A Cross-linguistic Investigation on the Correlation between Functional Load of Tone and Tone-Melody Correspondence

Natsumi Taniguchi
University of Southern California

1 Introduction

Text-setting, or how lyrics are aligned to melodies, has been utilized in phonetic and phonological studies over the years because it can reveal the cognitive salience of certain linguistic properties in a language (e.g. Hayes and Kaun, 1996; McPherson, 2018; Starr and Shih, 2017). An intuitive example of text-setting for English speakers is shown below with regard to how lexical stress is set to musical rhythm.[1] The two lines of music below illustrate two ways to align the word *linguistics* to the name-insertion portion of the Happy Birthday song. The first shows what most native English speakers would produce, while the second is an alignment that most native English speakers would find unnatural for *linguistics*, but perhaps natural for a word like *melody*. The difference lies in whether the syllable with primary stress [g^{w}ɪs] is aligned with a metrically strong position (i.e., the first beat of the third measure) or a metrically weak position (i.e., the fourth sixteenth note within the first beat of the third measure) in the music.

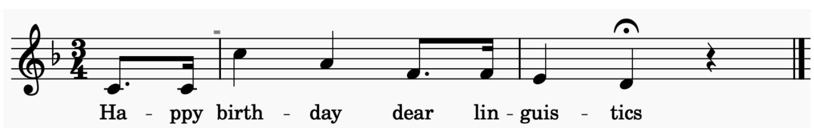

Figure 1: Happy Birthday song sung to "linguistics" with licit text-setting

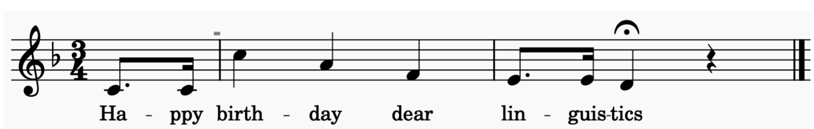

Figure 2: Happy Birthday song sung to "linguistics" with illicit text-setting

One might wonder, why do preferences like this exist? We can think about songs with lyrics as a balancing of sometimes-opposing musical and linguistic pressures. On the one hand, musical pressures, such as maintaining musical structure (phrase structures, metrical rhythm structures, chord progressions, etc.) and aesthetics (melodic choices, syncopation choices, etc.), demand control of acoustic dimensions (F0, duration/timing, etc.). On the other hand, linguistic pressures, such as the pressure to get the lyrics across to the listeners, either tweak the acoustics (especially spectral aspects, but F0 and duration/timing/etc. as well)

* Many thanks to Stephanie Shih, Canaan Breiss, Louis Goldstein, and colleagues at USC for their guidance and support. I'm also grateful for all of the great comments, questions, and suggestions I received at AMP 2025. All mistakes are mine.

1 This pattern in English (and other stress languages) has been studied extensively in verse more than in song (e.g. Hayes et al., 2012; Kiparsky, 1977)

© 2026 Natsumi Taniguchi
Proceedings of *AMP 2025*

or take advantage of musical events (such as utilizing musical prominence to encode lexical stress), resulting in the encoding of linguistic properties in the song. Preferences for text-settings such as those in Figure 1 can be thought of as the result of these sometimes-opposing pressures finding an acceptable balance.

This framing of text-setting lets us consider what linguistic properties are prioritized in their encoding and why. Given the specific linguistic pressure of getting the lyrics across to the listeners, we can posit a possible, partial answer to this question: if a phonological contrast is more useful for accurately recognizing speech in a language (in other words, useful for reducing confusability in speech in a language), then it is more likely to be encoded in songs with lyrics. The current study tests a narrower, cross-linguistic version of this hypothesis, focusing on tones: If tone is more important for reducing confusability in speech in language A than in language B, then tone would be more reflected in songs of language A than in songs of language B. This specific issue has been previously mentioned in the field. For instance, Schellenberg (2013) speculates on the effect of the functional load of tone within a language, particularly in explaining the difference in the degree of tone-melody correspondence between Cantonese and Mandarin. However, no quantitative investigation of this speculation was conducted within the paper.

The current study is a quantitative investigation of this issue. Functional load, or, intuitively, the work that a phonological contrast does in maintaining meaning contrasts within a language, was used to quantify the usefulness of tone in reducing confusability in speech. Degree of tone-melody correspondence and functional load of tone in Cantonese, Mandarin, and Japanese[2], were obtained and compared. Additionally, a meta-analysis of previous work on tone-melody correspondence and functional load for tone in Mandarin, Vietnamese, and Thai[3] was carried out. Overall, the results point towards a correlation between the functional load of tone and the degree of tone-melody correspondence across the five languages. This result supports the prediction that phonological features that are more important for accurately recognizing speech are more likely to be encoded in text-setting, inviting further investigation into whether other linguistic features also follow this pattern. More broadly, this finding can serve as additional evidence for the speech production process having some level of access to lexical knowledge, such as the functional load of phonological contrasts, and that listener-oriented communicative pressures may be a key motivating factor in the use of this knowledge.

