Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

L1 experience shapes the perception of intonational contours

Elaine Schmidt [1], *Carmen Kung*[1], *Brechtje Post*[2], *Ivan Yuen*[1], *Katherine Demuth*[1]

[1] Macquarie University, Australia
[2] University of Cambridge, UK

Elaine.Schmidt@cantab.net

Abstract

While the influence of L1 on the perception of segments is well established, the effect of L1 on the perception of suprasegmentals such as intonational contours is less known. Previous studies claim that suprasegmental processing is less sensitive to L1 experience because it is based mostly on general auditory mechanisms. Thus, falls and rises can be universally distinguished regardless of language background while different types of rises are processed similarly between language groups. However, often these studies have not included languages that actually use different types of rises linguistically. In this study we investigated the effects of L1 experience on the perception of rises by Australian-English listeners, for whom different rises signal the difference between interrogatives and declaratives in their L1, and Mandarin learners of English, who should be sensitive to tonal differences but not to the intonational differences tested here. Results demonstrate that the perception of suprasegmentals is indeed shaped by the L1 with a significantly better discrimination of rises by Australian-English listeners. Additionally, it appears that suprasegmental processing at the utterance level occurs independently from the lexical level since Mandarin listeners were not able to draw on tonal sensitivities for discrimination at the utterance (intonational) level.

Index Terms: intonation, speech perception, perceptual bias, second language learners

1. Introduction

Many studies have demonstrated that listeners are guided by perceptual biases stemming from their L1 which influence the perception of speech. Especially on a segmental level, the influence of the L1 is well documented (see [1] for a review). In an intonational perception study Grabe et al [2] demonstrated that all listener groups (English, Spanish, Chinese) could distinguish between falls and rises, but no differences were observable *within* different rises. Thus, the authors concluded that on a suprasegmental level perception is based on universal auditory processing mechanisms for slow moving frequency changes (i.e. the difference between falls and rises), while experience of the L1 only 'moulds' the perception of differences. Nonetheless, the fact that different rises could not be distinguished can be explained with the function of rises in the languages that were tested. Southern Standard British English typically uses rises for interrogatives and falls for declaratives. While the pitch height of the rise can differ, this does not result in a different interpretation of the linguistic function of the utterance. Similarly, in Mandarin, different intonational rises do not appear to signal different discourse functions. Hence, the linguistic experience with rises for both groups should be the same, and no differences should be expected for the perception of rises.

This leaves open the question of whether perceptual differences are observable if one of the listener groups uses rises contrastively in their L1. Australian English, for example, uses rises for both interrogatives and declaratives [3]. Subtle pitch differences in the height of the rise determine which linguistic function the utterance fulfils. Higher rises shift the perception toward an interrogative interpretation while lower rises are more likely to be interpreted as declaratives [4].

This raises the question of whether Australian English listeners are more attuned to perceiving fine pitch differences in rises. Additionally, by using Australian English in comparison to Mandarin, we can tease apart general auditory from linguistic effects in discrimination. If the sensitivities required to detect differences in different types of rises are purely auditory in nature, then both Australian and Mandarin learners should perform very well; for Australian listeners because intonational rises occur contrastively in their L1, and for Mandarin listeners because of their tonal language background. However, if suprasegmental processing is based on linguistic L1 experience, Mandarin learners should perform worse than Australian listeners since different rises only occur at the lexical level and do not form intonational contrasts that are linguistically meaningful.

Thus, in this study we investigate how far the perception of intonational contours of a language does indeed depend on L1 experience when the differences are smaller than a dramatic fall/rise difference, and whether the discrimination abilities are based on general auditory or linguistic mechanisms. By using sentences that could be either interrogative or declarative rises, and manipulating them at different positions in the utterance, we provide a linguistic and auditory contrast for Australian listeners, but only an auditory contrast for Mandarin listeners.

