



Prosodic Word Stress in Textsetting in Thai Pop Songs

Komtham Domrongchareon^{1,2}, Pittayawat Pittayaporn^{2,3}

¹Faculty of Music, Silpakorn University

²Department of Linguistics, Chulalongkorn University

³Center of Excellence in Southeast Asian Linguistics, Chulalongkorn University

domrongchareon_k@silpakorn.edu, Pittayawat.P@chula.ac.th

Abstract

This research examined the relationship between Thai prosodic words and musical note pairs, focusing on prominence matching and quantity matching. We hypothesized that the iambic structure of Thai disyllables would result in matching with more prominent musical positions and longer note duration. An analysis of 40 most frequently used disyllables from a 4,078,300 word corpus of Thai pop lyrics however revealed that prominence is not the main factor determining Thai textsetting. The sampled disyllables were found to match with initial prominence and final prominence note pairs roughly equally (48.8% v. 51.1%). On the contrary, quantity matching played a crucial role, with disyllables preferring to map to short-long note pairs (72.8%) especially at the end of musical phrases (92.2% of all phrase-final cases). Nevertheless, prominence matching plays a secondary role, resulting in the tendency of even disyllables to align with the prominence-final note pairs (76.8% of all even pair). This study thus demonstrates that, in contrast to previously studied languages, final stressed syllables in Thai prosodic words match with extended note quantity in textsetting, making musical prominence secondary. The importance of note quantity over musical prominence suggests the iambic/Trochaic Law's role in the language-music interaction.

Index Terms: prosodic words, textsetting, prominence matching, quantity matching, word stress

1. Introduction

The predominant correspondences between text and music involve pitch level (*musica localis*) and temporal level (*musica temporalis*) [1], [2]. Research on the temporal correlation in textsetting (e.g. [3], [4], [5], [6], [7], [8]) has shown that prominence matching, where linguistic prominence matches musical metrical prominence, is a common constraint cross-linguistically. Yet, the findings are based on limited studies on typological aspects, most of which examines English, other major European languages and only few others.

This study explores the roles of prominence matching and quantity matching in mapping prosodic word stress in Thai with a musical melody. The former involves the correlation between prominent syllables in words and their positions in the musical metrical structure (specifically, the beat into which prominent syllables fall), while the latter focuses on the correlation between syllables and their note values in music (the terms 'duration' and 'length' are avoided due to their strong connotations in linguistics). Furthermore, the interaction between prominence and quantity matching will be explored.

Since the best-studied language English is characterized by the trochaic foot type [3], and although the Iambic/Trochaic Law [9] posits that iambic and trochaic systems demand different types of prosodic contrasts, the temporal aspect of textsetting in iambic languages, such as Thai [10], may be subject to different constraints than those in their better-studied trochaic counterparts. Given that the Thai prosodic word is characterized by binary feet [11], each consisting of two syllables, and favors disyllabic roots, typically containing a single foot [12] with a final stress that is claimed to be associated with contrast in duration, this characteristic of disyllabic words makes them a suitable subject for this research.

We hypothesized that Thai disyllables are preferably aligned with note pairs in which the final member has greater quantities than those with longer initial syllable. If the condition is not satisfied, the final syllable is expected to match musical prominence. Furthermore, a disyllabic word in a phrase-final position should always exhibit its final syllable in greater quantity because a relatively long temporal distance between two successive notes tends to serve as a decisive indication of a grouping boundary [3].

Importantly, this study concentrates on pop music because pop songs are typically composed in a less strict given form, exhibit a more spontaneous compositional structure in relation to their lyrics, and represent contemporary musical practices compared to the refrain form. In this respect, it differs from most previous studies which are based on traditional chanted verses, folk songs, and nursery rhymes, in which songs are based on the refrain form (with different alternating verses and the same melody).

