Volume 23
Issue 1 *Proceedings of the 40th Annual Penn Linguistics Conference*
Article 5

2017

Text Alignment in Japanese Children’s Song

Sunghye Cho
University of Pennsylvania

Text Alignment in Japanese Children's Song

Abstract

In this paper, I examine the text-setting in 43 Japanese Children's songs, called "warabe uta". By examining the text-tune alignment and the text-meter alignment, I show that accented positions and accentual tonal transitions are more important than the other positions and other tonal transitions in the text-setting in Japanese songs.

Text Alignment in Japanese Children's Song

Sunghye Cho*

1 Introduction

In the literature of prosodic typology, there is an ongoing debate about the nature of pitch accents. Even in Tokyo Japanese, which is one of the most widely cited pitch-accent languages, the pitch accent is analyzed in two distinct ways: an accentual approach and a tonal one (Haraguchi 1977, Hyman 2006, Hyman 2009, Kubozono 2012, McCawley 1968, Pierrehumbert and Beckman 1988, Poser 1984, among others). In the accentual approach (1), an accent marks a prominent position, where a certain melody (i.e., a pitch drop from High (H) to Low (L) in the case of Tokyo Japanese) falls within a word.[1] For example, since the accent falls on the first mora in (1a), *há.si-ga* 'chopstick-NOM', a falling contour (HL) is aligned with the first and the second morae, *ha* and *si* respectively, resulting in a surface HLL tonal melody. However, when there is no accented mora as in (1c), *hasi-ga* 'edge-NOM', only an initial boundary melody, LH, is aligned in the first two morae and the H tone of the boundary melody spreads to the end of the word, resulting in a LHH surface melody without a falling contour. Alternatively, in the tonal approach (2), tones are assumed to be lexical features as in other typical tonal languages, and they are directly aligned with morae without the notion of accent.

(1) Accentual approach
  a. há.si-ga 'chopstick-NOM'
  b. ha.sí-ga 'bridge-NOM'
  c. ha.si-ga 'edge-NOM'

(2) Tonal approach
  a. HA.si-ga 'chopstick-NOM' — H L
  b. ha.SI-ga 'bridge-NOM' — L H L
  c. ha.SI-GA 'edge-NOM' — L H

The accentual approach (1) predicts that an accented mora has a phonological (abstract) prominence similar to stress in English, whereas in the tonal approach (2), the position claimed to be accented is not necessarily prominent. This paper aims to investigate both approaches, asking whether a pitch accent in Tokyo Japanese marks a phonological prominence or not, and to what extent it is similar to traditional tonal languages. If an accent marks a prominent position within a word, this would suggest that the accentual analysis is on the right track. However, if Tokyo Japanese is more similar to other tonal languages without apparent evidence of accent, the tonal approach is a better analysis for Tokyo Japanese.

Examining an abstract prominence in non-stress languages can be challenging, because abstract prominence is not something that can be directly measured on a quantitative scale. For that reason, this paper uses an indirect way of exploring the nature of pitch accent in Tokyo Japanese: text-setting alignments in vocal music. It is well known that stressed syllables are likely to occur at strong beats of the musical meter in vocal music of stress languages, such as English and German (Halle and Lerdahl 1993, Palmer and Kelly 1992, Temperley and Temperley 2013). If an accent in Tokyo Japanese is an abstract prominence similar to stress in stress languages, we would expect that accented morae are likely to align with strong beats of the meter in Japanese vocal music. On the other hand, previous studies on tonal languages like Chinese and Vietnamese show that there is a strong mapping between tonal transitions and musical note transitions (Kirby and Ladd 2016,

---

*I would like to thank Mark Liberman, Eugene Buckley, Robert Ladd, David Temperley, and Jianjing Kuang for their feedback on this project. Also, my thanks go to reviewers and audiences of PLC 40 and LabPhon 15. All errors are my own.

[1]Accented morae are marked with an acute diacritic in (1) and H-toned ones are represented with capital letters in (2).