1.1. The linguistic function of rises in Australian English

While all varieties of English use intonation to signal the difference between interrogatives and declaratives, the intonational differences for these two discourse functions are often more subtle in Australian English. In contrast to many other varieties where interrogatives are signalled by rises and declaratives are signalled by falls, Australian English (but also New Zealand English [5], Belfast English [6], etc.) uses rises for both. However, it appears as if interrogative and declarative rises are perceptually different [4].

Especially at the boundary, pitch height seems to determine the interpretation of an utterance with higher pitch, shifting the probabilities toward a question interpretation [4].

 doi: 10.21437/SpeechProsody.2016-50

Even though the shift occurs gradiently, a pitch difference of 4 semitones (ST) already reverses listeners' interpretation of an utterance. This indicates that listeners might be sensitive to subtle pitch changes in rises since these are often linguistically relevant in their L1.

1.2. The role of rises in Mandarin

Unlike Australian English, Mandarin does not use rising intonational contours contrastively for interrogatives or declaratives. Instead, questions can either be marked by question particles, even though they are not compulsory, or by various syntactic cues [7]. Since question particles usually occur in utterance-final position, it is conceivable that listeners' attention is drawn especially to the last syllable in syntactically ambiguous sentences. In absence of question particles and syntax, intonation can also signal questions. However, in contrast to English, Mandarin does not employ boundary tones as a cue to interrogatives. Instead, to mark an utterance intonationally as an interrogative, a global F0 raise is applied to the entire utterance [8], [7]. This might mean that local pitch changes, which only affect the pitch height of the pitch accent but not the beginning or end of an utterance, might go unnoticed. Thus, even though the presence of tonemes in Mandarin means that listeners are sensitive to subtle pitch differences in lexical items, their L1 might not prove to be a sufficient vehicle to detect changes on an utterance level.

2. Hypotheses

Based on previous studies demonstrating the influence of L1 perception on segmental contrasts, we hypothesise that suprasegmental information is also influenced by L1 language experience so that even small pitch differences within intonational rises can be detected by Australian listeners. However, we assume that an auditory sensitivity is not enough to detect differences in rises if these differences are not linguistically meaningful. Hence, Mandarin listeners will show worse discrimination abilities than Australian English listeners. Specifically, we hypothesise that

H1. Australian English native speakers are sensitive to subtle differences in rises and are better at distinguishing between different rises than Mandarin learners of English who do not use different rises linguistically in their L1.

H2. Mandarin listeners will show an effect of position with better discrimination for rises that differ in final compared to pitch accent position since in their L1 their attention is on the end of utterances in syntactically ambiguous sentences.

3. Methodology

Forty monolingual Australian English adults (20 male, 20 female) participated in the experiment as well as an (initial) 18 Mandarin learners of English from mainland China (9 male, 9 female). At the time of testing the Mandarin subjects had lived in Australia for less than 3 years but as they were university students in Australia, and had obtained an overall IELTS score of at least 6.5. All subjects reported normal hearing.

The carrier sentence ("They often played with…") and target words were recorded by 1 female and 1 male native speaker of Standard Australian English with neutral almost monotone intonation. The carrier sentence was further flattened synthetically to remove any additional information listeners might be able to use. The pitch was only systematically manipulated at the target words in utterance-final position; the carrier sentence remained the same.

The target words were trisyllabic pseudowords with initial stress and a CVCVCV syllable structure consisting of nasals/ liquids and vowels (e.g. Milana). The pitch was manipulated using PSOLA in Praat [9] either at the *pitch accent*, i.e. the first syllable of the target word, or at the *boundary*, i.e. the last syllable of the target word (see Fig. 1). The pitch was shifted upward/downward in 2 semitone (ST) steps starting from a medial rise (labelled as -4ST) to make a total of 5 steps for *boundary* and 4 steps for *pitch accent* manipulations to ensure the naturalness of all stimuli. All *pitch accent* manipulations ended in the same boundary tone (-4ST), and the *boundary* manipulations all used the same pitch accent (see Fig. 1).

