in the Chinese tone condition, but not for the constant pitch stimuli. The authors hypothesized that the unexpected results for English listeners might be due to postlexical or paralinguistic functions of pitch, which are mostly signaled by dynamic changes in f0 and not by pitch levels. This hypothesis was not explicitly tested in their study (nor in subsequent work) and will be the focus of the present investigation.

To our knowledge, there are only few studies comparing postlexical and lexical pitch information. In an fMRI study, Gandour (2003) compared English and Chinese listeners cortical activation during the discrimination of tone and intonation in three-syllable utterances in Mandarin Chinese (the last tone was the target tone for tone discrimination). Results showed that tone processing resulted in left hemisphere activation, while intonation also activated regions in the right hemisphere. The authors argued that the left hemisphere is active during local pitch movements (on a single syllable) while the right hemisphere predominantly processes longer stretches of pitch (sentence intonation). Liang and van Heuven (2007) tested the classification of tone (four Chinese tones) and intonation (echo question vs. declarative) by native Chinese and L2 learners from tonal and non-tonal languages. They reported that L2 learners from a tonal language were faster and more accurate in classifying tones, while L2 learners from a non-tonal language were faster in classifying postlexical meaning.

Since linguistic function (tone vs. intonation) is confounded with stimulus length (one syllable vs. multiple syllables) in the preceding experiments, it is still not clear to what extent linguistic function guides the way listeners attend to the speech signal. The present paper provides new experimental evidence aimed at contributing to the debate over whether and how the processing of pitch information is influenced by language experience and by linguistic function. We tested two listener populations, native Dutch listeners and native Mandarin Chinese listeners. Importantly, we tested for the first time how listeners of an intonation language attend to postlexical contrasts (question vs. statement), comparing it to their perception of a non-linguistic contrast on the one hand and to Chinese listeners' perception of lexical contrasts on the other. We made use of the

fact that pitch movements that signal lexical contrasts for tone language speakers can signal either a postlexical or a non-linguistic contrast for listeners of an intonation language. Consequently, the effects of non-linguistic and postlexical contrasts on listeners' perception need to be tested with different materials (see below). On the other hand, both of these non-lexical contrasts can be compared to potential lexical contrast using identical materials, but with different participant groups. In the present paper, we compared three pitch contrasts (non-linguistic, postlexical, and lexical), manipulated as between-subjects, between-items factor. More specifically, we investigated (a) whether Dutch listeners attend to pitch information *at all* if it does not signal a linguistic contrast, (b) whether their tendency to attend to and encode pitch movements is increased when the contrast is linguistically meaningful (i.e., postlexical information such as declarative vs. echo question) and (c) whether such an increase in the likelihood of attending to linguistically relevant pitch contrasts is comparable to a tone language speakers' likelihood to attend to *lexical* pitch contrasts.

This forced-choice ABX nonword matching task does not require lexical access and is therefore ideally suited to compare the perception of speech stimuli by listeners from different languages. Another advantage of this task is that it is not as strongly influenced by low-level acoustic comparisons as a same-different task might be. The ABX task we employ in the current study allows listeners to freely choose which dimension of variation to attend to, segments or pitch. Dependent measures will be response type (classification along the segmental or suprasegmental dimension in incongruent trials) and reaction times in incongruent as compared to congruent trials. The type of response signals which information – segments or pitch – is considered more salient and qualifies better as classification criterion. Based on previous studies (Cutler & Chen, 1997; Ye & Connine, 1999) we expect that listeners will favor the segmental dimension over the suprasegmental one. Reaction times in the ABX task reflect difficulties in decision-making (Macmillan & Creelman, 2005). Listeners have to compare target X to standard A and to standard B before they can evaluate which of the standards is closer to the target. An increase in reaction times in incongruent relative to congruent trials therefore is an indication of a more difficult decision. The more segments and pitch compete with each other, the more difficult the decision is.

Experiment 1

Experiment 1 tested how Dutch listeners process consonants and non-linguistic pitch information. The non-linguistic pitch movement used in this study is the Mandarin Chinese Tone 2 (rising pitch), produced on the first strong syllable (see details below), which is not linguistically relevant for Dutch listeners. As listeners are free to classify incongruent stimuli along the segmental or suprasegmental dimension in the ABX task used, we do not explicitly draw their attention to the variations in pitch.

Methods

Participants

Eight native Dutch speakers from the subject pool of the Max Planck Institute for Psycholinguistics took part for a small fee. They were between 20 and 30 years old, had no experience with a tone language or tonal Dutch dialect.[3] Furthermore, they had not received musical training and did not play an instrument. They were unaware of the purpose of the experiment and had no known hearing problems.

Materials

Two pairs of segmentally similar, trochaic CVCV nonwords were selected, *mova (/ˈmɔuvɒ/) vs. noba (/ˈnɔubɒ/)* and *denu (/ˈdɛunʊ/) vs. zemu (/ˈzɛumʊ/)*. They were produced either with a pitch fall or with a pitch rise on the first syllable, resembling Tone 4 and Tone 2 in Mandarin Chinese respectively.

Pilot studies with 20 native speakers of Dutch were carried out to ensure that listeners perceived no postlexical contrast between these syllables. Results showed that nonwords with a rise on the first syllable were equally likely to be perceived as a statement as the nonwords with a fall on the first syllable (84.7% vs. 80.0%, $p > 0.5$). In addition, to determine whether the words with the rise versus the fall on the first syllable contrasted in any meaningful way, we asked listeners to rate the emotional intensity of the nonwords from 1 (weak) to 5 (strong). Those nonwords with a rise on the first

---

[3] Dutch dialects in the Southern part of the Netherlands (Roermond, Limburg) are described as tonal (e.g., Gussenhoven, 1999; Gussenhoven, 2000; Gussenhoven & Van der Vliet, 1998).

syllable were rated as more emotional than those with a fall on the first syllable (2.9 vs. 2.5, p < 0.001). However, listeners were very inconsistent in their classification of the emotion of these nonwords with an initial rise. The results of this pilot study suggest that the stimuli used in this experiment were at best contrastive at the paralinguistic level, and perhaps not even fully contrastive at this level. Importantly for the goal of the current experiment, however, the nonwords with the rise on the first and second syllable carried no postlexical or lexical information for Dutch listeners.