2 Background

2.1 *Tone in the five languages* Given that Tokyo Japanese is often described as a "pitch accent" language rather than a tone language, this section will start by stating that "tone" is defined very broadly in this study. Specifically, "tone" will be defined as the use of pitch as a way to make contrasts between lexical items. Whether that is a good or the best definition for phonological theory is beyond the scope of this study. What is crucial here is that pitch used to distinguish meaning at the lexical level may interact with musical pitch in the process of setting lyrics to a melody.

The five languages examined in this study have a range of tonal systems. First, Cantonese, Vietnamese, Mandarin, and Thai are often described as lexical tone languages where each syllable, often a morpheme, is specified for a tone out of a set of tone units. Thai is said to have more bisyllabic morphemes than the other three languages. Overall, songwriters of these four languages have relatively high flexibility when choosing tones that match or do not contradict the melody.

Cantonese and Vietnamese have the larger tonal systems in this set, both commonly described with six tones, including several contour tones. Each tonal system is laid out below in Tables 1 and 2.[4]

Mandarin and Thai also have fairly complex tonal systems, as shown in Tables 3 and 4. Note for Thai that obstruent-ending syllables are restricted in what tones they can occur with (low or falling with long vowels, low or high level with short vowels). Thai tones are also said to have limitations on the possible tonal combinations for bisyllabic morphemes.

Note that the tones in all four systems are differentiated mainly by F0, but with additional phonetic dimensions as well, such as phonation and duration.

[2] The varieties are Hong Kong Cantonese, Standard Mandarin (Putonghua), and Tokyo Japanese, respectively.

[3] Standard varieties were the focus for Vietnamese and Thai as well.

[4] All tonal system descriptions of Cantonese and Mandarin are based on Yip (2002) and Chan (1987). Yip (2002) is used for Vietnamese and Thai as well. Thai is also based on Ketkaew and Pittayaporn (2015). Descriptions for Japanese are based on Kawahara (2015). See these references and others for more detailed descriptions.

high level	[55]
mid rising	[35]
mid level	[33]
low falling	[21]
low rising	[13]
low level	[22]

Table 1: Cantonese (Hong Kong) tonal inventory

high level	[45]
high rising	[35]
low level	[2]
low falling	[21]
lower concave	[214]
higher concave	[415]

Table 2: Vietnamese (Northern) tonal inventory

high level	[55]
high rising	[35]
low falling/level	[21] [214] at utterance-final
high falling	[51]

Table 3: Standard Mandarin tonal inventory

high level	[45]
mid level	[33]
low level	[21]
falling	[42]
rising	[24]

Table 4: Standard Thai tonal inventory

Japanese is the odd one in this bunch, often categorized as a pitch accent language rather than a tone language. The language has minimal pairs of words that are identical in segmental material but are distinguished by their pitch contours over the word, meaning pitch does play a contrastive role. However, words are often multisyllabic, and the pitch contour of a word is largely rule-governed and predictable except for the existence and position of the high-low tonal sequence. An example minimal triplet is shown below in (1). Each tone ("H" for high, "L" for low) is associated with each mora (mora boundaries are marked with "."). The first two words are distinguished by the position of the high-low tone sequence (marked with an apostrophe) and the third by the absence of a high-low tone sequence. The range of possible tone contours is severely restricted, where sequences like HLH, LLH, or LLL are not possible.

(1) a. /ha.shi.+ga/ "chopsticks + NOM"
H' L L
b. /ha.shi.+ga/ "bridge + NOM"
L H' L
c./ha.shi.+ga/ "edge + NOM"
L H H

To summarize, the four more canonical tone languages have varying, but relatively high levels of flexibility in choosing lyrics that can tonally match the melody, while Japanese is quite limited despite pitch having a contrastive role in the language. Note that all five languages have tonal alternation processes to some capacity, whether it be phonologically or morphologically triggered. This means that underlying sequences of tones do not necessarily line up with the surfacing pitch contours.

2.2 *Tone-melody correspondence* A particularly pitch-focused sub-genre of the text-setting literature explores tone-melody correspondence: the matching, or lack thereof, between linguistic tones and melodies. An important finding has been that linguistic tones do not correspond to individual musical notes (absolute pitches or pitch classes) in systematic ways, but rather tend to match the direction of pitch change in the tones and the melody over multiple syllables/musical notes. Given this property of tone-melody correspondence, Ladd and Kirby (2020) suggests a descriptive framework for looking at pairs of tones/musical notes, as illustrated in Figure 3. Repurposing terminology from classical Western music theory, *contrary setting* refers to a consecutive pair of tones/notes, or a bigram, where the tones move in one direction (up or down) while the melody moves in the other direction (down or up). *Similar setting* refers to a bigram where the tones and notes move in the same direction (up, down, or no movement), regardless of interval. Lastly, *oblique setting* is a bigram where either the tone or melody moves up/down while the other stays level. Ladd and Kirby (2020) differentiates between the two oblique settings (one where the melody stays level and one where the tone stays level) to allow for more detailed analyses of linguistic and musical factors. For even more detail, one can categorize the bigrams into the nine possible pairings between rising, falling, and level melody bigrams and rising, falling, and level tonal bigrams. The nine pairings in relation to the categories in Ladd and Kirby (2020) are shown in Table 5. This paper will utilize this framework to discuss tone-melody correspondence in a quantitative way.

