Exploring Text-to-Tune Alignment of Japanese Special Moras in “Happy Birthday to You”

Natsumi Taniguchi
International Christian University

1 Introduction

Text-to-tune alignment, or text-setting, is the lining up of lyrics to a melody, and it has been utilized in linguistic research as a phenomenon that reveals the prosodic properties of a language. Japanese tunes have provided data for debates in prosodic structure (Starr & Shih 2017), as we will discuss in depth later. These studies are possible because text-setting is both productive, as in natives can align any given lyrics to a tune near-unanimously, and predictable, as in the alignment patterns are rooted in outside principles.

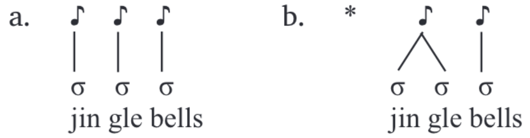
This research looks into the metrical structures of Japanese through the famous text-setting tune, "Happy Birthday to You", sung at birthday parties all over the world. In particular, Japanese names with three moras and a word-final heavy syllable show variation in outputs between two alignments: one where the last mora is aligned to the last beat and one where the last syllable is aligned. This paper will explore the motivations behind this variation through two experiments, production and perception. Section 2 will go over the basic background information regarding the relative factors in Japanese prosody. Then, section 3 will introduce the birthday song and the experiments conducted regarding the tune. Analyses for the alignments in focus will be discussed in section 4. The paper will conclude with section 5.

2 Background

Text-setting and the concepts surrounding the phenomenon will be explained first. The two aspects of Japanese prosody that influence the birthday song alignments the most are pitch accent and prosodic structure. 2.2 and 2.3 lay out the basics of the Japanese pitch accent system and prosodic structures. 2.4 briefs a Japanese text-setting phenomenon with output patterns comparable to the birthday song.

2.1 *Text-setting and language* Starr and Shih (2017) explain a few properties of text-setting fundamental to the linguistic discussion of the phenomenon. Text-setting is usable for research on prosodic structure because of an important characteristic. If a sequence of segments is sung to a single musical note, this can be taken as evidence to say that those segments construct a single prosodic unit. For instance, no more than a single syllable can be set to a single note in English. In the case of the chorus of the Christmas carol, “Jingle Bells,” all three syllables need a corresponding note to be sung. A construct like 1b would be almost impossible for most native speakers.

(1) Musical notes and syllables in “Jingle Bells” (Starr & Shih 2017)



It seems to be that Japanese maximally sets a syllable to a single musical note as well, as will be mentioned in section 2.3.

2.2 *Japanese pitch accent* There are two levels of relative tones in Japanese, high (H) and low (L) (Kawahara 2015). A lexical accent, which exists singularly per word, is marked by a falling pitch (an H-L tonal complex). Out of the two moras aligned with the accentual tonal complex, the mora in charge of the high tone is called the accent nucleus. The position of this pitch accent is contrastive, like the English stress accent. For instance, 2 shows

[*] I’d like to take a moment to thank Dr. Seunghun Lee, Dr. Haruo Kubozono, Dr. Shigeto Kawahara, and Dr. Hyun Kyung Hwang for answering my questions and connecting my ideas to existing theories so that I could further my thoughts. I am also indebted to my peers for their crucial support throughout the process. All remaining errors are mine.

Proceedings of *AJL7*

the contrast between three minimal pairs in Tokyo Japanese[1]. The particle *ga* is inserted to illustrate the pitch differences in their entirety. An apostrophe is inserted following the accent nucleus.

(2) Pitch accent patterns for "hashi-ga"

a. hashi-ga H' L L	"chopsticks"	b. hashi-ga L H' L	"bridge"	c. hashi-ga L H H	"edge"

Accent patterns like 2a have accent nuclei at the penultimate syllable, counting from the lexical edge, and are informally dubbed -2 accent patterns. Similarly, 2b patterns with accent nuclei at the word-final syllable are -1 accent patterns, and words like 2c with no pitch fall are called 0 accent or unaccented patterns. The tones aside from the H-L accent are determined automatically once the accent nucleus position is set (Kawahara 2015).

For languages that utilize pitch as a contrastive feature, it is logical to hypothesize that there may be some correlation between linguistic pitch patterns and the melodies that words are set to. There is literature that explores and confirms this thought. Mandarin Chinese and Vietnamese are found to have strong correspondences between musical and linguistic tone sequences (Kirby & Ladd 2016, Wee 2007). Cho (2017) builds off of these past studies to examine the Japanese pitch accent and its correlations with melody. In short, Cho (2017) found text-setting in Japanese children's songs does show correlations between accentual tonal transitions and the melody's pitch movement, though weaker than in Mandarin and Vietnamese. We will later see that the birthday song, a more spontaneous text-setting phenomenon than pre-written lyrics in children's songs, supports Cho (2017)'s findings in Japanese.

2.3 *The Japanese syllable* The necessity for the syllable (σ) has been a contention in Japanese prosody over the past few decades. Japanese is often described as a mora-language, where the mora (μ) is said to be the basic prosodic unit. However, many have argued for the need to include the syllable in the Japanese prosodic system. After a short explanation of the relation between syllables and moras in Japanese, let us quickly go over this debate.

Due to the simple syllable structures often without codas, the mora and syllable mostly coincide in Japanese. Exceptions occur when a "special mora", or sometimes called a dependent mora or non-syllabic mora[2], follows an independent or syllabic mora and creates a heavy syllable. In these cases, one heavy syllable equates to two moras: independent and dependent. There are four types of special moras in Japanese.

