ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

Rises and Falls in Dutch and Mandarin Chinese

Ao Chen [1], *Aoju Chen* [1,2], *René Kager*[1], *Patrick Wong*[3]

[1] Utrecht Institute of Linguistics, Utrecht University, the Netherlands
[2] Max Planck Institute for Psycholinguistics, The Netherlands
[3] Laboratory for Language, Learning, and the Brain, Chinese University of Hong Kong, China
a.chen@uu.nl, aoju.chen@uu.nl, r.w.j.kager@uu.nl, patrick.neuro@gmail.com

Abstract

Despite of the different functions of pitch in tone and non-tone languages, rises and falls are common pitch patterns across different languages. In the current study, we ask what is the language specific phonetic realization of rises and falls. Chinese and Dutch speakers participated in a production experiment. We used contexts composed for conveying specific communicative purposes to elicit rises and falls. We measured both tonal alignment and tonal scaling for both patterns. For the alignment measurements, we found language specific patterns for the rises, but for falls. For rises, both peak and valley were aligned later among Chinese speakers compared to Dutch speakers. For all the scaling measurements (maximum pitch, minimum pitch, and pitch range), no language specific patterns were found for either the rises or the falls.

Index Terms: alignment, scaling, lexical tone, pitch accent, tone language, non-tone language

1. Introduction

Pitch functions differently in different languages. In tone languages like Mandarin Chinese, pitch can function both lexically, i.e. distinguishing word meaning, and post-lexically, i.e. expressing sentence-level meaning. Post-lexical pitch variation does not change the identifiability of lexical tones [1]. In non-tone languages, such as Dutch and English, pitch mainly serves the post-lexical function. Regardless of the different roles of pitch in tone versus non-tone languages, some pitch patterns are common in both types of languages. Rises and falls are cases in point. These two patterns exist in the inventory of lexical tones in Mandarin Chinese, Cantonese, Thai and various other Southeast Asian languages [2]. Both can occur as pitch accents in non-tone languages [3].

The commonality of falls and rises however does not entail that they are realized in the same way phonetically in different languages. Tonal categories similar in shape can differ in alignment of pitch peak and pitch valley. For example, the peak of pre-nuclear rising accent in German has been found to align later compared to Greek, English and Dutch [4]. Further, languages differ in their standard pitch range, i.e. pitch range within which speakers of a language habitually speak [5]. For example, British English has a larger standard pitch range than Dutch [6][7] and German [8]. Such differences can lead to differences in the scaling of tonal categories. In the current study, we examine similarities and differences in the phonetic realization of rises and falls in Chinese and Dutch.

Tonal alignment of Dutch rising and falling accents has been extensively studied. For example, [9] found that for rises, though the low turning point consistently aligns to the onset of the accented syllable, the location of the peak is more variable. [10] showed that the peak of prenuclear fall is aligned to the offset of the vowel if the accented syllable ends with a long vowel, but during the following consonant if the vowel is a short vowel. Nuclear peaks are aligned earlier than prenuclear accents, and their alignment is not influenced by syllable structure but by whether the accent occurs in utterance-final position [11]. Peak and valley alignment of Chinese lexical tones seem to be of a different nature. Syllable boundaries have been argued to be the anchoring points of Chinese lexical tones [12][13]. That is, the lexical tones approach their pitch targets towards the offset of the syllable [12][13]. More specifically, the low turning point of the rises occurs at the center of the syllable and the peak tends to be synchronized to the end of a syllable [14]. For falls, the peak occurs around the center of the syllable, but closer to the offset, and the valley tends to occur closer to syllable offset [12]. Regarding scaling, [13] argued that the pitch range of non-level lexical tones is around 6 semitones in Chinese. Tonal scaling of pitch accents in Dutch has been studied in the context of focus marking (e.g. [15]) and dialectal differences [3]. There appear to be differences in tonal scaling across varieties of Dutch everything else being equal.