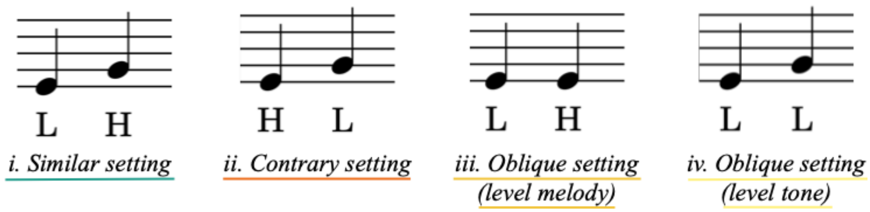

Figure 3: Bigram pair settings in tone-melody correspondence as categorized in Ladd and Kirby (2020). L and H represent low and high linguistic tones, respectively.

		melody		
		up	down	level
	up	**similar**	*contrary*	oblique
tone	down	*contrary*	**similar**	oblique
	level	oblique	oblique	**similar**

Table 5: The nine possible pairings of melody and tone bigrams, categorized into the settings defined by Ladd and Kirby (2020)

Ladd and Kirby (2020) report that the overall pattern found in tone-melody correspondence studies has been to "avoid contrary settings." In other words, decreasing the number of contrary settings is prioritized over increasing the number of similar settings, although some languages, such as Cantonese, are able to utilize both strategies (Chan, 1987; Ho, 2010; Wong and Diehl, 2002). The pattern is observed in numerous tone languages across many song cultures, including traditional Zulu dance songs and Xhosa songs (Rycroft, 1959), traditional Mixtec songs (Pike, 1946), and Tommo So women's folk songs (McPherson and Ryan, 2018), indicating that tone-melody correspondence cannot be attributed to particular cultures or areas. Rather, it seems to be a common characteristic among songs in languages that utilize linguistic tone in one way or another.

So, how do the five languages focused on in this study behave? Cantonese songs are well known to have high levels of correspondence, especially in the "Cantopop" genre (e.g., Chan, 1987; Ho, 2010; Li, 2021; Schellenberg, 2013; Wong and Diehl, 2002). Vietnamese pop songs are also said to have a fairly high rate, as reported in Kirby and Ladd (2016). Thai pop songs are less successful, especially in maximizing similar settings, where Ketkaew and Pittayaporn (2015) reports 57.7% similar and 11.5% contrary settings. Japanese songs are similar to Thai, though differences between genres have been observed. Traditional children's songs examined in Cho (2017) show 54.2% similar and 5.77% contrary settings, following the avoid contrary pattern. Pop songs are reported to have lower correspondence, however (Kitayama et al., 2024). Lastly, Mandarin has been known to have minimal correspondence, exemplified in Kirby (2022) where pop songs had 38% similar and 25% contrary settings.

Where these cross-linguistic differences come from is the focus of the present study. The stark difference in numbers between Mandarin and Cantonese has been a well-known puzzle in the literature, given their many linguistic and cultural similarities. There have been some suggestions of Information-Theoretic reasons for these cross-linguistic differences, including Schellenberg (2013), which speculates that the functional load of tone in Mandarin and Cantonese is a factor. There have also been implications of communicative pressures in the literature, notably Li (2021) reporting that Cantonese songs allow for contrary settings when the minimal pair word created by the melody does not make sense in the linguistic context. The current study, as we will see in later sections, finds that this communicative pressure may have cross-linguistic effects as well.

2.3 *Functional load* Intuitively, functional load can be thought of as how much work a phonological contrast is doing for meaning contrasts within a language. The concept of functional load in linguistics emerged in the early 20th century from the Prague School (e.g., Mathesius, 1929; Trubetzkoy, 1939), and many researchers have explored its possible relation to the diachronic loss of a phonological contrast (e.g., Wedel, Jackson, and Kaplan, 2013; Wedel, Kaplan, and Jackson, 2013). Wedel, Kaplan, and Jackson (2013) provides statistical evidence supporting this relation. Because meaning contrasts are crucial for accurately recognizing words and, ultimately, understanding speech, the present study utilizes functional load as a proxy for how much work tones do for the accurate understanding of speech.

