



Influence of pitch dimensions on non-native tone perception by Dutch and Mandarin listeners

Shuangshuang Hu¹, Ao Chen^{1,2}, René Kager¹

¹Utrecht Institute of Linguistics, Utrecht University, the Netherlands

²School of Communication Science, Beijing Language and Culture University, China

s.hu1@uu.nl, a.chen@uu.nl, r.w.j.kager@uu.nl

Abstract

Previous studies have shown that listeners of different language backgrounds attend to different pitch dimensions when perceiving non-native tones by acoustic approaches. The present study investigated which dimensions of pitch, namely, pitch contour, pitch level, together with the position where it occurs in a word, influenced non-native tone perception by tone language (Mandarin) and non-tone language (Dutch) listeners at the phonological level. A sequence-recall task with memory load and high phonetic variability was applied in the study. Language specific perceptual patterns were found for the two groups. Mandarin listeners outperformed Dutch listeners on encoding non-native pitch contour and pitch level contrasts on each position. Overall Mandarin listeners' perception of non-native tones was independent on the position. However, they needed contextual tonal references when encoding non-native pitch level contrasts. Dutch listeners showed perceptual difficulties in encoding pitch contrasts phonologically and were found partially "tone deaf" due to the lack of representations of contrastive tonal categories in their native language. They showed an overall preference for the word final position than word initial and word middle position when perceiving non-native tones.

Index Terms: cross-linguistic tone perception, phonological perception, pitch dimensions

1. Introduction

Languages show much variation in how meaningful pitch contrasts are signaled phonetically: by pitch levels (high versus low), pitch contours (rise versus fall), or positions (earlier or later in a word). In Mandarin Chinese, a typical lexical tone language, word meaning changes as pitch pattern (level or contour) changes. Previous studies have shown that Mandarin listeners form phonologically contrastive categories of native tones [1] [2] [3]. Compared with non-tone language listeners, they were found to have an advantage when perceiving non-native tonal contrasts (Thai tones), due to the important function of pitch in their native language [4].

Different from Mandarin where pitch variations alone can determine word meaning, pitch in Dutch serves as one of the acoustic correlates of word stress. For instance, *VOORnaam* "forename" and *voorNAAM* "distinguished" differ in terms of the position where the prominent syllable occurs, carrying higher pitch (H tone) among other acoustic cues such as longer duration and greater loudness [5]. However, such pitch marking only occurs in nuclear position in an intonation contour. Presumably due to a rich inventory of intonation types and the occurrence of word stress in the native language,

the Dutch listeners were found to be able to detect non-native tonal differences, but they failed to construct contrastive tone categories [3].

Previous studies have revealed that tone language listeners and non-tone language listeners show different preference for different pitch dimensions. For instance, tone language listeners such as Mandarin listeners assigned more weight to pitch contour contrasts, while non-tone language listeners such as English listeners attended more to pitch level changes when perceiving non-native tones [6]. However, these studies with regard to non-native tone perception have mainly investigated pitch processing at the acoustic level using AX/ABX task or multidimensional scaling method. Little is known about non-native tone perception at the phonological level. Non-tone language listeners such as Dutch listeners were found to be sensitive to some tonal contrasts acoustically [3], making it interesting to investigate whether they are able to maintain such sensitivity when encoding tonal contrasts at a more abstract level, the phonological level. Furthermore, they were found sensitive to the position of lexical stress [7] [8], which leads to the question whether their perception of other suprasegmental information, such as pitch contrasts, depends on the position. The present study attempts to investigate which pitch dimension, together with the position where the pitch contrasts occur, affect non-native tone perception at the phonological level by tone language (Mandarin) and non-tone language (Dutch) listeners.

2. Method

2.1. Participants

Thirty native speakers of Dutch (mean age: 24 years old, standard deviation (SD): 5 years, 5 males) and thirty native speakers of Mandarin (mean age: 24 years old, SD: 5 years, 9 males) participated in the study. All participants reported normal hearing without language impairment. All Mandarin participants speak Mandarin or a northern dialect of Mandarin Chinese as their native languages. None of the Dutch participants had exposure or knowledge of Mandarin or any other tone or pitch accent languages. None of the participants had received professional musical training for more than 3 years.

2.2. Stimuli

Cantonese tones were used as the stimuli in that it has a rich inventory of pitch level and pitch contour contrasts, as shown in Figure 1.