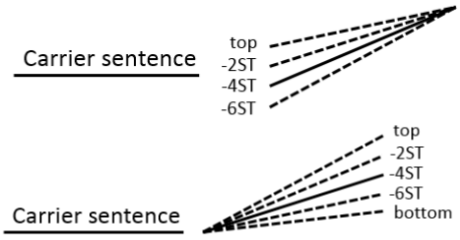

Figure 1: *Schematic representation of pitch accent (top) and boundary tone manipulations (bottom), in 2 semitone (ST) steps*

Four phoneticians judged independently whether the resynthesised sentences sounded natural. Cases in which there was disagreement were discarded with all associated pitch steps for both conditions. Eight target words (for each speaker) were resynthesised to make 152 items.

Listeners were presented with two target sentences occurring consecutively, and were instructed to decide as fast as possible whether the second sentence was the same as the first. In contrast to many discrimination studies, participants heard all possible pairings instead of just neighbouring steps, i.e. 1-2, 1-3, 1-4, 2-3, 2-4, etc. Additionally, the order of pairings, i.e. 1-2 vs. 2-1 was randomised since the perception of rises might be affected by the preceding intonational contour [10]. *Pitch accent* and *boundary tone* manipulations occurred independently of one another and were presented in separate blocks, as were stimuli produced by the male and female speaker.

4. Results

Since there were only four pitch steps in pitch accent position, but five steps in boundary position, we ran analyses for each position separately. Due to the uneven number of participants per language group, we used Type III Sum of Squares in a univariate ANOVA which leads to equal weighting of factors and thus controls for uneven sample size and variance. Results in the **boundary condition** revealed a significant main effect for *language background* ($F(1, 2788) = 7.93$, $p=.005$), *distance* ($F(4, 2788) = 313.85$, $p<.001$), and *speaker gender* ($F(1, 2788) = 11.29$, $p=.001$). This demonstrates that language background does affect the perception of intonational changes. In particular, in boundary position Mandarin learners were less accurate in the discrimination of intonational contrasts (Fig. 2). Additionally, we found two-way interactions between

language background and speaker gender (F(1, 2788) = 5.37, p<.05), and between language background and distance (F(4, 2788) = 3.44, p<.01).

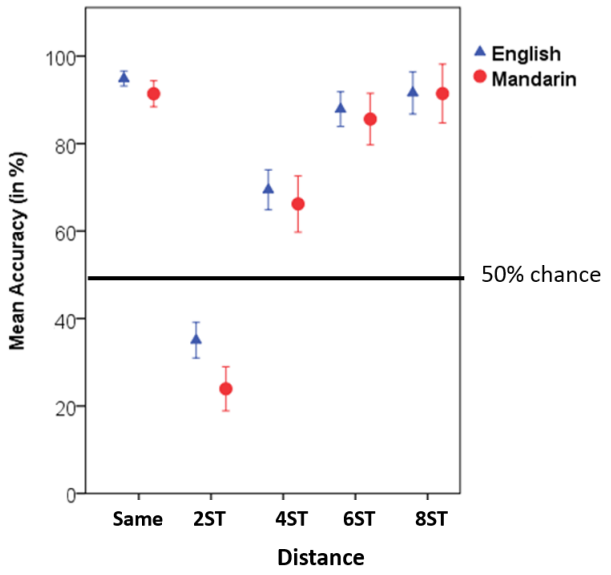

Figure 2: ***Boundary position*** *- Mean accuracy scores (in %) by Mandarin (red circle) and Australian English listeners (blue triangle) for discrimination of same/ different items.*

Further analyses to tease apart the interactions between language and speaker gender show that there are no speaker gender differences in English listeners. In contrast, Mandarin learners' accuracy is affected by speaker gender. In particular, if the pitch difference is only 2ST, they perform worse for the male than the female speaker.

Results in the **pitch accent condition** revealed significant main effects for *language background* (F(1, 2129) = 239.52, p<.001), and *distance* (F(3, 2129) = 294.58, p<.001) (Fig 3).