Methodology for calculating functional load has been debated over the years, ranging from simple minimal pair counts to system entropy-based methods. Each method picks up on certain properties of a language. The corpus study detailed in the next section utilizes a minimal pair count methodology. This method calculates functional load by simply counting the number of minimal pairs a phonological contrast differentiates in a language and dividing that number by the total number of minimal pairs in that language. This methodology reflects the lexical inventory of a language with little regard for token frequency in actual speech or sentential context. Wedel, Kaplan, and Jackson (2013) reports that this measure predicts diachronic phoneme contrast loss better than other measures. The meta-analysis of the literature, described in Section 4, takes values calculated with the system entropy-based method from Surendran and Niyogi (2003, 2006). This method takes a corpus and calculates how much redundancy there is overall for a version of the corpus with the phonological contrast in question and a version without the contrast. The relative difference between the redundancy levels of the two versions is the functional load value, which can be intuitively understood as how much work the phonological contrast is doing for the predictability of words in the language. This methodology reflects actual usage of the lexical inventory, meaning token frequency and sentential context are accounted for. The current study cannot conclude whether either methodology better predicts tone-melody correspondence.

3 Corpus study: Cantonese, Mandarin, and Japanese

This portion of the study obtains tone-melody correspondence proportions from songs in Cantonese, Mandarin, and Japanese, calculates the functional load of tones in the three languages using speech corpora or dictionaries, and, lastly, compares the two measures to test for a correlation.

3.1 *Methods* First, tone-melody correspondence proportions were obtained. For Cantonese and Mandarin, annotated data of nine popular music songs from 1997 to 2013 were pulled from Kirby (2022). Baseline and observed proportions for similar, contrary, and oblique settings were newly calculated by the author. Here, baseline proportions are the numbers one would get if they randomly paired the melodic and tonal bigrams in the songs analyzed. The total number of bigrams used for analysis was 1720 for Cantonese and 1644 for Mandarin. Note that contour tones were dealt with by only focusing on the offset pitch for both languages, following Chan (1987). For Japanese, 46 songs from 12 Disney animation movies (ranging from 1989 to 2016 US release) were annotated by the author. This amounted to 8365 total bigrams used for analysis. Melodic transcriptions were based on existing scores, but any alterations made by the singers were included. Linguistic tones of the lyrics were transcribed based on the author's intuitions as a native Tokyo Japanese speaker. To keep the tonal annotations free of intonational intuitions, judgments were made at the lexical level (including compounds). Once all musical notes and linguistic tones were transcribed, the direction of pitch change was categorized as rising, falling, or level in each melodic or tonal bigram. Pairs of corresponding melodic and tonal bigrams were then categorized into the 9 possible combinations and the three setting categories (similar, contrary, oblique) as defined in Ladd and Kirby (2020).

Next, functional load values for each tone system were calculated. This first study used a simple, minimal pair-count based methodology, where a list of words in each language was obtained either from a dictionary or corpus, and the proportion of tone-based minimal pairs was calculated. The dictionaries/corpora used were the CC-CEDICT dictionary for Mandarin and the Hong Kong Cantonese corpus (Luke and Wong, 2015) for Cantonese. PyCantonese (Lee et al., 2022) was used to parse the Cantonese corpus. The functional load value for Japanese was taken from Sibata and Shibata (1990).

Lastly, the two dimensions were plotted against each other across the three languages.

3.2 *Results* First, we will briefly look through the tone-melody correspondence proportions found in the three languages. The baseline and observed proportions of similar, contrary, oblique (level melody), and oblique (level tone) settings are reported in Figure 4. All languages show a significant difference between baseline and observed proportions in Chi-squared goodness-of-fit tests (Cantonese: χ^2(3, N=1720) = 1520.883, p $\leq$ 0.001, Mandarin: χ^2(3, N=1644) = 14.455, p $\leq$ 0.01, Japanese: χ^2(3, N=8365) = 46.936, p $\leq$ 0.001)). We can see that, although all languages follow a trend towards more similar settings and fewer contrary settings, Cantonese is the only one that clearly maximizes similar settings and minimizes contrary settings.