Recording

Following Dupoux et al., (1997), our speech materials were not recorded by a native speaker of neither of the two languages tested. Instead, we had a German speaker produce the recordings in order not to favor either Dutch or Chinese participants. This speaker had a university degree in Chinese and had a good working knowledge of Dutch. Every item was recorded 15 times in a sound attenuated cabin at the Max Planck Institute for Psycholinguistics and directly digitized onto PC using Adobe Audition (44.1 kHz, 16 bit).

Selection of stimuli

Nine items that were segmentally and suprasegmentally closest to each other (in terms of duration, f0-excursion, first and last f0-value) were chosen for the experiment. The mean f0-excursion of the pitch rise was 104.0 Hz (SD 16.4 Hz) compared to 156.1 Hz (SD 18.3 Hz) for the pitch fall $t(71) = 11.4$, $p < 0.0001$). Example pitch tracks are shown in Figures 2 and 3.[4]

Please insert Figure 2 about here.

Please insert Figure 3 about here

---

[4] In tone languages, pitch excursions are usually larger for falls than for rises (for discussion see also Bent, et al., 2006; Xu, 1994). To match f0-excursions between rises and falls better, the speaker tried to increase the f0-excursion of the pitch rises in a subsequent recording. However, this sounded very exaggerated and unnatural to native Dutch listeners.

Procedure

Participants were tested individually in sound-attenuated cabins. Each trial began with a Dutch lead-in sentence preparing them for the next trial (*De volgende komt heran*, 'The next trial is coming'). This sentence was recorded by a female Dutch speaker and adjusted in loudness to the experimental materials. After a one second pause, they heard the first stimulus, then a 600ms pause, the second stimulus, a 900 ms pause, and the critical word. Reaction times were recorded relative to the end of the critical word for a period of 2 seconds. The experiment was controlled using NESU (Nijmegen Experimental Setup).

Participants received written instructions to decide whether the third stimulus was more similar to the first or to the second stimulus and to press either the left or the right button of a two-button box. They were told that only a button press after the end of the third word would be recorded and that the experiment would not proceed if they pressed the button too early.

The main experiment consisted of 32 ABX nonword trials. Half of the trials were *congruent* trials in which the critical target X was of the same type as either the standard A or B (i.e. matching along the segmental *and* suprasegmental dimension, e.g., rising *mova* – falling *noba* – rising *mova*). The other half were *incongruent* in the sense that standard A matched target X along the suprasegmental but not segmental dimension, whereas standard B matched word X along the segmental but not suprasegmental dimension or vice versa (e.g., rising *mova* – falling *noba* – falling *mova*). The order of the standards A and B was counterbalanced for each participant. The target X contained all four segments (8 times *mova*, 8 times *noba*, 8 times *denu*, 8 times *zemu*) and all intonation contours (16 times a rise and 16 times a fall). Type of trial (congruent or incongruent) was manipulated as a within-subjects variable with 16 items per condition (8 of each segment pair). Four randomized lists were constructed. Finally, three familiarization trials (all congruent trials) were added at the start of each list. They consisted of pairs that had the same characteristics as the experimental stimuli: *nizu (/ˈnizʏ/), midu (/ˈmidʏ/), and boma (/ˈbɔumɒ/)*. Participants were randomly assigned to one of these lists.

Results

Twelve trials could not be analyzed because participants failed to respond within the allotted response period. Trials with reaction times (RTs) that were beyond three standard deviations of the grand mean of all participants were excluded as outliers (5 out of 244 trials). Remaining RTs ranged from 15ms to 1848 ms; these data points were converted to square-root to normalize the distribution of the raw RTs. (Anderson, 2001, chapter 12; Baayen, 2008, p 323f).

Statistical Analyses

Since reaction times in the ABX task correspond to decision difficulty, all remaining trials are included in the reaction time analyses. These RTs are analyzed using mixed-effects regression models (R version 2.10.10) which have been shown to be less dependent on normality and sphericity assumptions and more robust with respect to missing data than traditional separate subject and items analyses (e.g., Baayen, 2008; Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000; Quené & Van den Bergh, 2008). All models in this paper include *Items* and *Participants* as crossed random factors (both with 8 levels), and *Response button* (A or B) and *Trial type* (congruent or incongruent) as predictors. The initial models included all predictors and all interactions. Predictors were removed when they were not significant at $p = 0.1$ but only if they did not appear in significant interactions and if this did not deteriorate the fit of the model (as estimated by the Akaike Information criterion, a measure of the relative goodness of fit of a statistical model, cf. Akaike, 1974). Trials with residuals larger than 2.5 of the mean were removed to make sure the model was robust and not driven by outliers. The estimates (β), lower and upper bounds and *p*-values reported here are based on a Markov Chain Monte Carlo simulation with 10000 runs (posterior means and high posterior density interval and p-value).

Analysis of response type (classification along the segmental or suprasegmental dimension) was performed using binomial logistic regression models with *Response button* (A or B) and *Trial type* (congruent or incongruent) as predictors and *Items* and

Participants as crossed random factors. A logistic linking function was applied to account for the categorical nature of the dependent variable.

Analyses of responses

Participants classified the target along the segmental dimension in 98.2% of the congruent trials and in 94.1% in incongruent trials. Results showed no effect of *Response button* ($z = 1.25, p > 0.2$) and no effect of *Trial type* ($z = 1.15, p > 0.2$) on the odds[5] of responding along the segmental dimension.

