CHAPTER 2

Literature Review

Many researchers have used linguistic methodology to analyze music (Jackendoff and Lerdahl 1982, 1983, 1989, Gilbers and Schreuder 2002, Sundberg and Lindblom 1991, Van der Werf and Hendriks 2004), investigating the application of general grammar in music, including Optimality Theory. In this chapter, I will review some theoretical background, such as "Optimality Theory" and "Metrical Phonology", as well as some previous studies about the interaction between language and music.

2.1 Optimal Theory

2.1.1 Optimal Theory

Optimality Theory (OT) is a linguistic model originally proposed by the linguists Alan Prince and Paul Smolensky in 1993. OT views grammars as a set of ranked violable constraints. These constraints are assumed to be universal, and the differences between child phonology and adult phonology, or between the grammars of different languages, lie in differences of the ranking of the constraints. The operation in OT mainly involves two systems: Generator (GEN) and Evaluator (EVAL). Given an input form, a generator, GEN, generates a list of candidates which are assessed by the constraints. EVAL chooses the candidate that best satisfies a set of ranked constraints; this optimal candidate becomes the output. Here, two characteristics of this theory emerge: strict ranking and violability. The former refers to the principle that a candidate violating a high-ranking constraint cannot redeem itself by satisfying low-ranked constraints. The latter means that the optimal candidate does not have to satisfy all constraints. It, however, is the one which violates not necessarily the fewest, but the lower-ranked constraints. The following tableau shows the operation of Optimality Theory.

(1)

Input	Constraint A	Constraint B	Constraint C
Candidate 1	＊！		
☞Candidate 2		＊	＊
Candidate 3		＊＊！	

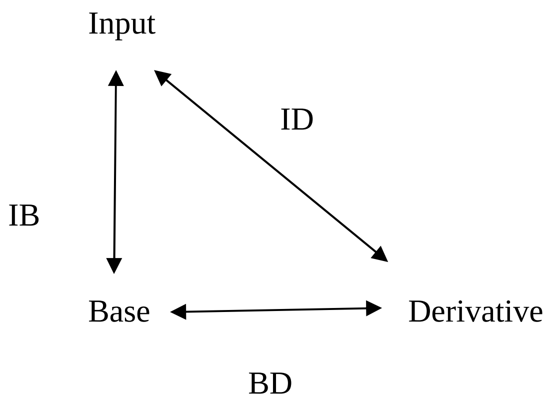
As the above tableau shows, constraint A is ranked higher than constraint B, followed by constraint C in a certain language. The violation of higher-ranked constraint is fatal, thus Candidate 1 is eliminated though it satisfies the following constraints. Candidate 2 has the minimal violation of the secondary constraint compared with Candidate 3, so it is chosen as the optimal output (with the symbol☞ indicating the best candidate.)

The constraints of EVAL are of two types. On the one hand, markedness constraints enforce well-formedness of the output, prohibiting structures that are difficult to produce or comprehend, such as consonant clusters, or some marked segments that are not common in the target language. On the other hand, faithfulness constraints require identity between the input and the output in various ways. Markedness and faithfulness constraints conflict in essence; consequently, the constraints' rankings, which differ from one language to another, determine the output forms.

2.1.2 Correspondence Theory

Influenced by Optimality Theory, McCarthy and Prince (1995) raise the Correspondence Theory, which focuses on the similarity between the output base and the reduplicated forms. Being elaborated by many researchers, (Ito, Kitagawa and Mester 1996, Kager 1999, Nelson 2003), this theory not only examines the correspondence between input and output (IB and ID), but explores the correspondence between output forms (BD), as shown in (2).

(2)



In Correspondence Theory, the identities of input-to-output and output-to-output relationships are examined by three types of faithfulness constraints, as shown in (3).

(3) Faithfulness constraints in Correspondence Theory

A. MAX-IO: Every input element must have its correspondent one in the output.

B. DEP-IO: Every output element must have its correspondent one in the input.

C. IDENT-IO: Every element in the input and in the output must be correspondent.

2.1.3 Generalized Alignment

McCarthy and Prince (1993) suggest that some languages require certain features appear at the left edge or the right edge of a domain. Such a distribution can be achieved by the postulation of alignment constraints. Developed from the edge theory of Selkirk (1986), and Inkelas (1989), McCarthy and Prince (1995a) propose the idea of Generalized Alignment, in which the phonological representation of a sentence is constructed through mapping the edges of syntactic constituents to the corresponding edges of phonological constituents. The definition of Generalized Alignment is shown in (4).