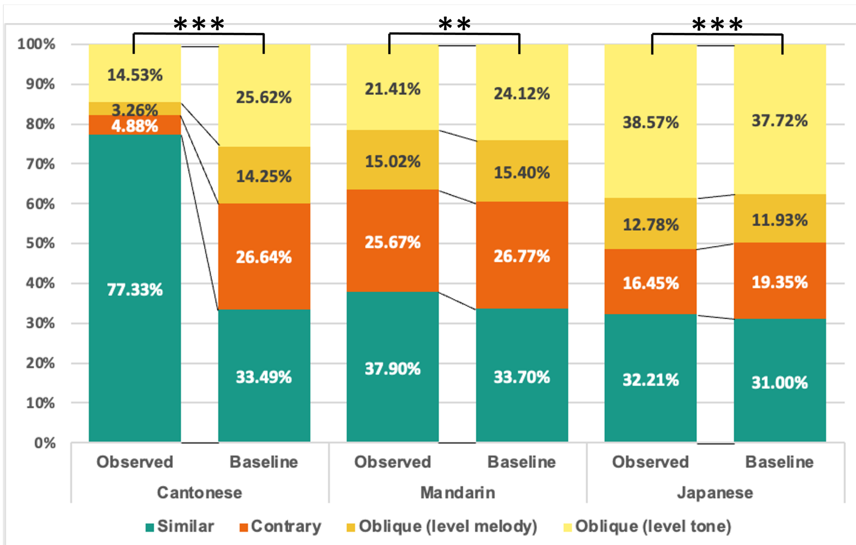

Figure 4: Tone-melody correspondence proportions in Cantonese, Mandarin, and Japanese. Cantonese and Mandarin data are taken from Kirby (2022) and recalculated. Asterisks signify a significant difference between baseline and observed proportions in a Chi-squared goodness-of-fit test (**: p $\leq$ 0.01, ***: p $\leq$ 0.001).

Next, we examine the functional load values for tones in each language. As illustrated in Figure 5, with a minimal-pair-based method using tones at the word level, Cantonese has the highest functional load of tone, followed by Mandarin, and finally Japanese. It is notable that Cantonese and Mandarin have this difference despite both being commonly categorized as lexical tone languages.

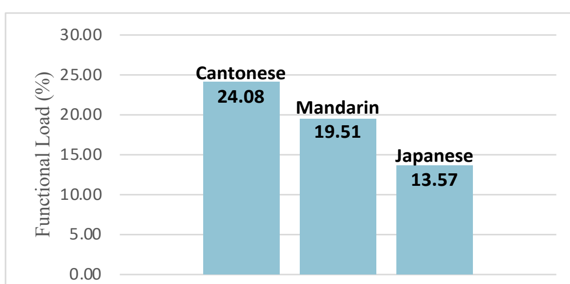

Figure 5: Functional load (minimal-pair based) of tone in Cantonese, Mandarin, and Japanese. Tones are all at the word level (post tone sandhi, compound rules, etc.).

Lastly, we will combine the above two graphs to test their correlation. The ratio between similar setting bigrams and contrary setting bigrams is used here to represent the degree of tone-melody correspondence with a single value: the higher the number, the more tone-melody correspondence. This value reflects the effects of both "avoid contrary" and "maximize similar" strategies available in tone-melody correspondence. A scatter plot with functional load of tone on the x-axis and similar-to-contrary ratio on the y-axis is shown in

Figure 6. The three data points show a positive correlation, suggesting the predicted correlation between the functional load of tone and the degree of tone-melody correspondence. To make sure that this correlation is not coincidental for these three data points specifically, three other languages were compared using a similar methodology.

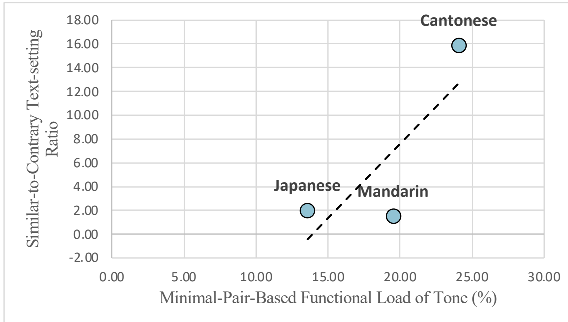

Figure 6: Functional load of tone and the ratio of similar-to-contrary settings across Cantonese, Mandarin, and Japanese. The dotted line is a linear trend line.

4 Meta-analysis of literature: Mandarin, Vietnamese, and Thai

This portion of the study obtains tone-melody correspondence proportions and the functional load of tones in Mandarin, Vietnamese, and Thai from existing literature, and compares the two measures. Note that functional load values used here are calculated using the system-entropy based method laid out in Surendran and Niyogi (2003, 2006) rather than a minimal-pair based method. This discrepancy in methodology makes the numbers in this study and the previous study not directly comparable.

4.1 *Methods* Tone-melody correspondence proportions in Mandarin, Vietnamese, and Thai were obtained from Kirby (2022), Kirby and Ladd (2016), and Ketkaew and Pittayaporn (2015), respectively. Each source uses a fairly large dataset, with nine popular music songs from 1997 to 2013 in Kirby (2022), 20 *tân nhạc* or 'new music' songs from 1946 to 1957 in Kirby and Ladd (2016), and 40 popular music songs in Ketkaew and Pittayaporn (2015).

Functional load values were taken from Liang and Levow (2025), which calculates the word-level functional load of tone in the three languages following an entropy-based method laid out in Surendran and Niyogi (2003, 2006) using the Common Voice database. Intuitively, this measure of functional load represents how much work tones are doing in the predictability of words in a language, rather than for the meaning contrasts in a language's list of words with no context or token frequency values.