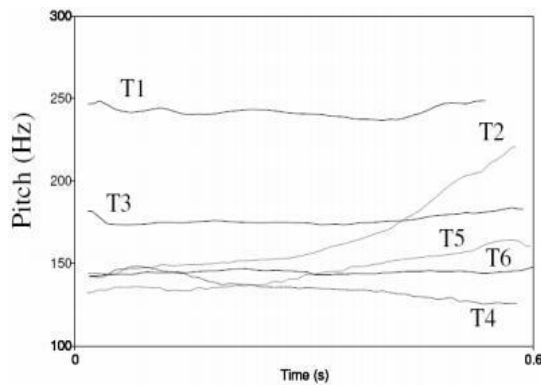


Figure 1. *Pitch patterns in Cantonese tones* [9]

Three contrasts were used as stimuli: T3 vs. T6 (mid level vs. low level tone, level contrast), T4 vs. T6 (low falling vs. low level tone, contour vs. level contrast) and T4 vs. T5 (low falling vs. low rising tone, contour contrast). T1 (high level tone) and T2 (high rising tone) were eliminated in that most pairs with T1 and T2 were salient to distinguish from due to their distinctive acoustic spaces [10].

Monosyllables of /ba/, /ta/, /ga/, /ka/, /da/, /pa/, /gu/, /ku/, /bu/, /bi/, /pi/ were produced for six times by three female and three male native speakers of Cantonese carrying each of the lexical tone mentioned above. For each speaker, three items of the best quality of each tonal syllable were selected, and each syllable was manipulated to 400 ms. The monosyllables were concatenated to the trisyllabic non-words, with an interval of 25 ms between each syllable in a word, making the duration of each token 1250 ms. The loudness of all the tokens were manipulated as 70 db. T3 (mid level tone) was used as the companion tone in the tri-syllabic nonwords in tonal contrasts. Tri-syllabic non-words in segmental contrast only used T3 to eliminate any tonal differences. All the contrasts occurred in word-initial, word-middle and word-final positions, respectively, as shown in Table 1. Segmental contrasts were not only used as a baseline in comparison to the tonal contrasts but also served as a baseline for memory ability.

Table 1: *Stimuli*.

Contrast	Pair	Token (carrying tonal pattern)	Position
Segmental	1	tapibu-gapibu (T3T3T3-T3T3T3)	initial
	2	gutapi-gukapi (T3T3T3-T3T3T3)	middle
	3	pibuda-pibuga (T3T3T3-T3T3T3)	final
T4 vs. T5	4	budapi (T4T3T3-T5T3T3)	initial
	5	pabigu (T3T4T3-T3T5T3)	middle
	6	kupiga (T3T3T4-T3T3T5)	final
T4 vs. T6	7	tabupi (T4T3T3-T6T3T3)	initial
	8	gubapi (T3T4T3-T3T6T3)	middle
	9	bidagu (T3T3T4-T3T3T6)	final
T3 vs. T6	10	pibaku (T3T3T3-T6T3T3)	initial
	11	kapubi (T3T6T3-T3T3T3)	middle
	12	bigupa (T3T3T3-T3T3T6)	final

2.3. Procedure

The experiment was adapted from the sequence-recall task from [11]. The experiment involved 12 pairs of non-words in total, each pair taking 7-8 minutes. The participants were required to come to the phonetic lab at UiL-OTS twice. They were instructed that they were going to learn six pairs of new words each time. The learning order of the 12 pairs was counterbalanced across participants. However, all the participants learned one of the three segmental contrasts at first so that they could get an idea of the experiment.

For each pair, there were 5 phases. The participants were instructed that they were going to learn new word A and B only differing in “melody”, associated with button A and B, respectively. In Phase 1, they firstly listened to 12 tokens of word A and then 12 tokens of word B. The 12 items were produced by six speakers (2 items per speaker). After they learned word A and B, Phase 2 started, where they could press button A or B to hear the words repeatedly as many times as they wished so that they could make sure they memorized the two words. In Phase 3, they did a practice where they should respond to the word they heard by pressing the corresponding button. There were 12 trials in the practice, in which they were required to reach criteria of making 7 correct responses in a row. If they failed in the practice, they would go back to Phase 2 to learn word A and B again until they could pass Phase 3. After they reached the criteria, they moved on to Phase 4 where they listened to word A and B in a two-word, three-word sequence and recalled the sequence by pressing the corresponding buttons. For instance, if they heard A-B, they should press button A firstly and B secondly in an A-B order. There were four trials in Phase 4, two trials of two-word sequence, and two trials of three-word sequence. A feedback would appear in the screen after they responded. Phase 5 was the test phase, containing 3 blocks. Block 1, 2 and 3 were words A and B in two-word, three-word and four-word sequence, respectively. Each block had 8 trials. No feedback was given in the test phase.