Wee 2007, Wong and Diehl 2002). For example, Wong and Diehl (2002) find over 90% mapping between tonal transitions and musical note transitions in contemporary Cantonese songs, and similarly, Kirby and Ladd (2016) show that there is about 77% correspondence between musical note transitions and linguistic tone sequences in Vietnamese popular songs. If Tokyo Japanese is a tonal language with a restricted tonal system, we will be able to find a strong mapping between musical and linguistic tonal transitions. In this study, I examine this question with traditional Japanese children's songs.

2 Methods

The songs analyzed in this study are all drawn from a single collection, *Nihonno Warabeuta: Sitsunai Yuugika Hen* (Obara 1932). The book contains many traditional children's songs, called *warabeuta*, from various areas of Japan. *Warabeuta* is a specific type of song that children sing while they are playing traditional games. Since the songs have been orally transmitted for a long time, the composers of the songs are usually unknown and there are subtle variations in both melody and lyrics depending on the region where the songs are sung. For this study, I selected 43 songs sung in the Tokyo area, which include 1,645 morae in total. As there were cases where two morae are aligned with the same musical note (e.g., *rai* in the first measure of Figure 1), the number of tokens is 1,456 mora and musical note pairs.

For the meter annotation of the songs, I classified every mora in 2 ways, firstly by accent type and secondly by metrical strength (beat) of the musical meter. For the accent types, accentedness was marked with either *u* (unaccented) or *a* (accented). Morae in content words were coded as *c* and those in function words were coded as *f*. This created four accent types in total: *ac*, *af*, *uc*, and *uf*. For example, in *hanà-ga* 'flower-NOM' in the third measure in Figure 1, the first mora *ha* is an unaccented one in a content word, it is labeled as *uc*, and the second mora *na*, which is accented, is labeled as *ac*. The last mora, *-ga*, which marks a nominative case, is labeled as *uf*. All unaccented words are defined as either *uc* or *uf*. Words with two or more possible accent positions are excluded from the analysis. For metrical strengths of the beats where each mora is aligned, I adopted the method used in Temperley and Temperley (2013) as follows: The tactus level—the main beat of each time signature, at which most people clap their hands and tap their feet while singing—was defined as strength level 2 (the quarter-note level in 2/4, 3/4, and 4/4 time signatures), and one level above the tactus was level 3 (the one-measure level in 2/4 and 3/4, and the half-measure level in 4/4). The one-measure level in 4/4 was defined as level 4, and all beats below the tactus were given level 1. For example, in a measure with four quarter notes in a 4/4 time signature, the first quarter note receives level 4, the second level 2, the third level 3, and level 2 is assigned to the last one. Figure 1 provides an example of the assignments of metrical levels.

Figure 1: The first line of *hiraita hiraita* '(A flower) bloomed, bloomed' showing the encoding of lyrics, accent types, and metrical levels.

For musical and linguistic tone transitions, I classified the direction of both transitions into three groups: rising, falling, and level. For example, the musical note transition from the first mora *hi* to the second mora *ra* of the second measure in Figure 1 is coded as level (L), since there is no change in musical note from *hi* (A note) to *ra* (A note). The second mora *ra* is classified as falling because there is a falling transition from *ra* (A note) to *i* (G note). Linguistic tone transitions are also computed in a similar way. For instance, since the word *hi.RA.i-ta* 'open-PAST' in the second measure in Figure 1 shows a LHLL surface melody, there is a rising transition from *hi* (L) to *ra* (H), a falling transition from *ra* (H) to *i* (L), and a level transition (no change) from *i* (L) to *ta* (L). There were cases where two morae are aligned to a single musical note as in *rai* of the first measure in Figure 1. In such cases, the linguistic tone of the first mora was used to determine the linguistic tone transition with respect to the preceding mora. The linguistic tone of the second mora was used in deciding the tone transition with respect to the following mora.

