REGISTER COMPLEX TRANSFER IN THE LEXICAL TONE-TUNE INTERFACE IN V-POP.

Khang ĐINH

Department of Linguistics, University of California, Santa Barbara.

<khang DOT ucsb AT gmail DOT com>

Abstract

A growing amount of research in recent years has sought to understand the interaction between lexical tone and melodic contour of song lyrics in tone languages. The question is relevant for understanding linguistic tonal perception, behavior, and text-setting. In Vietnam, the recent proliferation of V-pop ("Vietnamese pop") presents an opportunity to explore the behavior of Vietnamese tones channeled by the melodic conventions of twenty-first century Western-influenced pop music. A corpus of 45 V-pop songs was analyzed via tone-note pairs to show the distribution of contrapuntal contours between transitions of melodic pitch and lexical tone in the lyrics. Though 69.32% of the pairs demonstrate a favorability for tone-note pitch direction agreement, acoustic analysis reveals that V-pop performers conserve tonal realization by referencing other prosodic cues when lexical tone pitch is canceled by melodic constraint. The study broadens our conception of Vietnamese tone by observing the register complex, and demonstrates the value in leveraging the musicolinguistic interface to isolate aural language phenomena via speech-song resource transfer.

Keywords: Phonetics, phonology, pop music, text-setting, tone, Vietnamese

---

1. Introduction

If singing is the musical vocalization of speech, then how does it behave in a language that has an explicitly songlike quality to its speech? This is the case with tone languages, in which the interface of lexical tone in a singing context results in a musicolinguistic governance of the language. That is, if both singing and tone language speech separately require specific demands of the vocal tract, then these demands must be compromised in order for the two streams to simultaneously exist.

Though this question has been explored in a small but growing amount of research, the speech-song compromise has long been an intrinsic consideration for the music of tone languages, including Vietnamese. Master of Vietnamese music Nguyễn Vĩnh Bảo (1970) aptly noted:

> The strong tonality of the language has had a deep effect on Vietnamese music. A word with a high rising tone cannot be sung with a falling melody, and vice-versa. As a result, melodic forms were developed that could accommodate improvised changes of notes to fit the tones of the words used.

However, the emergent music genre of V-pop ("Vietnamese pop") is predicated on a music theoretic system (i.e. Western pop music tradition) not initially designed to accommodate linguistic tonal realization, posing potential problems for Vietnamese tonal production in V-pop songs. Though previous research on this "tone-tune" interface has underpinned tone as an inflection of pitch, Vietnamese tone features various tonal cues (Nguyễn Văn Lợi & Edmondson 1998; Phạm 2001, 2003; Kirby 2011) that may manifest uniquely in song.

The current study operationalizes the interaction between lexical tone and musical pitch in Vietnamese in order to analyze how the tonal register complex inherits and adapts the musicolinguistic interface. Using a largest-to-date corpus for tone-tune research extracted from V-pop songs, the current study explores how the register complex leverages tonal cues in the tone-tune interface. The results lead a discussion about how tone in Vietnamese encompasses a prosodic network beyond pitch inflection.

2. Background

2.1 Vietnamese language

Belonging to the Austroasiatic language family, Vietnamese is the official language of the country of Vietnam, spoken by roughly 90 million people in Vietnam and 5 million people in overseas diasporas. Orthographically, one of the most salient features to non-users of the language is its usage of the Latin script, which was first described in the 17th century by Jesuit missionaries, and refined in the early 20th century as the official alphabet (*chữ Quốc ngữ*) of

Vietnam under French colonization, replacing *chữ Nôm,* the Vietnamese writing system that utilizes Chinese characters. Vietnamese is often mistakenly categorized as a monosyllabic language due to the fact that its written form is visually segmented into individual morphemes; however, Vietnamese words may consist of polysyllabic words (e.g. “khang trang”, “nhà khoa học”).

Central to Vietnamese is its complex tone system, which is orthographically represented by five accents /◌́/, /◌̀/, /◌̉/, /◌̃/, /◌̣/; note that a sixth tone (level tone /◌/) exists but does not possess an accent to represent it—therefore, the visual absence of a tone accent in a word indicates that the word is to be pronounced with a level tone. Tones are used to contrast lexical meaning in Vietnamese, with each individual morpheme possessing its own independent tone marked about the vowel. The example below illustrates how tone changes the meaning of a word in the context of a minimal pair in Vietnamese:

/ɓɔ + ◌̉/ bỏ “discard”

/ɓɔ + ◌̀/ bò “cow”

2.2 Vietnamese tonal variation

The issue of whether to represent the Vietnamese tone system according to phonetic, phonological, or cognitive abstract units continues to be a debated topic in the literature on Vietnamese tone. The current literature observes phonetic, phonological, and perceptual mismatches, intonational influence on tone, and the size of the tonal inventory. The matter is complicated primarily due to tonal variations that exist phonologically between dialects of Vietnamese. The standard categorization delineates three general Vietnamese dialects that correspond to each of the three main regions (i.e. North, Central, South) of Vietnam. Differences in tonal inventories between these dialects can exist for the tones’ phonological realizations, for instance in the case of most Southern varieties of Vietnamese in which the /◌̃/ tone merges phonologically with the /◌̉/ tone. Auditory differences have also been shown to exist between listeners of different dialects, leading to the postulation that abstract cognitive cues better explain the tonal processes of Vietnamese than generalizable articulatory and acoustic parameters (Brunelle 2009). Another popular study in the literature on Vietnamese tone proved that

pitch-height was predicted by phonation, thereby suggesting that phonation rather than pitch was the primary phonetic cue of tone in Vietnamese (Pham 2003). Though studies on Vietnamese prosody remain relatively sparse (Lê et al. 2011, Mac et al. 2011, Ngo & Bui 2012, Mac et al. 2015), nascent observations are beginning to reveal the presence of dynamic tone behaviors in discourse-dependent contexts (Phan 2022). It has also been suggested that the number of distinct tones in Vietnamese be reflected in the representation of an eight-tone inventory rather than the standard six-tone (Pham 2003, Kirby 2011) system adopted by the orthography; the eight-tone system has been used by linguists to reflect the difference in the behavior of the /◌́/ and /◌̣/ tones when followed by an unreleased stop-final consonant. Though Vietnamese tonal behavior leaves much to be studied, it is evident that commonly held formalizations about Vietnamese possessing a universally inter-dialectical representation of tone continue to be challenged.

2.3 Selection of tonal inventory

Though Vietnamese has a diverse and robust selection of varieties, this study is framed through the tone system that appears in the general Northern variety. This decision was made because the overwhelming majority of V-pop samples in the corpus feature lyrics sung using Northern Vietnamese. The pitch hierarchy adopted in this study was organized based on the offset of the tones' frequency contours (Figure 1.1). Additionally, the eight-tone representation—rather than the orthographically standard six-tone representation—of Vietnamese is utilized in this study to capture the most fine-grained behaviors of the tonal inventory. This system includes two checked tones that occur only in final unreleased voiceless oral stops. A rising checked tone has a higher frequency than a rising tone, and a low checked tone has a lower frequency than a low glottalized tone (Figure 1.2).

For the rest of this study, Vietnamese tones will be referred to with a combination of their individual orthographic spellings and a hierarchical alphanumeric notation adapted from (Chao 1930) to indicate both the pitch and phonation characteristics of each tone. This decision was made to include a fuller representation of Vietnamese tone in the text for readers to better comprehend the register complex of the tones. The table below outlines the tonal inventory (Table 1):

Table 1

The tones are ordered according to a formalized pitch hierarchy of Northern Vietnamese tones. The hierarchy was determined by the pitch offset of the tonal contour. For the alphanumeric code, the letter represents phonation type (m = modal, b = breathy, c = creaky), and the number represents pitch relative to the tonal hierarchy (e.g. 8 being the highest pitch). For the checked tones, a parenthetical describing their phonological environment is provided next to the diacritic. Tones will be referred to with their orthographic spelling and alphanumeric code (e.g. *ngã C8*).

Diacritic	Orthographic Spelling	English Description	Alphanumeric Code	Phonation Type	Pitch Contour
◌̃	ngã	high curve	C8	creaky	
◌́ (◌́C̲)	sắc đóng	rising checked	M7	modal	
◌́	sắc	rising	M6	modal	
◌	ngang	level	M5	modal	
◌̉	hỏi	low curve	B4	breathy	
◌̣	nặng	low glottalized	C3	creaky	
◌̣ (◌̣C̲)	nặng đóng	low checked	B2	breathy	
◌̀	huyền	falling	B1	breathy	