(3) The four types of special moras

a. moraic nasal	/N/	e.g. /niNgeN/	[niŋgeɴ]	"human"
b. second half of a long vowel	/R/	e.g. /kuRki/	[kuuki]	"air"
c. second half of a diphthong	/J/	e.g. /kikaJ/	[kikai]	"machine"
d. first half of a geminate	/Q/	e.g. /iQpai/	[ippai]	"a lot"

The syllable-less view of Japanese (Labrune 2012) proposes that these heavy syllables be treated as sequences of a 'regular' mora and a 'deficient' mora that make up a single foot. There are many arguments against this perspective, including phonological and experimental evidence (summarized in Kawahara, 2016). Text-setting data has also contributed to this side of the discussion. Corpus research on popular songs clearly shows the existence of syllables being aligned to single musical notes, though special moras were also aligned to single musical notes (Starr & Shih 2017). No more than a syllable was aligned (for instance, two independent moras or one foot), supporting the claim that the syllable is the way to go rather than a foot. The birthday song adds evidence to this side of the argument in the context of spontaneous text-setting.

2.4 *A comparable tune: Japanese baseball chants* A particular text-setting phenomenon in Japanese shows similar patterns to "Happy Birthday to You." Ito et al. (2019) and Tanaka (2008) offer data and ideas on the Japanese baseball chant illustrated in Figure 1 and the text-to-tune alignment in this tune. This chant, typically delivered by fans at baseball games, has a three-beat melody into which names are inserted. Though the number of beats differs from the birthday song, which has two beats, it presents some patterns to compare.

[1] Examples are written roughly in the Hepburn romanization format. Long vowels will be represented by double vowels, geminates by double consonants, and the moraic nasal by a Times New Roman noncapitalized "n".

[2] This paper will refer to these moras as "special moras'. We will not call them "dependent moras" because in the text-setting environment, these moras can act independently from an independent mora. They will not be referred to as "non-syllabic moras" either in order to leave the possibility of them creating pseudo-syllables when aligned alone to a musical note.

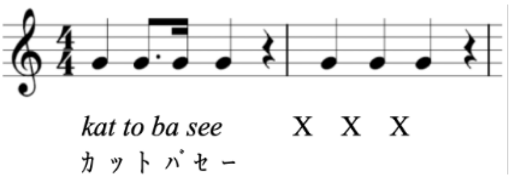


Figure 1. Japanese baseball chant (Ito et al., 2019)

Ito et al. (2019) summarize the rules that play into the text alignment of this chant by the input’s mora number. 4 are examples for each mora number[3]. The simplest is three mora names since the rule is to align each mora with each beat. For one or two mora inputs, spreading occurs from the left to make up for non-existing moras. For four mora inputs, the last syllable fills the last beat. The same occurs for five or more mora inputs. In other words, the last syllable is aligned with the last beat unless there are not enough moras to fill the first two beats. If aligning the last mora with the last beat leaves not enough to fill in the other two beats, spreading from the left occurs. Very little variation was reported. As we will see in the next sections, the birthday song alignments follow these rules with a few crucial deviations.

(4) Examples of alignments in the Japanese baseball chant (Ito et al., 2019)
a. 3 moras: ka-ke-fu → ka/ke/fu, ge-n-da → ge/n/da, e-to-o → e/to/o
b. 2 moras: ta-ni → ta/a/ni, ri → ri/i/i
c. 4 moras: ki-yo-ha-ra → ki/yoha/ra, i-chi-ro-o → i/chi/roo
d. 5+ moras: bo-o-cha-a-do → boo/chaa/do, ro-ba-a-to-so-n → ro/baato/son

Further on, Ito et al. (2019) put out an Optimality Theory analysis of the phenomenon with the goal of unifying the multiple rules briefed above into a single constraint ranking. They achieve this by choosing to use several constraints that are specific to the baseball chant, such as ALIGN-L(X_3,μ]) and ALIGN-L(X_3,σ]). These two constraints restate an intriguing rule that is shared by the baseball chant and birthday song; the last beat (X_3 for the baseball chant, X_2 for the birthday song) can be associated with a single syllable and no more[4]. Ito et al. (2019) do not dive into the independent principles that derive this rule and end the paper stating that more needs to be done to ground "the constraints better in the prosodic system of the language." In section 4.2, we will go through several hypotheses that attempt to explain this alignment rule.

3 Alignments in the birthday song

3.1 *"Happy Birthday to You"*

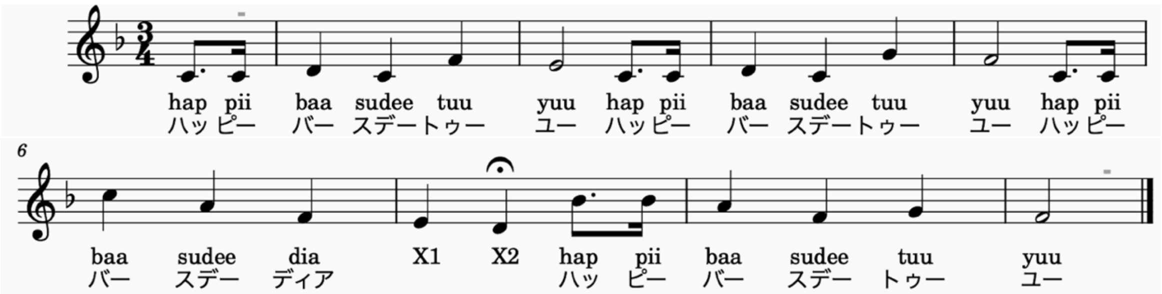

hap pii baa sudee tuu yuu hap pii baa sudee tuu yuu hap pii
ハッ ピー バー ステートゥー ユー ハッピー バー ステートゥー ユー ハッピー
6
baa sudee dia X1 X2 hap pii baa sudee tuu yuu
バー ステー ディア ハッ ピー バー ステー トゥー ユー

Figure 2. "Happy Birthday to You" with romanized Japanese lyrics

“Happy Birthday to You,” or simply the birthday song, is a classic celebrational song sung by friends and family to the birthday person. Though the original song is in English, many language cultures have incorporated this simple tune and adapted it to their language. In the case of Japanese, the “katakana” version of the English lyrics (original English lyrics conformed to Japanese phonotactics) has stuck as the most widely known and used rendition (Figure 2).