3 Results

3.1 Text-Tune Alignment

The first question I examine is whether there is a correspondence between musical and linguistic tonal transitions. This analysis includes all morae in both accented and unaccented words except the final mora of each song because the transition to the following mora cannot be computed. In total, 1,422 morae are included in this analysis. Table 1 shows the rates of the linguistic-musical tone correspondence. Similar motion, where tonal and musical transitions move in similar directions (Kirby and Ladd 2016), is shown in the bold-faced cells, and contrary motion, where tonal and musical transitions move in contrary direction (Kirby and Ladd 2016), is shown in the shaded cells. Oblique motion, where the transitions are neither similar nor contrary, is shown in the remaining cells. In my data, the overall degree of similar motion is 54.15%, that of oblique motion 40.08%, and that of contrary motion 5.77%.

		Musical transition		
		Rising	*Falling*	*Level*
Tonal transition	*Rising*	**170 (49.56%)**	39 (11.37%)	134 (39.07%)
	Falling	43 (13.48%)	**194 (60.81%)**	82 (25.71%)
	Level	177 (23.29%)	177 (23.29%)	**406 (53.42%)**

Table 1: Correspondence between melodic and tonal transitions. Bold-faced cells are similar motion, shaded cells are for contrary motion, and the remaining cells are oblique motion. Percentages are per row. Overall degree of similar motion: 54.15% (770 tokens out of 1422).

The result shows that falling tonal transitions correspond with falling melodic transitions (similar motion) about 60.81% of the time, whereas they are aligned with level melodic transitions about 25.71% of the time and with rising note transitions only 13.48% of the time. I conduct a chi-square goodness-of-fit test to see if this distribution is significantly different from the null hypothesis that a falling tonal transition randomly aligns with any melodic transitions (i.e., chance level: 33.33%). The result suggests that there is a significant difference between the distribution of the data and the one the null hypothesis predicts, confirming that a falling tonal transition is more likely to be aligned with a melodic falling transition than a level or a rising one ($\chi^2 = 115.57$, df = 2, $p < 0.0001$). While a falling tonal transition disfavors both level and rising melodic transitions, it is notable that contrary motion, where the tonal transition falls but the musical transition rises, occurs less frequently (13.48%) than oblique motion, where the tonal transition falls but the musical transition stays level (25.71%).

A similar result is found in the case of level tonal transitions. A chi-square goodness-of-fit test reveals that a level tonal transition is not randomly aligned, confirming the result that it favors the alignment with a level note transition (53.42%) more than with a rising or falling one ($\chi^2 = 138$, df = 2, $p < 0.0001$). However, unlike the case of falling tonal transitions, where contrary motion is more disfavored than oblique motion, level tonal transitions only show oblique motion, because the tonal transition stays level regardless of the melodic transition. Both rising and falling note transitions occur with the same frequency (23.29% for both falling and rising melodic transitions), suggesting that neither case is disfavored compared to the other, while both instances of oblique motion occur less frequently than level melodic transitions, which are in similar motion.

Lastly, rising tonal transitions also show a similar trend to falling transitions in that contrary motion is very disfavored (11.37%). However, a tonal rising transition is aligned with a melodic rising transition only about 50% of the time, not much different from the correspondence with a melodic level transition (about 40%). A chi-square goodness-of-fit test still finds that the alignment of a rising tonal transition is significantly different from that of the null hypothesis (χ^2 = 80.12, df = 2, $p < 0.0001$), but this result seems to be due to the large deviation of contrary motion ($|r|$ = -7.04). Lastly, it is notable that the correspondence between tonal and melodic transitions in Tokyo Japanese is not as strong as the ones found in other tonal languages, such as Cantonese and Vietnamese.

3.2 Text-Tune Alignment in Accent Positions

A falling tonal transition in Tokyo Japanese is found in two cases as in (3): one from an accented mora (H) to the following one (L) in accented words as in (3a), and the other from the final H boundary tone of unaccented words to the initial L of the LH boundary tones of the following word as in (3b) (unless the following word has a word-initial accent). I will refer to cases such as (3a) as accentual falling transitions and cases like (3b) as non-accentual falling transitions.

