THE ROLE OF PITCH DIMENSIONS IN NON-WORD LEARNING BY DUTCH AND MANDARIN LISTENERS

Shuangshuang Hu[1], Ao Chen[1,2], René Kager[1]

[1] Utrecht Institute of Linguistics, Utrecht University, The Netherlands
[2]School of Communication Science, Beijing Language and Culture University, China
s.hu1@uu.nl, a.o.chen@uu.nl, r.w.j.kager@uu.nl

ABSTRACT

Dutch listeners are known to employ positional stress for word recognition, while Mandarin listeners use pitch contour contrasts. The present study investigated the influence of pitch dimensions, namely, pitch level, pitch contour, together with pitch position in non-word learning by Dutch and Mandarin listeners. Both groups learned to identify disyllabic pseudo-words differing only in pitch dimensions in a picture selection task. Language specific perceptual patterns were found. Mandarin listeners were found to be able to encode non-native pitch contrasts for word identification, regardless of pitch position. They showed a preference for contour contrasts to level contrasts. Compared to Mandarin listeners, Dutch listeners encountered difficulties. Still, they showed a better performance when pitch contrasts occurred word-finally than -initially, which could be due to recency effect. The findings suggest the influence of phonological representations in the native language on mapping sound to meaning.

Keywords: Cross-linguistic perception, pitch dimensions, non-word learning.

1. INTRODUCTION

Languages show 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), entailing that our speech perception systems are selectively sensitive to pitch contrasts, depending on our native languages. For instance, in Mandarin Chinese, a typical lexical tone language, word meaning changes as pitch pattern changes. It is shown that Mandarin listeners form phonologically contrastive categories for native tones [1, 2]. The important function of lexical pitch benefits Mandarin listeners to have advantage over non-tone language listeners when perceiving non-native tones (Thai tones) [3]. They were found to be more sensitive to non-native pitch contour contrasts than level contrasts such as in Thai tones and Cantonese tones [4, 5].

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 only in terms of the position of the prominent syllable, carrying higher pitch (H tone) among other acoustic cues (longer duration and greater loudness) [6]. However, such pitch marking only occurs in nuclear position in a statement intonation contour. Presumably due to a rich inventory of nuclear tones in intonation and the occurrence of word stress in the native language, Dutch listeners were found to be able to discriminate some tonal contrasts in Mandarin, but they perceived non-native tones in a psychoacoustic manner [7].

Studies that show differences in sensitivity to pitch contrasts by Dutch and Mandarin listeners have been mainly conducted at the acoustic level [7, 8]. It is of interest to ask whether such different sensitivity would be retained or mediated at a higher cognitive level, word learning. Previous studies on word learning with non-native listeners are mostly focused on either the learning of phonetic features [9, 10] or the learning of suprasegmentals such as tones in nonlexical contexts [11, 12]. Not too much is known about the learning of mapping pitch dimensions to meaning. Moreover, Dutch listeners were found to exploit stress location for word recognition [13, 14], leading us to ask whether such sensitivity to positional marking would be employed in encoding pitch contrasts in learning words. Thus, the present study aims to investigate whether listeners from different language backgrounds, i.e., Mandarin listeners and Dutch listeners, are able to employ three pitch dimensions, namely, pitch contour, pitch level, together with pitch position in sound-to-meaning mappings. It attempts to investigate how native phonological knowledge is used in phonetic-phonological-lexical bridge in terms of different models.

2. METHOD

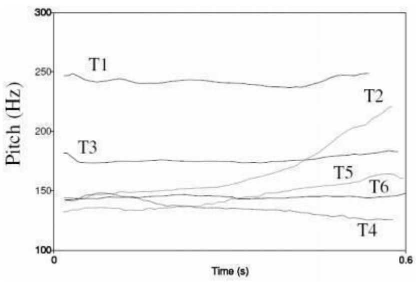
2.1. Subject

31 Dutch listeners (mean age: 23 years old, standard deviation (SD): 4 years, 9 males) and 27 Mandarin listeners (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 spoke Mandarin or a northern dialect of Mandarin Chinese as their native languages and have never been exposed to Cantonese. None of the Dutch participants had exposure or knowledge of Mandarin or any other tone or pitch accent languages. All the participants were non-musicians.

2.2. Stimuli

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

Figure 1: Pitch patterns of Cantonese tones [15].



Two contour tones, T4 (low falling) and T5 (low rising), and two level tones, T3 (mid level) and T6 (low level) were selected. T1 (high level) and T2 (high rising) were not used since these tones are highly salient to discriminate due to their distinctive acoustic spaces from the other tones [5].