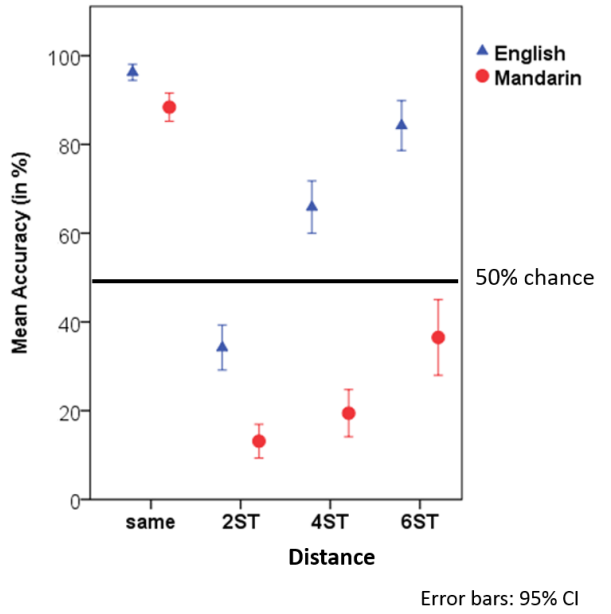

Figure 3: ***Pitch accent position*** *- Mean accuracy scores (in %) by Mandarin (red circle) and Australian English listeners (blue triangle) for discrimination of same/ different items.*

This suggests that the perception here is determined again by language background. Additionally, we found a two-way interaction between language background and distance (F(3, 2129) = 25.38, p<.001). Unlike in the boundary condition, Mandarin learners perform significantly worse for all different pairings (all below 50% chance level) in the pitch accent condition.

In the previous analyses, we grouped all pitch pairings together based on the relative distance, i.e. the pairing step 1 and 2, and the pairing step 3 and 4 both are just 2ST, i.e. 1 pitch step distance apart. In the following, we looked at the *actual distance* and ran further analyses to see whether the difference between step 1 and 2 is treated the same as the difference between step 2 and 3, etc. Results do not show any significant differences in the **boundary condition,** with regard to either language background, or the actual pitch step distance. However, in the **pitch accent** condition, *language background* F(1, 2129) = 100.61, p<.001) and *actual pitch step* (F(9, 2129) = 59.47, p<.001) are both significant. While the English native speaker group did not show significant differences in subsequent analyses, the Mandarin learners showed significant differences between steps 1-2 and 2-3 and 3-4 respectively, and between 1-3 and 2-4 (all at p<.001), see Fig. 4, for easier inspection we have excluded all 4 same-same pairings).

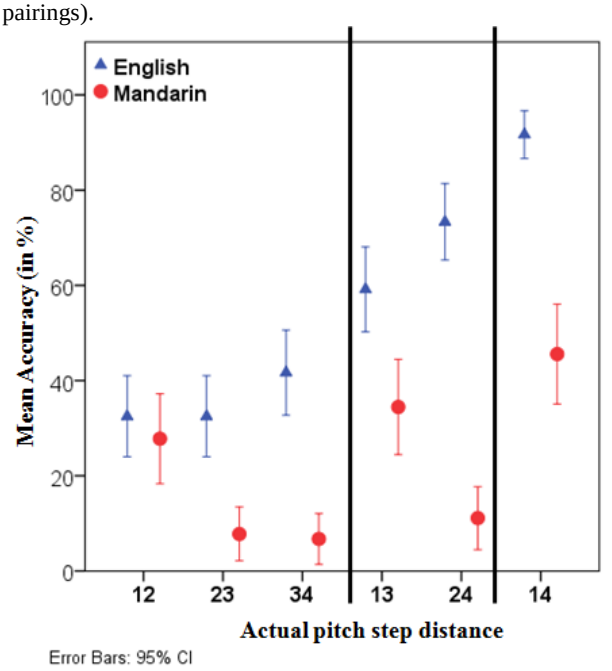

Figure 4: ***Pitch accent position*** *- Mean accuracy scores (in %) by Mandarin (red circle) and Australian English listeners (blue triangle) for discrimination of same/ different items per actual pitch step distance (leftmost box 2ST distance, middle box 4ST, rightmost box 6ST distance)*

This demonstrates that for Mandarin listeners, difference pairings containing the highest peak (step 1) and any other pitch step were easier to distinguish than pairings that did not contain the highest pitch step, even if the total distance was the same. For Australian English listeners there was no significant difference between any of the actual pitch step pairings.