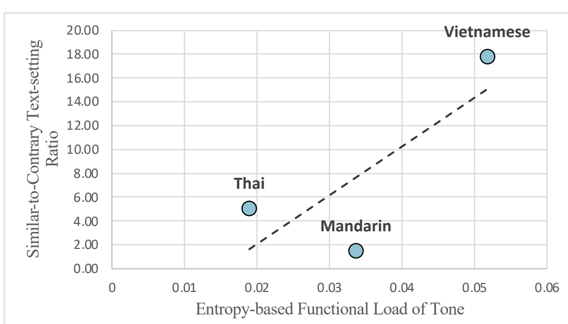

Figure 7: Functional load of tone and the ratio of similar-to-contrary settings across Mandarin, Vietnamese, and Thai. The dotted line is a linear trend line.

4.2 *Results* Figure 7 plots the entropy-based measures of functional load of tone and the similar-to-contrary ratio in Mandarin, Vietnamese, and Thai. Again, we see that the three languages show a positive trend line, suggesting a correlation between the functional load of tone and the degree of tone-melody correspondence. Together with the results of the corpus study represented in Figure 6, we can conclude that it is likely that the functional load of a tonal system and the degree of tone-melody correspondence correlate across the five languages examined.

5 Discussion

Both the corpus study on Cantonese, Mandarin, and Japanese, and the meta-analysis study on Mandarin, Vietnamese, and Thai indicate a possible correlation between the functional load of tones in a language and the degree of tone-melody correspondence that language shows in songs. To generalize, these results suggest that if tone is more important for the accurate understanding of speech in one language than in another, tone is more likely to be encoded in songs of that language. And more broadly, these results support the hypothesis that phonological contrasts that are more important for the accurate understanding of speech in a language are more likely to be encoded in songs of that language.

Within the text-setting literature, this likely correlation between functional load of tones and cross-linguistic differences in tone-melody correspondence points toward the existence and effect of communicative pressures in songwriting, opening up the possibility of other Information-Theoretic approaches to text-setting phenomena being fruitful. If we observe effects of communicative pressures across various text-setting phenomena, languages, and musical genres, then any explanatory model of text-setting would need to include these effects.

In the broader context of speech production and planning, the results of the present study are additional evidence that support the idea of communicative pressures being a significant factor in how speakers plan and produce their speech. Traditional phonology and generative linguistics in general have long shut down or ignored the possibility of "performance" factors affecting the shape of linguistic knowledge (e.g., Chomsky and Halle, 1968). Individuals have thought about this possibility, however, from at least Zipf (1932), seriously considering the communicative pressures in language use as things that may heavily guide production and perception of language and ultimately change the shape of the linguistic system. At this point in time, we have a substantial amount of evidence that points towards this idea. We have observed phenomena that demonstrate associations between phonetic variation and informativity, most famously the pattern of more informative units (words/syllables/vowels/etc) being pronounced with longer duration than less informative units (e.g., Aylett and Turk, 2004; Lieberman, 1963; Shaw and Kawahara, 2019). This particular phenomenon led to the proposal of the Smooth Signal Redundancy Hypothesis in Aylett and Turk (2004), which argues that there is an inverse relationship between language redundancy and acoustic redundancy (as manifested by phonetic dimensions such as duration). The possible correlation between the functional load of tones and the degree of tone-melody correspondence in languages as found in the present study may be understood as the manifestation of this inverse relationship in a musically-restricted environment. When language redundancy is high in terms of tones, or, to put it in other words, when tones do not offer much information in a language, it is more likely that tones are not prioritized in the encoding of linguistic information in songs of that language. Conversely, when tones do offer a lot of information in a language, such as in Cantonese or Vietnamese, tones are more likely to be encoded in songs.

It is exciting in itself that we have found another avenue of data where we may be able to find effects of communicative pressures, but it is further intriguing to consider the unique properties of text-setting compared to non-sung speech. First, text-setting is a very restrictive environment in terms of how and how much the acoustic signal can be enhanced. In sung speech, it is nearly always required that increasing acoustic redundancy be in line with the musical structure and/or aesthetics in some way, meaning there is very little flexibility in many acoustic dimensions, especially duration and pitch. Tones cannot be produced as they normally would be produced in non-sung speech, as they need to adhere to melodic aesthetics and the violable but robust rule in many musical genres of maintaining one pitch per syllable. Hence, songwriters utilize the relative flexibility they have in creating melody lines and choosing sequential words/morphemes by focusing on the direction of pitch transitions between sequential notes and tones. Although non-sung speech often has restrictions as well, such as background noise, sung speech has a unique set of acoustic dimensions that are heavily restricted, leading to creative solutions that are only observed in the medium.

