國立政治大學語言學研究所

博士學位論文

National Chengchi University
Graduate Institute of Linguistics
Doctoral Dissertation

英語與華語歌曲中語言與音樂之關係：音段、節奏與聲調

The Connection Between Language and Music in English and Mandarin Songs: Segment, Rhythm, and Tone

指導教授：蕭宇超 博士
Advisor: Yuchau E. Hsiao, Ph.D.

研究生：凌旺楨 撰
Student: Wang-Chen Ling

中華民國：一〇九年六月
June, 2020

DOI:10.6814/NCCU202001428

Acknowledgements

終於走到這一步了，感覺很不真實，但內心卻踏實。若不是指導教授蕭宇超教授的傾力指導，畢業對我而言遙不可及。感謝蕭教授一再鼓勵、引領。無論在什麼情況，心裡總是惦記著我的論文。老師的恩惠難以報答，學生銘記在心。也感謝口試委員黃慧娟教授、林蕙珊教授、歐淑珍教授、吳貞慧教授以及計劃書口試委員吳瑾瑋教授費心審閱我的論文，細心地指出需要修正之處，給予了相當寶貴的意見。

感謝政大教授們的教導與幫助。賴惠玲教授、徐嘉慧教授、黃瓊之教授、萬依萍教授、張郇慧教授、鍾曉芳教授以及戴智偉所長，您們課堂上的教導仍歷歷在目，而您們的鼓勵與關懷也成為我完成博士學位的動力。感謝惠鈴助教學姊在我博士班生涯內扮演的不可或缺的角色，只要有學姊在，我就能安心地完成每一件事。也感謝我的大學教授，陳純音教授、張妙霞教授以及丁仁教授。謝謝您們領我進入語言學的領域。

感謝為我加油打氣的好夥伴子權與怡臻，謝謝你們和我一起討論問題，為我解惑。也感謝欣蓉、德偉時常幫忙分擔事務。謝謝 Johnny 和侃彧熱心幫忙我解決問題。也謝謝明昌學長與凱琳學姊樹立了好榜樣。此外，也要感謝菘霖學長/教授與婉婷學姊的鼓勵與分享，幫助我能逐漸畫出博士學位的樣貌。

感謝公公、婆婆一直以來的付出與照顧，謝謝您們給予我溫暖，對您們的感謝之詞難以言盡。也謝謝先生的陪伴與鼓勵，讓我能順利完成論文，開啟人生新的階段。也要感謝乖巧到令人心疼的好女兒，妳總是默默地等待，犧牲與媽咪的相處時間，為了換來媽咪的博士學位。謝謝每一位家人多年來不離不棄的關懷與幫忙。

特別感謝父母親的全力支持，讓我無後顧之憂。您們是我心目中全世界最棒的父母。感謝您們費盡心力的養育之恩。您們不求回報的愛 讓我在快要放棄時，能爬起繼續往前行。這本論文收藏著你們為我的蹙眉著急以及為我終於能畢業的欣慰笑容。謹將此論文獻給您們。

最後也要感謝教會家長們願意讓孩子作為發音人，讓我能順利蒐集語料。感謝教會為我代禱的每一位弟兄姊妹，在我最慌亂無助時，幫助我尋得從神而來的平安。這一切的一切都要感謝神每一步的安排與帶領，讓我在寫論文的過程中，真真實實地經歷那不可思議的恩典。

DOI:10.6814/NCCU202001428

摘要

本論文以優選理論(Prince and Smolensky 1993/2004)分析英語及華語歌曲中音樂與語言的互動關係，並由音段、節奏與聲調三方面切入探討。本文首先研究語言到音樂的對應，檢視華語成人如何將自己說出的英語聲母子音串，以及韻尾子音串，感知進入音樂中唱出。研究結果顯示音樂節拍會導致唱的音段與唸的音段不同。此外，本文也討論作曲者如何將華語兒歌歌詞韻律結構，包含音步(foot)及語調詞組(intonational phrase)，感知進入音樂之中。本文發現韻律結構會與音樂旋律對整、並影響音樂節拍的指派。

音樂到語言的對應則是觀察華語兒童聽華語歌曲時，如何將音樂旋律感知為語言聲調，並說出其所聽到的音節聲調。結果顯示兒童所感知的語言聲調與歌曲旋律相近。然而所產出的聲調則因心理詞彙的連結與否，而可能造成產出與感知的聲調不同。

節奏的對應與音段的改變，顯示語言到音樂的感知語法具有重要角色。兒童產出聲調與感知聲調的不同，亦可驗證感知與產出係由獨立不同的語法所管理。

關鍵詞:

音樂與語言的對應、感知語法、詞彙連結、韻律結構、子音串、優選理論

DOI:10.6814/NCCU202001428

Abstract

This study aims to investigate the relationship between language and music through segment, rhythm, and tone in English and Mandarin songs. The language-music mappings are analyzed under the framework of Optimality Theory (Prince and Smolensky 1993/2004).