(4) Align(Cat1, Edge1, Cat2, Edge2) $=_{def}$

$\forall$ Cat1 $\exists$ Cat2 such that Edge1 of Cat1 and Edge2 of Cat2 coincide.

Where Cat1, Cat2 $\in$ PCat $\cup$ GCat

Edge1, Edge2 $\in$ {Right, Left}

Generalized Alignment is applied mostly in comparing two categories in prosodic hierarchy, such as syllable, foot, and PrWd. For example, the alignment constraint ALIGN-P_RW_D confines two prosodic categories, foot and PrWd:

(5) ALIGN- P_RW_D

Align (PrWd, L, Ft, L)

'The left edge of each prosodic word must match the left edge of some foot.'

2.1.4 Anchoring

To account for the preservation of contrasts in privileged position, Beckman (1998) claims that positions that are either acoustically prominent (stress) or psychologically prominent (morpheme-initial, part of the root) could be signaled by Positional Faithfulness constraint (such as FAITH ONSET). Anchor constraints are proposed to examine the correspondence between the targeted position in the base and the segment standing in the same position in the reduplicant. McCarthy and Prince (1995b) give the definition.

(6) LEFT-ANCHOR (Base, Reduplicant): The left edge of the reduplicant corresponds to the left edge of the base.

2.2 Metrical Phonology

2.2.1 Metrical Structure

Metrical structure is a hierarchical organization regarding the patterns of strong and weak beats in a linguistic piece. According to Lerdahl and Jackendoff (1983), metrical beats contribute to linguistic rhythm just as what musical beats do to musical rhythm. The preference of a final silent beat in a metrical line results in pausing or final syllable lengthening. The singer can choose to pause and articulate, or ignore the articulation mark with lengthening the note before it. Jakobson (1970) constructs a heptasyllabic line into different constituents based on footing, as decribed in (7).

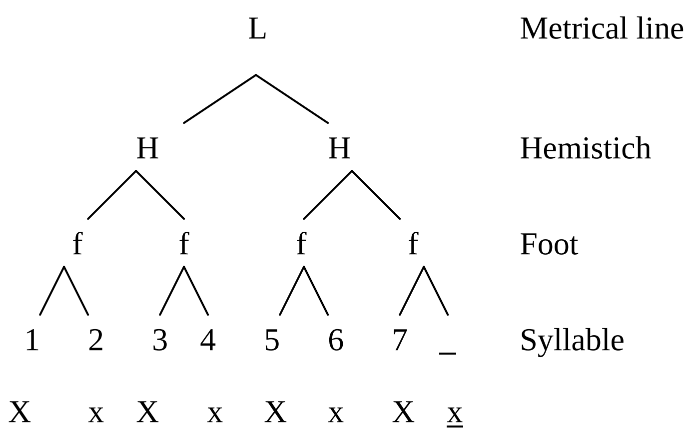
(7)

a. Heptasyllabic line: 12/34/567

b. Pentasyllabic line: 12/345

Applying the concept of beats to Chinese regulated verses, Chen (1979, 1980) proposes a metrical template as in (8)

(8)



As we can see from the template, a metrical line contains two hemistiches, which in turn contains two feet, respectively. A foot consists of two syllables, or one syllable followed by a silent beat. From the template, there are two types of regulated verse: The heptasyllabic pattern like (8) or the pentasyllabic pattern in which the first foot is dropped.

Duanmu (2004) conducts a corpus study of Chinese regulated verse and claims that a tonal language, such as Chinese, still displays a system of stress. He considers trochaic meter a part of Universal Grammar and reanalyzes Chinese regulated verses under trochaic meter. Agreeing with Duanmu, Hsiao (2006) applies the idea of metrical template and trochaic meter to investigate Changhua folk verse and provides an OT account of the rhythm of it. Some examples discussed in his research are shown below (Hsiao 2006).

(9) ‘The most worrying is being poor.’

*a.	X	x	X	x	X	x	X	x
	te -	it	huan - l o		si	ka-	lai	san
	first		worry		be	home		poor
b .	X	x	X	x	X	x	X	x
	te -	it	huan - lo si		ka-	lai	san	
	first		worry	be	home		poor	
c .	X	x	X	x	X	x	X	x
	te -	it	huan - lo	si	ka-	lai	san	
	first		worry	be	home		poor	

Each syllable in (9a) is assigned a demibeat, resulting in a feminine line; the reading is unacceptable. The be-verb *si* shares a demibeat with the preceding syllable in (9b) and with its following syllable in (9c); both readings are metrical. From this study, Hsiao shows the priority of a masculine rhythm and regards function words as a factor governing beat-sharing.