However, a direct comparison in the phonetic realization of rises and falls between Dutch and Chinese has not yet been conducted. [17] speculated that the pitch contours of the nuclear fall, rise and fall-rise are broadly similar to the lexical tones 2 (rise), 4 (fall) and 3 (dip tone). This speculation appeared to be based on a comparison in shapes as depicted in a figure illustrating the nuclear rise, fall, and fall-rise as realized in the proper name 'Lof' in Dutch [18] and a figure illustrating the shapes of time-normalized lexical tones in Chinese [17]. In this study, we address this issue phonetically by comparing the tonal scaling and alignment of the rising tone (tone 2) and falling tone (tone 4) in Chinese, and the rising (L*H) and falling (H*L) pitch accents in Dutch in monosyllabic-utterances elicited in pragmatically similar contexts.

2. Method

2.1. Participants

Nine native Chinese female participants and nine native Dutch female participants took part in the experiment. All the Chinese participants were raised in China, and were in the Netherlands for post-graduate studies. All the Dutch participants were raised in the Netherlands, and had no exposure to tonal languages or tonal Dutch dialects. All reported normal hearing and no language deficiency.

2.2. Stimuli

As mentioned above, we elicited the rises and falls from Chinese and Dutch speakers in the same pragmatic contexts. These contexts were originally composed by [19]. [19] was concerned with the appropriateness of the meanings assigned to four single-accent pitch patterns in the literature on Dutch intonation. The four pitch patterns were accent-lending rising, accent-lending fall; accent-lending rise and fall on one syllable; and accent-lending rise and a half fall on one syllable. The prototypical meanings that were suppose to go with these pitch patterns were “testing”, “selection”, “addition”, and “addition plus it was predictable”. In “testing”, speaker (“S”) leaves it up to hearer (“H”) to determine whether a variable (“V”) belongs to the background; in “selection”, S selects a V from the background (which means that V was present in the background at the moment of speaking); in “addition”, S adds a V to the background (which means that V was not present in the background at the moment of speaking); in “addition” plus “it was predictable”, V is added to the background, but that is a matter of everyday routine [19, p134]. To investigate the form-function mapping, [19] composed three contexts (A, B, C) for each meaning attribute in each of the two types of orientations, “default” and “vocative”. “Default” meant that the speaker referred directly to the referent of the focused information (the proper name), and “vocative” meant the proper name was produced to address the hearer. All contexts are situated in a school, and the speaker is a teacher, who is supposed to produce a proper name (e,g, Marina) in a certain pitch accent appropriate in that context. Take the meaning ‘testing’ for example. One of the contexts representing the meaning ‘testing’ was as follows: You are talking to a colleague about a possible party guilty of theft; you think that Marina is the offender, but you are not very sure; you want to verify whether your colleague shares your suspicion. The speaker should produce “Marina” that suited the context.

In the current study, the original written contexts were presented to Dutch speakers, and the contexts were translated into Chinese for Chinese speakers (listed in Appendix, please refer to [19] for Dutch contexts). Instead of using the proper name “Marina” as in [19], we used proper names Mi /mi/ and Nee /nei/. We chose these two syllables for the following reasons. First, the sonorant segments ensure continuous pitch contour in production and minimize the chance of truncation and compression in the phonetic realization of pitch accents due to a shortage of sonorant material [20]. Second, this study is part of a larger project that studies the subcortical processing of pitch, where tokens of /mi/ are used frequently as stimuli [21]. As Nee has a similar segmental structure to Mi, it might also be a good stimulus for subcortical processing studies. In addition, in Dutch, Nee is pronounced as the word ‘nee’, which is a high-frequency word, meaning “no”, whereas Mi is pronounced as the word ‘mi’, which is a low-frequency word, referring dish made of noodle-like material. By introducing the difference in frequency of the sounds, we have the chance to observe whether sound-frequency influences the realization of pitch patterns.

Prior to the production experiment proper, we did a pilot experiment with three Dutch speakers using all the 24 contexts used in [19]. The contexts were presented to the speakers randomly for several times, and they were asked to produce Mi and Nee proper for the contexts. We found that the speakers were more consistent in their choice of pitch contour (fall vs. rise) in some contexts than in other contexts. In order to minimize individual variation, in the experiment proper, we used three types of contexts that showed highest consistency in eliciting the pitch patterns: “default testing” which elicited L*H H%; “default addition” which elicitsed H*L L%, and “default addition plus it was predictable” which elicited H*L %. All three instantiations were used for each context type.