2. Literature review

Linguistic prominence, also referred to as "stress", is the property of a linguistic element standing out from its environment, typically due to its prosodic characteristics, such as intensity (amplitude), duration, and pitch [9]. Lexical stress is the prominence relation between the syllables of a phonological word, where a syllable occupies a relatively stronger position in the metrical constituent structure than other syllables of the same word [13]. Furthermore, a word that immediately dominates the foot is the lowest unit in the prosodic hierarchy that can manifest stress [14].

In the context of textsetting, the relationship between lexical stress, musical prominence, and quantity matching has been observed. On the one hand, it is generally agreed that there is a connection between linguistic stress and musical prominence (or strong beat), in which prominent linguistic elements tend to align with metrically prominent events in music. This phenomenon is also referred to as *prominence matching* or *stress-to-beat matching* [3], [4], [5], [6], [7], [8].

Prominence matching is a strong constraint in English (e.g. [4], [15]), whereas in Spanish [16], [17] and French [4], this rule has more flexibility, allowing mismatches to exist between prominence in speech and music [18], [19]. In Spanish and French, prominence matching is compromised when the distributions of syllables in two verses along the melody are identical, while English tends to redistribute note values.

On the other hand, quantity matching has a long history since ancient Greek where syllable weight has a direct correspondence with note quantity, regardless of prominence [20]. In Hausa, where chanted verses are constructed on a quantitative meter, it seems that quantity matching is reflected more faithfully in textsetting than stress meters are [21]. In a recent study on English, Hayes and Kaun [22] posit that the mapping of syllables to a discrete metrical grid is based on the phonetic durations of those syllables. Not confined to English, quantitative textsetting has been observed in Italian [7] and Urdu [5]. Moreover, final lengthening has also been observed to have a strong effect on quantity matching [3], [23].

Since tonal music is constructed on a stress meter (which is the root of European music and pop songs) rather than a quantity meter, there is an interplay between prominence matching and quantity matching. The former was never an absolute constraint, and the common practice of suppressing prominence matching to fulfill additional constraints is evident, such as in Spanish [16], [17] and French [4], as mentioned above. In English, where the degree of prominence matching seems stronger than in other languages, prominence matching is still pliable, interacting with the avoidance of sequences where no syllable is initiated, or STRONG-IS-LONG and preference for longer syllables in a strong beat [15].

However, distinct patterns may emerge when comparing languages and materials, where most languages (English and other European languages) and song forms (such as chanted verses, nursery rhymes, and refrains with fixed melodic phrases) maintain a robust association between linguistic stress and the salience of the musical metrical structure. In the case of disyllabic lexical stress in Thai, it adheres to an iambic foot structure [10], [12], wherein, according to the Iambic/Trochaic Law [9], “*elements contrasting in duration naturally form grouping with final prominence.*”

The existing literature reveals a significant research gap in the understanding of how a disyllable in Thai adheres to an iambic foot structure, particularly in relation to musical prominence and quantity in the textsetting in pop song. This gap presents an opportunity to investigate the relationship between linguistic stress and textsetting in iambic language, thereby contributing to a more comprehensive understanding of the universality of these patterns.

3. Method

3.1. Material

A corpus of Thai pop song lyrics, released from 2002–2019, was assembled by extracting text from the website Siam Zone [24]. This corpus comprises 4,078,300 tokens, representing a collection of 15,412 songs. Further, the 40 most frequently occurring disyllabic words were identified within the corpus, and 20 random samples of each of these words were selected from the corpus. Every target word was extracted from distinct songs. The final output consisted of 800 tokens.

For the consistency of the musical metrical structure, all songs for which the token was included must adhere to a simple binary meter, such as 4/4, 2/4, or 2/2, for their rhythmic structure (time signature). Furthermore, the target term should be employed within a melodic lyric (rap excluded). If a word recurs within the same song, only its initial occurrence will be taken into consideration.

3.2. Annotation

Due to the absence of automated tools for extracting the required information, all annotations were audibly and manually performed by the first author, who has been a university-level instructor in music and musicology for 15 years. Each token was labeled with its musical prominence and quantity profile.