This paper focuses on the particular two-note melody shown in the third to last measure in Figure 2. These two notes, notated as X_1 and X_2, are where singers insert the birthday person’s name. Like the baseball game chant the alignment depends heavily on the syllable structure of the input. Unlike the baseball chant, there is a significant

[3] Hyphens (-) represent mora boundaries, periods (.) syllable boundaries, and forward slashes alignment boundaries.

[4] The other beats (X_1 and X_2 for the baseball chant, X_1 for the birthday song) can be associated with two or more syllables. When this happens, the beat is split into multiple musical notes (for instance from one quarter note to two eighth notes) because, as explained in section 2.1, a single musical note can only be aligned with a syllable at maximum in Japanese. This still is counted as one beat, because these additional musical notes only appear when multiple syllables are associated with the beat.

variation in the alignment patterns of trimoraic and quadramoraic inputs with a final heavy syllable. The alignment rules for each input length are laid out in 5. These are based on the experiments reported in the next section.

(5) Alignment rules based on syllable structure in “Happy Birthday to You”

# of moras	Syllable structure	Alignment rule
1	1 light	Align the single mora to X_1 and elongate the vowel to fill in X_2. ex. ri → ri/i
2+	final light syllable	Align the final mora (light syllable) to X_2 and the rest to X_1. ex. ri-na → ri/na
2	1 heavy	Align each mora to X_1 and X_2. ex. shu-n → shu/n
3, 4	final heavy syllable	(i) Align the final mora (special mora) to X_2 and the rest to X_1. ex. ka-ri-n → kari/n, sa-bu-ro-o → saburo/o (ii) Align the final syllable (two moras) to X_2 and the rest to X_1. ex. ka-ri-n → ka/rin, sa-bu-ro-o → sabu/roo
5+	final heavy syllable	Align the final syllable (two moras) to X_2 and the rest to X_1.[5] ex. do-ra-e-mo-n → dorae/mon

This paper focuses on the variation between aligning the final mora versus syllable to X_2 among three and four-mora inputs. Aligning the last syllable is in accordance with the baseball chant as well as the five+ mora inputs. However, for reasons we will discuss in later sections, three and four-mora inputs have a second option. Let us name each of these alignment patterns mora-based alignment and syllable-based alignment. The following section will dive into the details of these two variations.

3.2 *Mora-based vs. syllable-based alignment* Let us start with some examples. Example 6 shows the possible alignments of trimoraic and quadramoraic inputs ending with /N/ (moraic nasal) or /R/ (long vowel). /J/ (diphthongs) are out of the scope of this research, although they likely have similar patterns[6].

(6) Possible alignment for three and four-mora inputs

# of moras	Input	Mora-based	Syllable-based
3	ka-ri-n	kari/n	ka/rin
	ta-ro-o	taro/o	ta/roo
4	su-zu-ra-n	suzura/n	suzu/ran
	sa-bu-ro-o	saburo/o	sabu/roo

One thing to note is that the mora-based alignment is the majority output among trimoraic inputs, while it is the minority among quadramoraic inputs. This distribution can be seen in the results of both Experiments 1 and 2.

This preference for the mora-based alignment among three and four-mora inputs is peculiar, not only because it does not line up with the baseball chant data or the birthday song alignment with 5+ mora inputs, but also because the motivation behind choosing to align the last mora over the last syllable is not obvious.

The mora-based alignment is actually common among two-mora monosyllable inputs. This is because the text-setting tune consists of two beats, and since special moras in Japanese are often aligned to single beats in text-setting (Starr & Shih 2017), it seems natural for an input with two basic prosodic units to distribute those to fill up each of the beats. This alignment pattern is also congruent with the alignments of trimoraic heavy-syllable-ending names in the baseball chant (which has three beats). Associating each beat with a prosodic unit is prioritized over not dividing heavy syllables.

This paradigm cannot be applied to the mora-based alignment among three-mora or four-mora names. It is only applicable when the number of moras is equal to the number of beats in the text-setting tune. For both three and four-mora inputs, aligning the last syllable to X_2 leaves enough moras to associate with X_1.

[5] This table illustrates the image that 5+ mora-inputs never align mora-based. Note that this has not been confirmed though experiment. It may well be that the proportion of mora-based alignments in relation to syllable-based alignments is gradient among inputs ending with heavy syllables, with the highest at 2-mora inputs (100%) and decreasing as the mora count increases.

[6] The existence of diphthongs and their types in Japanese is still debated upon. To keep things simple, diphthongs will not be discussed in this paper.

So what is causing this peculiar alignment to happen? As a first step to finding an answer to this question, two experiments were conducted to confirm the alignment distributions among trimoraic and quadramoraic inputs based on syllable structure and pitch accent.

