PERCEPTION OF JAPANESE PITCH ACCENT BY DUTCH LISTENERS: FROM ACOUSTIC TO PHONOLOGICAL PROCESSING

Shuangshuang Hu[1], René Kager[2]

[1] Shanghai Normal University, China [2] Utrecht University, The Netherlands
[1] shuangshuanghu@shnu.edu.cn, [2] r.w.j.kager@uu.nl

ABSTRACT

Both Dutch and Japanese employ positional marking for lexical contrasts, i.e., word stress position in Dutch and pitch accent position in Japanese. The present study investigated whether naive Dutch listeners could transfer a native positional sensitivity to perceive the position of Japanese lexical pitch accents. An AX discrimination task and a modified sequence recall task were applied to examine Dutch listeners' perception of Japanese pitch accents at the acoustic level and the phonological level, respectively. It was found that naive Dutch listeners showed as good performance as Japanese natives when perceiving Japanese pitch accent contrasts at the acoustic level. However, they showed difficulty when perceiving Japanese pitch accent contrasts at the phonological level while the natives remained good performance. The findings suggest that the positional sensitivity in the native language may facilitate the perception of non-native positional contrasts acoustically, but it may not be reliable enough at the phonological level.

Keywords: perception of Japanese pitch accent, phonological processing, positional marking.

1. INTRODUCTION

Languages differ in the use of prosodic cues to contrast lexical meanings. For instance, Dutch uses word stress to differentiate words. One syllable of any multi-syllabic word is typically marked with higher stress than the other syllable(s) [1, 2]. The meanings of words with identical segments can then be signaled by the position of stress, e.g., *VOORnaam* (stress on the first syllable) means "first name" while *voorNAAM* (stress on the second syllable) means "distinguished".

Tokyo Japanese (henceforth Japanese) uses pitch accent for lexical contrast. The meaning of a word in Japanese is determined by the presence or absence of an abrupt pitch fall (termed "pitch accent") and, if present, by the position of pitch accent in any multi-moraic word [3, 4]. For instance, the disyllabic (bimoraic) word "hashi" /haʃi/ signals three lexical meanings: "chopsticks" (initial-accented), "bridge" (final-accented) and "edge" (unaccented).

Prima facie, the two languages employ different word prosodic cues in their lexicons. Nonetheless, they do share commonalities: Japanese and Dutch use the abstract feature of positional marking: position of pitch accent in Japanese and position of stress in Dutch.

Previous research has underscored that Dutch listeners are sensitive to perceive stress position and can exploit positional marking for word recognition in their native languages [5, 6]. It is of interest to ask whether Dutch listeners can transfer a native positional sensitivity to perceive a non-native positional cue, Japanese pitch accent.

Studies on the perception of non-native positional cues have mainly focused on word stress. It is found that stress language listeners can perceive the positional contrasts of a non-native word stress at the acoustic level, i.e., via acoustic approaches such as AX and AXB discrimination tasks. However, they have difficulty to process the non-native stress when the task taps into phonological representation of stress, namely in a sequence recall task which incorporated phonetic variability and memory load [7, 8]. These findings suggest that the ability to perceive positional cues at the acoustic level does not guarantee a success at the phonological level.

Thus, the present study aims to examine whether stress language listeners, Dutch listeners, can generalize their sensitivity to native positional marking to positional contrasts of Japanese pitch accent (including the positional marked vs. unmarked contrast - accented vs. unaccented), not only acoustically but also phonologically. Via a discrimination task and a sequence recall task, respectively, the study attempts to investigate to what extent the native word prosody influences perception of non-native positional marking at the acoustic level and the phonological level in terms of perceptual models.

2. METHOD

2.1. Subject

40 Dutch listeners (mean age: 22 years old, SD=3.8, 16 males) and 36 Tokyo Japanese listeners (mean

age: 24 years old, SD=5.2, 15 males) participated in the two tasks in the study. All the listeners reported normal hearing without language impairment. None of the Dutch participants have been exposed to any tonal or pitch accent languages. All the participants were non-musicians.

2.2. Stimuli

AX discrimination task and sequence recall task were applied to investigate perception of Japanese pitch accent contrasts at the acoustic level and the phonological level, respectively.

2.2.1. AX discrimination task

Given acoustic properties of consonants and vowels [9], /no/ and /jo/ were used for constructing disyllabic nonwords /nono/ and /jojo/. Both the two words do not exist in Dutch or Japanese[i]. According to [10], disyllabic words in Japanese carry three accentual patterns with corresponding surface tonal patterns (apostrophe ’ symbolizes the accent) as shown in Table 1.

accentual contrast (positional contrast)	surface tonal pattern
initial-accented vs. final-accented	H’L vs. LH’
initial accented vs. unaccented	H’L vs. LH
final-accented vs. unaccented	LH’ vs. LH

Table 1: Stimuli.