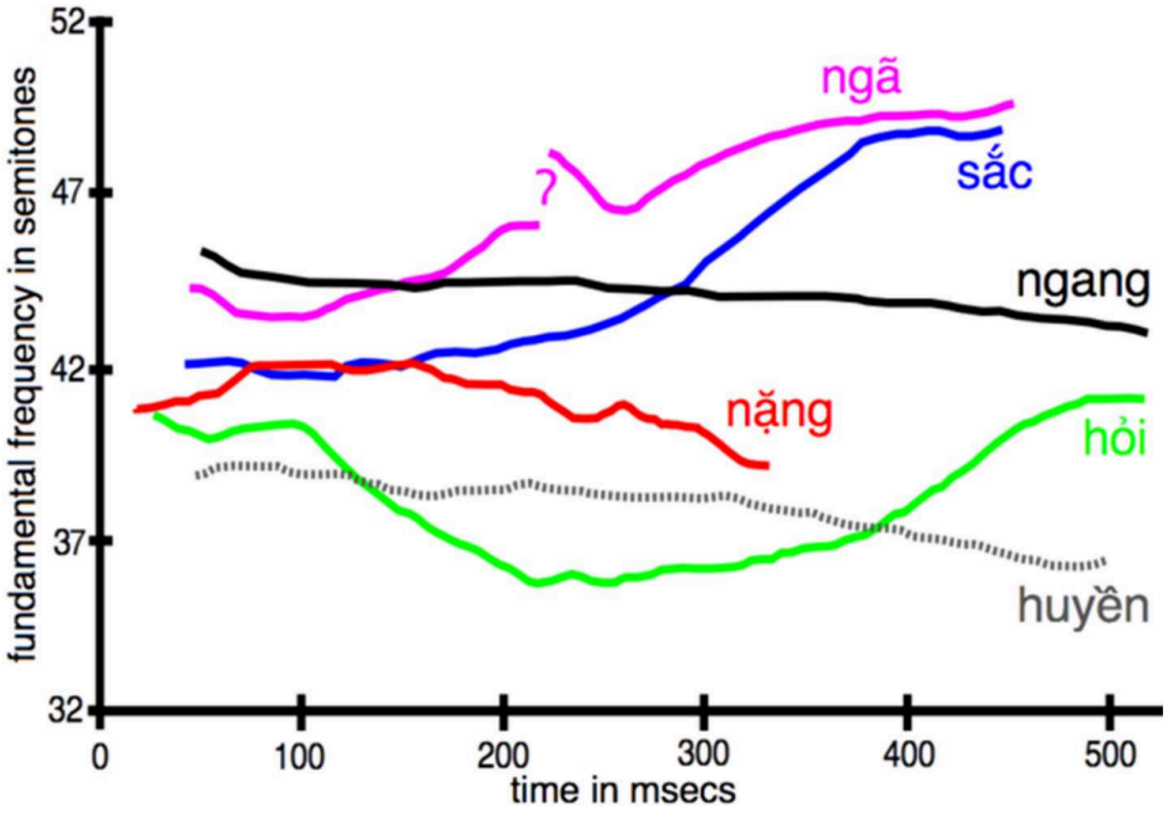

Figure 1.1. Fundamental frequency vs. time of Northern Vietnamese tones[1].

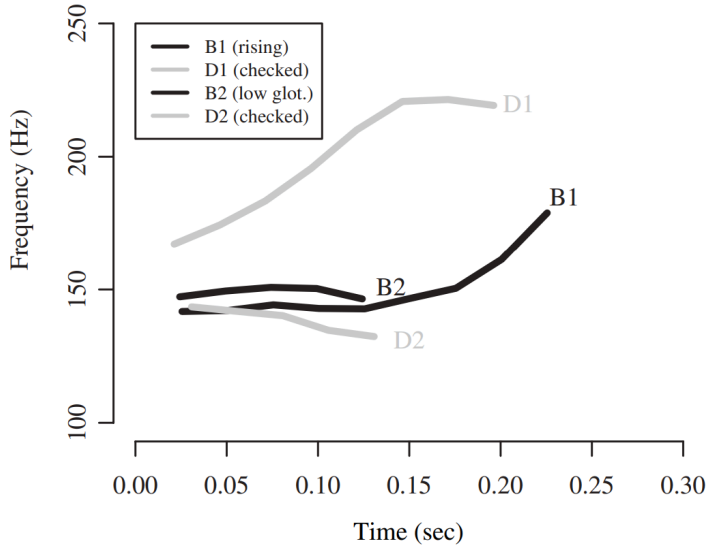

Figure 1.2. Fundamental frequency vs. time of Northern Vietnamese checked and open tones.[2]

[1] Six Northern Vietnamese (non-Hanoi) tones uttered by a male. From Nguyễn & Edmondson (1998).
[2] Northern Vietnamese (Hanoi) open tones and their checked counterparts uttered by a male. From Kirby (2011).

2.4 Previous Vietnamese tone-tune research

The question of how tones in tone languages interact with song has a sizable existing body of research. Refer to Ladd & Kirby (2020) for a literature review, and (https://www.musikologie.de/Literatur.html) to obtain a user-friendly purview of the question organized by discipline. However, the question of tone-melody matching for Vietnamese is quite sparse with only two studies existing (Ngô & Phan 2016, Kirby & Ladd 2016).

In Kirby & Ladd (2016), the researchers concluded that an avoidance of contrary motion, rather than an adherence to similar motion, was the key objective of tone-melody text setting in Vietnamese music. A corpus study was conducted from a convenience sample of 20 songs from 1946-1957 of Western-influenced Vietnamese popular music. They employed a methodology that analyzed the pitch transitions between 3,545 tone-melody pairs. The pitch contours of the tones were textually represented with two pitch targets (e.g. '[33]' for *ngang* level tone). Different hierarchies of tone based on their formalized pitch heights were compared to see which organization of the tonal inventory yielded the highest rate of tone-melody matching. Their best performing tone hierarchy had 77% similar, 19% oblique, and 4% contrary tone-melody pairs. They also found that the pitch offset between the transition of one lyric tone to the pitch onset of the following lyric tone was the most important text-setting constraint in their corpus.

In Phan & Ngô (2016), the researchers concluded that the salience of Vietnamese vibrato in song—and its comparable absence in speech—presents opportunities for analysis of musical modalities in Vietnamese folk music. An acoustic analysis measuring pitch and intensity was conducted on one Southern Vietnamese folk song to compare to what degree and what directional correspondence the pitch of linguistic tones would behave in sung versus spoken contexts. The comparative analysis was made between 7 recordings of singing and 6 recordings of speaking for the same Southern Vietnamese folk song. The pitch contours of the tones were textually represented with three pitch targets (e.g. '[444]' for *ngang* level tone). The researchers obtained sample speech recordings of the folk song from males and females from the various general regions of Vietnam (i.e. North, Central, South). They found that certain tones possessed congruence between their spoken and sung forms, while others reflected a greater liberty in their realization in each of the elicitation environments. The researchers also noted idiolectal

differences between singers as well as the utilization of the vibrato as a salient feature that inflected changes in the realization of linguistic tones.

In general, the current study extends the corpus methodology presented in Kirby & Ladd (2016) and observations about singer idiolect in Phan & Ngô (2016) to advance the empirical study of tone-melody matching. This approach helps with understanding text-setting in tone languages, and demonstrates the value of leveraging the musicolinguistic interface for prosodic research.

3. Method

3.1 Defining the tone-tune interface

It must be established that though the term "tone-melody" has been used to delineate the interface between lyric lexical tone and note pairs, this study will hereafter adopt the term "tone-tune" to denote this relationship. For this study's purposes, "tone" refers to one of the eight lexical tones represented in Table 1. Tone here does not refer to tonality in the schema of music theory, nor does it imply intonation, a prosodic constituent that affects phrasal and sentence-level connotation. It must also be noted that the current corpus study analyzes the phonemic rather than phonetic behavior of Vietnamese tones; that is, the tones are assessed by how well they adhere to a formalized hierarchy of the Vietnamese tonal inventory ordered by pitch height, and not how the pitches of the tones are acoustically realized in the audio stream (i.e. their frequency). For the current study, tune refers to a single line of pitches that are sung; tune here does not encompass rhythm nor intensity so as to isolate one musical parameter in the context of its interaction with linguistic tone.[3]

3.2 Methodology for tone-tune analysis

The approach of measuring tone-tune interaction in this study is derived from the methodology outlined by Kirby & Ladd (2016). The tone-note pairs are examined by their transitions between two successive pairs of lyric lexical tone and musical note. These tone-note pairs measure the transition from the pitch offset of the first pair to the pitch onset of the second pair; this transition can be called a bigram. The outcomes for a bigram are categorized according

[3] Tune, though very similar in meaning to melody, is the term used here for convenience and convention. While "melody" denotes a more formal association of pitches rhythmically organized in succession, "tune" connotes a more intuitive relation with song that is aligned with its roots in popular music.

to interactional principles adapted from musical counterpoint. In music theory, counterpoint loosely defines a compositional methodology derived from Western classical music. Counterpoint is concerned with the pitchwise relationship between two or more independent musical lines (i.e. a succession of notes across time) in order to create a copacetic layering of melodies. The interactional movement between these independent musical lines is described as "contrapuntal motion", of which the current study adapts three kinds: *similar motion*, *oblique motion*, and *contrary motion*. However, instead of describing the contrapuntal motion between multiple musical lines consisting of notes, the current study delineates this interaction between two lines or streams; the first stream is the tune of the given song, and the second stream consists of the lyrics of the tune.

For each bigram, an interaction can be classified as either matching or mismatching. A matching bigram can either be stationary or similar. Stationary is a unique delineation to this study, denoting a bigram in which both the phonemic pitch and note are repeated on the same tone and note, respectively. Similar bigrams are tone-note pairs moving in the same pitchwise direction. In a mismatching bigram, an interaction can either be oblique or contrary. An oblique interaction is classified further as being note stable or tone stable, meaning either the note stream holds for a given number of repetitions while the lyric stream changes tones, or the lyric stream repeats on a given tone while the note stream rises or falls in pitch. The other kind of mismatching bigram involves a contrary interaction, in which the note and tone move in a pitchwise direction opposite each other. In the question of tone-tune research, these interactions represent a measure of the "goodness" of a tone-note bigram under the presumption that the pitchwise behavior of the tone-tune interface can impact the semantic realization of lyrics in tone languages. When a lyric lexical tone and its corresponding note have a matching pitchwise relationship, then the relationship can be described as being "in tune". For a mismatching pitchwise relationship, the description "out of tune" is used instead.

The tune stream is operationalized through musical notes, while the lyric stream contains the tones under observation. Therein, their possible pitchwise interactions are ordered from left to right in Figure 2.1 from theoretically possessing the best text-setting decision to the most transgressive one. Figure 2.2 gives examples illustrating the contrapuntal pitch motion between the two streams.