5. Discussion

In this study we aimed to tease apart effects of linguistic versus auditory experience in the perception of pitch contours. By contrasting the perception of rises in Australian English listeners, who use rises linguistically to differentiate between interrogatives and declaratives, with that of Mandarin listeners who are sensitive to pitch differences as tonal speakers but only use rises contrastively in tonemes, not in intonational contours, we were able to disentangle the effects of general auditory sensitivity and linguistic experience. The results demonstrate that L1 experience does indeed shape the perception of suprasegmental information, just as it does in segments. Native speakers of Australian English seem to be overall more perceptually attuned to subtle pitch differences in rises than Mandarin learners of English. This suggests firstly that differences beyond the slow moving frequency changes between rises and falls can be perceived by listeners provided they occur contrastively in their L1. Secondly, the differences between Australian and Mandarin listeners demonstrate that pitch processing at the utterance level occurs independently of pitch processing on the lexical level.

In particular, **pitch accent** manipulations show very strong effects of L1 perceptual biases, and thus lend support to H1. While Australian English listeners consistently misinterpret a 2ST difference as 'same', they are above chance level (above 50%) for 4ST, and 6ST differences. In contrast, Mandarin learners of English cannot perceptually distinguish any of the difference pairings (Fig. 3). Thus, the results suggest that it is not a general auditory sensitivity that is at the bottom of the discrimination differences since Mandarin and Australian listeners performed very differently. Instead, it appears to be the linguistic experience of the L1 which shapes the perceptual sensitivities to suprasegmental information. Because rising intonation contours are used linguistically in Australian English, listeners attune to the different rises. As Schmidt et al. [4] demonstrated, Australian English listeners use pitch-rise differences to distinguish perceptually between interrogatives and declaratives in the absence of contextual cues. Since these rises are a common phenomenon in Australian English, it is conceivable that listeners are better attuned to pitch in order not to miss subtle cues that provide them with information about the intended function of an utterance. Additionally, since declarative and interrogative rises usually span more than one syllable, Australian listeners need to pay attention to various points in the utterance. This might then also explain why they show similar accuracy scores for pitch accent and boundary manipulations.

In contrast, it appears that the better sensitivity to tonal differences that Mandarin listeners have as a function of their language background, does not facilitate discrimination of intonational contrast. Perhaps the fact that their native tonal contrasts span a syllable while here the intonational contrasts span all 3 syllables of the target word might explain the poor performance. As such, the pitch differences in this experiment might not be perceived because they occur more globally while the listeners' attention might be focussed on local differences within syllables. Furthermore, we assume that Mandarin listeners pay special attention to the beginning and end of sentences in order to capture F0 raises of the whole utterance. However, in the pitch accent manipulations here, the pitch of the carrier sentence remained unaltered, and the boundary tone always ended in the same rise. The differences thus only occurred in the middle of the utterance.

This then explains the poorer performance for pitch accent over boundary tone manipulations. In the **boundary condition** Mandarin listeners performed similarly to Australian listeners for 4ST and 6ST steps. Only the 2ST step shows an effect of language background with better discrimination by Australian listeners (Fig. 2). Thus, H2 is also borne out by the data. The position of the manipulation determines Mandarin listeners' sensitivity, with better discrimination for rises that differ in boundary tones. We assume that this is related to the attention that native Mandarin listeners pay to the final segments of an utterance in expectation of a question particle [7], [8].