3. Analysis & Results

A Generalized Linear Mixed Model (GLMM) was conducted in SPSS 25. Contrast (4 levels), Position (3 levels) and Group (2 levels) were taken as fixed factors into the model. Contrast ($F(3, 17.26) = 43.70, p < 0.001$), Position ($F(2, 17.26) = 7.08, p = 0.001$) and Group ($F(1, 17.26) = 101.89, p < 0.001$) had significant main effect. A three-way interaction among Contrast, Position and Group ($F(17, 17.26) = 21.85, p < 0.001$) was significant as well. Mandarin listeners outperformed Dutch listeners in all contrasts and in all positions, as shown in Figure 2 and 3.

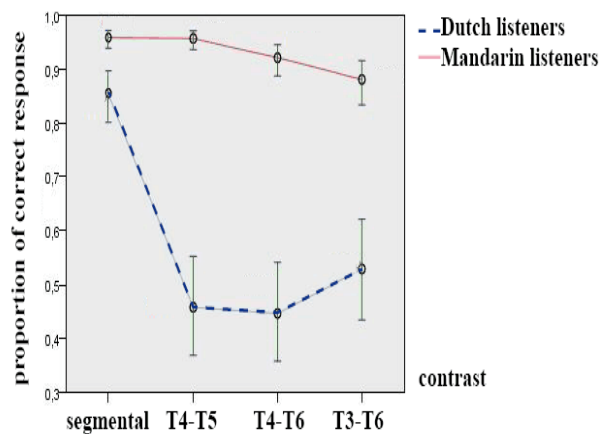


Figure 2. Dutch and Mandarin listeners' performance on each contrast

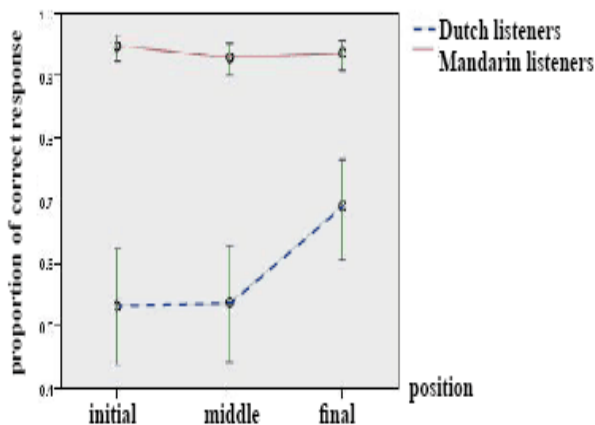


Figure 3. Dutch and Mandarin listeners' performance at each position

A separate GLMM analysis was conducted for each language group. For Dutch listeners, both the fixed factors Contrast $F(3, 8.63) = 89.13, p < 0.001$ and Position $F(2, 8.63) = 23.41, p < 0.001$ were significant, indicating Dutch listeners responded to both contrasts and positions differently. Contrast and Position had a significant interaction $F(6, 8.63) = 7.39, p < 0.001$. Post-hoc pair-wise analysis showed that Dutch listeners had a significant better perception in segmental contrast than in pitch contrasts $F(3, 8.63) = 55.71, p < 0.001$, indicating that they had more difficulties encoding contrastive pitch than segmental contrast phonologically. Moreover, regarding position, Dutch listeners had significantly better perception, regardless of tonal or segmental, occurred on word-final position than on word-initial $F(2, 8.63) = 23.55, p < 0.001$ and word-middle positions $F(2, 8.63) = 23.55, p < 0.001$. Particularly, when perceiving T4 vs. T6, they achieved significantly better performance (over 70%) on the word final position than on the word initial $F(2, 8.63) = 47.11, p < 0.001$ and word middle positions $F(2, 8.63) = 47.11, p < 0.001$, as shown in Figure 4.