For the prominence profile, each syllable in the disyllable was recorded according to its position (beat fall) in the music metrical structure and then compared. Since prominence in music is not a binary distinction but rather involves multiple metrically relevant levels, this study employed a relative system of prominence assessment [15], examining the distinctions between the ‘stronger’ and ‘weaker’ positions to which each syllable aligns. Finally, the token was annotated in two distinctive categories: in cases where the initial syllable of the word matches the stronger musical event, it was tagged as ‘initial prominence’, whereas when the final syllable matches the stronger musical event, it is classified as ‘final prominence’.

For the quantity profile, the note value of each syllable in a disyllable is observed using relative measurements according to its musical rhythmic structure. In the melodic–phrase–initial and melodic–phrase–middle contexts, syllables are encoded based on their inter-onset intervals (IOIs) [23]. Meanwhile, syllables in melodic–phrase–final positions are coded from their onset to their coda. In cases where the coda is ambiguous, such as in a long fadeout, it is coded with a maximal value of four beats. The values of each note, for which the syllable in a word was coded, are compared across three distinct categories: a) the first syllable being lesser in value than the second ($1^{st} < 2^{nd}$); b) the first syllable being equal in value to the second ($1^{st} = 2^{nd}$); and c) the first syllable being greater in value than the second ($1^{st} > 2^{nd}$).



Figure 1: Example of musical notation. แก้ม วิชญาณี. 2014. “ที่หนึ่งในหัวใจเธอ”. GMM.

Table 1: Example of extracted information from the music in Figure 1.

Token	Prominence matching	Quantity matching
หัวใจ [huə caj] ‘heart’	1, 2 ‘initial prominence’	1, 1.5 a) $1^{st} < 2^{nd}$

For a recheck, an additional annotator was enlisted. 80 tokens were randomly drawn and re-annotated. Cohen's kappa coefficient was measured for inter-rater reliability between the

two annotators, yielding values of $\kappa = 0.82$ (95% CI [-0.70, 0.94]) for prominence and $\kappa = 0.79$ (95% CI [0.65, 0.94]) for quantity. This indicates good agreement in both profiles.

3.3. Procedure and data processing

The frequencies of prominence matching and quantity matching categories are tabulated. Given the categorical nature of both variables, an initial step involves the statistical significance measurement of the relation between prominence and quantity matching. The application of the chi-squared (χ^2) and Cramér's V statistical tests were performed. Subsequently, the correlation coefficient was calculated via the multinomial regression model, examining the relationship between the two variables. All statistical analyses were performed utilizing MATLAB[®] 2023a.

4. Results

4.1. Prominence matching vs. quantity matching

Table 2: Frequency count of prominence matching and quantity matching ($n = 800$).

	1st < 2nd	1st = 2nd	1st > 2nd	Total
Initial	333	42	16	391
Final	250	139	20	409
Total	583	181	36	800

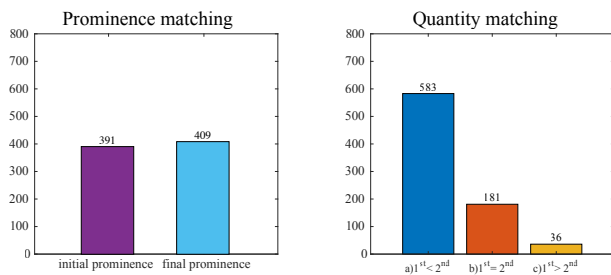


Figure 2: Frequency count by prominence matching and quantity matching.

Our results show that quantity matching is an important factor in Thai textsetting; however, prominence matching also plays a supportive role. For quantity matching, 583 tokens (72.87%) are mapped to note pairs with a greater quantity in the final syllable (1st < 2nd), while only 181 tokens (22.62%) are mapped to pairs of an equal quantity (1st = 2nd) and 36 tokens (4.5%) are mapped to those with a lesser quantity in the final syllable (1st > 2nd). Given the iambicity of Thai prosodic words, this pattern strongly suggests a preference for lexical stress to be matched with a note with a greater quantity and to avoid disproportionate pairing. As for prominence matching, 391 (48.87%) and 409 (51.12%) tokens were matched with note pairs with initial and final prominence, respectively, indicating an absence of accordance between lexical stress and musical prominence.

