TEXT-TO-TUNE ALIGNMENT IN JAPANESE "HAPPY BIRTHDAY TO YOU"

Haruo KUBOZONO[1] and Ai MIZOGUCHI[1,2]

NINJAL[1], Maebashi Institute of Technology[2]
kubozono@ninjal.ac.jp, aimizoguchi@maebashi-it.ac.jp

ABSTRACT

This paper examines how text is aligned with musical tunes in Japanese "Happy Birthday to You". Although this song is sung with the same melody as its English version, words are segmented in different ways from English. Based on a production experiment involving twenty native speakers of Tokyo Japanese, this paper demonstrates that (i) four-mora or longer words are segmented based on the syllable rather than the mora, while shorter words are parsed by the mora, (ii) unlike lexical stress in English, lexical pitch accent in Japanese plays no role in the alignment, resulting in tonal neutralizations, (iii) vowel length is also neutralized in most cases, and (iv) vowel devoicing exerts a certain effect.

This paper also considers implications of these findings for Japanese syllable structure, with main focus on the questions of what vowel sequences constitute diphthongs and how (much) superheavy syllables are avoided in the language.

Keywords: text-to-tune alignment, Japanese, syllable and mora, lexical pitch accent, word segmentation

1. INTRODUCTION

How text is aligned with musical tunes has attracted serious attention in linguistics since the pioneering work by Liberman [1], illuminating not only the relationship between text and tune, but also the roles of the syllable and lexical stress as well as the strategies involved in word segmentation. Specifically, lexical stress in English and German plays an important role in the alignment, where words with non-initial stress behave differently from initially-stressed words.

The effect of lexical stress manifests itself in vocative intonation or 'calling contour', too, where the stressed syllable is aligned with the nuclear pitch accent H* before a downstepped ꜜH ([2]):

(1) Calling contour in English

H* ꜜH H* ꜜH vs. L H*ꜜH
Johnny! Jona than! Re bécca!

(2) Calling contour in German

H* ꜜH H* ꜜH vs. L H*ꜜH
Lisa Ursula! Hans-Jürgen

Text-to-tune relationship has been studied in Japanese, too, centering around the roles of the mora and the syllable. Linguistic analyses of Japanese songs have revealed the dominant role of the mora over the syllable (e.g. [3]), which is in line with the well-known fact that the mora is the basic prosodic unit in traditional Japanese poetry such as *haiku* and *tanka*. More recently, [4] has presented statistical evidence that syllabic alignment is used along with moraic alignment, especially in translated (vs. native) songs.

More interesting facts have been revealed by the series of work on the baseball chant sung by baseball fans to cheer on their favorite players ([5, 6, 7]). This chant consists of three musical notes followed by a pause as shown in (3), to which long names as well as short ones are to be aligned.

(3) | ♩ ♩ ♩ 𝄽 |
| | |
i chi roo

Linguistic analyses of this chant have clarified that the syllable as well as the mora plays a pivotal role ([5]) and that distinctions in lexical pitch accent and vowel length are both neutralized in some restricted contexts ([7]). A phonological test based on the role of the syllable has further revealed that Japanese has only three tautosyllabic vowel sequences, or diphthongs: i.e. /ai/, /oi/, and /ui/ (as well as /ei/ if it is recognized as a diphthong rather than a monophthong /eː/). It has also shown that what appear to be trimoraic syllables such as /ain/, /oin/, and /uin/ are actually processed as a sequence of two syllables, e.g. /a.in/, /o.in/, and /u.in/ ([8, 9]).

2. PRODUCTION EXPERIMENT

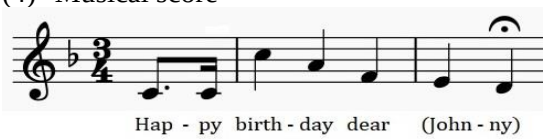
While the text-to-tune relationship in the baseball chant has revealed many things about Japanese phonology, it remains unanswered how general the alleged relationship is in the language, how words are segmented in other chants in Japanese, and how the language is different from English and other languages. To answer these questions, this current study has conducted a production experiment with 'Happy Birthday to You', a song that is sung with the same melody across languages. It is expected that this

experiment will reveal how Japanese speakers segment words and how their strategies are different from those used by English speakers.