This research first examines the language-to-music mapping through Mandarin-accented English in reading and singing. This study discusses how onset and coda clusters are produced in the linguistic output, and how the linguistic output is perceived into music. Segmental changes are shown in the language-to-music mapping, where beat assignment takes place and affects the linguistic output form. The language-to-music mapping is also explored through the composing of Mandarin children's songs. This study finds that the prosodic structures of foot and intonational phrase in Mandarin children's songs are aligned with music structures. Footing and intonational phrasing also affect the musical beat assignment.

The music-to-language mapping is examined through children's perception of the musical melody in songs. The issue lies in how native Mandarin-speaking children map the music pitches to Mandarin tones, and how they produce what they hear. In the perception of the musical pitches, singing words or phrases are perceived into the linguistic input forms with similar tonal values. However, the children may produce syllables whose tones are different from what they perceive because of lexical association.

This research provides an extensive study on the interaction between language and music. Segmental changes, rhythmic correspondences, and tonal adjustments in language-music mapping reveal that perception and production are governed by independent phonological grammars.

Keywords: Language-music mapping, Perception grammar, Lexical association, Prosodic structure, Consonant cluster, Optimality Theory

DOI:10.6814/NCCU202001428

Table of Contents

DOI:10.6814/NCCU202001428

DOI:10.6814/NCCU202001428

DOI:10.6814/NCCU202001428

Chapter 1

Introduction

1.1 Research Issue

This research aims to investigate the relationship between language and music through perception and production grammars in English and Mandarin songs. Three aspects are examined, namely, segment, rhythm, and, tone.

This study first examines the language-to-music mapping through Mandarin-accented English in reading and singing. I investigate how onset clusters are produced in the linguistic output, and how the linguistic output is mapped into music. Segmental changes are shown in the linguistic mapping and the language-to-music mapping where beat assignment takes place and affects the linguistic output. This study continues to discuss the segmental changes in coda clusters. The language-to-music mapping is also explored via the composing of Mandarin children's songs. The research examines how foot is formed in the linguistic output, how the structure of the foot affects the musical beat assignment, and how the musical structure surfaces in the musical output. The mapping of intonational phrase (IP) and musical structure is also discussed.

The music-to-language mapping is investigated through children's perception of the musical melody in Mandarin songs. The issue lies in how native Mandarin-speaking children map the musical pitches to linguistic tones, and how they produce what they hear. In the perception of the musical pitches, singing words or phrases are mapped to the linguistic input forms with similar tonal values. In production, toned strings are interpreted in the output through lexical association. This study examines the output tone of successful lexical association, which is followed by the discussion of the output

DOI:10.6814/NCCU202001428

tone with lexical access failure.

The outputs of the language-to-music mapping and the music-to-language mapping are governed by perception and production grammars, which are schematized in section 1.2.

1.2 The Proposed Model

This research proposes a model for the language-to-music mapping, and that for the music-to-language mapping.

The model in (1) predicts the language-to-music mapping, where segmental changes and rhythmic mappings are accounted. The linguistic input maps to the output through the relevant production grammar. The linguistic output then maps to the musical input by the perception grammar. Finally, the musical input maps to the output by the production grammar.

(1) Language-to-music mapping: segment and rhythm

Linguistic input
| *Production grammar*
Linguistic output
| *Perception grammar*
Musical input
| *Production grammar*
Musical output

The model in (2) predicts the music-to-language mapping, where tonal changes are accounted. The musical output maps to the linguistic input by the perception grammar, and then the linguistic input maps to the output by the production grammar.

(2) Music-to-language mapping: tone

Musical output
Perception grammar
Linguistic input
Production grammar
Linguistic output

This research provides an extensive study on the connection between language and music by positing an independent perception grammar. The independence of the perception grammar is revealed from segmental changes, rhythmic correspondences, and tonal adjustments. The language-music mappings are analyzed under the framework of Optimality Theory (Prince and Smolensky 1993/2004). Different constraints and rankings are respectively posited in the production and perception grammars.