Two monosyllables /ku/ and /pi/ carrying each of the lexical tones mentioned above were produced six times by a female native speaker of Cantonese. Three tokens of the best quality of each tonal syllable were selected and manipulated to have equal durations of 400 ms. The monosyllables were concatenated to the disyllabic non-word /kupi/, with a 25 ms interval between the syllables. T3 was used as a companion tone. In total, /kupi/ carried seven pitch patterns: T4T3, T5T3, T6T3, T3T4, T3T5, T3T6, T3T3, contrasting either in tonal pattern or in position, as shown in Table 1.

Table 1: Stimuli.

word /kupi/	Pitch pattern	Tonal contrasts	Positional contrasts
Word 1	T4T3	T4T3 vs. T5T3	T4T3 vs. T3T4
Word 2	T5T3	T4T3 vs. T6T3	T5T3 vs. T3T5
Word 3	T6T3	T6T3 vs. T3T3	T6T3 vs. T3T6
Word 4	T3T4	T3T4 vs. T3T5	
Word 5	T3T5	T3T4 vs. T3T6	
Word 6	T3T6	T3T6 vs. T3T3	
Word 7	T3T3		

2.3. Procedure

The experiment was programmed and conducted in ZEP [16] on an experiment laptop. It contained three phases: a learning phase, a practice phase and a test phase. In the learning phase, participants were instructed to learn seven "new words" (non-words) that differed only in "melody". Each non-word was represented by a picture of a phantasy object. They were instructed to learn the words and the paired pictures together. Seven buttons, from 1 to 7, representing the seven words, were shown on the screen. Each time they clicked on a button, the corresponding picture popped up. They could listen to the seven words as many times as they wanted until they were certain that they had learned the words with the paired pictures and were ready to proceed to the practice.

In the practice phase, the participants heard a word while simultaneously seeing four pictures on the screen. The four pictures contained one target picture, one pitch contrast competitor, one positional competitor and one contrast-unrelated distractor. The participants were required to choose the correct picture that corresponded to the word they heard by pressing the corresponding button on a four-button button box. Feedback on correctness of the response appeared right after they responded. There were seven trials in the practice phase. After participants made a response, they automatically proceeded to the next trial. After the practice phase, participants could either proceed to the test phase or go back to the learning phase to learn the words again if they wished.

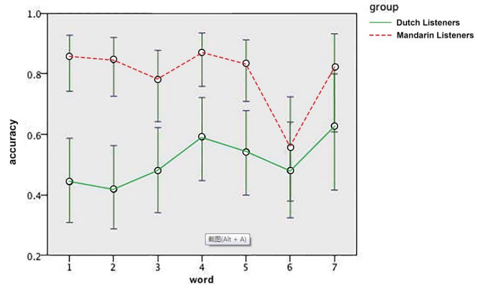
The procedure in the test phase was the same as in the practice phase. However, no feedback was provided. The seven non-words were tested three times in a randomized order, resulting in 21 trials in total.

A working memory test, as a baseline for memory capacity, was conducted immediately after the experiment. Participants listened to two sets of numbers in the native language from a sequence of 2 digits to 8 digits. They were required to repeat the numbers they heard in reversed order. The same numbers in each set were recorded by native speakers of Mandarin and Dutch, respectively. The interval between each number was 1s.

3. RESULTS

To compare the performance on word identification between the two groups, a generalized Linear Mixed Model (GLMM) was conducted in SPSS 25. Word (7 levels) and Group (2 levels) were taken as fixed factors into the model. Both Word ($F(6, 1.030) = 2.722$, $p=0.013$) and Group ($F(1, 1.030) = 17.699$, $p<0.001$) had significant main effect. Group had a significant interaction with Word ($F(6, 1.030) = 2.115$, $p=0.049$). As shown in Figure 2, Mandarin listeners showed significantly better performance than Dutch listeners in learning all the words (all p values <0.05) except for Word 6 (T3T6) ($F(1, 1.030) = 0.404$, $p=0.525$) and Word 7 (T3T3) ($F(1, 1.030) = 2.191$, $p=0.139$). For Mandarin listeners, Word ($F(6, 560) = 3.353$, $p=0.003$) was found significant. Mandarin listeners showed a significant disadvantage in identifying Word 6 (T3T6), with an accuracy of 56.1%, compared to their performance of other words (all p values <0.05). For Dutch listeners, Word ($F(6, 644) = 1.356$, $p=0.230$) had no significant main effect. To take a closer look at the errors they made, Mandarin listeners showed an error rate of 70.2% of identifying Word 6 (T3T6) as Word 7 (T3T3), while Dutch listeners had no error preference for all the words.