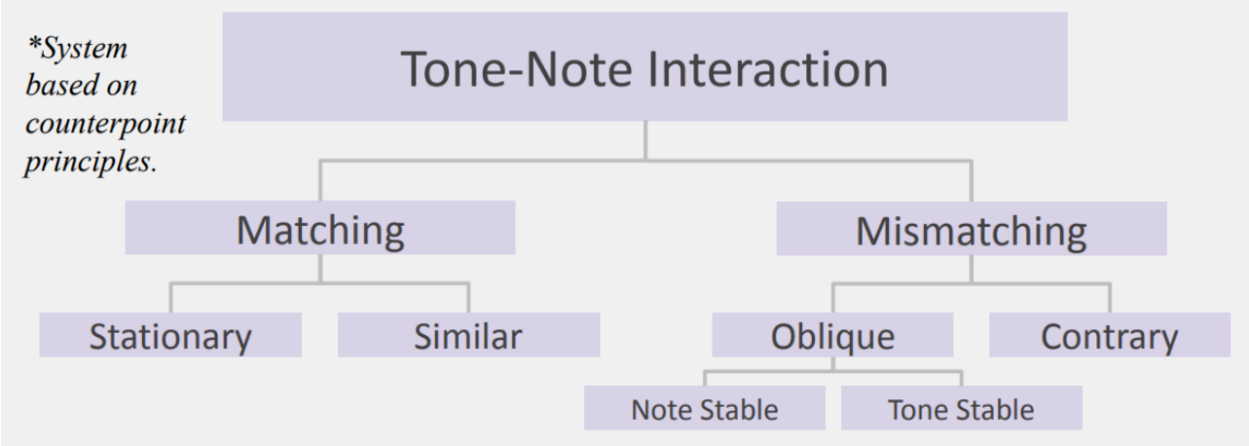

Figure 2.1. Tone-note interaction outcomes based on pitchwise contrapuntal motion.

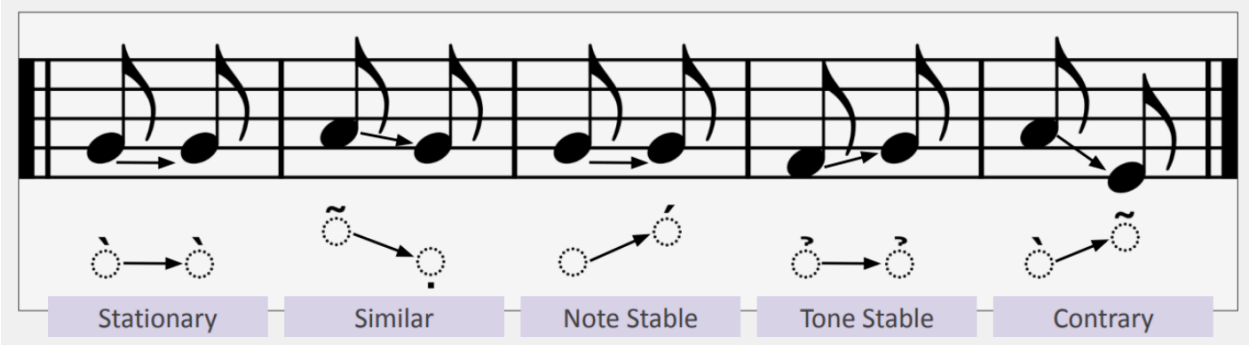

Figure 2.2. An illustration of all five possible tone-note interactions laid out on a musical staff. Arrows indicate the direction of the pitch.

Musical parameters for time are not accounted for in the current study. Music operates across the time axis; the speed of the music is quantified by tempo, the pulse is organized by meter, and the melody is horizontally spaced through rhythm. None of these temporal parameters are represented in the data. Rhythm was normalized such that each note of a bigram is represented with a time-unit of an eighth note (Figure 2.2); each lyric thus corresponds to one eighth-note. If a lyric is sung melismatically across multiple notes, then the lyric is simply repeated with separate eighth-notes in the data. This rhythmic normalization was done to control pitch as a parameter in the tone-tune interface.

4. Data

4.1 Corpus selection

The current study presents the largest corpus to date used for analyzing tone-tune interaction in music. The corpus consists of 45 V-pop songs released between the years 2016 to 2024.[4] The starting year 2016 because it includes the song "BỐNG BỐNG BANG BANG" by 365daband, which has the fourth most views for a Vietnamese music video (MV) on YouTube, and the most views for a V-pop song of all time on YouTube. For each year, the top five V-pop songs with the most views on their MV uploaded to YouTube were selected for the corpus.[5] 187 minutes and 59 seconds (3.1265 hours) of total footage from V-pop MVs was analyzed during the corpus selection process.

The question of what constitutes "V-pop" is quite interpretable. V-pop is characterized in this study as twenty-first century Vietnamese pop songs influenced by Western music theory, and typified by a very highly produced musical product. Some songs include folk and traditional Vietnamese music elements, noticeable influence from Korean, Japanese, Chinese, or other pop music cultures, and a dancelike quality. Although they may be considered as V-pop, slow ballads were not added to the corpus; the slow ballad is a massively popular genre of Vietnamese music, possessing its own specific lineage to older Vietnamese popular music, and thus is considered in this study as distinct from the kind of V-pop featured in the corpus. Vietnamese electronic music designed for the club scene and rap are also other subgenres categorized as V-pop that were not included in the present corpus; the former for lacking lyrical variance and the latter for its rhythmic rather than melodic focus. Remixes or covers of songs were also not included, even if they surpassed the view count of the original song. Most songs in the corpus do feature elements of the slow ballad, electronic dance music, and rap, but the importance is that these elements do not dominate the generic style of the song.

Demographic observations can be made from the V-pop corpus. Virtually all of the songs in the corpus are sung with a general Northern Vietnamese accent. Though there are instances of Central and Southern accents, the overwhelming usage of Northern Vietnamese indicates a predilection for this dialect in V-pop. The corpus was compiled to include a diversity of music artists. If one music artist had multiple songs featured in the top five most viewed V-pop songs

[4] A playlist of the songs in the corpus can be found at: https://www.youtube.com/watch?v=abPmZCZZrFA&list=PLA-2sqDgQeGmKdnUb28Q5dnWK-jlFQqtc.
[5] Updated as of October 10, 2024, the YouTube view count for all MVs in the corpus totals 5.488 billion views.

for that given year, then only their most popular song from that year was chosen. There are 23 songs released by a female artist, and 23 songs released by a male artist (note that this does not total to 46 songs, since one song was co-released by both a female and male artist). There are 52 unique performers (i.e. singers and rappers) across the corpus, with 20 females and 32 males.

4.2 Corpus cleaning

Three tiers were extracted for each song in the corpus: *lyric*, *lyric lexical tone*, and *note pitch*. The lyrics were extracted by a mix of listening and consulting multiple lyric transcriptions online; manual listening was used as the final judgment if transcriptions of a lyric disagreed across multiple official and unofficial lyric transcriptions of the song. The lyric lexical tone tier examined the toneme of each lyric to extract just the tone of that lyric. Unfortunately, official sheet music did not exist for many of the songs, so the note pitch tier was primarily collected with assisted manual ear transcription using *SonicVisualizer* to estimate a pitch trace of the vocals in the song (Figure 3); note that the author is a musician and has formal training in Western classical musicianship. The pitches were then represented with their corresponding MIDI (Musical Instrument Digital Interface) note value.

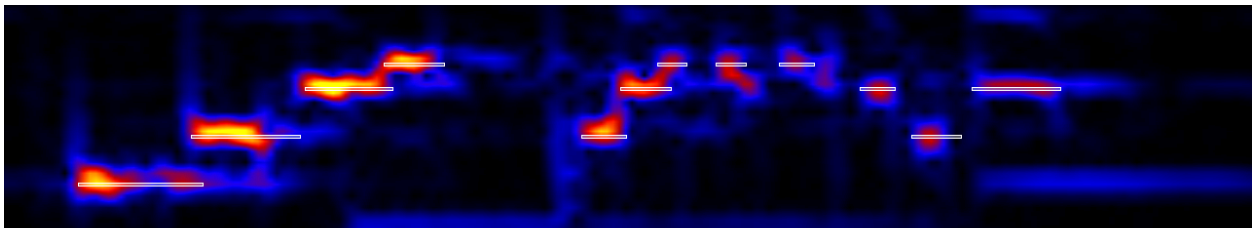

Figure 3. A *SonicVisualizer* spectrogram visualization of a tune from a V-pop excerpt. The singer's vocal frequencies are indicated by the clear lines.

Any phrases repeated with the exact same lyrics and tune were removed from the song. Standalone clauses entirely in English, background vocals, onomatopoeia, short exclamations (e.g. “Trời ơi!”), chorus repetitions with minor vocal inflections, and repeated sections with simple half-step/whole-step key changes were also cleaned from the data. For melismas lasting more than three distinct pitches, only the last note was transcribed to capture the pitch offset. Only perceptually prominent melismas spanning two to three notes were fully transcribed to include both the onset and offset pitches. Additionally, bigram analysis was not conducted on

bigram interactions that extended across boundaries or pauses in musical phrases. A boundary is determined as being a significant gap in time between two adjacent lyrics, or a clear start of a new syntactic clause in the lyrics. Pauses are marked according to whether they present a salient perceptual suspension of vocalization between two adjacent lyrics. Since repeats of choruses were omitted from the corpus, a boundary was created on the lyric after an omitted repetition of the chorus so that the transition from the last pre-chorus lyric to the first post-chorus lyric was not analyzed as a bigram. Lastly, English words and phrases were labeled in the corpus, but their pitchwise interaction with a subsequent Vietnamese lyric did not constitute a bigram.