3.3 *Experiment 1: production* The first experiment was conducted to examine the variation in alignments among Tokyo Japanese speakers. More specifically, the aim was to confirm that the mora-based alignment really is the majority output for trimoraic inputs ending with heavy syllables, as before this experiment, we only had the intuitions of a couple of native speakers to work off of. In the process, we wanted to get output data for the alignments for other syllable structures among two to four mora inputs.

Production was chosen to collect the possible outputs for each syllable structure. The results of this experiment were later referenced to create items for Experiment 2.

As this experiment was focused on collecting elicitations, (i) was the only hypothesis the experiment was controlled for.

(i) The mora-based alignment is the majority when the input is trimoraic and heavy syllable-ending (/N/ and /R/).

There was a less formal hypothesis for the alignments of light-syllable ending items, based on the author’s intuition, that they would align their last mora/syllable to X_2 no matter their syllable structure.

3.3.1 *Method*

3.3.1.1 *Participants* 19 Tokyo Japanese native speakers were recruited. 16 were university students, and three were in their 40s or 50s. A different set of 16 participants identified as female and the rest as male. Participants were separated into two groups randomly. Each group was given a different set of items.

3.3.1.2 *Design* Participants were instructed to record themselves singing along with guide audio while inserting the names they were given into the tune. Guide audio was given in order to control for inconsistent tempo or key. They were allowed to sing a name twice if they felt the need to do so.

3.3.1.3 *Materials* Items were chosen to represent each syllable structure for two to four mora words. Items are primarily common Japanese first names, but some are loanword names (for instance, *mikaeru*) or nouns (like *doragon*, which signifies “dragon”) because some syllable structures are hard to find among words that fulfill the first criteria. The item list includes heavy syllables with /J/ (diphthongs), but these are not considered in the analysis of results as they are out of the scope of this paper, as aforementioned.

Mora #	Syllable Structure		Set 1	Set 2
2	LL	-	shino	taku
	H	/N/	ren	shun
		/R/	yuu	ryou
		/J/	kai	tae
3	LLL	-	nobita	tsumugi
	HL	-	souta	rinka
	LH	/N/	karen	shion
			aran	karin
		/R/	tarou	miyuu
		/J/	misae	sakai
4	LLLL	-	hirofumi	nadeshiko
	HLL	-	kaoruko	kousuke
	LHL	-	iriina	mikaeru
	LLH	/N/	suzuran	doragon
		/J/, /R/	masadai	saburou
	HH	/J/, /N/	ryuusei	goemon

Table 1. Experiment 1 Items

3.3.1.4 *Analysis* Alignments in the recordings were judged based on pitch change and timing in comparison to the guide audio. Average percentages of alignments for each syllable structure category were calculated.

3.3.2 *Results* The data confirmed the hypothesis that the mora-based alignment is the majority output when the input is trimoraic and has a light-heavy syllable structure. 81.58% of the alignments for items ending with /N/ (moraic nasal) and 75.00% for those ending with /R/ (long vowel) were mora-based.

Light-syllable ending items also were aligned as predicted. 100% of the light-syllable ending items, no matter the number of moras, aligned their last mora/syllable with X_2. Similarly, monosyllabic bimoraic inputs aligned each of their moras to the two beats as estimated.

A crucial data point was four-mora items with a light-light-heavy syllable structure. These items also showed some mora-based alignments, though as a minority, unlike the trimoraic items. This result offers evidence for mora-based alignment among inputs with more moras than three, which could suggest that inputs with even more moras are allowed to align their final special mora with X_2 even if less preferred. Analyses for the mora-based alignment need to consider this data because now we know that this alignment is not exclusively for trimoraic words.

3.4 *Experiment 2: perception* A perception experiment was done with two objectives in mind. The first was to observe differences in alignment acceptance between production and perception. Experiment 1 did not force participants to produce as many possible alignments as possible. In fact, the nature of the experiment only called for a single production for each item. Experiment 2 made participants consider each alignment option independently to check the acceptability of secondary alignments,. There is also the possibility that speakers hear an alignment that they did not think of but find acceptable.

The second goal of Experiment 2 was to test the effect of pitch accent pattern on alignment acceptability. As described in section 2.2, the Japanese pitch accent involves a high-low tonal complex. As illustrated in Figure 3 (section 3.1), the birthday song also consists of a high-low pitch fall between the two beats, X_1 and X_2. It is not surprising if the lining up (or the failure to line up) of these pitch falls influence the acceptability of alignments. More specifically, the accentual and melodic pitch falls do not line up in unaccented and penultimate-accent inputs with a syllable-based alignment. It would make sense that these alignments are less acceptable. This estimation matched the intuition of a few native speakers. If this intuition was shared among speakers, it would be evidence for pitch accent patterns affecting alignment choices in the birthday song.[7]

The two hypotheses for comparing production and perception are:

(i) Among heavy-syllable-ending three and four-mora inputs, mora-based alignments are more acceptable than syllable-based alignments.

(ii) Syllable-based alignments among three and four-mora inputs will still be more acceptable than the control group alignments (alignments that were never produced in Experiment 1).

It is natural to hypothesize that the alignment that was more produced (among three and four-mora inputs, the mora-based alignment) would be more acceptable and the less produced (syllable-based alignment) less. Similarly, the minorly produced alignment (syllable-based alignment) would still have more acceptability than alignments that were never produced in Experiment 1.

For the pitch accent patterns, there are two main hypotheses:

(iii) For penult-accent (-2) and unaccented (0) trimoraic inputs ending with a heavy syllable, Tokyo Japanese speakers will perceive syllable-based alignments as unacceptable.

(iv) For antepenult-accent (-3) trimoraic inputs ending with a heavy syllable, Tokyo Japanese speakers will perceive mora-based and syllable-based alignments as acceptable more equally than for penult-accent and unaccented inputs.