Figure 2: Dutch and Mandarin listeners' performance on each word (accuracy)

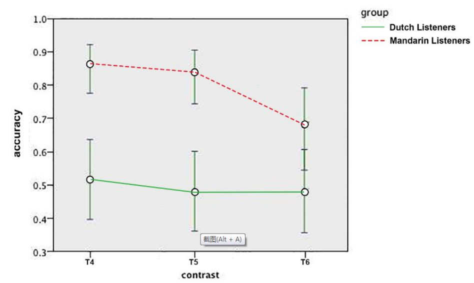


To investigate the influence of position on the performance between the two groups, the words were categorized into 3 contrasts (words carrying T4, T5 and T6). Note that T3T3 was not taken into the contrast due to the same tone on both positions. Contrast (3 levels), Position (2 levels: initial vs. final), Group (2 levels) were taken as fixed factors in GLMM. Contrast ($F(2, 1.030) = 3.917$, $p=0.009$) and Group ($F(1, 1.030) = 17.635$, $p<0.001$) were found significant. Position had no significant main effect ($F(1, 1.030) = 0.004$, $p=0.950$). Interaction between Contrast and Group ($F(2, 1.030) = 2.653$, $p=0.045$), and Position and Group ($F(1, 1.030) = 5.089$, $p=0.024$) were found significant. No three-way interaction among Contrast, Position and Group ($F(4, 1.030) = 1.458$, $p=0.213$) was found.

A separate GLMM analysis was conducted for each language group. For Mandarin listeners, Contrast ($F(2, 560) = 4.794$, $p=0.003$), but not Position ($F(1, 560) = 2.340$, $p=0.127$), were found significant. There was not significant interaction between Contrast and Position ($F(2, 560) = 2.470$, $p=0.086$). Mandarin listeners showed a significant better performance on words carrying contour tones T4 ($F(2, 560) = 3.111$, $p=0.016$) and T5 ($F(2, 560) = 3.111$, $p=0.041$) than words carrying level tone (T6), as shown in Figure 3. For Dutch listeners, Position ($F(1, 644) = 4.136$, $p=0.042$), but not Contrast ($F(2, 644)=0.432$, $p=0.730$) was found significant. They showed a better performance on word-final (with an accuracy of 56.3%, than on word-initial (with an accuracy of 44.9%) ($F(1, 644) = 5.456$, $p=0.021$). There was no significant interaction between Contrast and Position ($F(2, 644))= 0.859$, $p=0.424$).

In addition, working memory test was analyzed using an Independent T-test in SPSS 25. Mandarin listeners (with an accuracy of 94.97%) outperformed Dutch listeners (with an accuracy of 82.21%) ($F(1, 56)= 1.771$, $p<0.001$). This could be due to the fact that some of the randomized numbers used in the test are monosyllables in Mandarin but are disyllables in Dutch, which may cost more cognitive load.

Figure 3: Dutch and Mandarin listeners' performance on each tone (accuracy)



4. DISCUSSION

In general, Mandarin listeners achieved an overall success in learning non-words, while Dutch listeners encountered difficulties. This can be explained in terms of the function of pitch by the Feature Hypothesis [17] predicting that the more prominent a certain phonetic or phonological dimension (pitch) is in the native language, the easier it will be to learn to discern and use that dimension for non-native phonological processing of pitch. Pitch variations alone can distinguish lexical meanings in Mandarin while pitch is only one of the acoustic correlates of lexical stress in Dutch. Contrastive pitch at the word level in Mandarin may enable Mandarin listeners to encode non-native pitch contrasts more easily than Dutch listeners.

Mandarin listeners were able to learn non-words regardless of pitch position. It is the pitch contrast solely that determined their learning of non-words. Mandarin listeners, specifically, manifested strength in learning words contrastive in pitch contour (T4 vs. T5) and pitch contour versus pitch level (T4 vs. T6), with an overall accuracy of 85%. However, compared to their performance on contrasts involving pitch contour, they were vulnerable in learning words contrastive in pitch level (T6T3, T3T3). Different from Mandarin listeners, Dutch listeners did not show any preferred dimension for pitch contrasts. They were unable to encode any pitch contrasts (with an overall accuracy of 50%, around chance level), regardless of contour or level, in non-word learning. Still, they showed a better performance when the tone occurred word-finally (with an accuracy of 56.3%) than -initially (with an accuracy of 44.9%), suggesting a preference for word-final over word-initial position. This could be due to recency effect [18], rather than a transfer from the native language, where stress is dominantly prefinal.