Analyses of response times

Results showed no effect of *Response button* ($p > 0.8$) and no interaction between the two predictors ($p > 0.3$). There was only a main effect of *Trial typ*e ($\beta = 1.59$, Lower bound= 0.06, Upper bound = 3.03, $p < 0.05$). Responses to matching trials were on average 64 ms faster than responses to incongruent trials, see left bars in Figure 5.

Discussion

Dutch participants almost uniformly classified stimuli along the segmental dimension. The primacy of segmental information was expected based on previous experiments (Cutler & Chen, 1997; Ye & Connine, 1999). Interestingly, participants were slower in responding to incongruent trials (mismatching segmental and pitch information) than to congruent trials in this experiment even though the pitch contrast did not signal a linguistic contrast. There are two explanations for this finding. First, participants might be slowed down by any physical manipulation, irrespective of its function in the language, in line with earlier reports of an interference from task-irrelevant suprasegmental information, such as pitch level or loudness (Miller, 1978; Wood, 1974). Second, listeners might have been slowed down in incongruent trials because they reacted to the difference in emotional content between rises and falls.

This interference effect based on non-linguistic pitch contrasts serves as a baseline against which postlexical pitch information can be evaluated. In Experiment 2,

---

[5] Odds are the probability of an event occurring divided by the probability of the event not occurring.

we created stimuli in which the pitch manipulation signaled a postlexical contrast (that between a declarative statement and an echo question).

Experiment 2

In Experiment 2 the pitch rise was produced on the second syllable of the disyllabic stimuli. In contrast to the rise in Experiment 1, this rise on the utterance-final syllable signals an interrogative contour in Dutch (e.g., Haan, Pacilly, & van Bezooijen, 1997; Van Heuven & Haan, 2002). If attention to pitch information is modulated by linguistic function in the native language (much like attention to segmental information is modulated by linguistic function in the native language), then we predict that this kind of postlexical pitch contrast will be more closely attended to by Dutch listeners than the non-linguistic pitch contrast used in Experiment 1. If our prediction is correct, we expect to see even longer response times to incongruent trials than to congruent trials compared to Experiment 1. We do not, however, expect Dutch listeners to change their responses so they match along the suprasegmental rather than the segmental dimension.

Methods

Participants

An additional set of eight native Dutch listeners, selected from the same pool with the same criteria, participated for a small fee.

Materials

Materials with a pitch fall were identical to those in Experiment 1. Materials with a pitch rise were recorded in the same recording session as the materials described in Experiment 1 with the same speaker. Furthermore, the same selection criteria were used. In Experiment 2, the rise was realized on the second syllable, instead of on the first, resembling a Tone 2 in Mandarin Chinese (see example pitch track in Figure 4).

Please insert Figure 4 about here

The average F0-excursion of the rise in Experiment 2 was 145.1 Hz (SD 17.6 Hz), compared to 156.1 Hz (SD 18.3 Hz) for pitch falls ($t(71) = 2.5, p < 0.05$). A pilot test with 20 native Dutch listeners revealed that the pitch rise was interpreted as question in 76.2% of the cases. The emotional intensity of rises and falls did not differ (2.6 vs. 2.5 $p > 0.1$).

Procedure

The procedure and the experimental lists were identical to the ones used in Experiment 1. Only the nonwords produced with a pitch rise were replaced.

Results

Participants responded after time-out in four trials. Trials with RTs beyond three standard deviations of the grand mean of all participants were excluded as outliers (5 out of 252 trials). Remaining RTs, which ranged from 75.9 ms to 1978 ms, were converted to square root.

Analysis of responses

Participants classified along the segmental dimension in 100% of the congruent trials and in 89.9% of the incongruent trials. There was an effect of *Response button* on the odds of classifying along the segmental dimension ($\beta = 2.2$, z = 2.45, $p < 0.5$) but no effect of *Trial type* ($z = 0.01, p > 0.9$) and no interaction ($p > 0.4$). There were more responses along the segmental dimension (irrespective of *Trial type*) when the segmentally matching stimulus appeared as standard B rather than standard A (98.4% vs. 91.8%).

Reaction time analyses

Results showed no effect of *Response button* ($p > 0.6$), but a highly significant main effect of *Trial type* ($\beta = 3.00$, Lower bound: 1.91, Upper bound: 4.11, $p < 0.0001$), as depicted in Figure 5. There was no interaction between the two factors ($p > 0.5$). Response times to incongruent trials were on average 145.2 ms longer than response times to congruent trials.

Please insert Figure 5 about here

Combined RT analyses for Experiments 1 and 2

In order to investigate whether participants had a harder time classifying incongruent stimuli in Experiment 2 than in Experiment 1, data sets were merged and RTs were analyzed as before but with *Position of the rise* (first or second syllable) as between-experiment factor. Results showed no effect of *Response button* ($p > 0.2$) and no interaction between *Response button* and the other predictors (all p values > 0.5). The effect of *Trial type* approached significance ($\beta = 1.17$, p = 0.07). Moreover, there was a main effect of *Position of the rise* ($\beta = 4.85$, Lower bound: 1.50, Upper bound: 8.05, $p < 0.05$) and an interaction between *Position of the rise* and *Trial type* ($\beta = 1.83$, Lower bound: 0.05, Upper bound: 3.64, $p < 0.05$). As can be seen in Figure 5, participants were slowed down more in incongruent trials compared to congruent trials when the rise was produced on the second syllable (Experiment 2) compared to when it was produced on the first syllable (Experiment 1).

Discussion

Dutch listeners were more attentive to the linguistically meaningful pitch contrasts in Experiment 2 (declarative vs. echo question) than to non-linguistic pitch contrasts in Experiment 1. This evidence is not based on response types (in both experiments, listeners classified the stimuli overwhelmingly along the segmental dimension) but on reaction times. Compared to Experiment 1, in which reaction times to incongruent trials were slowed down with respect to congruent trials by 64 ms, they were slowed down by 145 ms in Experiment 2[6].