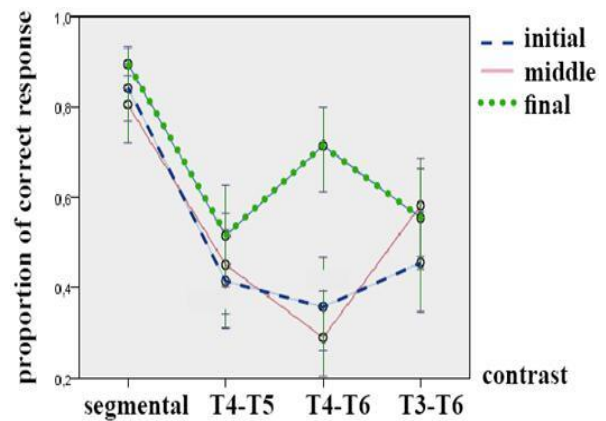


Figure 4. Dutch listeners' performance

For Mandarin listeners, Contrast was found significant $F(3, 8.63) = 17.79, p < 0.001$, indicating that they responded differently to contrasts. A post-hoc pair-wise analysis showed that Mandarin listeners performed better in encoding contour contrasts (T4 vs. T5) than in contour vs. level (T4 vs. T6) $F(3, 8.64) = 7.79, p = 0.011$ and pitch level contrast (T3 vs. T6) $F(3, 8.64) = 7.79, p < 0.001$. Their perception in segmental contrast was significantly better than that in T4 vs. T6 $F(3, 8.64) = 7.79, p = 0.007$ and T3 vs. T6 $F(3, 8.64) = 7.79, p < 0.001$. However, Mandarin listeners perceived T4 vs. T5 as well as they did in segmental contrast ($p = 0.735$), with an accuracy at approximately 96%. Position was insignificant $F(2, 8.63) = 2.05, p = 0.13$, indicating that overall, position didn't play a role in Mandarin listeners' perception. However, a significant interaction between Contrast and Position was found $F(6, 8.63) = 12.29, p < 0.001$, implying that Mandarin listeners' perception on some contrast depended on some position. Post hoc pair-wise analysis showed that when perceiving pitch level contrasts (T3 vs. T6), Mandarin listeners had significant better perception when the contrast occurred at word middle position, compared with either word initial $F(2, 8.63) = 10.042, p = 0.006$ or word final position $F(2, 8.63) = 10.042, p < 0.001$, as shown in Figure 5.

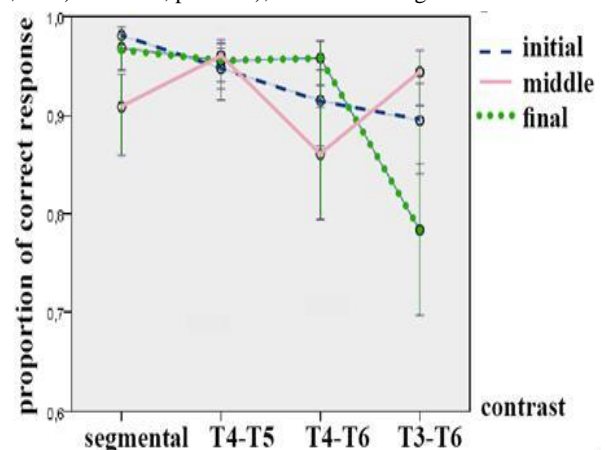


Figure 5. Mandarin listeners' performance

4. Discussion

In general, Mandarin listeners showed better performance than Dutch listeners in encoding all pitch contrasts, which could be explained by the function of pitch used in their native languages. According to the Feature Hypothesis [4] [12], the

more prominent a certain phonetic or phonological dimension is in the native language, the easier it might be to learn to discern and use that dimension for non-native phonological processing. Pitch variations distinguish lexical meaning in Mandarin while pitch is one of acoustic correlates in lexical stress in Dutch. Having pitch at the word level in Mandarin may enable Mandarin listeners to have an advantage in processing non-native pitch contrasts, compared with Dutch listeners.