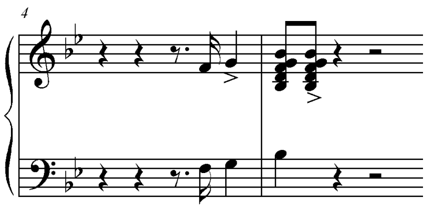
2.3 Language and Music

2.3.1 Language and Musical Structure

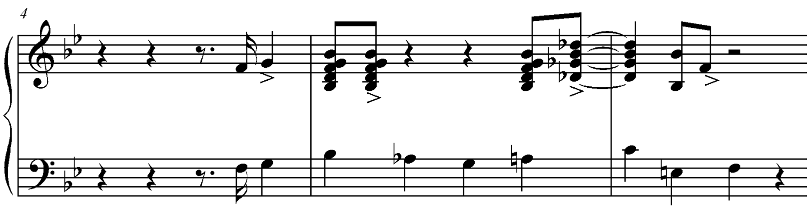
Liberman (1975) proposes that principally every form of temporally ordered behavior is structured the same way. Jackendoff and Lerdahl (1980) detect the resemblance between the ways how linguists and musicologists structure their research objects and further develop generative theories of musical grammars based on Western tonal music (Lerdahi &Jackendoff 1983). Sunberg and Lindblom (1991) claim that language and musical structures are constructed in a hierarchical way, so they can be decomposed into smaller parts. In terms of language, a sentence consists of several phrases, which consists of several words. Words are formed by syllables, which are composed of segments. In music, a musical stream of sound is hierarchically divided into several domains, which are composed of smaller domains. According to Gilbers and Schreuder (2002), the smallest domain in music is the motif, which is built up by at least two notes. Several motifs form themes, which are always recognizable for listeners. Several themes form a phrase, which is a kind of musical sentence. Several phrases then form larger musical sections. To show the decomposability of music structure, Gilbers and Schreuder (2002) give an example of the construction in the jazz original “Tuxedo Junction,” as provided in (10).

(10) Tuxedo Junction

a. Motif



b. Theme (phrase)



c. Section (verse)

2.3.2 Language and Musical Beats

Jackendoff (1989) discovers an interesting parallel between language and music: In both language and music, metrical weight is intimately related to length. To show the similarity clearly, he compares these two structures in his research about tonal music, as is shown in (11): a is adopted from Hayes (1995:38) and b is from Jackendoff (1989:21).

(11) a.

```
                    x
x                   x
x       x           x
x   x   x   x       x
Miss iss ipp i     mud
```

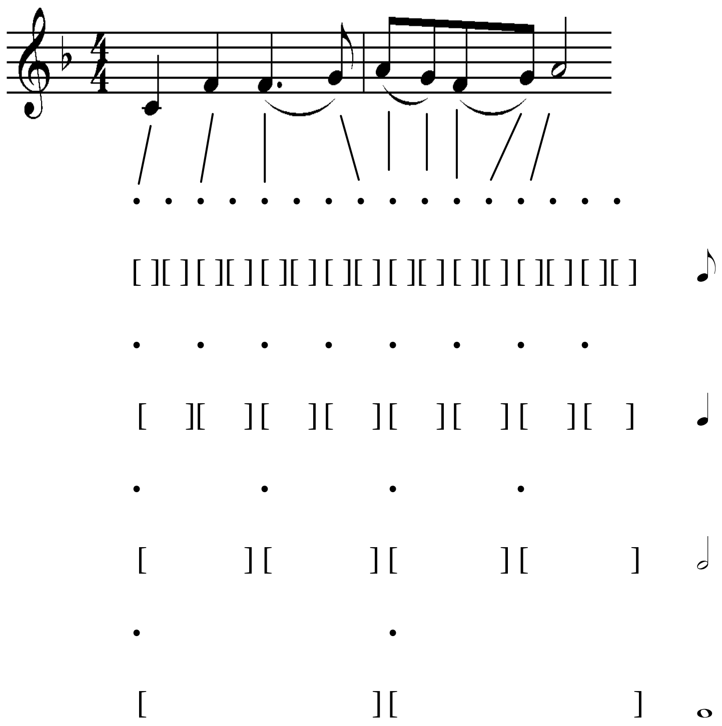
b.