2.1. Method

A total of 20 native speakers of Tokyo Japanese (10 males and 10 females) participated in the experiment. They are in two age groups, one aged between 19 and 46 years old, and the other in their seventies. The elderly group participated in a face-to-face experiment, while the younger group participated in an online experiment using Zoom. They were presented with a list of 109 names and were asked to sing the following portion of the Happy Birthday song twice with each name in the slot of *John-ny*. The list includes real Japanese names, English names, and new names coined from them.

(4) Musical score



To see the effects of various linguistic factors, the test words were selected in such a way that they range in phonological lengths (one mora to seven moras), syllable structures (different types of syllable weight), accent structures (presence/absence of a lexical accent as well as its position), morphological structure (monomorphemic vs. polymorphemic), and the nature of the final vowel (voiced vs. devoiced).

The subjects' responses in the online experiment were recorded using the recording function of the Zoom as well as a digital recorder placed near the speaker on the experimenter's side. Those in the face-to-face experiment were recorded in a silent room using a digital recorder. The data thus obtained consist of 2,180 tokens (109 stimuli x 20 subjects). They were analyzed auditorily by the experimenters as to how the test words are aligned with the tune in (4), that is, how they are segmented into two parts.

2.2. Results

The analysis has revealed the following pattern as the basic text-setting pattern in the Japanese song. Namely, the second musical note in the two-note slot in (4) is basically aligned with the last mora or syllable of the word, with the remaining syllables/moras all linked with the first note. This is illustrated in Table 1 with words ending in a light, i.e., monomoraic, syllable. Dots and hyphens indicate syllable boundaries and musical note boundaries, respectively.

Input		Segmentation pattern	Gloss
length	Word		
1μ	ri	ri-i	Li
2μ	ma.ma	ma-ma	mama
3μ	pa.me.ra	pa.me-ra	Pamela
4μ	ya.ma.ga.ta	ya.ma.ga-ta	Yamagata
5μ	fe.ru.nan.de	fe.ru.nan-de	Fernande
6μ	ma.ku.do.na.ru.do	ma.ku.do.na.ru-do	McDonald's

Table 1: Basic segmentation pattern.

Since words with a final monomoraic syllable like those in Table 1 are ambiguous between the two strategies, i.e. syllable-based vs. mora-based segmentation, we focused on stimuli that end in a heavy, i.e. bimoraic, syllable. A typical example is given in (5), which actually exhibits two segmentation patterns, a syllable-based pattern as in (5a) and a mora-based one as in (5b).

(5) te.he.ran 'Tehran' → (a) te.he-ran, (b) te.he.ra-n

This analysis has shown that the choice between the syllabic and moraic segmentation patterns in (5) is determined primarily by the length of the word. This can be confirmed by Fig. 1, where the proportion of the mora/syllable-based segmentation patterns is shown for each word length, for all speakers as well as for each age group.

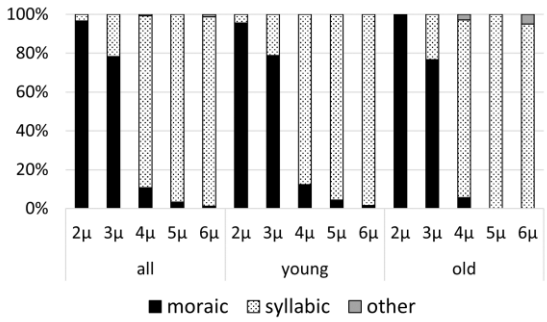


Figure 1: Segmentation strategy as a function of word length by the mora.

A crucial boundary exists between three-mora and four-mora stimuli: words consisting of two or three moras are aligned with the tune basically by the mora, with the last mora of the name associated with the second musical note on its own, whereas four-mora or longer words favor the syllable-based segmentation as shown in (5a). This result is in line with the result of the baseball chant reported in the literature ([5]). Moreover, the two age groups (young and old) show no noticeable difference, suggesting that word length plays a pivotal role independently of the speaker's age.