The first equates to hypothesizing that when the accent and melody pitch falls do not align, native speakers find the alignment unacceptable. The second is on the antepenult-accent inputs in relation to the penult-accent and

[7] Pitch accent was not considered in Experiment 1 because it was more focused on the effects of syllable structure on elicitations. Additionally, Experiment 1 used names rather than general nouns as items. Names in Japanese are known to take limited accent patterns in comparison to general nouns, especially when the last syllable is heavy. Nearly all of the heavy-syllable-ending names used in Experiment 1 are antepenultimate accent, which is the accent pattern that is not as influenced by accent-melody mismatch. A production experiment that controls for pitch accent would be ideal for a proper comparison of production and perception among inputs of all accent patterns.

unaccented inputs. Antepenult-accent inputs do not face accent-melody mismatch in syllable-based alignments. Therefore, it seems natural to hypothesize that they would not be felt as unnatural as the other accent patterns.

3.4.1 *Method*

3.4.1.1 *Participants* 21 native Tokyo Japanese speakers, out of which 19 were university students and two were middle-aged (30-49), were recruited for this experiment. There were 18 women and three men. None had participated in Experiment 1.

3.4.1.2 *Design* The experiment asked participants to listen to an audio clip of the birthday song with various items aligned and judge the naturalness of the alignment by picking a position on a rating scale. Google Forms was the medium. A rating scale from 1 to 5 was used, 1 being 'unnatural' and 5 being 'natural.' The 'naturalness' that each number represents was not explained in detail to decrease the possibility of manipulation by the researcher. Information on the participant's linguistic background, including proficiency levels in languages other than Japanese, places they have lived, and their caregivers' accent, was collected.

3.4.1.3 *Materials* A total of 23 words were each sung with two alignments, creating 46 items. Out of the 23, nine were target words. For each of the three accent patterns of words with a light-heavy syllable structure, 0, -2, and -3, three words were chosen. This experiment limited the heavy syllables to ones ending with a moraic nasal. Nouns were used instead of names because Japanese names do not have as much accent variety, and this experiment called for words of certain accent positions. The words were considered for their frequency of use, morpheme structure, and Sino-Japanese origins. Light-syllable ending nouns were chosen as the control group. The first three control words in Table 2 all have the same morpheme structure as the target items[8]. The others were chosen without morphemic specifications. Again, they do not differ greatly in their frequency of use. Lastly, there were eight filler words, all loanwords of either three or four moras. These were included, first, to distract the participants from the true motivations of the experiment and second, to see how much native speakers accept an alignment where the last vowel is dropped. The latter is a curiosity that stemmed from Experiment 1, where a loanword name ミカエル *Mikaeru* was sung like mika/er(u), treating the last two syllables like one syllable by dropping the last vowel, by 3 participants. Though the fillers used in this experiment do not provide sufficient data for formal analysis, they offer questions for future explorations.

[8] A trimoraic Sino-Japanese word structured X/XX morpheme-wise and LLL syllable-wise can only have either an antepenultimate accent or no accent at all. Hence, there are no -1 or -2 accent items among the first three controls.

	Accent Pattern	Romanization	English meaning
Target group	0	jikan	time
		toshin	city center
		kabin	vase
	-2	ehon	picture book
		gehin	coarse
		gosen	five-thousand
	-3	kadan	flower bed
		jiken	incident
		shimin	city resident
Control group	0	kiritsu	rule
	-3	bigaku	aesthetics
	0	kazari	decoration
	-1	otoko	man
	-2	okashi	snack
	-3	kakugo	determination
Fillers	4 moras	akuseru	car accelerator
		arubamu	album
		kompasu	compass
		kurafuto	craft
	3 moras	doriru	workbook (originally from 'drill')
		gurafu	graph
		ofisu	office
		tesuto	test

Table 2. Experiment 2 Items

Each word was sung twice with a different alignment (7). Target words were sung with the mora-based (ex. jika/n) and syllable-based alignments (ex. ji/kan). Control group words were first sung as they are normally sung: aligning the last syllable to X_2 (ex. biga/ku). They were then sung aligning the last two syllables with X_2 (ex. bi/gaku). More accurately, X_2 was separated into two eighth notes, and the two syllables were matched to those two notes, as illustrated in Figure 3. This alignment is hardly ever seen among native Tokyo Japanese speakers. It was added to this experiment intended to be extremely unnatural. In addition to giving participants a reference point for their judgments, this negative extreme offers analyzers a reference point. Fillers were sung aligning the last syllable (ex. akuse/ru) and the last two syllables with X_2, as explained in the previous paragraph (ex. aku/ser(u)).

(7) Example alignments of Experiment 2 stimuli

Item group	Example item	Alignment names	Alignments
Target group	jikan	Mora-based	jika/n
		Syllable-based	ji/kan
Control group	bigaku	last syllable to X_2	biga/ku
		last two syllables to X_2 (unnatural)	bi/gaku
Filler group	akuseru	last syllable to X_2	akuse/ru
		last two syllables to X_2 (last vowel dropped)	aku/ser(u)

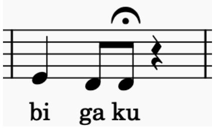


Figure 3. Example control item aligning two syllables to X_2

All items were sung by the author in a consistent key alongside a metronome at 70 bpm. They were recorded on Audacity. For each recording, it was checked that the alignments were in line with the metronome.

3.4.1.4 *Analysis* Raw scores were normalized through z-transformations per participant. Once normalized, averages of each item went through t-tests to compare the different alignments and item categories.