(3) Examples of falling tonal transitions in Tokyo Japanese
  a. Accentual: ha.SI-ga (L**HL**) 'bridge-NOM'
  b. Non-accentual: ha.NA-GA i.TA.i (LH**H L**HL) 'nose-NOM painful'

Here, I examine the falling tonal transitions to see if there is a difference between accentual falling transitions (3a) and non-accentual falling transitions (3b) in the text-tune alignment. This analysis examines only the morae that are aligned with a falling tonal transition. Table 2 shows the percentages of melodic transitions with accentual and non-accentual falling transitions.

The result shows that the degree of alignment with a melodic falling transition increases when only the accented morae are considered. The degree of correspondence between falling tonal and melodic transitions is about 60.8% in Table 1, but it increases to 68% in Table 2. Also, the degree of correspondence with a falling melodic transition is much greater for accentual falling transitions (about 68%) than non-accentual falling transitions (47%). Moreover, it is interesting that accentual rising tonal transitions highly disfavor melodic rising transitions (9.13%), whereas non-accentual falling transitions are less likely to do so (21.62%). I conduct a chi-square test to examine the difference between accentual and non-accentual falling transitions. The result shows that the difference between accentual and non-accentual falling transitions is significant, suggesting that accentual ones are more likely to favor a melodic falling transition (and disfavor a melodic rising transition) than non-accentual ones are (χ^2 = 16.08, df = 2, p = 0.0003).

	Musical transition		
	Rising	*Falling*	*Level*
Accentual falling	19 (9.13%)	**142 (68.27%)**	47 (22.59%)
Non-accentual falling	24 (21.62%)	**52 (46.85%)**	35 (31.53%)

Table 2: Percent of musical note transitions for accentual falling transitions (3a) and non-accentual falling transitions (3b). Percentages are per row. Bold-faced cells are similar motion, shaded cells are for contrary motion, and the remaining cells are oblique motion.

Like falling transitions, there are two distinct types of rising tonal transitions in Tokyo Japanese as in (4):

(4) Examples of rising tonal transitions in Tokyo Japanese
  a. Accentual: ha.SI-ga (**LH**L) 'bridge-NOM'
  b. Non-accentual: ha.NA-GA (**LH**H) 'nose-NOM'

Although both cases are realized with the initial LH boundary tones, previous studies show that the pitch expansion of the accentual rise is larger and more abrupt than that of the non-accentual rise (Pierrehumbert and Beckman 1988). This suggests a possibility that we may observe a difference between accentual and non-accentual rising tonal transitions in the text-tune alignment. Thus, I examine all rising tonal transitions to see if accentual rising tones (the transition from pre-accented morae to accented morae) and non-accentual rising tones (the other cases of rising tonal transitions) behave differently from one another in the text-setting of Japanese songs. Table 3 shows the percentages of melodic transitions with accentual and non-accentual falling transitions.

	Musical transition		
	Rising	*Falling*	*Level*
Accentual rising	**90 (56.6%)**	17 (10.69%)	52 (32.7%)
Non-accentual rising	**80 (43.48%)**	23 (12.5%)	81 (44.02%)

Table 3: Percent of musical note transitions for accentual rising transitions (3a) and non-accentual rising transitions (3b). Percentages are per row. Bold-faced cells are similar motion, shaded cells are for contrary motion, and the remaining cells are oblique motion.

The result of rising tonal transitions shows a similar trend to that of falling tonal transitions: accentual rising transitions are more likely to correspond with melodic rising transitions (56.6%) than non-accentual ones (43.48%). To see if this result is statistically significant, I conduct a chi-square test. The chi-square test reveals that the accentual and non-accentual rising tonal transitions are significantly different ($\chi^2 = 6.02$, df = 2, *p*-value = 0.049), confirming the observation that accentual rising transitions favor melodic rising transitions more than non-accentual ones do. Both results of rising and falling tonal transitions show that accentedness plays an important role in the textsetting in Japanese children's songs.

3.3 Text-Meter Alignment

The last question I examine is whether accented morae are more likely to fall on strong beats than unaccented ones in accented words. This analysis investigates only accented words. Figure 2 shows the mean metrical strength of each accent type. The mean values of accented syllables are higher than those of unaccented ones in both content and function words (*ac* = 2.36, *af* = 2.32, *uc* = 2.12, *uf* = 1.85). To see if this difference is statistically significant, I conduct a linear mixed-effects model with the metrical strength as a dependent variable and the accent category as a fixed effect predictor. Each song is included as a random effect. Table 4 summarizes the output of the analysis.