---

[6] One might be tempted to attribute this increase in reaction times for Experiment 2 to purely acoustic differences since the pitch rise in Experiment 2 had a larger f0-excursion than the one in Experiment 1 (145 Hz vs. 104 Hz). A larger f0-excursion naturally is more salient and might draw listeners' attention more to pitch variation than the less pronounced f0-excursion in Experiment 1. To address this issue, we replicated Experiment 2 but used PSOLA resynthesis to reduce the f0-range of the rise to match that of Experiment 1. Results of another group of eight native Dutch listeners were identical to those in Experiment 2. This additional set of data indicates that the differential results in Experiments 1 and 2 are not simply caused by acoustic salience of the pitch manipulation and corroborates our interpretation that linguistically meaningful pitch contrasts are attended to more closely than non-linguistic pitch contrasts.

Furthermore, the ordering of the standards A and B did not affect response times but the type of response (whether target X was matched to one of the standards along the segmental or suprasegmental dimension). However, participants were more likely to classify along the segmental dimension when the segmentally matching standard was 'close-by' (i.e. when it appeared as standard B) than when it was further away (i.e. when it appeared as standard A). To phrase it differently, the likelihood to match along the *suprasegmental* dimension was increased when the suprasegmentally matching standard was just heard before. One possible explanation for this finding is that postlexical pitch information is difficult to retain in memory and is therefore not used as classification criterion when it appears as standard A. When the suprasegmentally matching stimulus appears as standard B, however, pitch information is still active enough in working memory to qualify as classification criterion. We will come back to this issue in the General Discussion.

The linguistically meaningful pitch movements in Experiment 2 signaled a postlexical contrast, mapping onto the difference between declarative statements and echo questions. Naturally, we cannot compare how the processing of postlexical pitch information compares to the processing of lexical pitch information with Dutch listeners because Dutch does not have lexical tone contrasts. Therefore, in Experiment 3, we used tone language speakers to address this question.

Experiments 3a and 3b

Experiments 3a and 3b were replications of Experiments 1 and 2 with native Mandarin Chinese listeners. For this listener group, both the pitch rise on the first and the pitch rise on the second syllable signal potential lexical information. Based on the findings by Ye and Connine (1999) as well as Cutler and Chen (1997), we expect Chinese listeners to classify preferably along the segmental dimension, reflecting a language-independent general bias. Furthermore, we expect a similar - or even stronger - increase in response times for incongruent trials with respect to congruent trials as in Experiment 2 since mismatching *lexical* pitch information might be more difficult to ignore than mismatching postlexical information.

Methods

Participants

Sixteen native listeners of Mandarin Chinese participated for a small fee, eight were
assigned to Experiment 3a, using the materials of Experiment 1 (pitch rise on first
syllable), eight to Experiment 3b, using the materials of Experiment 2 (pitch rise on the
second syllable). Participants were between 20 and 30 years old and were all were
residents in Nijmegen where they had been living for one year on average at the time of
the study (range: 3 months to 1.5 years). They all knew English but were still dominant in
Chinese (they spoke Chinese regularly at home or with friends or family).

Materials

The materials were identical to the ones described in Experiment 1 and 2. One token of
each stimulus (12 in total, 4 nonword sequences and three pitch realizations) was
pretested with another group of five native Mandarin Chinese listeners. They heard the
stimuli and transcribed them using pinyin, a Latin-based alphabet that includes tonal
information. Analyses of the transcriptions showed that all rises were transcribed as Tone
2, while all the falls were transcribed as Tone 4. The non-Chinese phonemes /z/ and /v/
were consistently mapped onto /ts/ and /w/ in the transcriptions showing that Chinese
listeners distinguish them from /d/ and /b/ used as a segmental contrast in the nonword
pairs (*denu-zemu* and *mova-noba*).

Procedure

The procedure, the instruction for participants, and the experimental lists were identical
to Experiments 1 and 2. The lead-in sentence was translated into Mandarin Chinese
(*qing3 ting1 xia4 yi zu3 ci2,* 'Please listen to the next group of utterances) to simulate a
Chinese setting. They were recorded by a native speaker of Mandarin Chinese and
adjusted in loudness to the experimental materials.

Results

Participants responded after time-out in 13 trials of Experiment 3a and in 40 trials of
Experiment 3b. As before, only trials with RTs within three standard deviations of the

grand mean of all participants were analyzed. In Experiment 3a, this excluded 4 out of 243 trials; remaining RTs ranged from 26.3 ms to 1977.9 ms. In Experiment 3b, 6 out of 216 trials were excluded; remaining RTs ranged from 10.4 ms to 1949.7 ms. The remaining data points were converted to square-root and analyzed using a single model as the pitch manipulation in these two experiments all conveyed a potential lexical function. Position of the pitch rise (produced on the first or second syllable) was included as an additional factor in the analyses.

Analysis of responses types

When the rise was produced on the first syllable (Experiment 3a), Chinese participants responded along the segmental dimension in 99.2 % of the cases in congruent trials and in 58.0% of the cases in incongruent trials. Note that the high percentage of responses along the suprasegmentally matching dimension - which was not what we expected - were distributed equally over participants and items. When the rise was realized on the second syllable (Experiment 3b), participants responded along the segmental dimension in 99.0% of the congruent trials and in 86.7% of the incongruent trials. Combined analyses showed main effects of *Response button* ($\beta = 1.25$ $z = 2.90$, $p < 0.005$), *Position of the rise* ($\beta = 1.96$, , $= 2.76$, $p < 0.01$) and *Trial type* ($\beta = 4.94$, $z = 4.31$, $p < 0.0001$) but no interaction between any of the factors (all p values > 0.2). As for Dutch participants, there were significantly more classifications along the segmental dimension when the segmentally matching standard was positioned in B than in A of the ABX task (89.2% vs. 81.1%). Furthermore, classifying along the segmental dimension was more frequent in Experiment 3b compared to Experiment 3a (92.8% vs. 78.7%) and in congruent vs. incongruent trials (see percentages above).