3.4.2 *Results* In short, all four hypotheses were supported by the data of this experiment.

Figure 4 shows the averages of raw scores speakers gave to the two alignments for each category of input. As explained in section 3.4.2.2, 1 was the lowest score, signifying 'unnatural,' and 5 was the highest, signifying 'natural.' Figure 5 compares the averages of the normalized scores. 0 indicates the average score participants gave (i.e., the conceptual equivalent to 3 on the raw score scale). Positive scores are relatively acceptable scores and negative scores the opposite.

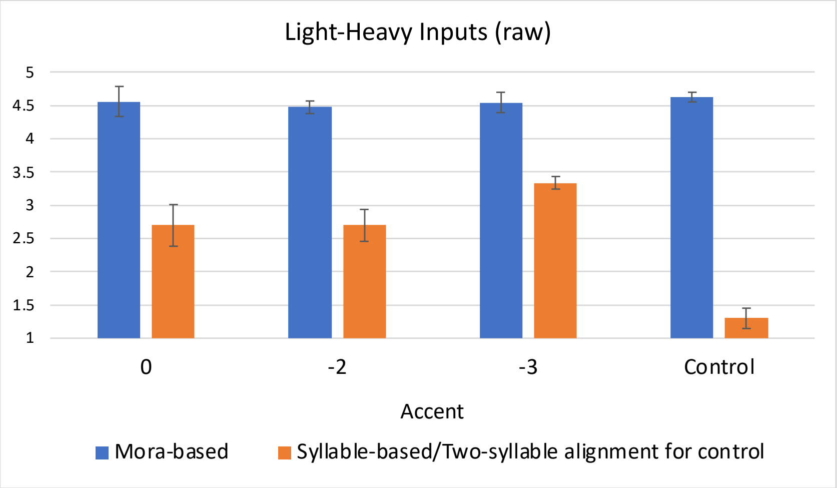


Figure 4. Raw averages of acceptability levels for trimoraic light-heavy items

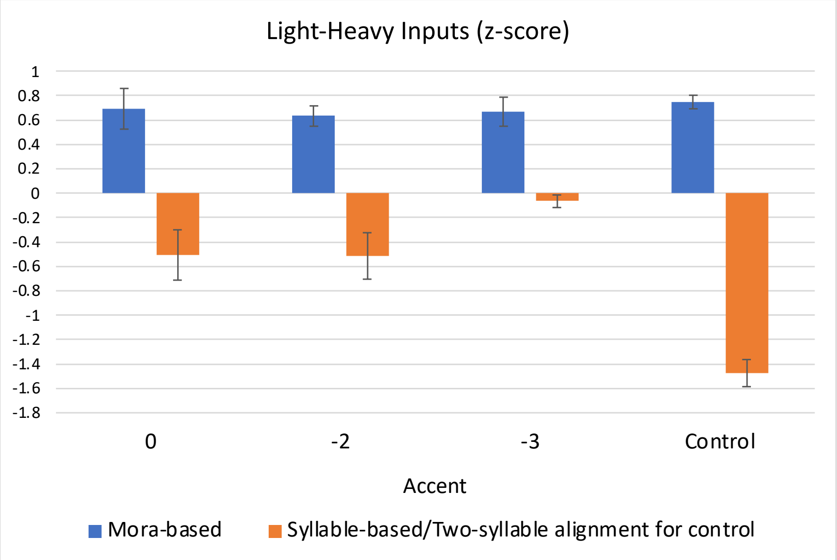


Figure 5. Normalized averages of acceptability levels for trimoraic light-heavy items

First, the mora-based alignments were significantly more acceptable than the syllable-based alignments across all categories (0 accent: t(4)=7.86, p<0.001, -2 accent: t(4)=9.49, p<0.001, -3 accent: t(4)=9.51, p< 0.001). This supports hypothesis (i).

Next, the syllable-based alignments are significantly more acceptable than the negative control group (0 accent: t(3)=7.57, p<0.01, -2 accent: t(3)=7.99, p<0.01, -3 accent: t(3)=25.70, p<0.001). This aligns with hypothesis (ii).

Hypothesis (iii) and (iv) are also mostly supported by the data. As can be read from the graphs, the syllable-based alignment for unaccented(0) and penult-accent(-2) inputs is much less acceptable than the mora-based alignment counterparts. Additionally, the syllable-based alignment for unaccented(0) and penult-accent(-2) inputs is significantly less acceptable than for antepenult(-3) inputs (0 and -3 accents: t(4)=-3.61, p<0.05, -2 and -3 accents: t(4)=-3.88, p<0.05). In other words, the syllable-based alignment among 0 and -2 accent inputs is significantly less acceptable than their mora-based alignment counterparts and their -3 accent counterparts.

An additional point to be noted is that the acceptability among the mora-based alignments of each category, including the control group, do not differ significantly (control and 0 accent: t(2)=0.41, p=0.72>0.05, control and -2 accent: t(3)=1.89, p=0.16>0.05, control and -3 accent: t(2)=0.91, p=0.46>0.05). This confirms that the mora-based alignment is an acceptable option for three mora items.

4 Analyses

In order to answer the research question, what are the motivations behind the mora-based alignment among three and four mora inputs, we will explore two possible approaches. The two primary factors that explain the alignment distribution best are the output's prosodic structure and the input's pitch accent pattern. We will first discuss the issue surrounding the indivisibility of X_2. Then, we will scrutinize the two approaches based on prosodic structure and pitch accent.