The findings of the different patterns between the two groups can be accounted for by several models. According to Perceptual Assimilation Model (PAM) [19], Mandarin listeners may map non-native T4 vs. T5, T4 vs. T6 onto T4 (falling) vs. T2 (rising), and T4 (falling) vs. T1 (level) tone, respectively, in the native tonal categories, which led to a good discrimination between these contrasts. They misidentified T3T6 as T3T3, which suggests that they regarded T6 and T3 as two allotones of T1 (high level tone) in Mandarin, which resulted in poor discrimination. However, it seems that PAM fails to explain the findings of Dutch listeners' performance on non-word learning. PAM predicts that contrastive pitch patterns should be perceived to be dissimilar from the native language and should not be mapped to the native phonological category since Dutch does not use contrastive pitch lexically, which would result in good discrimination. However, this contradicts the findings.

Another account can be based on models in terms of the influence of native phonology [20], predicting that if the native grammar lacks the phonological feature that differentiates a particular non-native contrast, listeners should be unable to perceive the contrast and hence have trouble to acquire the novel phonological representations. For Mandarin listeners, the presence of phonological tonal representations in the native language facilitates the learning of non-native pitch contrasts. Compared to Mandarin listeners, the failure of Dutch listeners in word learning may lie in difficulties to establish phonological representations of non-native pitch contrasts. The lack of phonological representations of lexically contrastive pitch, regardless of pitch level or contour, in the native language may hinder them to link pitch patterns to lexical meaning.

Also based on the influence of native language, [21] proposed a 'multi-store' model that tone and non-tone listeners store long-term memory representations of pitch-based phonological categories. The stored representations in long-term memory influence listeners' perceptual weight given to non-native pitch features. When perceiving novel pitch contrasts, the surface phonetic forms of pitch contrasts are stored in short-term memory [22]. However, in a more cognitively demanding task such as non-word learning, the phonetic forms at the acoustic level may decay, and the phonological representations stored in the long-term memory may kick in. Mandarin listeners store long-term memory representations of tonal categories, i.e. categories of contrastive contour tones, but not contrastive level tones. Such phonological representations in the long-term memory may help them to encode non-native pitch contour contrasts in word learning. Unlike Mandarin listeners, Dutch listeners may lack long-term memory representations of lexical pitch contrasts, which adds to their difficulties in mapping pitch contrasts to lexical meaning.

The current findings have shown the influence of native language on the availability of pitch dimensions for sound-to-meaning mappings by tone and non-tone language listeners.

5. REFERENCES

[1] Wang, W. S.-Y. 1976. Language change. *Annals of the New York Academy of Sciences* 28, 61-72

[2] Hallé, P. A., Chang, Y., Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics 32*(3), 395-421.

[3] Schaefer, V., Darcy, I. 2014. Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology* 5 (4), 489-522.

[4] Gandour, J. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11, 49-175.

[5] Qin, Z., Mok, P., 2014. Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics* 34, 1–25.

[6] Gussenhoven, C. 2003. *The Phonology of Tone and Intonation*. Cambridge University Press.

[7] Liu, L., Chen, A., Kager, R. 2017. Perception of tones in Mandarin and Dutch listeners. *Language and Linguistic*s 18(4), 622-646.

[8] Chen, A., Liu, L., Kager, R. 2015. Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences* 48, 62-69

[9] Bradlow, A., Akahane-Yamada, R., Pisoni, D. B., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61, 977–985.

[10] Jamieson, D. G., Morosan, D. E. 1989. Training non-native speech contrasts in adults: Acquisition of the English /ð/–/θ/ by francophones. *Perception and Psychophysics* 40, 205–215.

[11] Gottfried, T. L., Suiter, T. L. 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics* 25, 207–231.

[12] Wang, Y., Spence, M. M., Jongman, A., Sereno, J. A. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106, 3649–3658.

[13] Cutler, A., Donselaar, D. van. 2001. *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44, 171–195.

[14] Heuven, V. J. van. 1988. Effects of stress and accent on the human recognition of word fragments. In Ainsworth, W. A., Holmes, J. N. (Eds.), *Spoken context: Gating and shadowing, Proceedings of Speech 88, 7th FASE symposium Edinburgh*, 811–818.

[15] Mok, P., Zuo, D., Wong, P. 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change* 25, 341-370.

[16] Veenker, T. J. G. 2012. The Zep Experiment Control Application. Utrecht Institute of Linguistics OTS, Utrecht University.

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

[18] Demany, L., Semal, C. 2007. The role of memory in auditory perception. In: Yost, W. A., Popper, A. N., Fay, R. R. (Eds.), *Auditory perception of sound resources*, Springer, 77-113.

[19] So, C. K., Best, C. T. 2010. Cross-language perception of nonnative tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53, 273-293.

[20] Brown, C. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136–193.

[21] Xu, Y., Gandour, J., Francis, A. L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America* 120(2), 1063–1074.

[22] Xu, Y. 1991. Depth of phonological recoding in short-term memory. *Memory and Cognition* 19(3), 263-273.