Analyses of response times

There was no effect of *Response button* ($p > 0.5$) and no interaction between *Response button* and the other predictors (all p values > 0.4). Results showed a main effect of *Trial type* ($\beta = 6.37$, Lower bound: 5.20, Upper bound: 8.74, $p < 0.0001$), but no effect of *Position of the rise* ($p > 0.8$) and no interaction ($p > 0.6$). Also, the fit of the model did not deteriorate after removing *Position of the rise* and the interaction term. The fit of the

simpler model, as estimated by the Akaike Information criterion was 2994.4 compared to 2996.7 for the full model[7] ($\chi^2(2) = 1.71$, $p > 0.4$).[8] As Chinese participants' RTs did not differ across experiments, the reaction time data for Chinese listeners are pooled (see right bars in Figure 6). Responses to incongruent trials were on average 322 ms slower than responses to congruent trials.

Combined analysis of Experiments 2 and 3b

To test whether lexical and postlexical pitch contrasts are processed differently, we compared Dutch and Chinese listeners' performance in stimuli with a pitch rise on the second syllable. To this end, trials of Experiments 2 and 3b were merged into one data set. *Language* was entered as additional predictor in the analysis.

Combined analysis of response types showed no effect of *Language* and no effect of *Trial type* (all ps > 0.4).

To test whether participants were slowed down more in incongruent trials compared to congruent trials when pitch information was lexically contrastive as compared to postlexically contrastive, we also compared reaction times. Results showed no effect of Response button ($p > 0.8$) and no interaction with the other factors. There was a main effect of *Trial Type* ($\beta = 3.11$, Lower bound: 1.84, Upper bound: 4.39, $p < 0.0001$), no effect of *Language* ($p > 0.6$) but an interaction between *Trial Type* and *Language* ($\beta = 2.55$, Lower bound: 0.47, Upper bound: 4.28, $p < 0.01$). The difference in RTs between congruent and incongruent trials was larger for Chinese than for Dutch listeners. Results of Experiments 2 and 3b are combined in Figure 6.

Please insert Figure 6 about here

Discussion

Results of Chinese listeners show that potential lexical pitch contrasts are attended to more strongly than postlexical pitch contrasts by Dutch listeners. First, with respect to

---

[7] In comparing two statistical models, a significantly lower value of the Akaike Information criterion indicates a better fit.
[8] Excluding responses along the suprasegmentally matching dimension results in very similar models.

response type, Chinese listeners were significantly more likely than Dutch listeners to classify the target along the suprasegmental, pitch dimension (7.1% for Dutch incongruent stimuli overall compared to 13.3% for Mandarin Chinese listener when the rise was produced on the second syllable and 42% when the rise was produced on the first syllable). The difference in responses for Chinese listeners between Experiment 3a and 3b can be explained by Chinese phonology. In Chinese, toneless syllables do not usually occur in word-initial position (e.g., Chen & Xu, 2006). Classifying stimuli along the pitch dimension therefore appears to be more frequent when stimuli conformed to this pattern than when they did not (this had no effect on response times, however). More relevant for the present study is the fact that Chinese listeners considered pitch information as criterion for classification, which was not the case for non-linguistic or even postlexical information. This suggests that potential lexical pitch information competes with segmental information more strongly than postlexical pitch information.

Like Dutch responses in Experiment 2, Chinese responses were also influenced by the position of the segmentally matching standard. If standard B was the segmentally matching one, then there were significantly more responses along the segmental dimension than when standard A represented the segmental match. In the discussion of Experiment 1 we speculated that postlexical pitch information might be difficult to retain in memory and that listeners therefore relied more on the comparison between standard B and target X. Since there is no difference between Chinese and Dutch listeners, potential lexical information may be equally hard to remember over some distance as postlexical information. This hypothesis needs further experimental support since the current study was not designed to test this possibility.

Also, in terms of reaction times, potential lexical pitch contrasts were processed differently than postlexical pitch contrasts. As predicted, Mandarin Chinese listeners showed significantly slower responses in incongruent trials than in congruent trials, reflecting difficulties in decision-making. Importantly, Chinese listeners' responses in incongruent trials were even slower than those of Dutch listeners in Experiment 2 (in which a pitch rise signaled postlexical information). Thus, when it comes to the relevance of pitch information for nonword matching decision speed, lexical pitch information is attended to more than postlexical pitch information.

General Discussion

This study examined Dutch and Chinese listeners' processing of segmental and suprasegmental (pitch) information under three conditions: non-linguistic pitch contrasts (Experiment 1), postlexical pitch contrasts (declarative vs. echo question, Experiment 2), and potential lexical pitch contrasts (Tone 2 vs. Tone 4, Experiments 3a and 3b). In summary, by differentiating these three functions of pitch information instead of just the two extremes focused on by most past studies (lexical versus non-linguistic), we have found a pattern of results that (a) confirms language-specific differences in pitch perception, and (b) helps us to understand obvious differences in the perception of individual tones across language populations (via recourse to postlexical functions of pitch).