1.3 Organization

This dissertation is composed of nine chapters. Chapter 1 introduces the core research issue, the theoretical proposal, and the organization of this dissertation. Chapter 2 discusses previous studies on the language-music connection, perception, prosodic phonology, classic OT, and Stratal OT. Both Chapter 3 and Chapter 4 compare the linguistic mapping and the language-to-music mapping through Mandarin-accented

English in reading and singing. Chapter 3 discusses how musical beat assignment affects the segmental changes in relation to the onset clusters and how the prosodic word conditions musical beat assignment in the musical input. Chapter 4 continues to discuss the Mandarin-accented English in reading and singing, with a focus on coda clusters. Chapters 5 and 6 examine rhythmic correspondences in the composing of Mandarin children's songs. Chapter 5 investigates the mapping from foot structure to musical structure while Chapter 6 continues to probe into the mapping between intonational phrase (IP) and musical structure. Chapter 7 investigates children's perception of the musical melody in the singing of Mandarin songs. Chapter 8 continues to examine children's perception of tones in Mandarin songs and the music-to-language mapping that produces unassociated terms. Chapter 9 offers the conclusion and proposes the remaining issue for further studies.

DOI:10.6814/NCCU202001428

Chapter 2

Literature Review

This chapter introduces the current researches of music-language connection as well as the theoretical background of perception grammar, prosodic phonology, and Optimality Theory (Prince and Smolensky 1993/2004).

2.1 Music-Language Connection

Previous studies demonstrate the parallel feature between language and music. Lindblom (1978) proposes that lengthening of final elements is not only shown speech but also in music. Lerdahl and Jackendoff (1983) observe the similarity between metrical beat in linguistic rhythm and musical beat in musical rhythm. Sunberg and Lindblom (1991) propose that both language and musical structures are structured hierarchically and can be parsed into smaller sections. Schreuder (2006) also explores the resemblance between linguistic rhythm and musical rhythm.

Studies show the reconciliation between language and music in structure and stress mapping. Halle and Lerdahl (1993) discover that when singers encounter novel stanza for a song they know, they have the consistent ability to set the stanza into the song. For example, singers tend to match stressed syllables to strong positions in music. Following the study of Halle and Lerdahl (1993), Hayes (2005) investigates the textsetting intuition under the framework of OT, in which constraints can be violated for a more important purpose. For instance, stressed syllables are placed in weaker rhythmic positions in order to avoid long lapse, where long sequence is without syllable.

DOI:10.6814/NCCU202001428

The linguistic and metrical mappings also show various rhythmic correspondences. Selkirk (1984) proposes the silent grid positions, which may correspond to pausing or syllable lengthening. Hsiao (2006, 2007) observes the silent demibeat in Taiwanese nursery rhymes and Changhua folk verse. Metrical lines with silent demibeat in the final position are regarded as the masculine lines. In the Changhua folk verse in (3), the final demibeat is silent so it is regarded as a masculine line.

(3) X x X x X x X x
o tsiao peh tsiao lai thao tsia
black bird white bird come secretly eat
'Black birds and white birds come to steal food.'

(Hsiao 2006: 10)

Hsiao (2006) also observes that immediate constituents (ICs) share a demibeat to create masculine lines. As in (4), the pair of ICs *tshut-lai*, share one demibeat.

(4) X x X x
tsao tshut-lai khuanN
Run DIR-DIR look

(Hsiao 2006:17)

On the other hand, Huang (2007) builds a corpus and examines the alignment of prosodic structure and the movement of the finger rhymes. Sung (2012) also investigates the structure alignment, and syllable-to-musical beat mapping between Chinese verse line and music. For example, the last syllable in a stanza corresponds to the longest beat.

DOI:10.6814/NCCU202001428

Linguistic mora is found to be influenced by music. Ito, Kubozono, Mester & Tanaka (2019) examine the rhythmic adaptation of batters' names into baseball chants. Baseball fans set batters' names into three beats (X). The base chant shows structural mapping between rhythm and the linguistic mora. For example, the mapping principle for the 3-mora names requires aligning the initial mora to the initial beat (X_1), the final mora to the final beat (X_3), and the medial mora to the medial beat (X_2). Therefore, *ba-a-su* surfaces as *baa-aa-suu* instead of **baa-suu-uu*.

For musical pitch and linguistic tonal mapping, Wong and Diehl (2002) investigate Cantonese songs and discuss how the lyrics of a song in a tone language are understood. Wee (2007) also discusses how listeners of Mandarin songs identify the lyrics from the musical melodies. Wee (2007) proposes that, when preserving contrast between musical heads and linguistic heads, which are at prominent positions, listeners are able to reconstruct lyrics. The tone-tune correspondence is also observed in Shona, a Bantu language spoken in Zimbabwe. Schellenberg (2009) finds that sung melodies in Shona correspond to the spoken melodies.