2.3. Procedure

The recording took place in a sound-attenuated booth. The speakers were recorded by means of the software Audacity with a sampling rate of 48 kHz. The nine contexts were printed out on a piece of paper. The Dutch speakers were given the original texts in [19], and the Chinese speakers were given a Chinese translation of the Dutch contexts. Before the recording started, the participants had time to familiar themselves with the contexts and they could ask the experimenter if they had any questions. For each context, each speaker was asked to read and imagine the context, and when they finished reading, they should produce Mi or Nee in a way that was appropriate in that context. For each target utterance, the production was repeated three times. The speakers all first produced Mi in all contexts, followed by the production of Nee in all contexts. For the Chinese speakers, the recording was done in the following order: Mi carrying the rising tone (tone 2), Nee carrying the rising tone, Mi carrying the falling tone (tone 4), and Nee carrying the falling tone.

3. Analysis and results

3.1. The measurements

The production of the Dutch speakers was first presented to two phoneticians, who judged each utterance as rise, fall, or neither. For the Dutch speakers, three of the utterances (in “default addition”) were not included in the analysis due to discontinuous pitch contours. Among the remaining 483 utterances, 154 were produced with a rising pitch contour, and 282 were produced with a falling pitch contour. Among the rises, most occurred in the three instantiations of “default testing”, which is consistent with [19] in terms of context-pitch accent matching. The falls were frequently produced in the rest of the contexts, also consistent with [19]. The number of rises and falls produced in each instantiation of each meaning by the Dutch speakers is given in Table 1.

As the falls produced in the three contexts of ‘addition + predictable’ did not have a low boundary tone, we assumed that there were not comparable to tone 4 in Chinese as the falling patterns produced in the three contexts of ‘addition’, we decided to focus on the utterances produced in “default testing”, and “default addition” contexts in both languages.

Table 1. *The frequency of rises and falls in each context.*

Contexts	Number of rises	number of falls
Default testing A	**48**	4
Default testing B	**43**	2
Default testing C	**46**	6
Default addition A	3	**50**
Default addition B	1	**52**
Default addition C	9	**42**
Default addition+pre A	0	33
Default addition+pre B	2	43
Default addition+pre C	2	50

The utterances were subsequently annotated phonetically. For each utterance, we first found the syllable boundary. For each speaker, a proper setting for pitch floor and ceiling was selected, where the range of pitch floor was 75 to 100 Hz, and the range of pitch ceiling is 500 to 650Hz. Within the syllable boundary, three landmarks were annotated, which were pitch peak (H), pitch valley before H (L1), and pitch valley after H

(L2).

For each tonal pattern, we obtained two sets of measurements, one for tonal alignment, and one for tonal scaling. The measurements within each category are listed below:

Tonal alignment measurements:

- Location of H
- Location of L1
- Location of L2
- Peak alignment (LoMax)
- Valley alignment (LoMin)

Tonal scaling measurements:

- Pitch height of H
- Pitch height of L1
- Pitch height of L2
- Pitch range

For the rises, LoMax was calculated as absolute time between the onset of the syllable and the location of pitch peak. LoMin alignment was calculated as the absolute time between the offset of the syllable and the location of the pitch minimum. Pitch range of the rises was calculated as the difference between H and L1. For falls, as previous literature has suggested that the pitch pattern is mainly determined by the fall after the peak [15], valley alignment was calculated as the distance from the location of L2 to the offset of the syllable. Pitch range of calls was calculated as difference between H and L2.

3.2. Analysis

We analyzed the realization of rises and falls separately. For each pitch pattern, and for each phonetic measurement, we carried out a mixed-effect model analysis with 'language' as the fixed factor, and 'speaker' and 'contexts' (A, B or C as listed in Table 1) as the random factors in SPSS to establish to what extent the language of the speakers could explain the variation in the phonetic realization of each pattern. Table 2 lists the effect of 'language' for rise and fall for each measurement.