As for shorter stimuli, the most interesting case is that of light-heavy bisyllables, which favor segmentation at a syllable-internal mora boundary:

e.g. *a.ran* → *ara-n* 'Alan', *ta.roo* → *taro-o* 'Taro', *ji.mii* → *jimi-i* 'Jimmy', *ma.sai* → *masa-i* 'Masai'. This segmentation is strikingly different from the one used by native speakers of English and German, who respect syllable boundaries faithfully: e.g. *a-lan* 'Alan', *ma-sai* 'Masai'. The latter segmentation pattern is marginally observed in our Japanese data, too, but the mora-based segmentation overwhelms the syllable-based one in light-heavy bisyllables (72.5% vs. 27.5%).

The role of the mora can also be confirmed by the fact that bimoraic monosyllables are segmented at the mora boundary, with the coda consonant (moraic nasal) aligned with the second musical note on its own: e.g. *jon* → *jo-n* 'John'. This contrasts with the pattern usually observed in English and German, where coda consonants are aligned with the second musical note together with the preceding vowel: e.g. *jon* → *jo-on* 'John'.

While words consisting of one to three moras follow a moraic segmentation pattern, most four-mora or longer words are subject to the syllable-based segmentation shown in (5a): e.g. *wa.shin.ton* → *wa.shin-ton* 'Washington', *ku.ro.ma.tii* → *ku.ro.ma-tii* 'Cromartie', *an.do.rei* → *an.do-rei* 'Andrei'.

A careful analysis of the data further reveals that the syllabic vs. moraic proportion varies slightly depending on the type of heavy syllable. In four-mora stimuli, for example, the syllabic segmentation in (5a) accounts for 93% of stimuli ending in a long vowel but 85% of those ending in a moraic nasal.

3. DISCUSSION

3.1. Pitch accent's role

While our analysis has demonstrated the roles of the mora and the syllable in the text-to-tune alignment, it shows no evidence for the role of the lexical pitch accent. Our stimuli actually include many pairs of words like those in (6), which contrast either in the presence (') or absence (^{0}) of the pitch accent or in its position. However, none of these pairs showed a difference in the segmentation pattern: all test words simply followed the regular segmentation pattern in Table 1 and the musical melody shown in (4) above, hence showing no difference in tonal melody.

(6) a. ma'.ma → ma-ma 'mom'
ya.no^0 → ya-no 'Yano'
b. be.ru.gi'i → be.ru-gii 'Belgium'
be.ru.rin^0 → be.ru-rin 'Berlin'
c. a'.ki.ko → a.ki-ko 'Akiko'
a.ki.o^0 → a.ki-o 'Akio'
d. te'.he.ran → te.he-ran 'Tehran'
te.he'n.ra → te.hen-ra 'Tehenra (new name)'

It is worth adding here that not just the contrasts in lexical pitch accent but also those in vowel length are often neutralized. Thus, the pairs or triplets given in (7) are temporally neutralized as short vowels are often lengthened in the song.

(7) a. ma.no → maa-noo 'Mano'
maa.no → maa.noo 'Maano'
maa.noo → maa-noo 'Maanoo'
b. o.ta.ni → oo.ta-ni 'Otani'
oo.ta.ni → oo.ta-ni 'Ohtani'

3.2. Morphology's role

Our analysis has also shown that morphological structure plays no role. The test stimuli include both morphologically complex words as in (8a) and simplex words as in (8b). However, the text-to-tune alignment is not affected by the presence or absence of a word-internal morpheme boundary (#) or its position.

(8) a. ya.ma'#ga.ta → ya.ma.ga-ta 'Yamagata'
oo#ta.ni → oo.ta-ni 'Ohtani'
b. pi.no'.ki.o → pi.no.ki-o 'Pinocchio'
i.ta.ri.a^0 → i.ta.ri-a 'Italy'

3.3. Devoiced vowels

On the other hand, devoiced vowels in word-final position exerts a certain effect on the segmentation pattern. This can be seen from the examples in (9): words usually pronounced with a final devoiced vowel favor the seemingly irregular segmentation pattern in (i), where the final two syllables are aligned with the second musical note, over the pattern in (ii), where only the final syllable (with a full vowel) is aligned with the note. This irregular behavior of devoiced vowels is consistent with the finding from modern songs ([4]), as well as the well-known fact that moras with a devoiced vowel behave like a heavy syllable together with their preceding mora in the assignment of lexical pitch accent (e.g. [10]).