4.3 Rap transcription

Rap has been virtually unresearched in studies of the tone-tune interface. However, twenty-seven (60%) of the forty-five songs in the current V-pop corpus feature some kind of rap section. This possibly indicates that rap is somewhat of a convention in V-pop songs, or at least is familiar to the population of V-pop listeners. Two kinds of rap styles are delineated from observation of the corpus: sung versus spoken rap (Figure 4.1). Instances of sung rap had a clearer defined melodic structure and were thus transcribed into the note pitch tier. Sung rap still possessed a cadential focus and rhythmic propulsion, but retained a somewhat fixed and predictable melodic contour similarly found in tune (Figure 4.2) Spoken rap, conversely, had a speech-like quality that abandoned any semblance of distinct musical pitch, but strongly resembled the elevated musical cadence of rap (Figure 4.3). Due to its intractably dynamic pitch contour and organic speech-like nature, spoken rap was left untranscribed in the note pitch tier. In general, any lyrics that did not have their pitches transcribed due to difficulties in identifying a salient musical pitch were represented and discounted from bigram analysis with a pitch of 0 in the corpus.

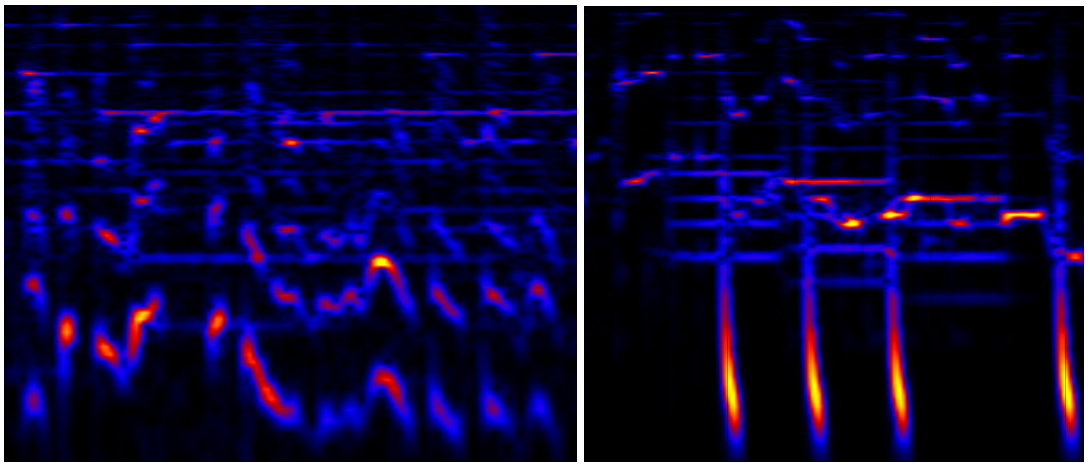

Figure 4.1. On the left: A spectrogram of spoken rap. The contour of the pitches extracted from vocal frequencies possesses no clearly defined pitch targets in the words. On the right: A spectrogram of sung rap. The vocal frequencies are clearly defined by horizontal pitch lines.

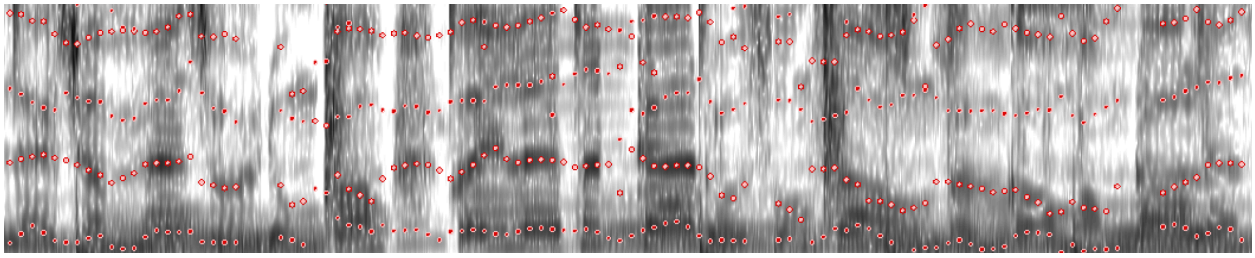

Figure 4.2. A *Praat* spectrogram slice at 1 minute and 33 seconds from "Lỡ Say Bye Là Bye" by Lemese and Changg (2021) resembling sung rap. The five formants indicated by red dots of the audio stream follow a clear horizontal formant structure due to the definite pitch targets of the tune.

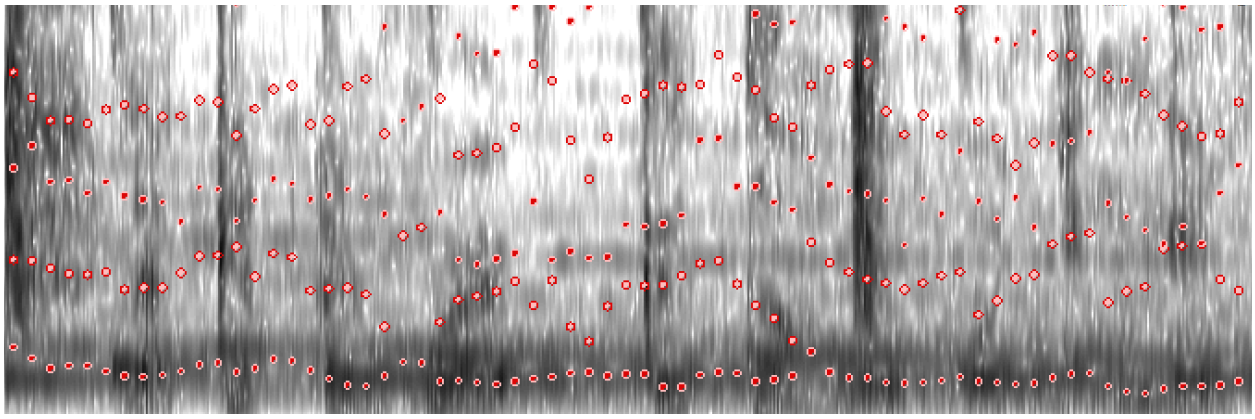

Figure 4.3. A spectrogram slice at 2 minutes and 1 second from "2 4" by W/n (2024) resembling spoken rap. It is difficult to track the contour of the first five formants (displayed by red dots) due to the dynamic tonality of spoken Vietnamese.

5. Results

5.1 Pitchwise interactions

Tone-tune analysis yielded the largest amount of tone-note interactions present in the literature. A total of 12,934 tokens for each of the three tiers (i.e. lyric, lyric lexical tone, note pitch) were extracted from the corpus for pitchwise analysis. Of the null interactions—that is, tone-note pairs labeled as boundaries, English word interactions, indistinct pitches, or left unlabeled—there were 50 unaccounted and 2,784 accounted for in the analysis, leaving a total of 10,100 analyzable bigrams describing the behavior between Vietnamese lyric lexical tones and musical notes across 45 of the most popular V-pop songs from 2016 to 2024.

The distribution of the tones appears in Table 2. The total count consisting of 12,934 tokens adhere to the general expectations of tone frequency in Northern Vietnamese. *Ngang M5* appeared most frequently consisting of 40.44% of the lyric lexical tones, which is expected given that it is a level tone considered to be in the middle of the pitch hierarchy of the Vietnamese tonal inventory. Tones with a simple contour followed next in frequency, with *huyền B1* occurring 21.96% of the time, and rising tones (i.e. M6 and M7) occurring 18.40% of the time. Of the rising tones, *sắc M6* appeared more than twice as frequently at 12.77% compared to its checked variation *sắc đóng M7* at 5.70%. *Nặng C3* appeared 4.67% of the time, with its breathy checked variant *nặng đóng B2* appearing slightly less frequently at 3.29%. The complex contour tones decidedly appeared least frequently in the corpus, which is expected given their complex pitch contours; *hỏi B4* occurred 5.96% of the time, while *ngã C8* occupied just 4.75% of the tone distribution. Due to the dynamic change of pitch direction in *hỏi B4* and *ngã C8*, these tones not only appear less frequently in the language, but may also be generally unfavored in text-setting due to the fact that they could further constrain the tune to accord with a particular contour.

Table 2

All of the counted Vietnamese tones extracted from lyrics in the corpus with their percentages.

	Ngã	Sắc Đóng	Sắc	Ngang	Hỏi	Nặng	Nặng Đóng	Huyền	All Tones
Count	615	737	1652	5230	771	604	426	2840	12934
Percentage	4.75%	5.70%	12.77%	40.44%	5.96%	4.67%	3.29%	21.96%	100.00%

The tone-tune interface in V-pop maintains a general favorability for being in tune between lyric tones and their corresponding note pitch (Table 3). However, the current model demonstrates a greater amount of bigrams being out of tune than has been shown in research on older Vietnamese popular music (Kirby and Ladd 2016). Of the 10,100 analyzable bigrams, the current model counts 7,001 (69.32%) matching bigrams and 3,009 (30.68%) mismatching bigrams. Similar bigrams occur most frequently (52.28%), affirming that an adherence to the tonal pitch hierarchy continues to be an important text-setting constraint in Vietnamese popular music. However, stationary bigrams (17.04%) occur much less frequently, indicating that tone-tune correspondence does not require a strict relational pitchwise exactitude. Oblique interactions occur 2,516 (24.91%) times, which is more frequent compared to the 19% of oblique bigrams in (Kirby and Ladd 2016). Of the oblique bigrams, 1,373 (13.59%) are note stable while 1,143 (11.32%) are tone stable. Contrary bigrams still appear least frequently, occupying just 583 (5.77%) of the tone-note interactions. However, the increased presence of oblique interactions reflects the musical behavior of V-pop, thereof characterized by repetitive tunes and formally constrained musical structures intended to enhance the predictability of the groove and catchiness to the listeners.

Table 3

Counts of all the tone-note bigrams in the corpus, including null interactions.