Table 2. *Effects of language for each measurement for rise and fall.*

	Pitch alignment		Pitch scaling			
	LoMin	LoMax	H	L1?	L2	range
Rises	p=0.068	p<0.01	n.s.	n.s.		n.s.
falls	n.s.	n.s.	n.s.		n.s.	n.s.

3.2.1. Analysis of the rises

For rises, we found a significant effect of 'language' on LoMax, $F(1, 17.47) = 11.92$, $p<0.01$; a marginal significant effect of 'language' on LoMin $F(1, 17.93) = 3.79$, $p=0.068$. No significant effect of 'language' can be found for pitch maximum $F(1, 17.46) = 0.10$, n.s., pitch minimum, $F(1, 18.00) = 3.19$, n.s., or for pitch range $(1, 17.49) = 0.58$, n.s.. Figure 1 plots the alignment measurements of the two groups of speakers, and Figure 2 plots the scaling measurements. It can be seen that Chinese and Dutch differed in peak alignment, where Chinese speakers had a later peak alignment compared to Dutch speakers. Chinese speakers were also slightly later in terms of valley alignment. In contrast, Dutch speakers and Chinese speakers did not differ significantly in scaling measurements. Not only did they have the same degree of pitch excursion in rises, but also comparable maximum pitch and minimum pitch. For all the measurements, including 'speaker' as a random factor significantly improved the fitness of the model (all $p<0.05$). However, no significant improvement was found by including 'context' as a random factor.

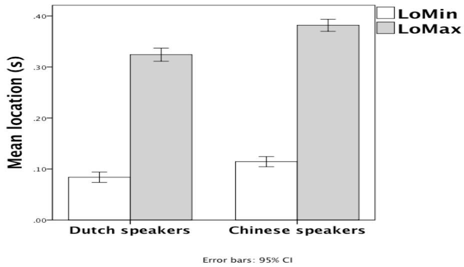

Figure 1. *Peak and valley alignment of the rises by Dutch and Chinese speakers.*

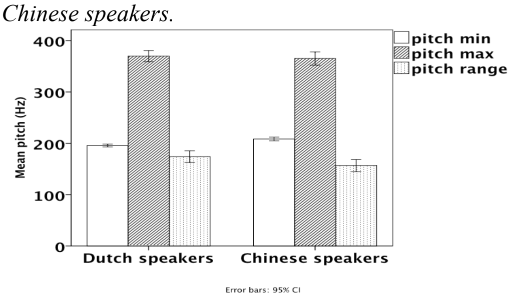

Figure 2. *Pitch minimum, pitch maximum, and pitch range of the rises by Dutch and Chinese speakers.*

3.2.2 Analysis of the falls

For fall, no significant effect of language can be found for any of the measurements: LoMax $F(1, 17.84) = 2.38$, n.s.; LoMin2 $F(1, 14.82) = 2.48$, n.s.; Maximum pitch $F(1, 17.92) = 0.53$, n.s.; Minimum pitch after peak $(1, 17.86) = 1.22$, n.s.; pitch range $F(1, 17.97) = 0.28$, n.s. Figure 3 plots the alignment measurements of fall of both groups and Figure 4 plots the scaling measurements. Same as for rises, for each measurement, by including participants as a random factor, the fitness of the model was significantly improved (all $p<0.05$), whereas including contexts as random factor did not improve the model. As can be seen from the figures, for both groups, the peak and valley of fall were aligned to similar locations. With regard to the scaling measurements, it can be seen that Dutch speakers and Chinese speakers showed comparable degree of pitch excursion, as well as low and high point.