All tokens were embedded in a carrier sentence with declarative intonation:

kinoo ______ (target nonword) ga kita.

“Yesterday______(target nonword) nominative marker came”.

The nominative marker “ga” serves to differentiate the unaccented and the final-accented word [11]. All sentences were produced by a female Japanese phonetician in Praat in a sound-proof room. The targeted tokens were extracted from the carrier sentences and were normalized as 96 ms, 70 dB.

The overall design of AX discrimination task is: contrast (3) x the identity of X (2: XA, XA) x token set (2: /nono/, /jojo/) x repeat times (2) with fillers (12 AA pairs), containing 36 trials in total. The inter-stimulus interval (ISI) in each trial was 400 ms. Note that the same trials did not show up consecutively.

2.2.2. Sequence recall task

The sequence recall task integrated a memory load and phonetic variability [12]. It aims to eliminate listeners’ reliance on fine-grained acoustic details and taps into phonological representation.

In this modified task, four contrasts were constructed: one segmental contrast, i.e., nonword /nogo/ vs. /nopo/ as the baseline and three accentual contrasts as shown in Table 1, carried by nonword /nono/. All the accentual tokens were produced six times each by three female and three male Japanese phoneticians and were extracted and manipulated in the same fashion as those in AX discrimination task. Segmental tokens were produced in isolation with a flat tone by the same six speakers and normalized as same as the accentual ones. For each speaker, three items of the best quality were selected.

Two-word, three-word and four-word sequences were used, with ISI set to 80 ms as in [12]. In each sequence, the nonwords were produced by different speakers. The order of the voices was counterbalanced over the sequences.

There were eight trials of each sequence length, as follows:

Two-word length: AA, BB, AB, BA, AA, BB, AB, BA

Three-word length: AAA, AAB, ABA, ABB, BBB, BBA, BAB, BBB

Four-word length: AABA, ABAA, ABBA, BAAB, BABB, BBAB, ABAB, BABA

The overall design is: contrast (4) x trials (24), yielding 96 trials in total.

2.3. Procedure

The two tasks were programmed and conducted in ZEP [13] on an experiment laptop.

2.3.1. AX discrimination task

All the participants were instructed to listen to two tokens from a foreign language. They were required to judge as soon as possible whether the two tokens were the same or not by pressing the corresponding buttons “Same” or “Different” on the button box. The trial proceeded only after the participant made the response.

The task consisted of a practice phase (four trials) and a test phase. Feedback was only provided in the practice phase. All the trials were counterbalanced in the test phase.

2.3.2. Sequence recall task

All the participants were instructed that they were going to learn four pairs of new words in a foreign language. They learned the segmental contrast first, and then the three accentual contrasts which were counterbalanced across participants.

Each contrast contained 5 phases. The participants were instructed that they were going to learn new words A and B, associating with buttons

A and B, respectively. In Phase 1, participants first listened to 6 tokens of word A followed by 6 tokens of word B. After learning words A and B, they proceeded to Phase 2 where they could press button A or B to listen to the words repeatedly as many times as they wished to make sure they had memorized the two words. In Phase 3, they did a practice in which they heard a word and judged whether the word was word A or B by pressing the corresponding button. There were 20 trials in the practice, in which they were required to reach a criterion of 75% accuracy. If they failed in the practice, they would go back to Phase 2 to learn the two words again until they could reach the passing criterion for Phase 3. They then proceeded to Phase 4 where they listened to words A and B in two-word and three-word sequences. They were required to recall the sequences 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. After finishing Phase 4, they proceeded to Phase 5, the test phase, containing 3 blocks (8 trials per block). Blocks 1, 2 and 3 were words A and B in two-word, three-word, and four-word sequences, respectively. A feedback message "OK" would appear on the screen, simultaneously presented with a sound message "okay" once they responded to each trial to eliminate the possible use of echoic memory by the participants [14].

3. RESULTS

3.1. AX discrimination task

A correct response made by a participant was marked as "1" while an incorrect response was marked as "0". Any missing response was regarded as an incorrect response. To compare the performance between the two groups, a generalized Linear Mixed Model (GLMM) was conducted in SPSS 26. Contrast (3 levels) and Language (2 levels) were taken as fixed factors and intercepts for participant and item were taken as random effects into the model. Neither Contrast (F (2, 1818)= 2.523, $p=0.051$) nor Language (F (1, 1818) = 0.036, $p=0.849$) were found significant. No interaction between Contrast and Language was found as well (F (2, 1818) = 0.382, $p=0.683$).