However, when considering the interactions between the two factors, the effect of prominence matching can be observed. Among the 391 initial prominence pairings, a substantial majority, accounting for 333 tokens (81.41%), are matched with note pairs with a greater quantity of the final syllable (1st < 2nd) and 16 tokens (3.91%) to those characterized by a greater quantity in the initial syllable (1st > 2nd). Importantly, only 42

tokens (10.26%) are mapped to pairs with an equal note quantity (1st = 2nd). This pattern is noticeably different from that of the final prominence pairings, which are more tolerant of equal quantity. Among the 409 final prominence pairings, 250 tokens (63.93%) are matched with note pairs with a greater quantity in the final syllable (1st < 2nd) and a minority of 20 cases (5.11%) with those displaying a greater quantity in the initial syllable (1st > 2nd). Importantly, 139 cases (35.54%) are mapped with note pairs of equal quantity (1st = 2nd), a much higher figure than in the initial prominence cases. This differential pattern suggests that prominence matching helps reinforce iambicity only when quantity matching fails to do so.

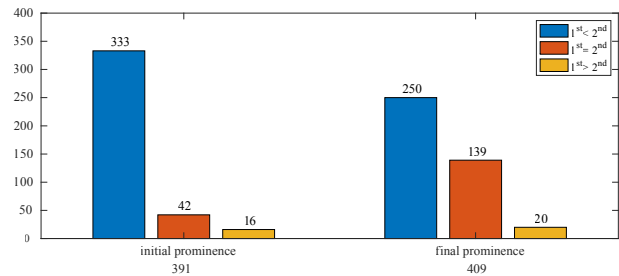


Figure 3: Frequency count of note quantities grouped under prominence matching.

A further analysis was performed to test whether there was a significant association between prominence matching and quantity matching. The chi-squared test was performed between both variables. There was a statistically significant relationship ($\chi^2 = 66.7$; $p < 0.001$; Cramer's $V = 0.28$). In addition, multinomial regression analysis was employed to assess the statistical significance of quantity matching in predicting prominence matching. The resulting regression model is expressed as follows: $Prominence = -0.28 + 1.48X + 0.50Y$, where X represents notes equal in quantity and Y signifies the scenario wherein the initial syllable exceeds the final syllable in quantity. The overall regression was statistically significant ($\chi^2 = 66.7$; $p < 0.0001$).

The model indicates that when the note quantity of the initial syllable is less than that of the final syllable, there is a weak tendency for the musical prominent to match with the initial syllable (-0.28; 95% CI [-0.45, 1.64]; $p < 0.001$; Although statistically significant, this observed tendency is characterized by a modest effect size. However, when a word has an equal note quantity for both the initial and final syllables, there is a noticeable increase in the trend of musical prominence to match with the final syllable (1.48; 95% CI [1.10, 1.86]; $p < 0.001$). Regardless, when the note quantity of the initial syllable exceeds that of the final syllable, this difference is not statistically significant (0.50; 95% CI [-0.16, 1.18]; $p = 0.14$). This lack of significance is attributed to the limited number of tokens.

4.2. Word position

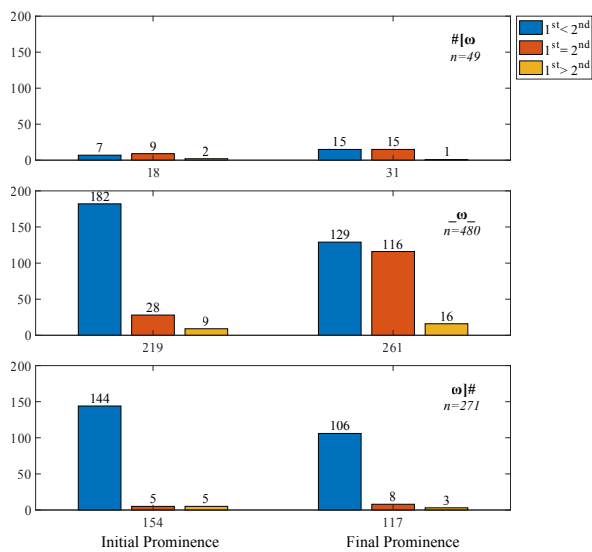


Figure 4: Frequency count by word position in quantity matching grouped under prominence matching.