(11a) is a linguistic metrical grid of a noun phrase and (11b) is the beat construction of one musical piece. The Xs in the bottom row of (11a) and the dots in the top row of (11b) serve as place markers. Linguistically, a place marker signals a syllable; musically, a place marker indicates a beat of the lowest level in the musical structure. The other Xs in (11a) are used to mark stressed syllables. Syllables endowed more Xs are more stressed. Similarly, the dots in (11b) are used to signify stressed beats at a given level, thus the beat with the most dots is the most stressed.

2.3.3 Time-span Reduction Theory

Lerdahi and Jackendoff (1983) propose the idea of “time-span”. In music, beats are idealized musical signals; they are analogous to points and represented by dots. Beats occur in time and an interval of time takes place between two beats. Such interval is termed “time-span.” Thus, it is time-spans that have duration, not beats. The segmentation of time spans is affected by metrical structures. Below is an example of the time-span segmentation. Note that the lines are used to associate each note with its place marker at the lowest metrical level and the time spans at each level are marked by brackets.

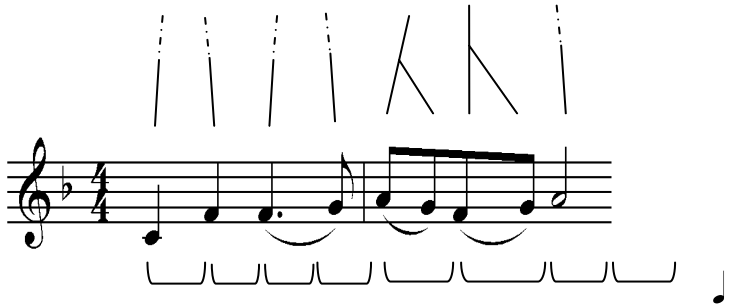
(12) *su kui ang*, ‘all seasons red’ Line 1



The time-span structure helps determining a head-elaboration relationship among pitches in a musical piece. In music, the listeners or performers have the intuitive understanding of important pitches, which are acoustically prominent or have relative structural importance. Musical heads refer to notes with relatively important pitches, while elaborations refer to those with less important pitches. Elaborations can be deleted without affecting the overall structure of the piece. Such deletion causes the reduction of time-spans, so the process is called “time-span reduction.”

Let us construct the tree step by step. The smallest level in this song is the eighth-note level, so no notes are deleted. At the quarter-note level (♩), if there is only one note in a span, its branch extends upward by a dotted line, to be connected at a later stage to another branch. If there are two notes in the span, the left one is the head and dominates the right one, as shown in the second measure of (13).

(13)



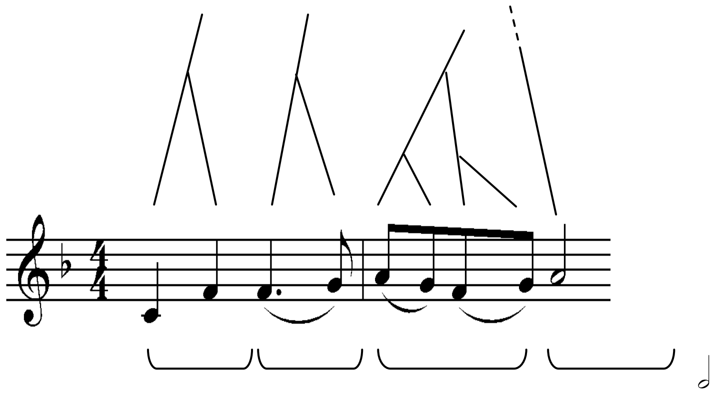
Here, the head is determined by TSRPR1 (Lerdahi and Jackendoff 1983)[1].

TSRPR 1 (Metrical Position):

> Of the possible choices for head of a time-span T, prefer a choice that is in a relatively strong metrical position.

At the next level, the half-note level (𝅗𝅥), the head-elaboration is also determined by TSRPR1. Notes at metrically stronger positions dominate their companions in the same span. The process of the half-note level is shown in (14).