(9) a. fe.ru.na'n.de.su̥ 'Fernandes'
→ (i) fe.ru.nan-de.su̥, (ii) fe.ru.nan.de-su
b. ban.gu.ra.de'.shu̥ 'Bangladesh'
→ (i) ban.gu.ra-de.shu̥, (ii) ban.gu.ra.de-shu
c. ma.ru.ti'.ne.su̥ 'Martinez'
→ (i) ma.ru.ti-ne.su̥, (ii) ma.ru.ti.ne-su
d. moo.ri'.sha.su̥ 'Mauritius'
→ (i) moo.ri-sha.su̥, (ii) moo.ri.sha-su

The effect of devoiced vowels can be confirmed with the data in Fig. 2, which shows the proportion of the two segmentation patterns in (9)—(i) and (ii)— as a function of the nature of the word-final vowels in

five-mora or six-mora words ending in a sequence of two light syllables like those in (9).

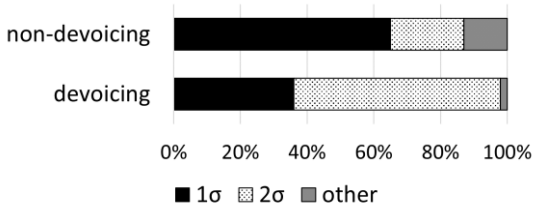


Figure 2: Segmentation patterns as a function of the nature of the word-final vowels.

4. IMPLICATIONS FOR SYLLABLE STRUCTURE

Having understood the basic principles underlying the text-to-tune alignment in the Happy Birthday song, let us finally consider the implications of these principles for the syllable structure of Japanese.

4.1. Diphthongs

One issue that is still debated in the literature concerns the inventory of diphthongs in the language ([8]). Since words longer than three moras predominantly favor the syllabic segmentation in (5a), we can use this strategy to test which vowel sequence actually belongs to the same syllable and which one does not in our alignment data. This analysis has revealed that the four types of vowel sequences in (10a) tend to be aligned with the second musical note, those in (10b) usually split into two musical notes, and /au/ in (10c) fluctuates between the two patterns.

(10) a. /ai/ o.chi-ai ‘Ochiai’
/oi/ ma.ki-roi ‘McIlroy’
/ei/ a.do-rei ‘Adlay’
/ui/ kin-sui ‘Kinsui’
b. /ae/ oo.ma-e ‘Oomae’
/ao/ a.sa.ga-o ‘Asagao, morning glory’
/oe/ ka.wa.go-e ‘Kawagoe’
/eo/ mai.ne-o ‘Mineo’
/ia/ i.ta.ri-a ‘Italy’
/ua/ ma.na.gu-a ‘Managua’
/ea/ e.ku.re-a ‘éclair’
/oa/ e.ka.no-a ‘Ecanoa’
/io/ pi.no.ki-o ‘Pinocchio’
/ie/ yo.ri.i-e ‘Yoriie’
/uo/ a.ka.u-o ‘Akauo, red fish’
/ue/ ta.me.su-e ‘Tamesue’
c. /au/ rin-dau ~ rin.da-u ‘Lindau’

(10b) includes six vowel sequences that are permitted in native morphemes: /ae/, /ao/, /oe/, /ie/, /uo/, and /ue/. It is interesting to find that not all vowel sequences that are permitted in native morphemes form diphthongs. Overall, the results in (10) accord with the claim by [11] and [9], who demonstrated with dialectal accent data that only vowel sequences ending in /i/ form diphthongs in Japanese. One thing that is not reported in these previous studies is the behavior of /au/ in (10c): in the assignment of lexical pitch accent, this vowel sequence behaves as a sequence of two syllables just like those in (10b).

4.2. Superheavy syllables

Our data also provide some evidence pertaining to the debate about superheavy, i.e. trimoraic, syllables. Japanese permits what appear to be trimoraic syllables in loanwords: e.g. *wain* ‘wine’, *ainshutain* ‘Einstein’, *saaroin* ‘sirloin’, *rinkaan* ‘Lincoln’. Accentual analyses in the literature have shown that these trimoraic sequences do not form one syllable, but actually a sequence of two syllables ([12, 9]).