One thing to note here is that the solution to the musical restrictions may differ across languages or musical cultures. The present study, as well as many other tonal text-setting studies, have focused on the matching of pitch transitions and compared languages or musical genres based on how much this solution in particular is utilized. However, it is possible that other strategies exist and are used in languages or musical genres that are better suited for those strategies than pitch transition matching. In Figures 6 and 7, we saw that Mandarin seemed to constantly have a mid-level functional load of tone with a low degree of tone-melody correspondence, not following the predicted correlation as well as the other four languages. Perhaps this pattern is explainable if we shift our attention to other possible strategies for encoding tonal information in songs, such as the use of phonation or duration instead of pitch. It may also be that Mandarin songs give up on encoding tonal information and, instead, make up for the lack of redundancy by enhancing other linguistic features such as segmental phonological contrasts or syntactic/semantic predictability.

Secondly, text-setting data consists of carefully chosen, categorical data rather than spontaneous and gradient productions. Much of the data examined in communication-oriented literature is spontaneous or read speech, where the speaker is fairly oblivious that the patterns of interest are, first of all, a thing, and, second, happening in their speech. The data is also very gradient. Text-setting data like the one we investigated in this study is made up of categorical decisions (e.g., choosing a specific phrasing to go with a certain melody, resulting in a specific tone-melody setting) that have been carefully chosen by the songwriter who has likely listened to their song many times. Despite these differences, text-setting still seems to show similar effects from communicative pressures as spontaneous or read speech. This leads to the intuitive conclusion that these effects will show up as long as the same communicative goal of transmitting a message through a noisy channel exists in the activity investigated, even when the effects are the result of a much more conscious process.

There are several caveats to the results of the current study, the largest being the inconsistency in musical genres and functional load calculations across languages and the limited variety of musical genres and languages included. In order to investigate the effect of a linguistic property like functional load, it would be ideal to keep the musical factors as consistent as possible between languages, such as by using the same songs or different songs from similar musical genres. Musical genre has been known to affect the degree of tone-melody correspondence within a language (List, 1961), and since we are looking for effects of communicative pressures, the communicative nature of the genre is a crucial factor to control for. The data is also limited in variety, both in terms of musical genre and language. Regarding language variety, although Japanese does add some variation, the inclusion of tone languages from geographical areas and cultures outside of Asia would be necessary to actually conclude that the correlation exists.

6 Conclusion

This study plotted the functional load of tone in five languages (Cantonese, Mandarin, Japanese, Vietnamese, and Thai) against the degree of tone-melody correspondence of songs in each language. We found that these two measures likely correlate across the five languages, supporting the hypothesis that phonological contrasts that are more important for accurately understanding speech are more likely to be encoded in text-setting. More broadly, this suggests that the communicative pressure of wanting the listener to understand the lyrics may be a guiding motivation for songwriters in the creation of sung speech, analogous to communicative pressures observed in non-sung speech. Text-setting data like the present study's is a new addition to the study of communicative pressures on speech and has the potential to provide insights that non-sung speech data cannot. Additionally, this study demonstrates that Information-Theoretic approaches to text-setting can reveal characteristics of text-setting that have not been focused on in past research. Further investigations into text-setting phenomena of various types and in various languages and musical cultures would benefit both lines of research.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech [Publisher: SAGE Publications Ltd]. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Chan, M. (1987). Tone and melody interaction in cantonese and mandarin songs. *UCLA Working Papers in Phonetics*, *68*, 132–169.

Cho, S. (2017). Text alignment in japanese children's song. *University of Pennsylvania Working Papers in Linguistics*, *23*(1), 5.

Chomsky, N., & Halle, M. (1968). The sound pattern of english. [Publisher: ERIC].

Hayes, B., & Kaun, A. (1996). The role of phonological phrasing in sung and chanted verse. *The Linguistic Review*, *13*(3). https://doi.org/10.1515/tlir.1996.13.3-4.243

Hayes, B., Wilson, C., & Shisko, A. (2012). Maxent grammars for the metrics of shakespeare and milton. *Language*, *88*, 691–731. https://doi.org/10.2307/23357550

Ho, W.-s., Vincie. (2010). *A phonological study of the tone-melody correspondence in cantonese pop music* [Doctor of Philosophy]. The University of Hong Kong [Pages: 991035825399703414, b50899752]. https://doi.org/10.5353/th_b5089975

Kawahara, S. (2015). Chapter 12 the phonology of japanese accent. *Handbook of Japanese Phonetics and Phonology*.

Ketkaew, C., & Pittayaporn, P. (2015). Do note values affect parallelism between lexical tones and musical notes in thai pop songs?

Kiparsky, P. (1977). The rhythmic structure of english verse [Publisher: The MIT Press]. *Linguistic Inquiry*, *8*(2), 189–247. Retrieved October 14, 2024, from https://www.jstor.org/stable/4177986

Kirby, J. (2022). Comparative tonal text-setting in mandarin and cantonese popular song. *Proceedings of CLS*, *58*, 211–230.