Previous studies discuss the language-music mapping from the perspective of either perception or production grammar. The present study proposes that both perception grammar and production grammar are involved in the connection between language and music. The segmental change according to beat duration is examined in Chapter 3-4. The alignment between prosodic structures and the musical structures is examined through Mandarin children's songs in Chapter 5-6. The pitch-tone correspondence is discussed in Chapter 7-8.

DOI:10.6814/NCCU202001428

2.2 Perception Grammar

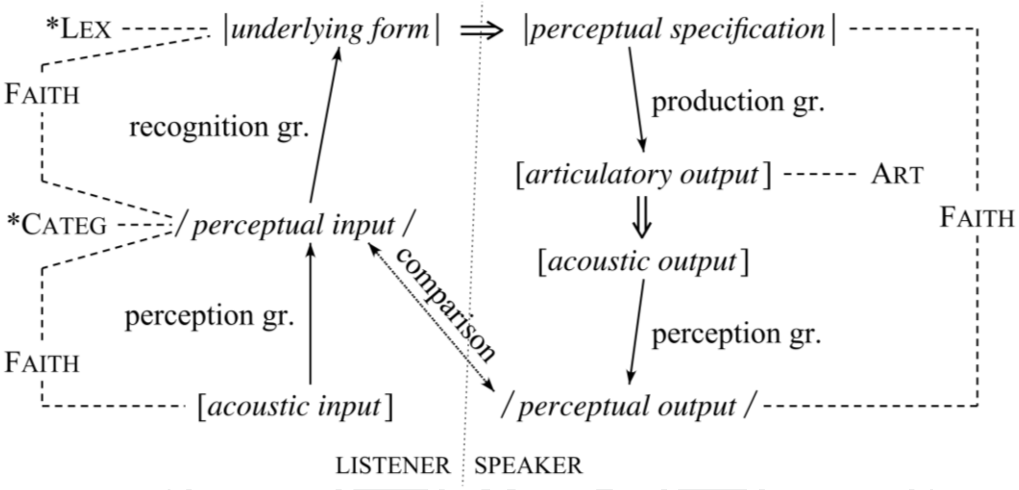
Previous researches have argued over the independence of perception grammar. Studies of loanword adaptation demonstrate that perception and production together contribute to the output form of loanwords. Silverman (1992) proposes that there are two stages in the adoption of loanwords. The first stage is the perceptual scan where some, but not all of the aspects are detected. For example, when Cantonese speakers perceive English words, they do not perceive English voicing contrast, which Cantonese lacks. The output of the perceptual scan becomes the input for Operative Level. Among the detected segments, more salient ones tend to be preserved in the output. For instance, in the English word, *place*, /s/ is more salient than /l/, so the Cantonese output is [p^heysi], which deletes /l/, whereas /s/ is preserved.

Yip (1993) follows the work of Silverman (1992) and proposes the constraint based-analysis of the Cantonese loanwords. Yip (1993) proposes that Cantonese loanwords are close to the perceived input. On the other hand, the output of the Cantonese loanword phonology must conform to surface well-formedness. The violation of faithfulness is for minimally bi-syllabic outputs and for preserving highly salient segments by vowel insertion.

Kenstowicz (2003) reviews Gbeto's (1999) cross-linguistic loanword in Fon. In the review, he proposes separate constraint rankings for perception and production grammar for French loanwords in Fon. In the perception process, word-final stops that are preceded by obstruents are diminished. Therefore, deletion or epenthesis would take place. This motivates a separate perception mapping in loanword adaptation. Take *post* for example. The constraint ranking for perception is Dep-V >> *stop/obstruent___# >> Max-C, which selects *pos* as the optimal output. On the other hand, the constraint ranking for production grammar is Max-C >> *stop/obstruent___# >> Dep-V, which selects *posu* as the output.

Boersma (2001) argues that the production of a word, involves not only perception and production grammar, but also recognition grammar. Boersma (2001) proposes a grammar model that illustrates the process of perception, production, and recognition grammar.

(5) The grammar model of functional phonology



(Boersma 2001:24)

The left side of the figure in (5) shows that the listeners perceive other speaker’s utterance and keep it as the underlying form after lexical recognition. The left side of the figure comprises the comprehension grammar which contains both perception and recognition grammar. The right-hand side of the figure shows how the listeners produce the sound they perceive.