Regarding language-specific differences in pitch perception, results of both response types and reaction times show that linguistic function guides how pitch information is processed. First, only Chinese listeners – for whom pitch information is lexical – classified targets in incongruent trials along the pitch dimension. Although classifications along the segmentally matching dimension still prevailed, it is noteworthy that pitch was considered as a classification criterion *at all*. Chinese listeners' high percentage of classifications along the suprasegmental dimension suggests that pitch variation signaling potential lexical information is difficult to ignore. Possibly, for speakers of tone languages, lexical tone information competes directly with segmental information for processing resources. Alternatively, segmental and tone information might simply be more integrated for speakers of a tone language than for speakers of a non-tonal language (see also Repp & Lin, 1990). The results of the current study do not allow us to differentiate between these two possible explanations. It is tempting to speculate that because we observe this effect with a nonword categorization task, this may indicate that lexical tone information is likely to be processed prelexically (for further evidence along these lines see Hallé, Chang, & Best, 2004; Lee, 2007). A recent study testing participants' lexical representation of pitch (Galts, Kabak, & Braun, 2011), compared the ability to learn novel sound-object correspondences containing lexical tone (e.g., *taman, sukai* with different lexical tones on the second syllable) for listeners from

different language backgrounds (German, Russian, French, and Chinese). After a learning phase, participants were presented with matching and mismatching sound-object pairs. Their task was to decide whether the sound matched the presented object or not (based on what they had learned in the training phase) in three conditions: tonal mismatch, segmental mismatch and control (matching tone and segments). As expected, Chinese participants outperformed the other three groups in the tonal mismatch condition, but were less sensitive to the segmental mismatch condition than the other groups. Regardless of why the Mandarin Chinese listeners in this study were more likely to choose to match nonwords along the suprasegmental dimension, it is clear that their behavior differs from Dutch listeners.

Second, reaction time analyses clearly showed that the more linguistically relevant pitch information is (lexical > postlexical > non-linguistic), the more strongly pitch information competed with segmental information as the decision criterion. Chinese listeners had a very hard time classifying the stimuli in incongruent trials (320 ms delay with respect to congruent trials), which suggests that segments and pitch were both very good criteria. In comparison, Dutch listeners were quicker in the classification of incongruent trials than Chinese listeners, possibly, because classifying along the pitch dimension did not really seem an option for them. It was not the case, however, that Dutch listeners ignored pitch information altogether. Even when pitch did not signal a linguistic contrast, responses to incongruent trials were slower than responses to congruent trials (by 64 ms). This corroborates earlier reports that even non-linguistic pitch information is processed and retained in short-term memory (Braun, et al., 2006 for imitation study; Hallé, et al., 2004 for categorization evidence; Lee & Nusbaum, 1993 for Garner-type studies; Miller, 1978; Repp & Lin, 1990; Wood, 1974). More importantly, decision-making became more difficult , when pitch movements conveyed a postlexical contrast (echo question vs. declarative statement). In that case, responses to incongruent trials were on average 154 ms slower than responses to congruent trials.

This intermediate status of postlexical pitch contrasts in speech processing (more important than non-linguistic pitch differences and less important than potential lexical pitch contrasts) helps us to understand some of the apparent differences in outcomes of earlier studies. On the one hand, even though all linguistically meaningful pitch contrasts

are attended to more than non-linguistic pitch contrasts, potential lexical pitch information is different from post-lexial pitch contrasts (e.g., Gandour, et al., 2003). This appears to be the reason why a number of neurophysiological studies show differential pitch processing by speakers from tonal and non-tonal languages (e.g., Chandrasekaran, et al., 2007; Gandour, 1983; Krishnan, et al., 2005; Wang, et al., 2001). Hence, although postlexical pitch information is processed very rapidly and efficiently (e.g., Dahan, et al., 2002 for intonational cues to reference resolution; Van Heuven & Haan, 2002 for intonational cues to declarative vs. echo question interpretation; Weber, Braun, & Crocker, 2004), it remains different from potential lexical pitch information. On the other hand, our study clearly shows that at least Tone 2 and Tone 4 can be mapped onto a postlexical contrast in Dutch (c.f. Broselow, et al., 1987). In that case, pitch movements are attended to more closely than when they have no relevant function. That these two linguistic functions of pitch need to be differentiated from non-linguistic pitch movements is reflected in some of the above-mentioned processing studies (e.g., Chandrasekaran, et al., 2007; Wang, Jongman, & Sereno, 2006). As discussed in the introduction, differences across language groups (tonal and non-tonal speakers) are larger for some tones than for others. Where there are larger differences across language groups, tones were probably treated as non-linguistic by listeners from an intonation language. When there are smaller differences across language groups, listeners from an intonation language might have interpreted the stimuli in a linguistically meaningful way.[9]

In combination, the results of our forced choice and response latency measures point to a step-wise increase in how strongly listeners attend to speech pitch, from non-linguistic to postlexical to lexical. We think it likely that this distinctive processing of functionally different pitch movements (both lexical and postlexical) has its roots in early infancy (e.g., Fikkert & Chen, 2011; Li & Thompson, 1977; Mattock, Molnar, Polka, & Burnham, 2008; Quam & Swingley, 2010), and is thus a fundamental property of human communication systems.

---

[9] As an anonymous reviewer pointed out, it might be interesting to test how flexible Chinese-Dutch bilinguals are in their attention to tonal information. Conceivably, attention to speech pitch in bilinguals is modulated by the dominant language and possibly also by the language setting they are in.

Acknowledgements

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control, 19*(6), 716-723.

Anderson, N. H. (2001). *Empirical Direction in Design and Analysis*. Mahwah: New Jersey: Lawrence Erlbaum.

Baayen, H. R. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Baumann, S., Grice, M., & Steindamm, S. (2006). *Prosodic marking of focus domains - categorical or gradient?* Paper presented at the 3rd International Conference on Speech Prosody, Dresden, Germany.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). Influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 97-103.

Best, C., T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Birch, S., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language, 47*, 571-588.

Bock, K. J., & Mazzella, J. R. (1983). Intonational marking of given an new information: Some consequences for comprehension. *Memory & Cognition, 11*(1), 64-76.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech, 49*(4), 451-493.

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down on-line speech comprehension. *Language and Cognitive Processes, 26*, 350-375.