Kirby, J., & Ladd, D. R. (2016). Tone-melody correspondence in vietnamese popular song. *5th International Symposium on Tonal Aspects of Languages (TAL 2016)*, 48–51. https://doi.org/10.21437/TAL.2016-10

Kitayama, Y., Kawahara, S., & Furusawa, R. (2024, January 13). (sada masashi no akusento kenkyuu "kosumosu"no junnkougyakkou bunseki kara saguru)[pitch accent study on sada masashi: Looking at similar and contrary settings in "cosmos"]. In *(sada masashi kaitai shinsho taaherusadatomia)*. Daiwa Shobo.

Ladd, D. R., & Kirby, J. (2020, December 31). Tone–melody matching in tone-language singing. In C. Gussenhhoven & A. Chen (Eds.), *The oxford handbook of language prosody* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.47

Lee, J., Chen, L., Lam, C., Lau, C. M., & Tsui, T.-H. (2022). PyCantonese: Cantonese linguistics and NLP in python, 6607–6611.

Li, E. K. C. (2021). Cantopop and speech-melody complex.

Liang, S., & Levow, G.-A. (2025, August). Tone in perspective: A computational typological analysis of tone function in ASR. In M. Hahn, P. Rani, R. Kumar, A. Shcherbakov, A. Sorokin, O. Serikov, R. Cotterell, & E. Vylomova (Eds.), *Proceedings of the 7th workshop on research in computational linguistic typology and multilingual NLP* (pp. 82–92). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.sigtyp-1.11

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech [Publisher: SAGE Publications Ltd]. *Language and Speech*, *6*(3), 172–187. https://doi.org/10.1177/002383096300600306

List, G. (1961). Speech melody and song melody in central thailand [Publisher: University of Illinois Press]. *Ethnomusicology*, *5*(1), 16–32. https://doi.org/10.2307/924305

Luke, K. K., & Wong, M. L. (2015). The hong kong cantonese corpus: Design and uses [Publisher: Journal of Chinese Linguistics Publication Office]. *Journal of Chinese Linguistics Monograph Series*, (25), 312–333.

Mathesius, V. (1929). La structure phonologique du lexique du tchèque moderne. *Travaux du Cercle linguistique de Prague*, *1*, 67–84.

McPherson, L. (2018). The role of music in documenting phonological grammar: Two case studies from west africa. *Proceedings of the Annual Meetings on Phonology*. https://doi.org/10.3765/amp.v7i0.4493

McPherson, L., & Ryan, K. M. (2018). Tone-tune association in tommo so (dogon) folk songs [Publisher: JSTOR]. *Language*, 119–156.

Pike, K. L. (1946). The Flea. Melody Types and Perturbations in a Mixtec Song [Number: 2]. *Tlalocan*, *2*(2), 128–133. https://doi.org/10.19130/iifl.tlalocan.1946.406

Rycroft, D. (1959). African music in johannesburg: African and non-african features [Publisher: International Council for Traditional Music]. *Journal of the International Folk Music Council*, *11*, 25–30. https://doi.org/10.2307/834851

Schellenberg, M. H. (2013). The realization of tone in singing in cantonese and mandarin [Publisher: University of British Columbia].

Shaw, J. A., & Kawahara, S. (2019). Effects of surprisal and entropy on vowel duration in japanese [Publisher: SAGE Publications Ltd]. *Language and Speech*, *62*(1), 80–114. https://doi.org/10.1177/0023830917737331

Sibata, T., & Shibata, R. (1990). –akusento wa douongo wo dono teido benbetsu shiuruka [to what extent can accent distinguish homophones?] [Publisher:]. = *Mathematical linguistics*, *17*(7), p317–327.

Starr, R. L., & Shih, S. S. (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

Surendran, D., & Niyogi, P. (2003, November 24). Measuring the functional load of phonological contrasts. https://doi.org/10.48550/arXiv.cs/0311036

Surendran, D., & Niyogi, P. (2006, October 25). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In O. Nedergaard Thomsen (Ed.), *Current issues in linguistic theory* (pp. 43–58, Vol. 279). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.279.05sur

Trubetzkoy, N. S. (1939). Grundzüge der phonologie (travaux du cercle linguistique de prague 7). *Prague: Cercle Linguistique de Prague*.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, *56*(3), 395–417. https://doi.org/10.1177/0023830913489096

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186. https://doi.org/10.1016/j.cognition.2013.03.002

Wong, P. C., & Diehl, R. L. (2002). How can the lyrics of a song in a tone language be understood? [Publisher: SAGE Publications Sage CA: Thousand Oaks, CA]. *Psychology of Music*, *30*(2), 202–209.

Yip, M. (2002, August 15). *Tone* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Zipf, G. K. (1932). Selected studies of the principle of relative frequency in language [Publisher: Harvard university press].