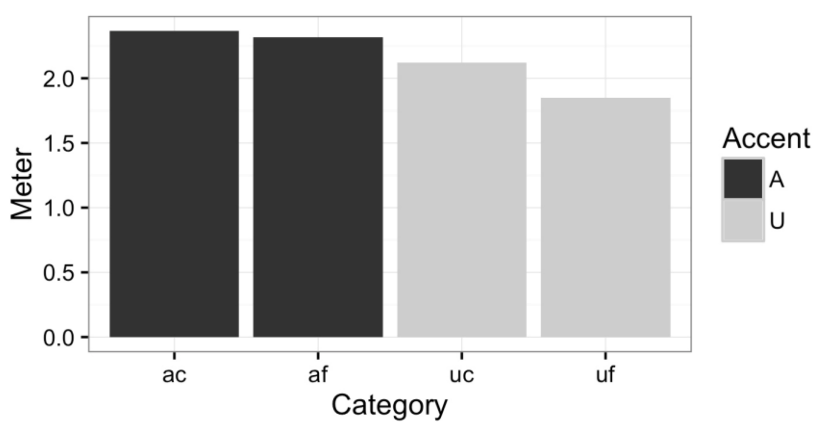

Figure 2: Mean metrical strength by accent type. Level 4 is the strongest beat and Level 1 is the weakest beat of the meter.

The result of the mixed-effect analysis shows that *ac* is not significantly different from *af* with regards to metrical strength (*p* = 0.806). The model estimates that the metrical strength of *af* is only 0.05 smaller than that of *ac*. However, the estimated metrical strengths of *uc* and *uf* are 0.23 and 0.53 respectively, both smaller than that of *ac*, and these differences were found significant (*p*

< 0.001 in both comparisons). The result suggests that accentedness plays a role in the text-meter alignment in Japanese children's songs.

	Estimated	Std. Error	*t*-value	Pr (>|*t*|)
(Intercept)	2.35	0.06	36.97	< 0.001 ***
af	-0.05	0.2	0.25	0.806
uc	-0.23	0.06	-3.65	< 0.001 ***
uf	-0.53	0.07	-7.19	< 0.001 ***

Table 4: The result of the linear mixed-effect analysis. The reference category is *ac*.

Subtracting the mean metrical strength of unaccented morae from that of accented morae results in a single number that represents the degree of text-meter alignments. Temperley and Temperley (2013) name this value *Stress-meter alignment value* (SMAV) in their study of French and English songs. I call this value *Prominence-meter alignment value* (PMAV) to refer to the degree of both stress-meter (as in English and French) and accent-meter alignments (as in Japanese). The PMAV value for Japanese songs in this study is 0.32 (= 2.36 – 2.04). This value is much smaller than that of English (1.15 = 3.08 – 1.93) and that of French (0.77 = 2.57 – 1.8) as reported in Temperley and Temperley (2013). The result suggests that Japanese children's songs have a tendency for accented morae to fall on strong beats of the meter, but this tendency is much weaker than the stress placement rule found in the songs of typical stress languages like English.

4 Discussion and Conclusion

In this study, the overall degree of correspondence (similar motion) between tonal and melodic transitions in this study was about 54%. This value was higher than the chance level (33.33%) but lower than the results found in other studies of tonal languages, such as Wong and Diehl 2009 (91% tone-tune correspondence in Cantonese contemporary pops) or Kirby and Ladd 2016 (77% correspondence in Vietnamese popular songs). However, contrary motion is avoided in the text-tune alignment in Tokyo Japanese (5.77% of the total), and this result is comparable to the result of Vietnamese (Kirby and Ladd 2016), which also showed a high avoidance of contrary motion (about 4% of the total).