Our data from the text-to-tune alignment support this previous finding to a large extent. They contain many words with a final moraic nasal preceded by a long vowel or diphthongal sequence as exemplified in (11). Of these, diphthong-like vowel sequences plus the moraic nasal in (11a) generally exhibit a bisyllabic behavior in our data, with the second element of the vowel sequences and the moraic nasal aligned with the second musical note. Here, it is interesting indeed that the vowel sequences in (10a)—/ai/, /oi/, /ei/, and /ui/—split into two syllables when they are followed by a coda consonant, although they otherwise behave as one syllable as shown in (10a). This reinforces the argument that superheavy syllables are disfavored, if not entirely prohibited, in Japanese (e.g. [8]).

(11) a. ko.ka-in, *ko-kain ‘cocaine’
ba.ren.ta-in, ?ba.ren-tain ‘Valentine’
kya.ro.ra-in, *kya.ro-rain ‘Caroline’
ma.ke-in, *ma-kein ‘McCain’
hi.ro-in, *hi-roin ‘heroine’
saa.ro-in, *saa-roin ‘sirloin’
b. rin-kaan > rin.ka-an ‘Lincoln’
ha.ro-wiin > ha.ro.wi-in ‘Halloween’
ha.ri-keen ~ ha.ri.ke-en ‘hurricane’
to.ron-boon > to.ron.bo-on ‘trombone’

On the other hand, monophthongal vowel sequences plus a moraic nasal in (11b) favor the monosyllabic pattern over the bisyllabic one (66.3% vs. 31.3%), suggesting that they resist being split to two syllables. This agrees with the result obtained from the analysis of the baseball chant ([9]), although it is different from what has been reported on the basis of a phonological test on the assignment of lexical accent ([8, 9]).

5. ACKNOWLEDGEMENTS

We would like to thank the subjects of our experiment. The work reported in this paper is supported by JSPS KAKENHI grants (19H00530 and 20H05617) to the first author.

6. REFERENCES

[1] Liberman, M. 1979. *The Intonational System of English.* Garland Publishing, Inc.
[2] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge University Press.
[3] Kubozono, H. 1999. Kayō ni okeru mōra to onsetsu [the mora and the syllable in song]. In: Speech and Grammar Study Group (ed.), *Bunpō to Onsei II [Speech and Grammar 2]*. Kurosio, 241–260.
[4] Starr, R. L., Shih. S. S. 2017. The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting. *Glossa*. 2(1), 93, 1–34.
[5] Tanaka, S. 1999. Nihongo no onsetsu to yonhaku no tenpurēto: Senryū to puroyakyū seien ni okeru ziamari no bunseki [Syllables and four-mora templates in Japanese: Analysis of 'ji-amari' in senryu poems and baseball chants]. In: Spoken Language Working Group (ed.), *Bunpō to Onsei II [Speech and Grammar 2]*. Kurosio, 261–290.
[6] Ito, J., Kubozono, H., Mester, A., Tanaka, S. 2019. Kattobase: The linguistic structure of Japanese baseball chants. *Proceedings of Annual Meeting of Phonology (AMP) 2018* (online). Washington, DC.
[7] Kubozono, H. 2020. Neutralizations in vowel length and word accent in Japanese. *Kobe Papers in Linguistics*. 12, 69–79.
[8] Kubozono, H. 2018. Prosodic evidence for syllable structure in Japanese. In: Guillemot, C., Yoshida, T., Lee, S. J. (eds.) *Proceedings of the 13th Workshop on Altaic Formal Linguistics. (MIT Working Papers in Linguistics 88)*. Tokyo, 35–50.
[9] Kubozono, H. 2022. *Word and Sentence Prosody: The Endangered Dialect of Koshikijima Japanese.* De Gruyter Mouton.
[10] Kubozono, H. 2002. Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Onsei Kenkyu*, 79–97.
[11] Kibe, N. 2000. *Seinanbu Kyushu Nikei Akusento no Kenkyū [Studies in the Two-Pattern Accent Systems in South-west Kyushu]*. Bensei Shuppan.
[12] Kubozono, H. 1999. Mora and syllable. In: Natsuko Tsujimura (ed.), *The Handbook of Japanese Linguistics*, Blackwell, 31–61.