	Stationary Interactions	Similar Interactions	Note Stable Interactions	Tone Stable Interactions	Contrary Interactions	Null Interactions
Count	1721	5280	1373	1143	583	2784
Percentage	13.36%	40.98%	10.66%	8.87%	4.52%	21.61%

The proportion of each interaction for a given tone (Table 4) reveals unique characteristics of each tone's gestalt. The tones *ngang M5* and *huyền B1* are overrepresented in stationary bigrams, making up 90.3% of these kinds of interactions. This indicates that V-pop singers potentially disregard the utility of pitch in level and falling tones, instead using them for rhythmically focused melodic phrases. *Ngã C8*, *hỏi B4*, *nặng C3*, and *nặng đóng B2* all possess more instances of contrary bigrams than tone stable bigrams, which does not follow the frequency distribution of tone stable interactions (11.32%) to contrary interactions (5.77%). This may indicate that the complex tonal contours of *ngã C8* and *hỏi B4* tend to possess a lower degree of faithful pitch realization in V-pop. Meanwhile, *nặng C3* and *nặng đóng B2* are perhaps

associated at the same level in the tonal pitch hierarchy as the falling tone. For oblique tone-note interactions, note stable bigrams occur more frequently than tone stable for all tones except for *ngang M5*, which goes against the fact that note stable bigrams (13.59%) appear more frequently than tone stable bigrams (8.87%) in the corpus. This indicates that the pitch height of *ngang M5* is somewhat disregarded, and may have a "toneless" impact on the corresponding toneme. The frequency and proportional distribution of *ngang M5* indicates that it is the most versatile tone. One salient anomaly appears for contrary *hỏi B4* bigrams, which account for 17.5% of all contrary bigrams. Despite making up only 5.96% of all tones, this overrepresentation indicates that *hỏi B4* adheres least closely to the model's tonal pitch hierarchy. In fact, the dominance in frequency of similar bigrams (52.28%) constituting over half of all tone-note interactions may indicate that the Vietnamese tones are organized according to a phonetic classification based on high register (*ngã C8, sắc đóng M7, sắc M6, ngang M5*) and low register (*hỏi B4, nặng C3, nặng đóng B2, huyền B1*). However, this phonetic classification is complicated as exemplified in the case of inconsistent realizations of the complex contour tones *ngã C8* and *hỏi B4*. This is in line with issues of the mismatch between phonetic and phonological realizations of the Vietnamese tonal inventory (Pham 2003). Thus, the results of the pitchwise tone-tune interface may indicate that other variables besides pitch may influence the realization of tone in V-pop.

Table 4

An association matrix between type of bigram interaction and Vietnamese tone extracted from each lyric.

	Stationary	Similar	Note Stable	Tone Stable	Contrary
Ngã	10	234	116	34	37
Sắc Đóng	21	339	123	42	42
Sắc	98	751	223	124	101
Ngang	1201	2043	254	681	103
Hỏi	20	258	125	32	102
Nặng	14	275	110	12	40
Nặng Đóng	4	174	89	7	29
Huyền	353	1206	333	211	129

5.2 Register complex transfer

Acoustic analysis revealed that V-pop singers referenced other tonal cues besides pitch in order to conserve tonal realizations. As mentioned earlier, the pitchwise behavior of tones in the corpus generally followed a phonetic classification according to high and low pitch registers. However, it has been argued that Vietnamese tones may reference phonation rather than pitch as a primary tonal cue (Pham 2003). At the very least, Northern Vietnamese tones possess a voice quality component to them that is especially salient in *ngã C8* and *nặng C3*. Indeed, in cases of both bigrams being in tune or out of tune, prosodic constituents including pitch, voice quality, prominence, and duration were observed in the acoustic signal. In bigrams involving contrary motion, the cancellation of tone pitch by the corresponding note in the tune stream would be a severe transgression for tonal realization according to a phonetic classification of the Vietnamese tonal inventory. However, the leveraging of other prosodic constituents to conserve tonal realization in these contrary bigrams demonstrates the complexity of the Vietnamese register complex. For each of the cases, a *SonicVisualizer* spectrogram will portray the pitch motion in the tune, while a *Praat* visualization will be used to describe all other tonal cues present in the same bigram interaction; each *Praat* spectrogram displays the MIDI pitch, lexical tone, and lyric tiers (from bottom to top in each annotation box).

5.3 Creakiness and breathiness

Creakiness[6] was used to conserve the voice quality of complex contour tones when lexical tone pitch was canceled in contrary tone-note interactions. Figure 5.1 demonstrates how the singer employs a voice quality distinction to maintain the creakiness of *nặng C3* despite the fact that the tune falls from 56 to 54 (in MIDI note values) on an interaction that is supposed to rise in tonal pitch from *huyền B1* to *nặng C3*. Frequency pulses become irregular in the waveform and the F0 drops during the creakiness on the lyric “vậy”. In both spectrogram visualizations, the intensity decreases as a result of posterior glottal closure in the larynx.

[6] This term is broadly to encompass creaky voice, laryngealization, vocal fry, and other variants of this voice quality.

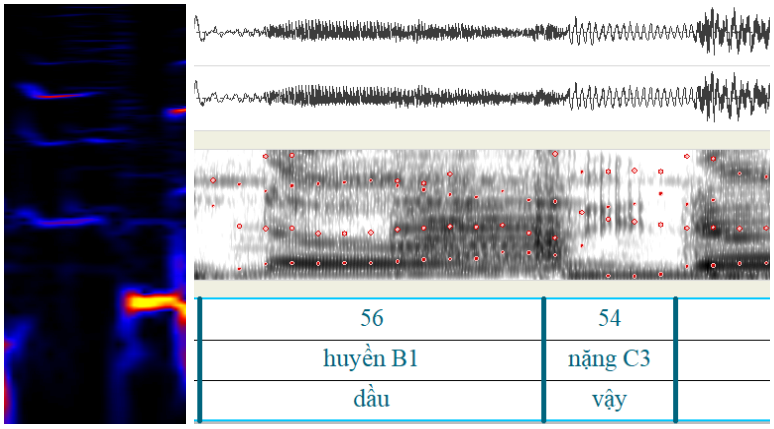

Figure 5.1. "Thích Em Hơi Nhiều" by Wren Evans (2021).

Acoustic visualizations of creakiness on a contrary bigram for the lyric "vậy". The red dots on the *Praat* spectrogram indicate the first five formants.

Creakiness also operates in register complex transfers from pitch to voice quality for tone pairs with a larger pitch height difference. In Figure 5.2, the note falls from 63 to 60 on a tone pair that is supposed to rise in pitch from *ngang M5* to *ngã C8*. The singer thus uses creakiness, as demonstrated by a decrease in intensity and irregularity in the pulse.

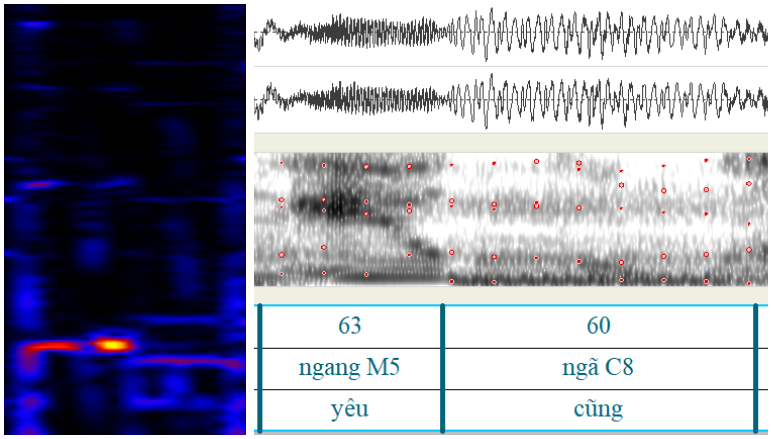

Figure 5.2. "See Tình" by Hoàng Thùy Linh (2022).

Acoustic visualizations of creakiness on a contrary bigram for the lyric "cũng". The red dots on the *Praat* spectrogram indicate the first five formants.

Figure 5.3 shows once more a contrary bigram featuring lyric lexical tones rising from *ngang M5* to *ngã C8* on a corresponding fall in pitch in the tune from 67 to 65. A steep decreasing contour at the onset of *ngã C8* is visible in the *SonicVisualizer* spectrogram. The intensity decreases as well, and vertical striations in the *Praat* spectrogram resulting from aperiodicity during creakiness appear.

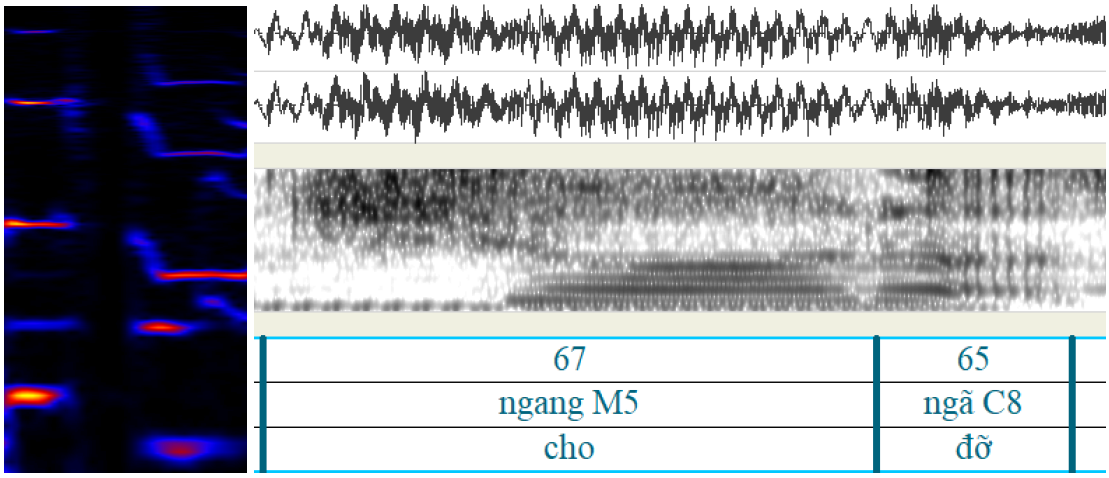

Figure 5.3. "Cà Phê" by Min (2022).
Acoustic visualizations of creakiness on a contrary bigram for the lyric "đỡ".