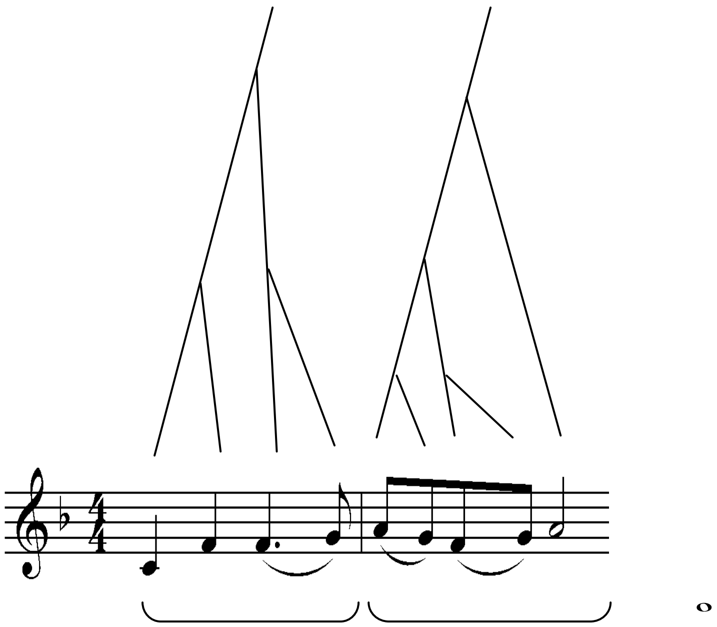
(14)



1 For more principles of selecting musical heads, please see Lerdahi and Jackendoff (1983).

Similarly, the tree at the whole-note level is constructed in (15).

(15)



The idea of head-elaboration relationship is often applied to analyze a song of different versions. Yen (2004) has compared different versions of a folk song *meng jiang nu* (Ho, 1986) under the Time-span Reduction Theory and she claims that different versions of a song can be seen as a musical theme and its variations. Examples are given in (16), where (16a) is considered the theme and (16b) its variation. The time-span reduction is displayed by the tree diagram above the score, where H stands for head and E refers to Elaboration.

(16) a.

b.

The stress pattern of the song is trochaic, thus the first beat of each measure is strong and the first note is the most stressed. So, in the first measure of (16b), the first note is preferred to be the head according to TSRPR1, which prefers a head that is in the relatively strong metrical position; the second note is heard as an ornament and can be reduced. As predicted, the reduced form will be the same as the first measure in (16a).

As for the second measure in (16b), the first note is chosen the head for the same reason explained above. The fourth note is head because of TSRWFR4 (Lerdahl and Jackendoff 1983).

TSRWFR 4: if a two-element cadence is directly subordinate to the head *e* of a time-span T, the final is directly subordinate to *e* and the penult is directly subordinate to the final.

In a similar way, the reduced form the second measure in (16b) is the same as that in (16a).

2.3.4 Application of Optimal Theory in Language and Music

Lerdahl and Jackendoff (1983) proposes the idea of preference rules in their musical theory in which some outputs of a musical piece are preferred than others. It is possible for a musical representation to violate a preference rule if the violation helps to satisfy a more important preference rule. The concept of violable preference rules is very similar to that of the well-formedness condition in Prince and Smonlensky's Optimality Theory. Some rules, concerning grouping structure, metrical structure, or time-span structure, restrict the selection of optimal musical representation. Examples are provided in (17).

(17) **GPR 1** (Grouping Preference Rule):

Strongly avoid groups containing a single event.

MWFR 1 (Metrical Well-Formedness Rule):

Every attack point must be associated with a beat at the smallest metrical level present at that point in the piece.

With the prevalence of constraint-based phonology, the Optimality Theory has been applied to analyze the relationship between music and language. Gilbers and Schreuder (2002) are the pioneers in showing how the Optimality Theory helps to explain the output of a musical representation. They discuss the findings in Lerdahl and Jackendoff's music theory from the perspective of OT, believing that the chosen outputs are determined by preference rules. Some examples are provided in (18).

(18) **TSRPR 1**: Choose as the head of a time-span the chord (or the note) which is in a relatively strong metrical position (positional markedness).

TSRPR 2: Choose as the head of a time-span the chord (or the note) which is relatively harmonically consonant (segmental markedness).

TSRPR 7: Choose as the head of a time-span the chord (or the note) which emphasizes the end of a group as a cadence (comparable to the boundary marking effect of alignment constraints in language).

The set of preference rules is hierarchical: TSRPR 2 is stronger than TSRPR 1; TSRPR 7 is stronger than both TSRPR 1 and TSRPR 2. With the above ranked constraints, the chord of a Mozart sonata, which is given in (19), is evaluated in (20) (Gilbers and Schreuder 2002).

(19) Mozart: Sonata K. 331,I

(20)

	TSRPR7	TSRPR2	TSRPR1
☞ E			*
A^6	* !	*	

In (19), the first four measures form a group. In determining the chord of the first group, although the A^6-chord[2] is metrically in a stronger position than E-chord, the dominant constraint TSRPR 7 prefers chord E as the cadence of this musical piece.