When examining the positioning of words within the phrase, one can discern their impact in the context of Thai textsetting. We found that word position influences the quantity. When the disyllabic word occurs within a phrase (ω), the word does not show a significant preference for prominence matching. While considering prominence matching by category, with musical prominence matching with the final syllable, the number of words with equal quantities is very close to those with a greater quantity of the final syllable (129 vs. 116 tokens). In the scenario where musical prominence matches the initial syllable, words tend to have greater quantities of the final syllable.

While considering words at the phrase-final (ω |#), they do not exhibit a preference for musical prominence. However, words in the phrase-final position show a strong and distinct preference for having the word end with a greater quantity, both in cases where prominence matches the initial and final syllables. When one controls for the final lengthening factor, the effect of the word's position in a phrase indicates that the distribution of an equal note quantity decreases significantly in phrase-final positions compared to phrase-middle positions (-2.14 ; CI $[-1.55, -2.73]$; $p < 0.001$). This indicates that final lengthening is a strong constraint in Thai textsetting. Further, phrase-final lengthening has been observed systematically in both language [25] and textsetting [3], [23]. Given that Thai manifests right-headedness at the phrase level [10], the phrase-final position appears to emphasize asymmetry in duration. Finally, the category of phrase-initial occurrences ($\#[\omega]$) yields limited information due to an inadequate sample size.

5. Discussion

Textsetting in Thai pop song melodies shows that disyllables exhibit a preference for note pairs where the final syllable carries greater quantity regardless of whether they show initial or final prominence. In instances where quantity was distributed equally across words, the final syllable tended to align with musical prominence. This alignment can be interpreted as a

compensatory mechanism aimed at satisfying the word-final stress condition.

It is observed that phrase-final-lengthening and prominence matching still exist in Thai pop song textsetting. Prominence matching is apparently influenced by quantity matching, which deviates from prior studies. This is consistent with the correspondence between word stress and quantity matching. However, this consistency is subject to additional factors when the asymmetry of quantity is not met. These findings diverge from studies in other languages that typically emphasize prominence matching. Such divergence supports the Iambic/Trochaic Law [9], where the grouping of elements with contrasting durations naturally leads to final prominence.

It is worth noting here that the syllable weight factor was not entirely overlooked. As the target words were retrieved by frequency, most of the frequent disyllables were compounded and exhibited the heavy-heavy structure. Only two target words out of 40 exhibited a light-heavy structure. Regardless, an analysis of weight was conducted. However, the analysis did not yield results significantly different from the general findings when all tokens were analyzed.

6. Conclusion

In conclusion, this study reveals that, in connection to the Thai language, as an iambic language, quantity matching plays a primary role in textsetting, while prominence matching serves as a supporting factor when quantity-related conditions are not met. This is significant because it demonstrates that, in Thai as an iambic foot type, the textsetting realization was found to be different from trochaic languages.