Breathiness was also observed in the vocal stream as a way of conserving the voice quality of a tone. A breathy phonation will introduce noise in the high frequency range of the spectrum due to turbulent airflow pushing through the loosely compressed vocal folds. Lyrics with breathy voice in contrary bigrams are sometimes marked with a salient exhale at the end of lyric as visible in the right panel of the *Praat* spectrogram in Figure 5.4.

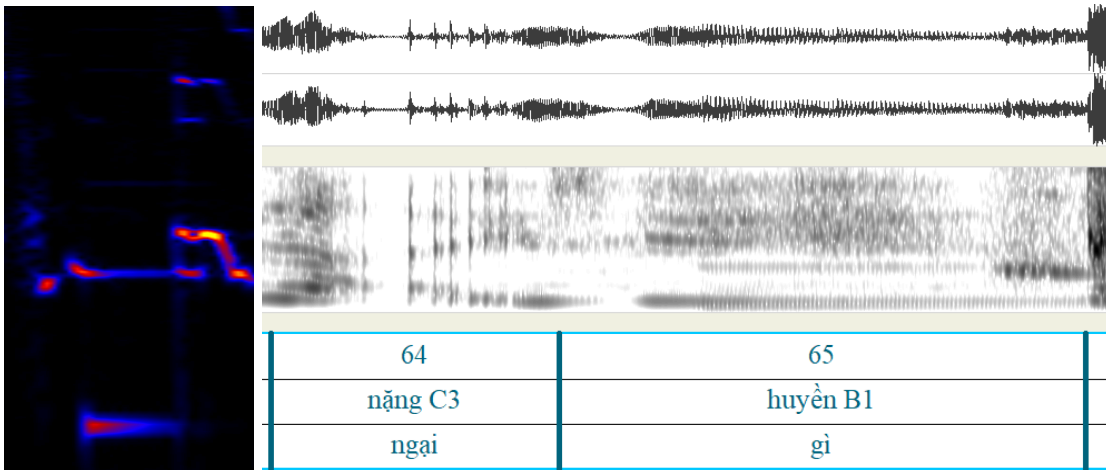

Figure 5.4. "Bùa Yêu" by Bích Phương (2018).

Acoustic visualizations of creakiness and breathiness on a contrary bigram for the lyrics "ngại gì". Aperiodic pulses indicate creakiness during the word "ngại". The right-edge of "gì" features a prominent exhalation indicated by the dark frequency band. This cue enhances the breathiness of *huyền B1*, and reflects the singer's artistic usage of a sighing musical gesture to complement tonal realization.

5.5 Hỏi variation

As mentioned in section 5.1, the *hỏi B4* tone does not behave consistently in the pitch hierarchy. That is, it may sometimes function as a low tone or a high tone (Pham 2003). This matter is further complicated due to the differences in the way *hỏi B4* is realized in different Vietnamese dialects. In the Hanoi variety of Northern Vietnamese, *hỏi B4* may behave as a low falling tone (Kirby 2011), but other varieties of Northern Vietnamese pronounce it with a clear initial pitch fall and rise at the end (Nguyễn & Edmondson 1998). In some Southern Vietnamese varieties, the pitch contour of *hỏi B4* is merged with *ngã C8* (Phan 2022). Thus, the analysis confirms that there are many ways *hỏi B4* references other tonal cues besides pitch in contrary bigrams.

The pitch contour of *hỏi B4* can be conserved by vibrato. Figure 6.1 shows a pitch increase 56 to 58 from *ngang M5* to *hỏi B4*. The second formant in the spectrogram traces a light but noticeable dip in frequency before rising again during the lyric "đổi". In other cases, the curve of the pitch contour in *hỏi B4* is exaggerated even if sung quietly relative to the tune; the contour may show up with a "V" contour in the spectrogram (Figure 6.2). In any of these cases, the singer conserves the general pitch contour of *hỏi B4* while employing a vocal technique for stylistic flair.

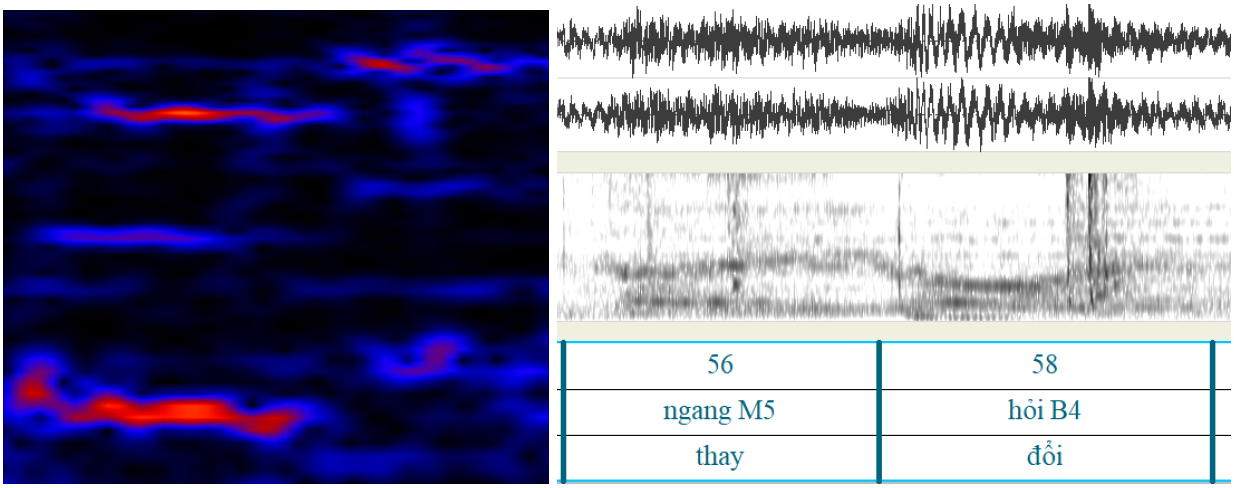

Figure 6.1. "Đếm Ngày Xa Em" by Only C ft. Lou Hoàng (2016).

Acoustic visualizations of vibrato being used to emulate tonal contour on a contrary bigram for the lyric "đổi".

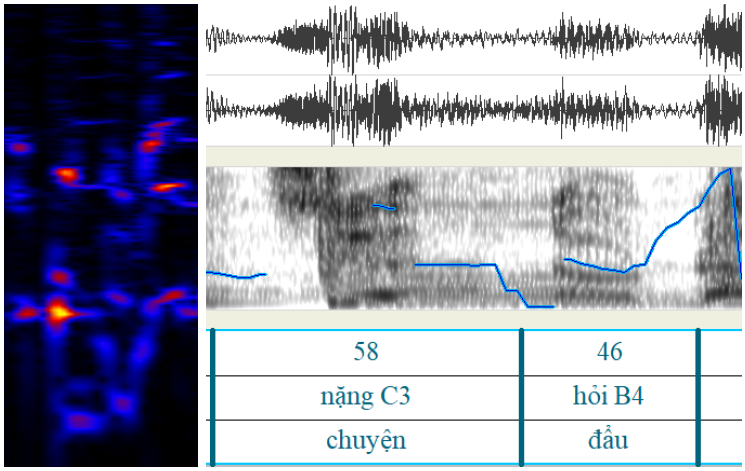

Figure 6.2. "Xoay Hè Khác Lạ" by Trúc Nhân ft. OSAD, Linh Cáo, and Bùi Công Nam (2024). Acoustic visualizations of a subtle pitch contour preservation on a contrary bigram for the lyric "đầu". The blue lines on the *Praat* spectrogram outline a pitch trace for frequencies between 50 and 100 dB.

Hỏi B4 may sometimes trade its breathy phonation for glottalization in contrary bigrams. Though not as drastic as in the case of creaky voice, glottalization includes a decrease in intensity, especially at the higher frequency range due to the closing of the glottis. In most cases, *hỏi B4* does not fully glottalize but rather pursues the principle of glottalization, resulting in a tensing of the vocal folds that appears as a subtle decrease in intensity in both the waveform and spectrogram (Figure 6.3). As noted in section 5.5, the *hỏi B4* tone merges with *ngã C8* in some Southern varieties of Vietnamese. In V-pop, this assimilation may even occur when singers

perform using a Northern accent as a way of conserving the complex pitch-phonation contours of *hỏi B4* and *ngã C8*, both of which are curve tones. Figure 6.4.1 displays a contrary bigram between *ngã C8* and *hỏi B4* tones. The ascending curly blue vocal traces in the *SonicVisualizer* spectrogram reveal the creakiness with which the corresponding lyrics "hãy" and "để" are pronounced in. A prominent white band that cuts across the *Praat* spectrogram results from the F0 lowering that occurs during creaky voice. In the very same song represented in the singer also sings *hỏi B4* with breathy and modal phonations (Figure 6.4.2). In this case, the pitch rises from 49 to 56 on a tone stable oblique bigram featuring the *hỏi B4* tone. On the first lyric "khỏi" the tone is realized as breathy and tensed as evidenced by the vertical band of noise followed by white space in the left panel of the *Praat* spectrogram. Immediately after, the singer differentiates the *hỏi B4* by singing the following lyric "ngẩn" with modal phonation and onset stress. This breathy-modal transfer is faintly captured by the *SonicVisualizer* spectrogram in which the three frequency bands headed by pink intensities have a sudden attack in the middle correlating with the onset stress of "ngẩn". By transferring different phonation types in the realization of *hỏi B4* in a tone stable bigram, the singer is being resourceful in distinguishing the tone's register versatility when constraints of the tune stream force pitch cues to be out of tune. The realization of *hỏi B4* as creaky, breathy, and modal in different cases may give insight into gradation in the register complex of Vietnamese.