The difference that is observable in the actual pitch step distance (Fig. 4) in the **pitch accent condition** in Mandarin learners is difficult to interpret. It suggests that the difference between the highest pitch step (1) and any other pitch step is perceptually more salient than the difference between say step 2 and 3, or step 3 and 4. Potentially this result is related to perceptual salience due to the pitch peak in relation to the carrier sentence. The pitch height of step 1 is higher than that of the carrier sentence, while step 2 is at the same level as the carrier sentence and steps 3 and 4 are below.

Finally, it is not clear why Mandarin, but not English listeners, show differences in accuracy depending on speaker gender. Potentially, this difference could be a function of overall pitch height. While the pitch steps are the same for both speaker genders, the overall pitch is higher for the female speaker, which might facilitate the discrimination process for Mandarin listeners. As the pitch is manipulated in semitones, the female intonation contours have a wider pitch range than the male ones which might make them perceptually more salient to Mandarin listeners.

6. Conclusions

Like segmental perception, suprasegmental perception seems to be strongly influenced by L1 experience. In this study we demonstrated that even subtle pitch differences between different types of rises can be discriminated by listeners if they occur with a linguistic function in their L1. Thus, it does not seem to be a general auditory sensitivity that drives perceptual differences but a more specified linguistic sensitivity based on L1 experience. Since Australian English uses different types of rises for interrogatives and declaratives, listeners attune more readily to pitch differences in sentences that are syntactically ambiguous. In contrast, Mandarin listeners do not seem to be as sensitive to rises in intonation despite a sensitivity to tonal contrast from their L1. These findings raise important questions for the production of yes/no questions by Mandarin L2 speakers of Australian English. Here, further research is needed to look at whether and how pitch is used in production to distinguish the two. Both the perception and the production of rises is important for effective communication between Australians and non-native speakers of Australian English if top-down contextual cues are unavailable for disambiguation.

7. Acknowledgements

We are very grateful to the Child Language Lab for their helpful comments and suggestions. This project is funded by the ARC Centre of Excellence in Cognition and its Disorders (ARC CE110001021) and the ARC Laureate Fellowship (ARC FL130100014) to Katherine Demuth.

8. References

[1] J. F. Werker and R. C. Tees, "The organization and reorganization of human speech perception," *Annual Review of Neuroscience*, vol. 15, pp. 377-402, 1992.

[2] E. Grabe, B. S. Rosner, J. E. Garcia-Albea and X. Zhou, "Perception of English Intonation by English, Spanish, and Chinese Listeners," *Language and Speech*, vol. 46, no. 4, pp. 375-401, 2003.

[3] J. Fletcher and J. Harrington, "High-rising terminals and fall-rise tunes in Australian English," *Phonetica*, vol. 58, no. 4, pp. 215-229, 2001.

[4] E. Schmidt, B. Post, C. Kung, I. Yuen and K. Demuth, "The effect of listener and speaker gender on the perception of rises in AusE," in *ICPhS XVIII*, Glasgow, 2015.

[5] P. Warren, "Patterns of late rising in New Zealand English: Intonational variation or intonational change?," *Language Variation and Change*, vol. 17, pp. 209-230, 2005.

[6] G. Garding and A. Arvaniti, "Dialectal variation in the rising accents of American English," *Papers in Laboratory Phonology IX: Change in Phonology*, vol. 9, pp. 547-576, 2004.

[7] F. Liu, Intonation Systems of Mandarin and English: A functional approach, Chicago: Unpublished Doctoral Thesis, 2009.

[8] J. Yuan and C. Shih, "Confusability of Chinese intonation," in *Speech Prosody 2004*, 2004.

[9] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 5.1.12) [Computer program]," 2007. [Online]. Available: http://www.praat.org/. [Accessed 18 November 2009].

[10] C. Gussenhoven and A. C. Rietveld, "Fundamental frequency declination in Dutch: testing three hypotheses," *Journal of Phonetics*, vol. 16, no. 3, pp. 355-369, 1988.