Braun, B., Kochanski, G., Grabe, E., & Rosner, B. S. (2006). Evidence for attractors in English intonation. *Journal of the Acoustical Society of America, 119*(6), 4006-4015.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The Acquisition of second language sound system* (pp. 350 - 362). Cambridge, MA: Newbury House.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics 7*, 207 - 239.

Chandrasekaran, B., Krishnan, A., & Gandour, J. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148-156.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech: Evidence from neutral tone in Standard Chinese. *Phonetica, 63*, 47-75.

Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge, UK: Cambridge University Press.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59*, 165-179.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1-10.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28*(5), 746-755.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34(2)*, 269 - 284.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*(2), 292-314.

Fikkert, P., & Chen, A. (2011). *The role of word-stress and intonation in word recognition in Dutch 14- and 24-month-olds.* Paper presented at the 35th Annual Boston University Conference on Language Development, Boston.

Flege, J. E., Bohn, O.-S., & Jan, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*, 268-294.

Galts, T., Kabak, B., & Braun, B. (2011). *Stress typology of the L1 matters in the lexical encoding of novel tonal contrasts.* Paper presented at the 10th International Symposium of Psycholinguistics, San Sebastián, Spain.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 149-175.

Gandour, J. T., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience. *Brain and Language, 84*, 318-336.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist, 25*, 350-358.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics, 25*(2), 207-231.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht, Holland; Cinnaminson, N.J., U.S.A.: Foris Publications.

Gussenhoven, C. (1999). *Tone systems in Dutch Limburgian dialects.* Paper presented at the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics, Tokyo: Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Languages.

Gussenhoven, C. (2000). The lexical tone contrast of Roermond Dutch in optimality theory. In M. Horne (Ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers.

Gussenhoven, C., & Van der Vliet, P. (1998). The phonology of tone and intonation in the Dutch dialect of Venlo. *Journal of Linguistics, 35*, 99-135.

Haan, J., Pacilly, J. J. A., & van Bezooijen, R. (1997). An anatomy of Dutch question intonation. In J. Coerts & H. de Hoop (Eds.), *Linguistics in the Netherlands 1997* (pp. 97-108). Amsterdam: John Benjamins.

Hallé, P., A., Chang, Y.-C., & Best, C., T. (2004). Identification and discrimination of Mandrin tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*, 395-421.

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-olds: A crosslingusitic perspective. *Infancy, 13*, 440-455.

Kazanina, N., Philipps, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *PNAS, 103*, 11381-11386.

Kohler, K. (1991). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK), 25*, 115-185.

Krishnan, A., Xu, Y., Gandour, J., & P., C. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*, 161-168.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philisophical Transactions of the Royal Society B, 363*, 979-1000.

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*(2), 157-165.

Li, C., & Thompson, S. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*, 185-199.

Liang, J., & Van Heuven, V., J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes, Volume 2: Experimental studies in word and sentence prosody* (pp. 27-61). Berlin/New York: Mouton de Gruyter.

Liberman, M. (1975). The Intonational System of English. Unpublished doctoral dissertation, MIT.

Liberman, M., & Sag, I. (1974). *Prosodic form and discourse function.* Paper presented at the Papers from the Tenth Regional Meeting of the Chicago Linguistics Society, IL.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide (2nd edition)*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*, 1367-1381.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on L2 acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229-258.

Miller, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics, 24*(2), 1175-1180.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of Acoustical Society of America, 102*(3), 1864-1877.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica, 46*, 181 – 196.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America, 90*(6), 2956-2970.

Quam, C., & Swingley, D. (2010). Phonological knowledge guides 2-year-old's and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language, 62*, 135-150.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects with binomial data. *Journal of Memory and Language, 59*(413-425).

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

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*, 1346 – 1356.

Singh, L., Morgan, J., & White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51*, 173-189.

Snedeker, J., & Trueswell, J. (2002). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103 – 130.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguity. *Journal of Psycholinguistic Research, 25*, 247 – 268.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry, 31*(4), 649-689.

Terken, J. M. B., & Nooteboom, S. (1987). Opposite Effects of Accentuation and Deaccentuation on Verification Latencies for Given and New Information. *Language and Cognitive Processes, 2*, 145-163.

Van Heuven, V. J., & Haan, J. (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 61-86). Berlin: Mouton de Gruyter.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic Perception of Mandarin Tones by Chinese and American Listener. *Brain and Language, 78*(3), 332-348.

Wang, Y., Jongman, A., & Sereno, J. A. (2006). Second language acquisition and processing of Mandarin tone. In E. Bates, L. H. Tan & O. J. L. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (pp. 250-257). Cambridge: Cambridge University Press.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681-712.

Weber, A., Braun, B., & Crocker, M. W. (2004). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49*(3), 367-392.

Welby, P. (2003). Effects of pitch accent type and status on focus projection. *Language and Speech, 46*, 53-81.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganizatino during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics, 15*, 501-508.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*, 2240-2253.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*, 609-630.

Figure captions

Figure 1: F0-contours of the four Chinese tones (reprinted from Moore & Jongman, 1997, p. 1865).
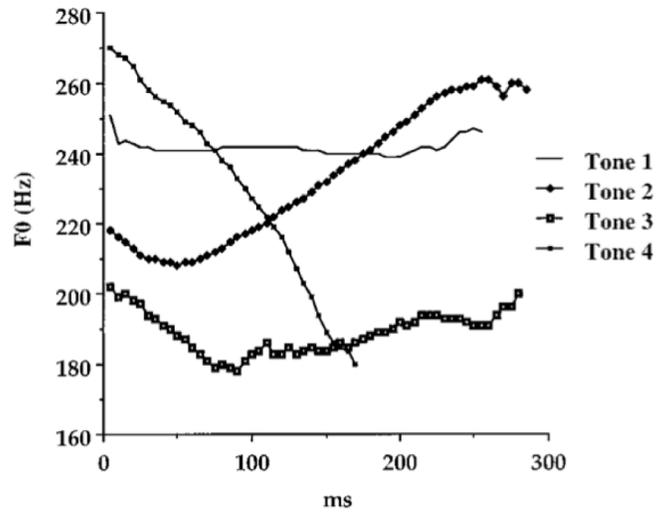