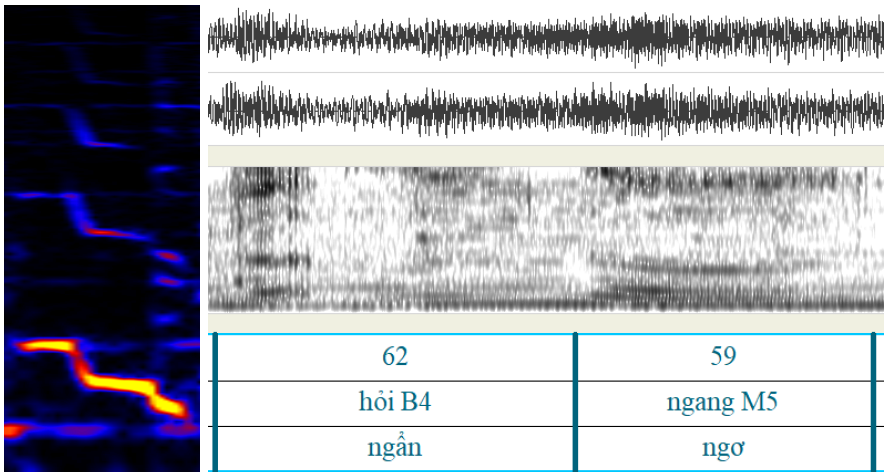

Figure 6.3. "Làm Người Yêu Em Nhé Baby" by Wendy Thảo (2016).
Acoustic visualizations of tensing on a contrary bigram for the lyric "ngẩn".

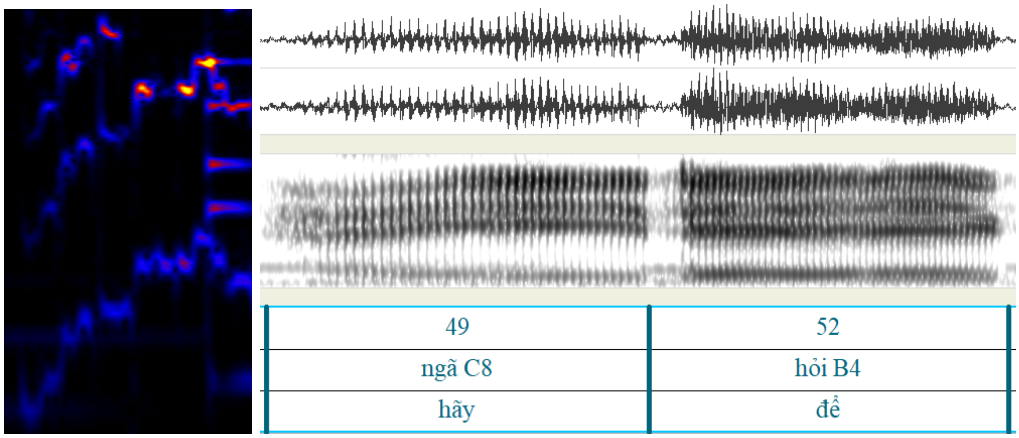

Figure 6.4.1. “Lời Yêu Ngây Dại” by K.H.A. (2019).

Acoustic visualizations of tonal creakiness assimilation on a contrary bigram for the lyrics “hãy để”.

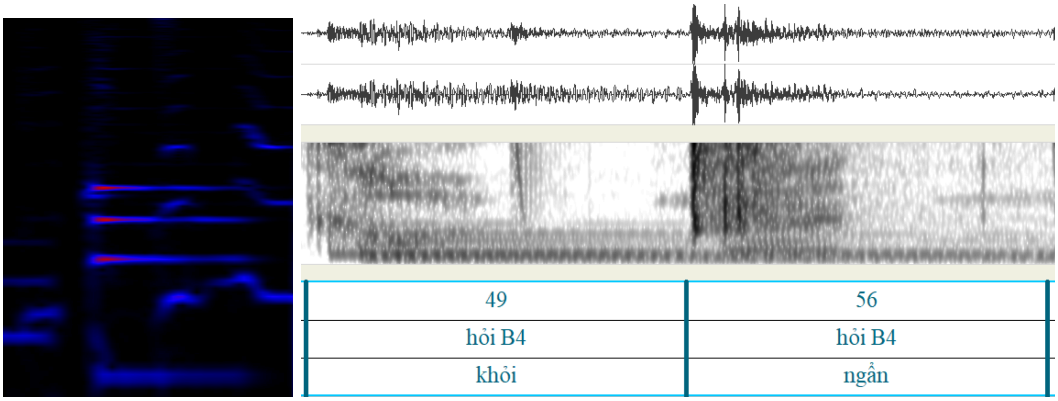

Figure 6.4.2. “Lời Yêu Ngây Dại” by K.H.A. (2019).

Acoustic visualizations of breathiness and modal voicing on a contrary bigram for the lyrics “khỏi ngẩn”.

5.6 Checked tones

Checked tones in Vietnamese occur on vowels preceding unreleased voiceless oral stops in the coda’s final position (Kirby 2011). It has been argued that instead of being allophonic tonemes of *sắc M6* and *nặng C3*, the checked tones *sắc đóng M7* and *nặng đóng B2* induce the manner of the final consonant (Pham 2003). Checked tones tend to possess a shorter duration than the other tones due to the oral closure of the final consonant. The argument to distinguish checked tones is convincing for *nặng C3* and *nặng đóng B2*, which feature entirely different phonation characteristics. Whereas *nặng C3* features glottalization, the phonation of *nặng đóng B2* can vary between the breathiness in *huyền B1* and the modal voice in *ngang M5*.

V-pop singers expressed the characteristics of the checked tones in contrary and oblique bigrams. The singer represented in Figure 7.1 conserved the stress accent on the *sắc đóng M7* for the lyric “bất”. The pitch rises from 67 to 69 on a tone pair that is supposed to fall in pitch (C8 → M7). To compound the pitchwise transgression occurring for this contrary bigram, the *ngã C8* tone of the first lyric “sẽ” fails to employ creakiness, as evidenced by its absence in the dynamically balanced frequency distribution in the *Praat* spectrogram; the waveform also appears modal and regular in its pulse. To rectify this potential tonal error, the singer gives prominence to the lyric containing *sắc đóng M7*, thereby emulating the quick durational stress onset and heightened intensity of the checked tone. This prominence causes the waveform to noticeably swell at the onset of the lyric.

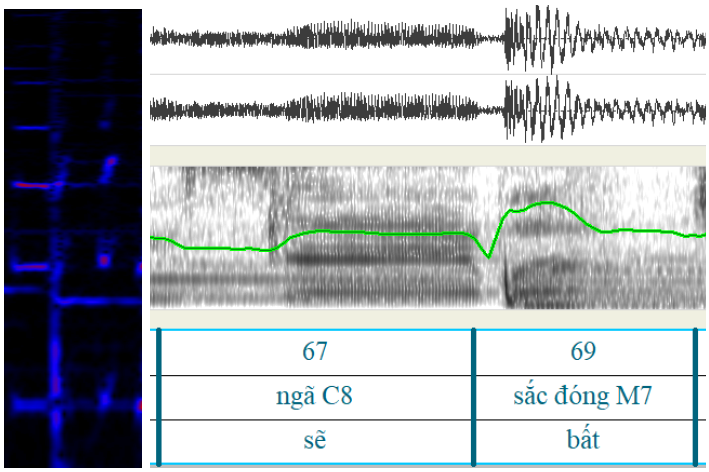

Figure 7.1. “Đã Lỡ Yêu Em Nhiều” by JustaTee (2017).
Acoustic visualizations of stress accenting on a contrary bigram for the lyric “bất”. The green line on the *Praat* spectrogram shows intensity.

The durational quickness of the checked tone is highlighted by the example of Figure 7.2.1 in which a contrary bigram (66 → 64, M6 → M7) contrasts *sắc M6* and *sắc đóng M7*. This example illuminates the immediate differences between the default tone and its supposed allophonic toneme. The lyric with *sắc M6* has a much longer vowel length than the subsequent lyric with *sắc đóng M7* that features an even formant structure and increased prominence at the onset of the syllable as seen in Figure 7.1 with the *nặng đóng B2* tone. Longer oblique sequences can also offer insights into the register complex variation in Vietnamese. Figure 7.2.2 features four

adjacent tone-note pairs interacting under the pitch constraint of a note stable sequence. Across all four pairs, a MIDI pitch of 59 is sustained across a sequence of *nặng C3*, *nặng đóng B2*, *huyền B1*, and *huyền B1* (C3 → B2 → B1 → B1). Tonal differentiation occurs in various ways in this excerpt. From the *Praat* visualization, it can be observed that *nặng C3* is glottalized; *nặng đóng B2* expresses the checked tone characteristics of short duration and increased intensity; and *huyền B1* pulsates quickly with lots of noise in the higher frequencies due to breathiness.

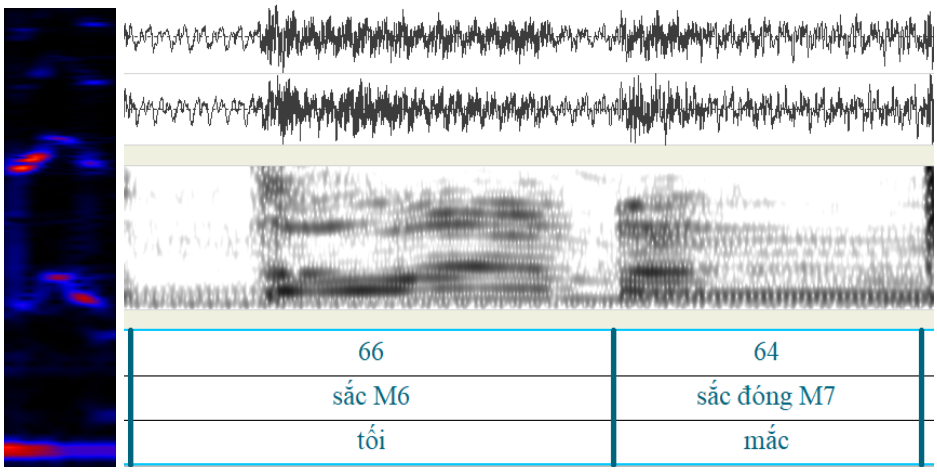

Figure 7.2.1. "Đừng Làm Trái Tim Anh Đau" by Sơn Tùng M-TP (2024).