Van der Werf and Hendriks (2004) further modify the preference rules and rename the constraints. For example, the GPR1 is renamed SINGLE constraint: groups never contain a single element. They conduct a computational experiment to investigate the similarities and differences between language and music. From their study, the grouping structure of a musical piece can be determined by a set of ranked constraints.

As for Chinese Phonology, more and more researchers pay attention to the interaction between language and music. According to Wang (1983), the harmony between language and music cannot be ignored in the literature of Chinese traditional music, especially in regulated verses or folk songs. The mappings between language

[2] In diatonic scale, the seven keynotes, Do, Re, Mi, Fa, So, La, Si, can be notated as C, D, E, F, G, A, B. thus, A-chord means that La is music's harmonic center and E-chord implies a piece of music is harmonically centered on Mi.

and music can generally be classified into two forms: one is to compose the tune for harmonizing with the existing verse; the other is to edit the lyrics according to the melodies of the tune. Hsu and Kao (2003) conduct an experiment to examine the harmony between melody and lyrics among Chinese songs. They propose the principles of concordance that defines the mapping between music tunes and linguistic tone value. Yen (2004) compares western classical music and Chinese folk songs under the framework of Optimal Theory and observes music universals across different music genres. She proposes some relevant constraints based on Lerdahl and Jackendoff (1983), as shown in (21).

(21) **MIN-BINARITY**: A group contains no fewer than two pitch-events.

PITCH(X)PITCH(X): Adjacent pitch-events that have the same pitch are grouped together.

ALIGN (Group, Left/Right, Slur[3], Left/Right): Align the left/right edge of a musical slur with the left/right edge of a group.

With the above constraints, the grouping structure of the first measure in a Taiwanese folk song "All Seasons Red (四季紅)" is evaluated in (22).

3 A slur is a musical notation, connecting notes that are to be sung or played without a break. It is marked by a curve above several musical notes.

(22) Analysis of the first measure in “All seasons Red (四季紅)”

 Input: 1 2 3 4	MIN-BINARITY	PITCH(X)PITCH(X)	ALIGN(Group, L/R, Slur, L/R)
a. [1] [2] [3] [4]	＊＊＊＊！	＊	
b. [1] [2 3] [4]	＊＊！		＊
c. [1] [2 3 4]	＊！		＊
d. [1 2] [3 4]		＊！	
☞ e. [1 2 3 4]			＊
f. [1] [2] [3 4]	＊＊！	＊	
g. [1 2] [3] [4]	＊＊！	＊	
h. [1 2 3] [4]	＊！		＊

The first constraint MIN-BINARITY requires that one bracketed group must have more than one note. The second constraint PITCH(X)PITCH(X) commands the second and the third notes to be grouped together. The lowest alignment constraint restricts the two edges of a group to be aligned with the domain of a slur, the third note ♩. and the fourth note ♪. From the evaluation, candidate e is chosen as the best.

Song (2006) investigates some verses from the Book of Songs, known as Shijing, under the framework of OT. She proposes some constraints that govern the mapping between lyrics and melodies in these verses, as provided in (23).

(23) **LONGΣ(R)**: The right syllable of a foot must be longer than the left one.

MAX-IO: Every syllable of the linguistic input must be retained in the musical output.

***𝄽**: Any rest mark is prohibited.

σ=♩: The beat of every syllable equals to a quarter note.

With the above constraints, a line of *Kuan Ju* is evaluated in (24).

(24)

關關雎鳩，在河之舟		LONGΣ(R)	MAX-IO	*𝄽	σ=♩
☺a.					*
b.		*!			
c.		*!		*	*
d.			*!		

From the tableau, candidate A wins out because it satisfies the upper three constraints. Her research suggests that the verse is composed melodically with correspondence to the phonological form of the metrical line. Furthermore, Song

(2008) employs the Optimality Theory to explore the interactions of language and music in Jiang Kuei's verses. She claims that harmony is the principle of composition no matter how language and music integrate, either tune-based or verse-based.

In brief, a song is the combination of melody and lyrics. The delicate mappings between music (note, beat and musical structure) and language (syllable, foot, tone and prosodic structure) in Western and Chinese verses have been explored. In the following sections, several Taiwanese folk songs will be investigated in detail to show how the melodies are affected by lyrics and how the lyrics are arranged under the influences of different musical styles.