On the contrary, Smolensky (1996) proposes a single grammar that works for production and comprehension. This is fought against by Boersma (2001), who indicates that only the maximally faithful candidate will win in comprehension grammar, which is not always true.

(6) *VoicedCoda >> MaxVoi

｜rɑd｜ 'wheel'	*VoicedCoda	MaxVoi
a. [rɑd]	*!	
☞b. [rɑt]		*

The coda in (6) is voiced so it is eliminated by *VoicedCoda. The production grammar chooses [rɑt] as the listener's optimal output in sacrifice of MaxVoi, which requires that underlying voicing feature have an output correspondent.

(7) *VoicedCoda >> MaxVoi

[rɑt]	*VoicedCoda	MaxVoi
*☞*a. ｜rɑt｜'rat'		
b. ｜rɑd｜'wheel'		*!

The constraints in (7) evaluate the top left cell, so each candidate does not violate *VoicedCoda. The constraint ranking in (7) which is identical to that in (6) will always choose ｜rɑt｜ as the underlying form of the listener even though the speaker may refer to 'wheel.' Boersma (2001) solves this problem by capturing the phonology-semantics interaction. In terms of the fact that words with lower frequency are less likely to be recognized, he proposes the constraint, *Lex, which evaluates the underlying form and can rule out words with lower frequency.

(8) * Lex (｜rɑd｜'wheel') >> *VoicedCoda >> MaxVoi >> * Lex >> (｜rɑt｜'rat')

[rɑt]	* Lex (｜rɑd｜'wheel')	*VoicedCoda	MaxVoi	* Lex (｜rɑt｜'rat')
☞a. ｜rɑt｜'rat'				*
b.｜ rɑd｜'wheel'	*!		*	

As in (8), * Lex (｜rɑd｜'wheel') can be lowered during acquisition.

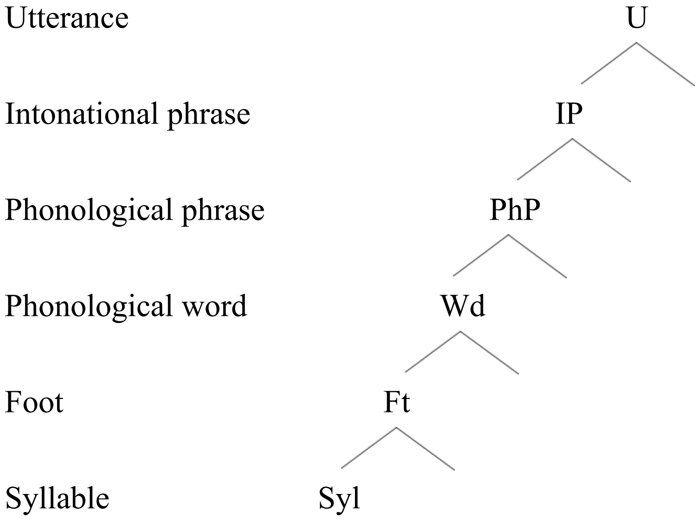
This study also proposes that perception and production are separate grammars. The evidence is shown in music-to-language mapping and language-to-music mapping. For example, when perceiving song lyrics, children perceive the musical pitch faithfully into the linguistic tone. However, when producing the lyrics, their production form will be influence by lexical association and surface well-formedness constraints. While lexicon recognition in Boersma (2001) takes place in recognition grammar, this study proposes that lexical association shows in the production grammar. More discussions will be shown in Chapter 7-8.

2.3 Prosodic Phonology

2.3.1 The Prosodic Hierarchy

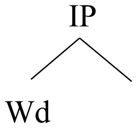
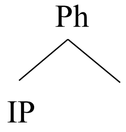
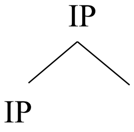
The prosodic hierarchy is proposed by Selkirk (1980), Nespor and Vogel (1986), and Inkelas (1989), among others. The prosodic hierarchy divides phonological structures into smaller constituents, as in (9).

(9) Prosodic Hierarchy



The prosodic hierarchy is subject to the strict layering hypothesis (Selkirk 1984; Nespor and Vogel 1986).

(10) Strict layering hypothesis

(a) *skipping	(b) *inverting	(c) *recursive
		

(10a) shows the violation of *skipping since IP directly dominates Wd, and skips the Ph level. The violation of *inverting is illustrated in (10b) where Ph is at a lower level than IP. (10c) violates *recursion, which prohibits prosodic structures from dominating themselves.

The present research discusses how the prosodic structures, which are IP, foot and prosodic word are aligned with musical beats and structures.