Figure 2: Example contour of a pitch fall in Experiment1. F0-is shown from 100 Hz to 400 Hz in all figures
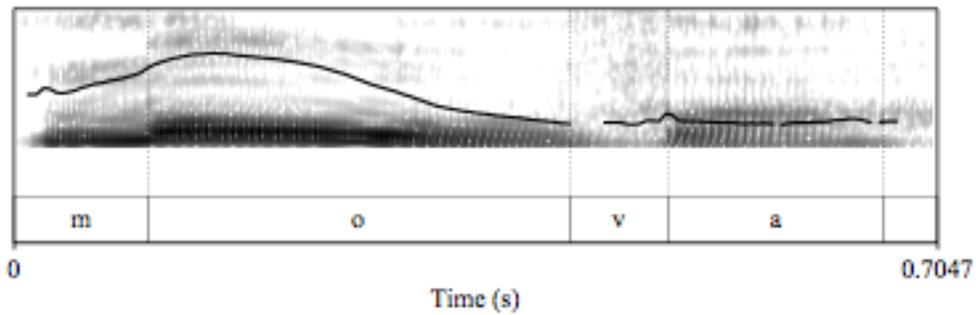
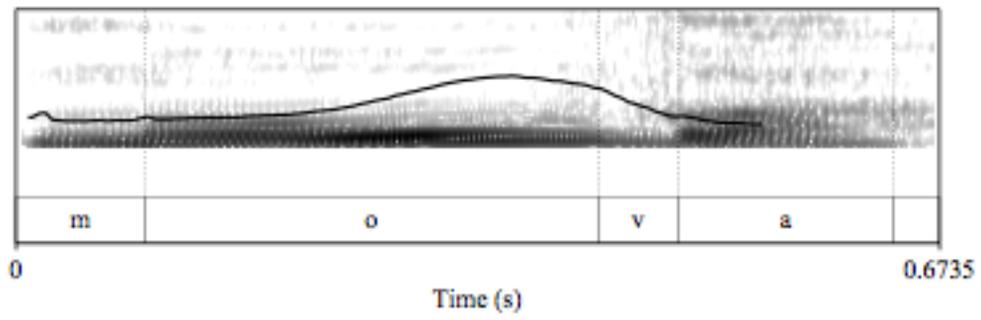Figure 3: Example contour of a pitch rise in Experiment 1.

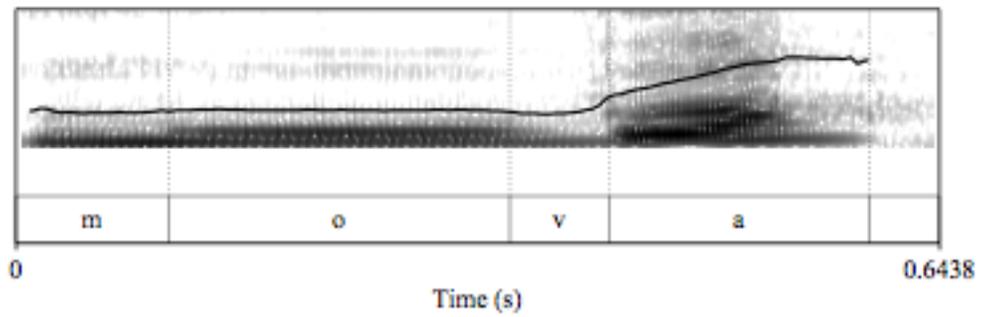Figure 4. Example contour with a pitch rise in Experiment 2.

Figure 5. Reaction times for Dutch listeners in Experiment 1 (left bars, non-linguistic pitch contrast) and Experiment 2 (right bars, postlexical pitch contrast). Mean values are estimated by the statistical model (posterior mean of MCMC sampling); whiskers represent the 95% high posterior density.
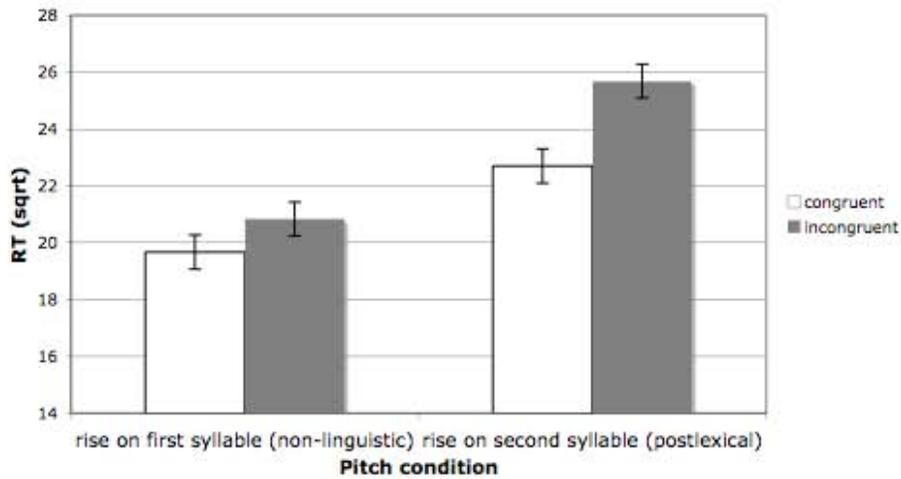
Figure 6. Reaction times comparing postlexical pitch contrasts (left bars, Dutch listeners) to lexical pitch contrasts (right bars, Chinese listeners). Mean values are estimated by the statistical model (posterior mean of MCMC sampling); whiskers represent the 95% high posterior density.