Compared with encoding segmental contrasts (nearly 90% accuracy), Dutch listeners had significant worse performance on processing pitch contrasts (around 50% accuracy on average), indicating that they had perceptual problems in encoding pitch contrasts phonologically. They showed better performance when the pitch contrasts occurred at the final position than at the initial and middle positions. This could be due to the recency effect, in which listeners might be more attentive to the offsets than to the onsets [13]. In particular, when perceiving pitch contour vs. level contrast (T4 vs. T6), Dutch listeners achieved better performance (over 70% accuracy) in word final position than in word initial and word middle position, indicating that they seemed to be able to encode pitch contour vs. pitch level contrast though they were confused by pitch contour contrast (T4-T5) and pitch level contrast (T3-T6). This could be due to the influence of the inventory of intonation contours in Dutch. According to Perceptual Assimilation Model (PAM) [14], when T4-T6 contrast was on the final position, Dutch listeners may map the overall pitch pattern T3-T3-T4 onto a nuclear pitch accent H* followed by the final boundary tone L% in intonation category, making it distinctive from the overall pitch level pattern T3-T3-T6.

Mandarin listeners were able to encode non-native pitch contrasts phonologically. Position overall didn't play a role in their perception. They achieved over 90% accuracy in perceiving pitch contour contrast (T4 vs. T5) and pitch contour vs. pitch level contrast (T4 vs. T6). In particular, they perceived the pitch contour contrast as well as the segmental contrast, both with an accuracy of around 96%. This can be accounted for by PAM [14], in which they may map the non-native tones T4, T5 and T6 onto the phonological tonal categories, falling tone, rising tone and level tone, respectively, in their native language. However, when perceiving pitch level contrast (T3 vs. T6), which is absent in their native language, their performance was significantly improved when the contrast was surrounded in a context (in the middle position). It could be in that lacking pitch level categories in their native language requires tonal references for them to perceive non-native pitch level contrasts correctly.

The current findings have interesting implications for research on cross-linguistic pitch processing of non-native lexical tones at the phonological level. Future study may involve pitch-accent languages such as Japanese for a more comprehensive view on how phonological knowledge is used in pitch processing.

5. References

- [1] P. A. Hallé, Y. Chang and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners", *Journal of Phonetics*, vol. 32, pp. 395-421, 2004.
- [2] Y. Xu, J. T. Gandour and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical

perception of pitch direction.", *Journal of Acoustic Society of America*, vol. 120, no.2, pp. 1063-1074, 2006.

- [3] L. Liu, A. Chen, and R. W. J. Kager, "Perception of tones in Mandarin and Dutch listeners.", *Language and Linguistics*, vol. 18, no.4, pp. 622-646, 2017.
- [4] V. Schaefer and I. Darcy, "Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones.", *Laboratory Phonology*, vol. 5, no. 4, pp. 489-522, 2014.
- [5] C. Gussenhoven, "*The Phonology of Tone and Intonation*", Cambridge University Press, 2004.
- [6] J. Gandour, "Tone perception in Far Eastern languages.", *Journal of Phonetics*, 11, pp.49-175, 1983.
- [7] A. Cutler and W. van Donselaar, "Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch", *Language and Speech*, vol. 44, pp.171-195, 2001.
- [8] Heuven, V. J. van.1988. Effects of stress and accent on the human recognition of word fragments in Spoken context: Gating and shadowing, *Proceedings of Speech 88, 7th FASE symposium*, 1988, pp. 811-818. Edinburgh.
- [9] P. Mok, D. Zuo, D and P. Wong, "Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese", *Language Variation and Change*, vol. 25, pp. 341-370, 2013.
- [10] Z. Qin, and P. Mok, "Discrimination of Cantonese Tones by Speakers of Tone and Non-tone Languages.", *Kansas working papers in Linguistics* 34, 2014.
- [11] E. Dupoux, S. Peperkamp and N. Sebastian-Galles, "A robust method to study stress 'deafness' ", *Journal of the Acoustical Society of America*, vol.110, pp. 1606-1618, 2001.
- [12] R. McAllister, J.E. Flege and T. Piske, T, "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian", *Journal of Phonetics*, vol.30, pp.229-258, 2002.
- [13] L. Demany, and C. Semal, "The role of memory in auditory perception" in *Auditory perception of sound resources*, pp 77-113, 2007.
- [14] C. K. So and C. T. Best, "Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences", *Language and Speech*, vol. 53, pp. 273-293, 2010.