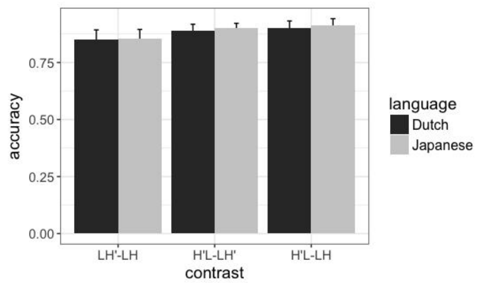

Figure 1: Dutch and Japanese listeners' performance in AX discrimination task.

As shown in Figure 1, Dutch listeners perceived all the three accentual contrasts as well as Japanese natives, both groups reaching around 85% accuracy for LH' vs. LH and around 90% accuracy for H'L vs. LH' and H'L vs. LH.

3.2. Sequence recall task

A response is correct if the entire sequence is recalled correctly. Correct responses were marked as "1" while incorrect responses were marked as "0". Any missing response was regarded as an incorrect response. A GLMM was computed in SPSS 26 to analyze whether the two groups differed in perception of the accentual contrasts at the phonological level. Contrast (4 levels: one segmental contrast and three accentual contrasts), Sequence Length (3 levels: 2-, 3- and 4-word length) and Language (2 levels) were taken as fixed factors, and intercepts for participant and item were taken as random effects into the model. The F-tests showed that Contrast (F (3, 7278) = 74.846, $p < 0.001$) and Language (F (2, 7278) = 20.050, $p < 0.001$) had main effect. The interaction between the two factors was found significant as well (F (3, 7278) = 13.998, $p<0.001$). Sequence Length (F (2, 7278) = 1.919, $p = 0.304$) was not significant, with no interaction with Language (F (2, 7278) = 1.027, $p = 0.358$), or Contrast (F (6, 7278) = 0.725, $p = 0.630$), indicating that the differences among language groups and contrasts were comparable in each sequence length.

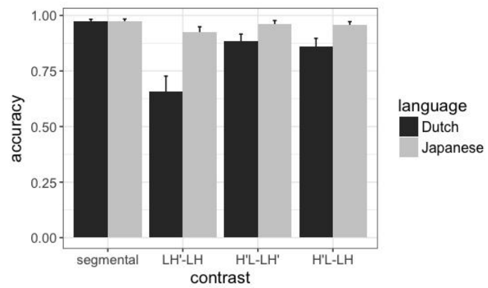

Figure 2: Dutch and Japanese listeners' performance in sequence recall task.

Figure 2 displays the performance of each language group in perceiving segmental contrast (control condition) and accentual contrasts (accentual condition) with sequences of different lengths collapsed. In perceiving segmental contrast, the two groups achieved nearly ceiling effect, both with accuracy around 97%. However, they showed significant differences in perceiving accentual contrasts. Japanese natives outperformed Dutch listeners in perceiving all the accentual contrasts (F (3, 7278) = 25.721, $ps<0.001$).

Dutch listeners performed the best when perceiving segmental contrast, compared with their perception of accentual patterns (F (3, 3834) = 27.739, $ps<0.001$). They achieved around 85% accuracy when perceiving H'L vs. LH' and H'L vs. LH, but their performance of perceiving LH' vs. LH dropping to 65.7% accuracy, significantly worse than that of perceiving the other two contrasts (F (3, 3834) = 27.739, $ps<0.001$). In contrast, Japanese natives remained good performance, with an accuracy around 95% when perceiving all the contrasts. Their perception of accentual contrasts was as good as that of segmental contrast (F (3, 3450) = 1.193, $ps>0.05$).

4. DISCUSSION

In the acoustic task (AX discrimination task), Dutch listeners were found as sensitive as Japanese natives to positional contrasts of Japanese pitch accent. This can be explained by the influence of native phonology [15], which predicts if the native grammar has the phonological feature that differentiates a particular non-native contrast, listeners are able to perceive the contrast. Dutch uses position of word stress, which may benefit them to transfer the native positional sensitivity to perceive Japanese pitch accented contrast initial-accented vs. final-accented (H'L vs. LH'). For accented vs. unaccented contrasts (H'L vs. LH, LH' vs. LH), although Dutch lexicon does not employ positional marked vs. unmarked, Dutch listeners may still distinguish the accented word from the unaccented one because one in the pair is positional marked. Another account is that Dutch listeners may resort to perceive the surface tonal patterns of the accentual contrasts acoustically. According to PAM [16], Dutch listeners may map H'L, LH' and LH onto the native nuclear pitch accents H*L, L*H and L*(H%) [17] in intonation categories, respectively, which led to good perception. However, such perceptual association between the non-native Japanese accentual patterns and Dutch intonation categories is in nature speculative.

When the task highlights the phonological representation of Japanese pitch accent, Dutch listeners' perceptual sensitivity seemed be mediated. They were not as advantageous as Japanese natives who remained good performance. The findings suggest that Dutch listeners' ability of perceiving Japanese pitch accents at the acoustic level does not entail the success at the phonological level. It is notable that Dutch listeners showed significant difficulty, with accuracy around 65%, when processing final-accented vs. unaccented (LH' vs. LH) at the phonological level. If according to the model in terms of the influence of native phonology [15], Dutch listeners should be able to employ the native positional marking to differentiate the marked final-accented from the unaccented word, which contradicted the finding.