4.1 *The indivisibility of X_2* There is one significant characteristic of X_2 that we yet to discuss in depth; A maximum of only one syllable can fit into the beat. This is very irregular compared to X_1 where any amount of syllables/moras can be aligned. We have seen this pattern among the examples in the previous sections. They can be observed very clearly in a Japanese staffing firm commercial using the birthday song tune (ManpowerGroup Co., Ltd, 2022).

(8) Alignment of a long input in the birthday song

	X_1	X_2
マンパワーディア	自分らしく 働きたいあな	た
manpawaa dia	jibunrashiku hatarakitai ana	ta[9]
Manpower dear	like yourself want to work	you

"Manpower dear you who wants to work like yourself "

8 illustrates the alignment. X_2 only aligns a single syllable [ta], with no regard to the length of the whole phrase or the morpheme boundary of *anata* "you". This maximum single-syllable alignment seems to be inviolable.

The same pattern is seen in Japanese baseball chants. As Tanaka (2008) and Ito et al. (2019) describe, the chant's X_3 maximally allows one syllable. X_2 differs in that it maximally aligns a quantitative trochee. The rest of the input goes into X_1, just like the birthday song.

As described in section 2.3, one prosodic unit, likely a syllable in Japanese, can be maximally aligned to one musical beat. This is a psychological rule that is not violable. With this in mind, aligning multiple syllables into one beat signifies that, musically, the beat is divided into multiple beats so that it can associate each syllable to one beat. Hence, the birthday song rule that X_2 only allows a single syllable maximum can be rewritten as X_2 cannot be divided into multiple musical beats. On the other hand, X_1 can be divided as much as necessary.

What is causing this restriction on the last beat? One idea is that the beat X_2 aligns with the end of a musical phrase. X_2 is often sung with a fermata, i.e., without continuing the rhythm. Right after X_2 is sung, at what timing to continue singing the rest of the song is not strictly determined. Perhaps this temporary loss of rhythm marks the end of a musical phrase. If the phrasal ending aligns exactly with X_2, it makes sense that no further musical notes are allowed to be added after X_2, because that would move the position of the phrasal ending.

One point to address is that this rule on X_2 provides evidence for the syllable-inclusive theory of Japanese prosodic structure. The alignments that set the final two moras (a regular and special mora) to X_2 can only be explained using the idea of a syllable. Feet cannot account for why two-mora two-syllable sequences cannot be aligned to X_2.

4.2 *Approach 1: prosodic structure* This approach will consider a few constraints regarding prosodic structure that could decipher why the alignment variations are distributed like we found through the experiments.

As we investigated in the previous section, the last beat of birthday song alignments maximally aligns one syllable. Let us accept this as a rule for now. In Ito et al. (2019), the constraints used related to this rule were ALIGN-LEFT(X_3, σ]) and ALIGN-LEFT(X_3, μ]). They explained the baseball chant alignments by ranking these two constraints among others. In our study, one option is to rank these two constraints reversely between rankings for three-mora inputs and four-mora inputs. This is just restating the phenomenon, however, and not scraping at the principles behind the variation. Therefore, we would like to utilize just ALIGN-LEFT(X_2, σ])[10], because the syllable-

[9] The alignment can be clearly identified by the pitch change from X_1 to X_2 rather than the rhythm.

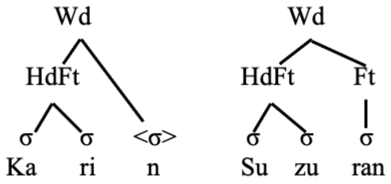
[10] X_3 is adapted to be X_2 to fit the birthday song. Either way, they are the last beat of the tune.

based alignment is likely the default alignment, and try to find a constraint that will induce a mora-based alignment when ranked above Align-Left(X_2, σ]).

There are currently two possible constraints that could be placed above Align-Left(X_2, *σ*]): IdentLength-V(HdFt), or Dep-μ(HdFt), and Binary- *σ* (HdFt). The first talks about the lengthening that happens in the first syllable when trimoraic inputs are aligned syllable-based. When a name like *Karin* is aligned syllable-based, an extra [a] (which we are assuming listeners perceive phonologically rather than as a phonetic effect[11]) is added onto the first syllable to fill up the first musical beat. The hypothesis is that this vowel epenthesis is disliked enough to pull out the mora-based alignment. The disliked epenthesis needs to be in the head foot (or what equates to X_1 in the case of trimoraic inputs), or else lengthening of the last special mora would be disfavored as well, making the mora-based alignment less favorable, which is not the situation. However, this constraint cannot explain the alignment distributions for four-mora inputs since the syllable-based alignment, the majority alignment among four-mora inputs, involves vowel lengthening while the mora-based alignment does not. If we included the constraint, the minority alignment would be chosen instead of the majority alignment.

The second possible constraint is on parsing binarity. One characteristic shared between the prosodic structures of three-mora inputs aligned mora-based and four-mora inputs aligned syllable-based (the majority alignments of three and four-mora inputs), is that the foot is parsed binarily at the syllable level. 9 illustrates this similarity.

(9) Prosodic structures for 3 and 4-mora inputs with their majority alignments[12]



Perhaps the constraint that can account for both inputs is that the head foot is syllabically binary. This constraint can explain the mora over syllable-based preference for trimoraic inputs because the syllable-based alignment creates a two-mora monosyllable head foot and the syllable level is not binary. Another constraint for ternary head feet at the mora level is necessary to rationalize the four-mora inputs because the mora-based alignment creates a trimoraic head foot. This analysis works so far for the three and four-mora input alignments found in the data.