Acoustic visualizations of checked and open tone stress contrasts on a contrary bigram for the lyrics "tối mắc".

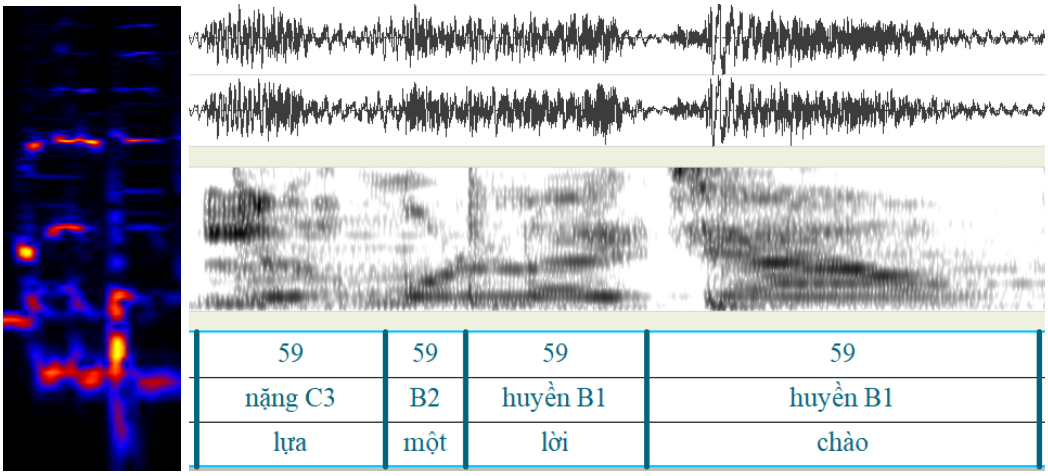

Figure 7.2.2 "Đừng Làm Trái Tim Anh Đau" by Sơn Tùng M-TP (2024).

Acoustic visualizations of a sequence of tones with phonation and durational conservation on a rap sequence of note-stable bigrams for the lyrics "lựa một lời chào".

As mentioned earlier, *nặng đóng B2* may be expressed with modal or breathy expression, while *nặng C3* is pronounced with creakiness. In Figure 7.3, the vocal trace spectrogram represents the process of glottalization in *nặng C3* with a clear fall in frequency and suppressed intensity. The stiffening of the vocal cords during glottalization shows up in the regular spectrogram as a dark vertical frequency band. The following *nặng đóng B2* is expressed with modal phonation as indicated by the regular periodicity of the waveform. Therein, even when pitch cancellation occurs in the example (52 → 54, C3 → B2), tonal cues such as prominence and voice quality prevail in distinguishing the tone.

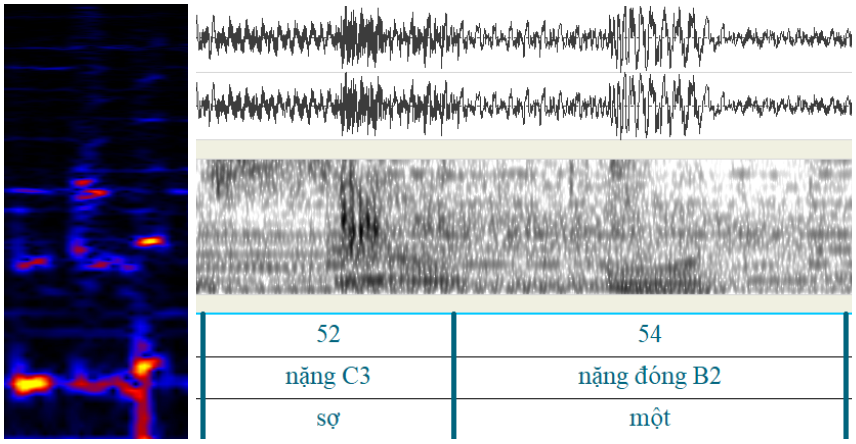

Figure 7.3. "Yêu Đương Khó Quá Thì Chạy Về Khóc Với Anh" by Erik (2022). Acoustic visualizations of checked and open tone phonation conservation on a contrary bigram for the lyrics "sợ một".

6. Discussion

There is clear evidence that there are multiple cues that contribute to the identity of a tone in Vietnamese. This network of cues may be referred to as the register complex. Though there is still much debate about whether Vietnamese is a tone language or register language, this study reveals that the constituents of the register complex do not all have to be active during the pronunciation of a tone for it to maintain its identity. This is evidenced by the fact that V-pop performers leveraged different tonal cues when pitch cancellation occurred in the tone-tune interface.

Though V-pop continues to demonstrate a strong favorability to be in tune for tone-note pairs, the increased number of oblique interactions reflects the stylistic ethos of pop music.

Catchy tunes, repetitive melodic structure, and rhythmic groove constrain the melody to occupy narrower vocal registers, at the extremes giving rise thus to the presence of oblique (note stable) streams in the music. Despite the fact that V-pop singers demonstrate resourcefulness in register complex transfer for both oblique and contrary interactions, elements of this musical trend perhaps broach a discussion into how globalized cultural phenomena may impact the behavior of a language in different environments, especially in the case of V-pop wherein a music theoretic system not specifically designed for tone language text-setting is fused with the language nevertheless. A sociolinguistic study can be extended to future research.

This study contributes toward our understanding of the tone-tune interface for Vietnamese and tone languages in general. The largest corpus to date for these purposes allowed the question to corroborate pitchwise patterns that may answer crucial text-setting questions for tone languages. The study is the first to investigate the role of tonal transfer in the register complex under the tone-tune interface, expanding the purview of tone beyond a centralization of pitch. Musicolinguistic data may offer a valuable perspective into understanding linguistic behaviors via the speech-song contrast. Future research may focus on the interpretability of the *hỏi B4* (low curve) tone in different prosodic contexts, the role of musical form on the tone-tune interface, the influence of syntactic dependency on the pitchwise interaction of tone-note pairs, the behavior of tone in rap, and the sociocultural relevance of music videos in V-pop.

7. Conclusion

A corpus study of 45 V-pop songs from 2016 to 2024 was conducted to observe the interaction of phonemic pitch and musical pitch on tonal realization. An eight-tone representation of Vietnamese with the inclusion of checked tones was employed in order to distinguish the fine-grained behaviors of each tone. An acoustic analysis was performed on vocal excerpts in which lexical tone pitch was suppressed by musical pitch. Results from the acoustic analysis reveal that various transfers between different tonal constituents of the Vietnamese register complex allow tones to negotiate pitch discrepancies between song and speech. The study advances Vietnamese phonological theory by accounting for what specific prosodic constituents comprise a tone gestalt. Broadly, probing the musicolinguistic interface to observe the behaviors of speech in different acoustic environments can help inform our understanding of prosodic phenomena.

8. Acknowledgements

I would like to thank Dr. Simon Todd, Dr. Andrew Watts, and Huy Phan for their useful input on devising my methodology. This project was supported by the John and Mary Kelly Summer Undergraduate Fellowship fund at the UC Santa Barbara College of Creative Studies.

References

Brunelle, Marc. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*. 37.79–96. Elsevier Ltd.

Chao, Y.R. 1930. A system of tone letters. *Le maître phonétique*, 45.24–27.

Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association*. 41.381–92. Cambridge, UK: Cambridge University Press.

Kirby, James, and Robert D. Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. *The 5th International Symposium on Tonal Aspects of Languages (TAL2016)*. 48–51.

Ladd , D. Robert , and James Kirby. 2020. Tone-melody matching in tone language singing. *The Oxford handbook of language prosody*. 676–87. Oxford, UK: University Press.

Lê et al. 2011. A study on Vietnamese prosody. *New Challenges for Intelligent Information and Database Systems. 63–73.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Mac et al. 2011. How Vietnamese attitudes can be recognized and confused: Cross-cultural perception and speech prosody analysis. *2011 International Conference on Asian Language Processing*. 220–3. IEEE.

Mac et al. 2015. Modeling Vietnamese speech prosody: A step-by-step approach towards an expressive speech synthesis system. *Trends and Applications in Knowledge Discovery and Data Mining*. 273–287. Cham: Springer International Publishing.

Ngo, Thi Duyen, and The Duy Bui. 2012. A study on prosody of Vietnamese emotional speech. *2012 Fourth International Conference on Knowledge and Systems Engineering*. IEEE.

Nguyễn, Bảo Vĩnh. 1970. Introduction to Vietnamese Music. https://www.namkyluctinh.org/a-nghethuat/vinhbao-introtovnmusic.pdf.

Nguyễn, Văn Lợi, and Jerold A. Edmondson. 1998. Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*. 28.1–18.

Phạm, Hoa T. (Andrea). 2001. Vietnamese tone: Tone is not pitch. Ph.D. dissertation, University of Toronto.

Phạm, Hoa T. (Andrea). 2003. Vietnamese tone: A new analysis. New York; London: Routledge.

Phan G.A.T., and Ngô T.N. 2016. An initial analysis on the interactions of Vietnamese linguistic tones & Vietnamese folk music. *5th International Conference on Vietnamese Studies: Sustainable Development in the Context of Global Change*. Hanoi, Vietnam: Vietnam National University.

Phan, Huy T. 2022. Aspects of Bến Tre phonology. *10th International Conference on Austroasiatic Linguistics (ICAAL10)*. California State University, Long Beach.