2.3.2 Intonational Phrase

Nespor and Vogel (1986) presume that the entire sentence is a single intonational phrase (IP), which can be restructured for physiological reasons or for ease of language processing. Nespor and Vogel (1986) also propose that a line that is too long is not preferred, so IP can be broken down into shorter ones, as in (11b-c).

(11)
(a) (My friend's baby hamster always looks for food in the corners of its cage)$_{\text{IP}}$
(b) (My friend's baby hamster)$_{\text{IP}}$ (always looks for food in the corners of its cage)$_{\text{IP}}$
(c) (My friend's baby hamster)$_{\text{IP}}$ (always looks for food)$_{\text{IP}}$ (in the corners of its cage)$_{\text{IP}}$

(Nespor and Vogel 1986:194)

DOI:10.6814/NCCU202001428

The domain of an intonational phrase (IP) can be syntactically or semantically defined. Some studies indicate that IPs are formed based on the syntactic configurations (Downing 1970, 1973; Bing 1979). Halliday (1967) and Selkirk (1984) recognize that IPs are semantically based. Consider the Sense Unit Condition proposed by Selkirk (1984) in (12).

(12) Sense Unit Condition

Two constituents C_i, C_j form a sense unit if (a) or (b) is true of the semantic interpretation of the sentence:

(a) C_i modifies C_j (a head),

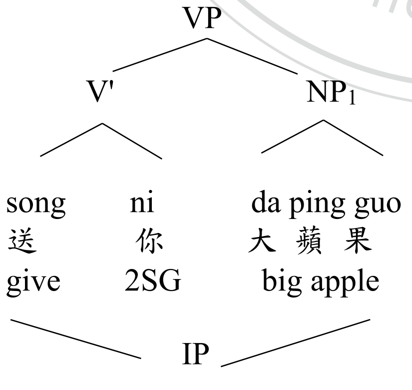
(b) C_i is an argument of C_j (a head).

(Selkirk 1984:291)

A sense unit is formed by modifier-modified relation and the head-argument relation between syntactic constituents. Consider the example in (13).

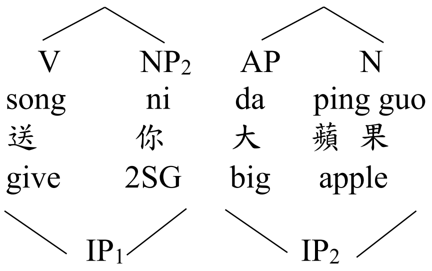
(13)

(a) ‘Give you a big apple.’



DOI:10.6814/NCCU202001428

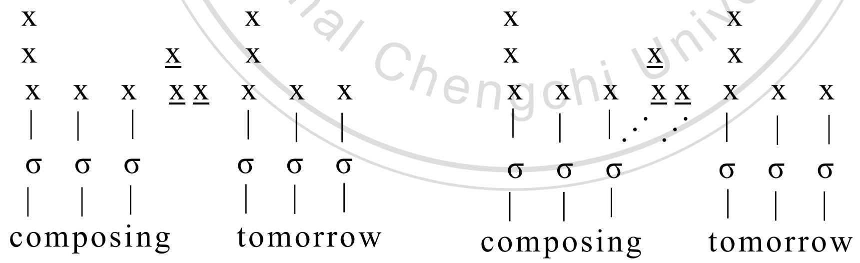
(b) ‘Give you a big apple.’



In (13a), NP$_1$ is taken as an internal argument by V' so V' and NP$_1$ together form a sense unit. At the lower level of the syntactic tree, *ni* ‘you’ is the indirect object of V *song* ‘give’, and the AP *da* ‘big’ is a modifier of the N *ping guo*, ‘apple’. Therefore, *song ni*, and *da ping guo* can be respectively parsed into intonational phrases, as in (13b).

The prosodic pause, which is among the factors that contribute to the parsing of intonational phrase, is illustrated in (14).

(14)


(a) pausing

(b) syllable-lengthening

The grid exhibits units of conceived time, which is termed demibeat by Selkirk (1984). As shown in (14), x-symbols stand for silent demibeat positions, which are regarded as pauses when they are unaligned. On the other hand, syllable-lengthening is represented by aligning x-symbols with syllables. Selkirk (1984) indicates that pausing and syllable-lengthening are actually the same phenomena.

The silent demibeat addition rule is proposed by Selkirk (1984), as in (15).

(15) Silent demibeat addition

Add a silent demibeat at the end (right extreme) of the metrical grid aligned with
(a) a word,
(b) a word that is the head of a nonadjunct constituent,
(c) a phrase,
(d) a daughter phrase of S.