7. References

- [1] P. Beda, *Musica quadrata seu mensurata*. C13.
- [2] C. Palmer and S. Hutchins, "What Is Musical Prosody?," in *Psychology of Learning and Motivation*, vol. 46, Elsevier, 2006, pp. 245–278. doi: 10.1016/S0079-7421(06)46007-2.
- [3] J. Halle, "Constituency matching in metrical texts." Department of Music, Yale University, 2004.
- [4] J. Halle and F. Delle, "Comparing musical textsetting in French and in English songs.," *Towards a typology of poetic forms*, pp. 63–78, 2009.
- [5] P. Kiparsky, "Stress, Meter, and Text-Setting," in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds., Oxford University Press, 2020, pp. 656–675. doi: 10.1093/oxfordhb/9780198832232.013.46.
- [6] F. Lerdahl and R. Jackendoff, *A generative theory of tonal music*, Repr. Cambridge, Mass.: MIT Press, 1983.
- [7] T. Proto and F. Dell, "The structure of metrical patterns in tunes and in literary verse. Evidence from discrepancies between musical and linguistic rhythm in Italian songs," *Probus*, vol. 25, no. 1, pp. 105–138, May 2013, doi: 10.1515/probus-2013-0004.
- [8] N. Temperley and D. Temperley, "Stress-meter alignment in French vocal music," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 520–527, Jul. 2013, doi: 10.1121/1.4807566.
- [9] B. Hayes, *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press, 1995.
- [10] J. F. Bennett, "Iambicity in Thai," *Studies in the linguistic sciences*, vol. 24, pp. 39–57, 1994.
- [11] T. Luangthomgkum, "Rhythm in standard Thai," Edinburgh, 1978.
- [12] M. Brunelle and P. Pittayaporn, "Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia," *DIA*, vol. 29, no. 4, pp. 411–433, Dec. 2012, doi: 10.1075/dia.29.4.01bru.

- [13] A. Cutler, "Lexical Stress in English Pronunciation," in *The Handbook of English Pronunciation*, 1st ed., M. Reed and J. M. Levis, Eds., Wiley, 2015, pp. 106–124. doi: 10.1002/9781118346952.ch6.
- [14] M. Nespors and I. Vogel, *Prosodic Phonology: With a New Foreword*. DE GRUYTER, 2007. doi: 10.1515/9783110977790.
- [15] B. Hayes, "Textsetting as constraint conflict," in *Language Faculty and Beyond*, vol. 2, J.-L. Aroui and A. Arleo, Eds., Amsterdam: John Benjamins Publishing Company, 2005, pp. 43–62. doi: 10.1075/lfab.2.02hay.
- [16] T. A. Morgan and R. D. Janda, "Musically conditioned stress shift in Spanish revisited: empirical verification and nonlinear analysis," *Studies in Romance linguistics*, pp. 273–288, 1989.
- [17] R. Rodríguez-Vázquez, "Text-setting Constraints: A Comparative Perspective," *Australian Journal of Linguistics*, vol. 30, no. 1, pp. 19–34, Jan. 2010, doi: 10.1080/07268600903134004.
- [18] T. Proto, "Singing in German: text-setting rules and language rhythm," *Atti del IX convegno AISV*, pp. 9–10, 2013.
- [19] T. Proto, "Reality and perception of stress-beat mismatches in German Volkslied," presented at the 9th International Conference on Speech Prosody, Póznán, Poland. Retrieved, 2018.
- [20] M. L. West, *Ancient greek music*. Clarendon Press, 1992.
- [21] B. Hayes and R. G. Schuh, "Metrical structure and sung rhythm of the Hausa rajaz," *Language*, vol. 95, no. 2, pp. e253–e299, 2019, doi: 10.1353/lan.2019.0043.
- [22] B. Hayes and A. Kaun, "The role of phonological phrasing in sung and chanted verse," *The Linguistic Review*, vol. 13, no. 3–4, 1996, doi: 10.1515/tlr.1996.13.3-4.243.
- [23] K. M. Ryan, "Syllable weight and natural duration in textsetting popular music in English," *English Language and Linguistics*, vol. 26, no. 3, pp. 559–582, Sep. 2022, doi: 10.1017/S1360674322000156.
- [24] "เนื้อเพลง เนื้อเพลงใหม่." Jan. 2021. [Online]. Available: <https://www.siamzone.com/music/thailyric/>
- [25] A. E. Turk and S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *Journal of Phonetics*, vol. 35, no. 4, pp. 445–472, Oct. 2007, doi: 10.1016/j.wocn.2006.12.001.