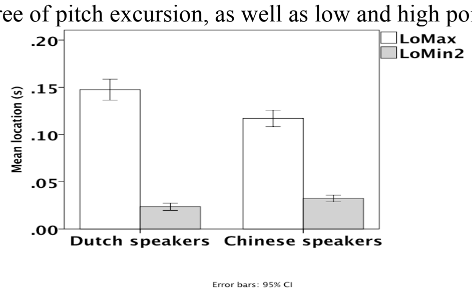

Figure 3. *Peak and valley alignment of the falls by Dutch and Chinese speakers.*

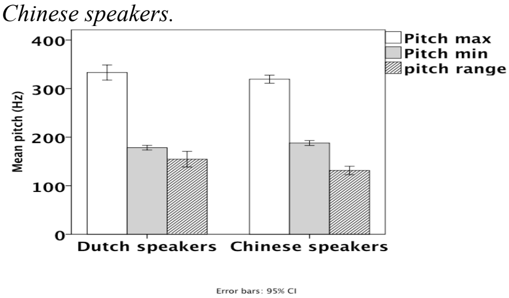

Figure 4. *Pitch minimum, pitch maximum, and pitch range of the falls by Dutch and Chinese speakers.*

One thing that is worth mentioning was the differences in scaling and alignment produced by the speakers in different contexts representing the same meaning attribute. For rises, peak and valley alignment could differ for 50ms among Dutch speakers, and the pitch range might differ up to 100Hz in different contexts. Similar cross-instantiations variation could be found for falls. Chinese speakers also exhibited cross-context variation in peak and valley alignment in rises. It was somewhat unexpected that adding the random factor 'context' did not improve the performance of the models including both the fixed factor. Due to the small number of contexts per meaning attribute, it is hard to quantify any potential differences between the contexts. It is however possible that the speakers had different interpretations of the degrees of the meaning in these contexts. This would in turn explain why the random factor 'speaker' did improve the performance of a model including only the fixed factor 'language'.

4. Conclusions

In the current study, we examined tonal alignment and scaling in the rises and falls in a tone language (Chinese) and a non-tone language (Dutch). We found that cross-linguistically, speakers differed in the phonetic realization of the rises but not in the phonetic realization of the falls. The differences concerning the rises were observed in the alignment measurements but not in the scaling measurements. This implies similarity in the standard pitch range between Dutch and Mandarin Chinese. The Chinese speakers aligned the peak of the rises much later relative to the syllable onset, compared to the Dutch speakers. This same difference held for valley alignment, albeit to a lesser degree. These differences are in line with [9][11][12] in the literature on Chinese prosody and the findings on the alignment of nuclear rises in Dutch. Our results thus show that Dutch nuclear rising accent L*H is phonetically different from tone 2 in Mandarin Chinese. The Chinese and Dutch speakers exhibited comparable tonal scaling and alignment of the falls, suggesting that the Dutch nuclear falling accent H*L is phonetically similar to tone 4 in Mandarin Chinese.

Our results have interesting implications for research on perception and processing of lexical tones by speakers of a non-tone language. Traditionally, speakers of a non-tone language have been found to be poor at perceiving lexical tones [22]. Their not speaking a tone language has been used as an explanation for their performance. Our results suggest that this may not be the whole picture, and raises the question how similarities and differences in the phonetic realization of similarly-looking patterns between languages may shape the perception of lexical tones.

Appendix

Default testing A: 你正在和同事谈论一次偷盗行为，你认为Mi(或者Nee)是偷盗的人，但是你不确定，你想跟你同事确定你的猜测。

Default testing B: 你的学生告诉你(Mi 或者 Nee)今天不来上课了，因为她要去议会参加永久教育委员会的会议。你感到非常惊讶，这是一个玩笑吗?

Default testing C: 你在参加教师会议。会议需要指派一个学生参加学校管理。你的同事提名了几名候选人，你自己头脑中有个人选，但是你完全不确定这个人能不能被其他的同事接受，你提出一个试探性的建议：

Default addition A: 你正在和一个同事商量哪个学生适合参与建立校报，你突然想到一个学生，她想成为一名记者：

Default addition B: 你正在和同事开会讨论食堂收银台的现金丢失问题。你们讨论了盗贼的几个方面，突然一切都清楚了，一定是她：

Default addition C: 你在参加一个教师会议。会议需要指派一名学生参加学校管理。你的同事提名了几个候选人，你自己头脑中有个人选，你十分确定其他人也会接受你想到的人选：