(Selkirk 1984)

In them of the fact that neither of sense unit and prosodic pause could by itself define the intonational phrase, Hsiao (1995) provides the parameter of intonational phrase, as in (16).

(16) 𝓘-Parameter

𝓘=< γ] ◦ , SU > where γ = boundary tone, ◦ = pause

] = right edge, SU = sense unit

The parameter exhibits that an intonational phrase is a sense unit that ends in a boundary tone followed by a pause. Hsiao (1995) indicates that a silent beat is obligatorily added to the end of an IP in a normal tempo.

In Chapter 5-6 of the present study, the composer composes musical melody based on given lyrics. The intonational phrase boundaries are thus defined by the provided punctuations such as commas, periods, or exclamations.

DOI:10.6814/NCCU202001428

2.3.3 Foot Formation Rule

Chen (1984) proposes the foot formation rules to account for Mandarin verses. The rules take syntactic information into account and operate in the order in (17):

(17) Foot formation rule

(a) Immediate Constituency (IC): Link immediate constituents into disyllabic feet.

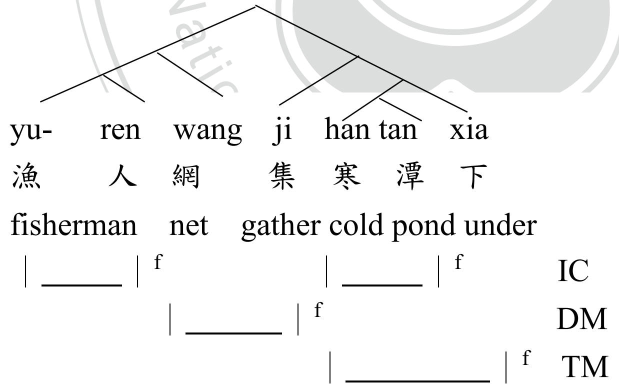
(b) Duple Meter (DM): Scanning from left to right, string together unpaired syllables into binary feet.

(c) Triple Meter (TM): Join any leftover monosyllable to a neighboring binary foot according to the direction of syntactic branching.

(Chen 1984:223)

The main focuses of the rules in (17) are ICs and the tree branching direction of the syntactic tree.

(18) 'Fishermen's nets gather under the cold pond.'



As exemplified in (18), ICs, *yu* and *ren*, *han* and *tan* have the priority to form into two feet. DM scans from left to right and strings *wang* and *ji* into a foot. As shown in (18), the branching of *wang* and *ji* is in the opposite direction. However, they can still be strung into one foot. Then TM parses *xia* to the neighboring foot, *han tan*.

Shih (1986) proposes the modified foot formation rule for Mandarin common speech, as in (19).

(19) Foot formation rule

(a) Immediate Constituency (IC):

Link immediate constituents into disyllabic feet.

(b) Duple Meter (DM):

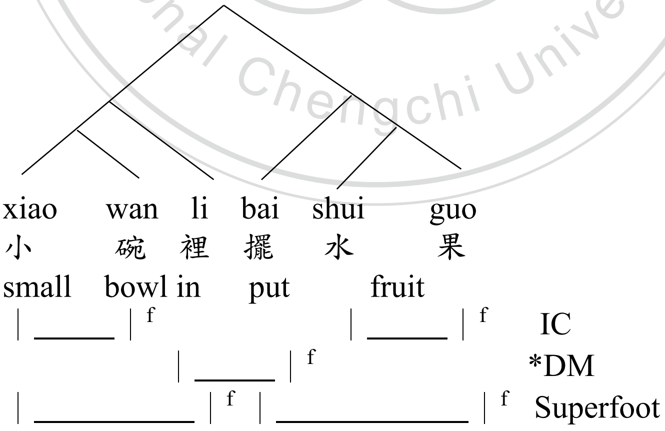
Scanning from left to right, string together unpaired syllables into binary feet, unless they branch in the opposite direction.

(c) Superfoot (f’): Join any leftover monosyllable to a neighboring binary foot according to the direction of syntactic branching.

(Shih, 1986: 110)

In Shih’s (1986) foot formation rule, DM cannot string syllables that belong to different branching direction.

(20) ‘In the small bowl is where the fruit is placed’



(20) shows that *li* and *bai* cannot form a DM since they have opposite branching direction. Therefore, *xiao wan li* and *bai shui guo* respectively forms two superfeet.