4.3 *Approach 2: pitch accent* As we saw in the results of Experiment 2, pitch accent is likely a factor that influences the preference in alignments, if only less crucial than syllable structure. For syllable-based alignments, it is much less acceptable when the pitch falls of the accent and the melody do not line up (0 and -2 accents in the experiment) than when the pitch falls do line up (-3 accent in the experiment). We could interpret this as either the mismatch of accent and melody is avoided or the lining up of accent and melody is allowed. The difference between these two interpretations is whether we consider the syllable-based alignment fairly acceptable or unacceptable at default. Either way, there is an influence of input pitch accent pattern on the output choice.

There is evidence for the correlation of accent pitch changes and melodies in non-spontaneous text-setting for Japanese (Cho 2017), as discussed in 2.2.1. Similar patterns are observed in tonal languages like Mandarin and Vietnamese (Kirby & Ladd 2016, Wee 2007).

The issue with considering pitch accent as a significant factor is that it does not account for why -3 accent inputs, like *Karin*, which accent pitch fall lines up with the melody fall with a syllable-based alignment, still prefer a mora-based alignment. 10 and 11 illustrate the situation. Notice that the second pitch in the mora-based alignment does not match the lexical accent of *Karin*. If the pitch accent and melody matching were critical, it would make sense for the syllable-based alignment to be produced the most. However, the majority output is the mora-based alignment where the pitch falls do not line up. From this, we must conclude that, although pitch accent has an influence over alignment preference, that influence is somewhat limited.

(10) Lexical accent of *Karin*
Ka-ri-n
H L L

[11] There are several instances of segment lengthening in the text-setting inside the birthday song, and it is debatable whether these are all phonologically significant to the speaker or listener. This is not our focus today, however, so, during this analysis, we will assume that the particular lengthening relevant to the constraint at hand is phonologically significant.

[12] Foot parsing assumes that the beats in the tune act as prosodic unit boundaries. The format follows Itô & Mester (1992).

(11) Melody alignment of *Karin* in the birthday song

Syllable-based alignment	Mora-based alignment (majority)
Ka / rin	Kari / n
H L L	H **H** L

5 Conclusion and future directions

In summary, this research explored the mora-based alignment among three and four-mora names set into the Japanese birthday song and attempted to explain the motivations behind alignment variation. A production experiment was done to confirm the variation distribution. Then, a perception experiment was conducted to test the effects of pitch accent on alignment preference. With the data from these two experiments, it is likely that the output’s prosodic structure is the main factor behind alignment choices, and the input’s pitch accent can secondarily influence preference.

There is plenty more to explore to further these analyses. For instance, the two experiments conducted were mainly focused on trimoraic inputs. It would be ideal to obtain production and perception data for inputs with four or more moras to construct more general analyses for the tune. Furthermore, we have seen several factors at play for these alignments. Some items are more influenced by pitch accent than others, while others are more influenced by prosodic structure. Multiple aspects are governing the phenomenon with varying strengths of influence depending on the situation. This kind of system may go well with a Maximum Entropy-style analysis.

The birthday song may also be usable to test issues in Japanese prosody such as the existence of a superheavy syllable (discussed in Kubozono, 2021). If superheavy syllables are allowed to be aligned to X_2, that would be evidence that supports their existence. Additionally, the behavior of loanwords in this text-setting phenomenon is also intriguing. Japanese loanwords often have an epenthetic vowel at the word-final position due to the lack of coda consonants in the language’s phonology. Interestingly, these epenthetic vowels are allowed to be deleted when text-setting to the birthday song. Exploring these items would be an exciting way to examine Japanese loanword phonology.

References

Cho, Sunghye. 2017. Text alignment in Japanese children’s song. *University of Pennsylvania Working Papers in Linguistics* 23(1). 5.

Ito, Junko, Haruo Kubozono, Armin Mester & Shin’ichi Tanaka. 2019. Kattobase: The linguistic structure of Japanese baseball chants. *Proceedings of the Annual Meetings on Phonology* 7. https://doi.org/10.3765/amp.v7i0.4470.

Itô, Junko & Armin Mester. 1992. Weak layering and word binarity. *Ms., University of California, Santa Cruz*.

Kawahara, Shigeto. 2015. 11 The phonology of Japanese accent. In *Handbook of Japanese phonetics and phonology*, 445–492. De Gruyter Mouton.

Kawahara, Shigeto. 2016. Japanese has syllables: a reply to Labrune. *Phonology* 33(1). 169–194. https://doi.org/10.1017/S0952675716000063.

Kirby, James & D Robert Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. In, vol. 10, 2016–10.

Kubozono, Haruo. 2021. *Ippangengogaku kara mita nihongo no purosodii: Kagoshima hougen wo chuushin ni [Japanese Prosody from General Linguistics Perspectives]*. Tokyo, Japan: Kurosio Publishing. https://go.exlibris.link/Dwh6LCQP.

Labrune, Laurence. 2012. *The Phonology of Japanese*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199545834.001.0001.

Starr, Rebecca Lurie & Stephanie S. Shih. 2017. The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting. *Glossa: a journal of general linguistics* 2(1). https://doi.org/10.5334/gjgl.355.

Tanaka, Shin’ichi. 2008. *Rizumu/akusento no “yure” to on’in/keitai-kouzou [Fluctuation in rhythm and accent and phonological and morphological structure]*. Tokyo, Japan: Kurosio Publishing. https://go.exlibris.link/dPGq9XCk.

Wee, Lian Hee. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics* 35(1). 128.

ManpowerGroup Co., Ltd. 2022. *Manpower to You “Uta yo hibike” hen [Manpower to You version “Uta yo hibike”]*. https://youtu.be/VdTUcr0B9r4.