Default addition+predictable B: 一个同事刚刚问你上午哪个学生没有来，你说Mi(或者Nee), 但是明显你的同事没有听清，又问"谁"，你有点不耐烦地回答说：

Default addition+predictable B: 你召集你的同事开会，因为现在食堂的收银台总丢失现金。你知道是谁干的，因为你早上抓到她偷钱，而且这不是第一次了：

Default addition+predictable C: 你正在和同事商量哪个学生适合参与筹建校报。突然你想到一个学生，她明显是个合适的人选，因为她在新闻稿比赛中获奖，你们在前一次教师会议中给她发了奖。

5. References

[1] Chao, Y.,-R., "Tone and intonation in chinese", Bulletin of the Institute of History and Philology, Academia Sinica, 4, 121-134, 1933.
[2] Yip, M., "Tone", Cambridge University Press, 2002.
[3] Gussenhoven, C., "The phonology of tone and Intonation Cambridge: CUP, 2004.
[4] Atterer, M. and Ladd, D. R., "On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German", Journal of Phonetics, 32(2), 177-197, 2004
[5] Chen, A., Gussenhoven, C. and Rietveld, A., "Language-specificity in perception of paralinguistic intonational meaning", Language and Speech, 47(4), 311-350, 2004.
[6] de Pijper, J., "Modeling British English Intonation", Dordrecht, the Netherlands: Foris, 1983.
[7] Willems, N., "English intonation from a Dutch point of view: An experimental-Phonetic investigation of English intonation produced by Dutch native speakers", H. I. Ambacht: Intercontinental Graphics, 1982
[8] Mennen, I., Schaeffler, F., and Docherty, G., "Pitching itdifferently: A comparison of the pitch ranges of German and English speakers", Proc. 16th ICPhS Saarbrücken, 2007.
[9] Caspers, J. and van Heuven, V., "Effects of time pressure on the phonetic realization of the dutch accent-lending pitch rise and fall", Language and Speech, 50, 161-171, 1993.
[10] Ladd, D. R., Mennen, I. and Schepman, A., "Phonological conditioning of peak alignment in rising pitch accents in Dutch", The Journal of the Acoustical Society of America, 107(5), 2685-2696, 2000.
[11] Schepman, A., Lickley, R. and Ladd, D. R., "Effects of vowel length and "right context" on the alignment of Dutch nuclear accents." *Journal of Phonetics, 34*(1), 1-28, 2006.
[12] Xu, Y. and Emily Wang, Q., "Pitch targets and their realization: Evidence from Mandarin Chinese", Speech Communication, 33(4), 319-337, 2001.
[13] Xu, Y., "Effects of tone and focus on the formation and alignment of f0 contours", Journal of Phonetics, 27(1), 55-105, 1999.
[14] Xu, Y., "Consistency of tone-syllable alignment across different syllable structures and speaking rates", Phonetica, 55, 179-203, 1998
[15] Hanssen, J., Jörg P. and Gussenhoven, C., "Prosodic effect of focus in Dutch declaratives", Proc. of Speech Prosody, Campinas, 609-612, 2008.
[16] He, X. "Mandarin-accented Dutch prosody", doctoral dissertation, Radboud University of Nijmegen, 2012.
[17] Xu, Y., "Contextual tonal variations in Mandarin", Journal of Phonetics, 25(1), 61-83, 1997.
[18] Hanssen, J., Peter, J. and Gussenhoven, C. "Frase-final pitch accommodation effect in Dutch", proc. of ICPhS XVI, 1077-1080.
[19] Caspers, J., "Experiments on the meaning of four types of single-accent intonation patterns in Dutch", Language and Speech, 43, 127-161, 2000.
[20] Graber, E., "Comparative intonational phonology: English and German", MPI Series 7. Wageningen: Ponsen et Looien, 1998.
[21] Wong, P., Skoe, E., Russo, N., Dees, T. and Kraus N., "Musical experience shapes human brainstem encoding of linguistic pitch patterns", Nature Neuroscience, 10, 420-422, 2007.
[22] Kiriloff, C., "On the auditory discrimination of tones in Mandarin", Phonetica, 20, 63-76, 1969.