The result showed that there is a difference between accentual tonal transitions and non-accentual ones. Accentual falling transitions highly favored the alignment with a falling melodic transition (about 68% of the time). However, non-accentual falling transitions were less likely to be aligned with a falling melodic transition than accentual falling (46% of the time). Also, while accentual falling transitions disfavored contrary motion (9%), non-accentual falling ones were more tolerable with contrary motion (21.6%) than accentual ones. A similar trend was also found in the difference between accentual and non-accentual rising transitions, but to a lesser degree. If Japanese were a tonal language with a limited tonal system, we would not expect to see such a difference between accentual and non-accentual transitions. The result of the study shows that accented positions and accentual tonal transitions are more important than the other positions and other tonal transitions in the text-setting in Japanese songs.

For the text-meter alignment, it was found that accented morae tended to align with stronger beats of the meter than unaccented morae. The difference between *ac* (accented mora in content words) and *af* (accented mora in function words) was not significant, but those categories were more likely to fall on strong beats than *uc* (unaccented mora in content words) and *uf* (unaccented mora in function words). The findings of the study can be compared to the results of the previous studies on the text-setting problems. While accented morae occur at stronger beats of the meter than unaccented ones, the PMAV of Japanese children's songs in this study (0.32) was much weaker than that of English (1.15) and that of French (0.77) in Temperley and Temperley 2013. This difference in PMAVs of the three languages suggests that the difference in metrical strength between stress and pitch-accent languages is gradient, in that phonological prominence in one language is realized more weakly than phonological prominence in another language.

Considering that accented morae are likely to occur on strong beats of the meter and to be aligned with musical falling transitions, it seems that the results of this study support the accentual approach in analyzing pitch accent in Tokyo Japanese. Accentual falling and rising transitions show a high degree of correspondence with accented morae, while non-accentual falling and rising tonal transitions (the initial LH boundary tones) showed a weak correspondence. Also, it seems that the mismatch between initial boundary tones and rising melodic transitions is more tolerable than the one between accentual falling and melodic falling transitions.

As the results of this study provided a way of investigating the nature of pitch accent in Japanese vocal music, this study demonstrates that the relationship between language and music is a fruitful way of investigating questions about prosody. For example, the text-setting in vocal music of intonation-only languages, like Seoul Korean, has not been studied in depth. Investigation of these languages may add valuable insights to the literature of prosodic typology. Also, different dialects of Japanese in various genres of music deserve further investigation in capturing the nature of pitch accent in the Japanese language.

References

Halle, John, and Fred Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55:3–23.

Haraguchi, Shosuke. 1977. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology.* Tokyo: Kaitakusha.

Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23:225–257.

Hyman, Larry M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31:213–238.

Itô, Junko, and Armin Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory,* ed. J. N. Beckman, L. W. Dickey, and S. Ubnanczyk, 181–209.

Kirby, James, and Robert Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. Paper presented at Tonal Aspects of Language: 5th International Symposium, Buffalo, NY. URL http://dx.doi.org/10.7488/ds/1434

Kubozono, Haruo. 2012. Varieties of pitch accent systems in Japanese. *Lingua* 122:1395–1414.

McCawley, James D. 1968. The accentuation of Japanese noun compounds. *Journal-Newsletter of the Association of Teachers of Japanese* 5:1–11.

Obara, Akio. 1932. *Nihonno Warabeuta: Shitsunai Yugika Hen.* Tokyo: Shakai Shisosha.

Palmer, Caroline, and Michael H. Kelly. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31:525–542.

Pierrehumbert, Janet, and Mary Beckman. 1988. *Japanese Tone Structure.* (Linguistic Inquire Monograph 15). Cambridge, MA: MIT Press.

Poser, William. 1984. The Phonetics and Phonology of Tone and Intonation in Japanese. Doctoral dissertation, Massachusetts Institute of Technology.

Shibatani, Masayoshi. 1990. *The Languages of Japan.* Cambridge: Cambridge University Press.

Temperley, Nicholas, and David, Temperley. 2013. Stress-meter alignment in French vocal music. *Journal of Acoustical Society of America* 134:520–527.

Wee, Lian-Hee. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35:128–144.

Wong, Patrick C. M., and Randy L. Diehl. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30:202–209.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104–6305
csunghye@sas.upenn.edu