Alternatively, the trouble Dutch listeners encountered in processing final-accented vs. unaccented (LH' vs. LH) phonologically could be due to the failure to perceive the surface tonal contrasts realized by the accentual patterns. Phonetically speaking, the H tone in the final accented position in a word is slightly higher than the H tone in the final position when the word is unaccented [11]. Dutch listeners were sensitive to perceive LH' vs. LH at the acoustic level, which could be in that they could rely on fine-grained acoustic information in AX task. Previous studies have reported that they were good at discriminating subtle pitch differences when perceiving non-native lexical tones acoustically [18, 19]. However, when it taps into phonological representation of lexical pitch accent, Dutch listeners may no longer rely on pitch differences. In Dutch lexicon, pitch is one of the acoustic correlates of stress, that is, the stressed syllable in a word is realized by a higher pitch together with longer duration and larger amplitude [20]. Pitch is not the exclusively acoustic cue to word stress in Dutch whereas tonal patterns in Japanese lexicon are realized and determined by the positional cue, the accent. The lack of use of lexical pitch in Dutch might lead to the failure of perceiving Japanese pitch accent contrasts by Dutch listeners at the phonological level.

To sum up, the current findings showed that Dutch listeners were able to perceive the positional contrasts of Japanese pitch accent acoustically. However, they had difficulty when processing the non-native positional cues at the phonological level. This seemed to indicate that when perceiving the positional contrasts of Japanese pitch accent at the phonological level, the abstract positional marking in Dutch may not be reliable enough.

5. REFERENCES

[1] Kager, R. 1989. *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris.

[2] Trommelen, M., Zonneveld, W. 1999. Word stress in English and Dutch. In: Hulst, H. van der (ed), *Word Prosodic Systems of the Languages of Europe*. Berlin: Mouton de Gruyter, 478-515.

[3] Poser, W. J. 1984. *The Phonetics and Phonology of Tone and Intonation in Japanese*. Ph.D. dissertation. Massachusetts Institute of Technology.

[4] Kubozono, H. 2008. Japanese accent. In: Miyagawa, S., Saito, M. (eds.), *The Oxford Handbook of Japanese Linguistics*. Oxford: Oxford University Press, 165-191.
[5] Cutler, A., van Donselaar, W. 2001. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44, 171–195.
[6] Jongenburger, W. 1996. *The Role of Lexical Stress during Spoken-word Processing*. Ph.D. dissertation, University of Leiden. The Hague: Holland Academic Graphics.
[7] Peperkamp, S., Dupoux, E. 2002. A typological study of stress "deafness". *Laboratory Phonology* 7, 203-240.
[8] Peperkamp, S., Vendelin, I., Dupoux, E. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics* 38, 422-430.
[9] Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
[10] Haraguchi, S. 1999. Accent. In: Tujimura, T. (ed) *The Handbook of Japanese Lingusitics*. Oxford: Blackwell, 1-30.
[11] Vance, T. J. 1995. Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics* 23(4), 487–499.
[12] Dupoux, E., Peperkamp, S., Sebastian-Galle, N. 2001. A robust method to study stress "deafness". *Journal of the Acoustical Society of America* 110 (3), 1606-1618.
[13] Veenker, T. J. G. 2012. The Zep Experiment Control Application. Utrecht Institute of Linguistics OTS, Utrecht University.
[14] Morton, J., Crowder, R., Prussin, H. 1971. Experiments with the stimulus suffix effect. *Journal of Experimental Psychology* 91, 169-190.
[15] Brown, C. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136 – 193.
[16] 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.
[17] Gussenhoven, C. 2005. Transcription of Dutch Intonation. In: S.-A., Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford: Oxford University Press, 118-145.
[18] Leather, J. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In: Leather, J., James, A. (eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech*. Amsterdam: University of Amsterdam, 72-89.
[19] Liu, L., Chen, A., Kager, R. 2017. Perception of tones in Mandarin and Dutch listeners. *Language and Linguistics* 18(4), 622-646.
[20] Sluijter, A. M. C. 1995. *Phonetic Correlates of Stress and Accent*. Holland Institute of Generative Linguistics Dissertations 15. The Hague: Holland Academic Graphics.

[1] Note that the disyllabic word /jojo/ (the approximant [j] and a short [o]) is a non-word in both the languages in that the Japanese word "jojo" in the animation "Jojo's Bizarre Adventure" is /dʒo.dʒo/, not /jojo/. "Jojo" (a toy) in Dutch is /ˈjəʊ.jəʊ/.