Based on Chen (1984) and Shih (1986), Hsiao (1991) proposes the beat counting device in terms of the discrepancy between lexical syllables and functor syllables. The metrical beat is assigned with a lexical syllable first and then the functor syllable is assigned with a beat in normal or slow speech, behaving like a lexical syllable and is left-adjoined to the nearest beat in fast speech. Hsiao (1991) proposes the following foot formation rule that is on the basis of beat counting device.

(21) Foot formation revisited

(a) Immediate Constituent Foot (ICF): Any adjacent beats which are assigned to ICs form an ICF.

(b) Adjacent Beat Foot (ABF): Any two adjacent beats which are not assigned to ICs are paired into an ABF.

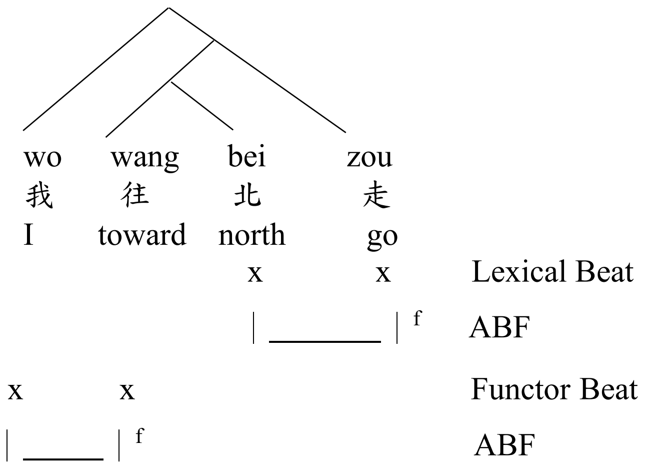
(c) Jumbo Foot (JF): Any unpaired single beat is recruited by a neighboring foot to form a Jumbo Foot if the beat c-commands the adjacent beat contained in the foot.

(d) Minifoot (MF): The leftmost single beat constitutes a Minifoot iff it is followed by an intonational phrase boundary %.

(Hsiao 1991:38)

Hsiao's (1991) foot formation rule is exemplified in (22).

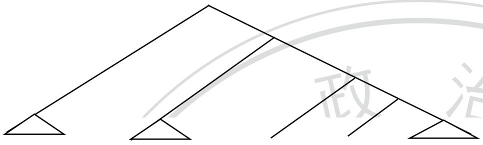
(22) 'I went toward the north.'



As shown in (22), *bei* and *zou* are assigned lexical beats, whereas *wo* and *wang* are assigned functor beats. The lexical beats, *bei* and *zou* are paired into an ABF first. Then the functor beats, *wo* and *wang* are parsed into another ABF.

The present study applies Shih's (1986) and Hsiao's (1991) foot formation rules for constructing foot in Mandarin children's songs. The lyrics of the children's songs are similar to common speech. Take (23) for example.

(23) 'Love to somersault when nothing to do whole day long.'



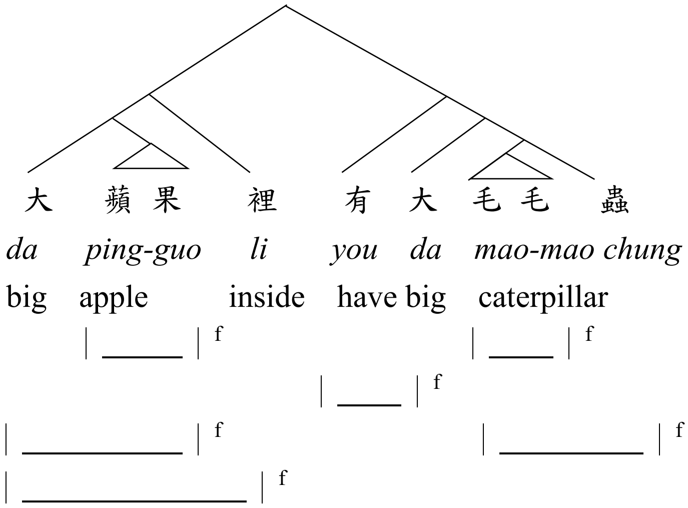
zheng tian *mei shi* *ai* *fan* *gen tou*
整 天 沒 事 愛 翻 跟 頭
whole-day nothing-to-do love turn somersault
| _____ | f | ______ | f | ______ | f
| ______ | f

ICs, *zheng-tian*, *mei-shi*, and *gen-tou* are first parsed into feet. Then *ai* and *fan* are strung into a foot since their branching directions are the same.

Example (24) is another example taken from the present study.

DOI:10.6814/NCCU202001428

(24) ‘There is a big caterpillar in a big apple.’

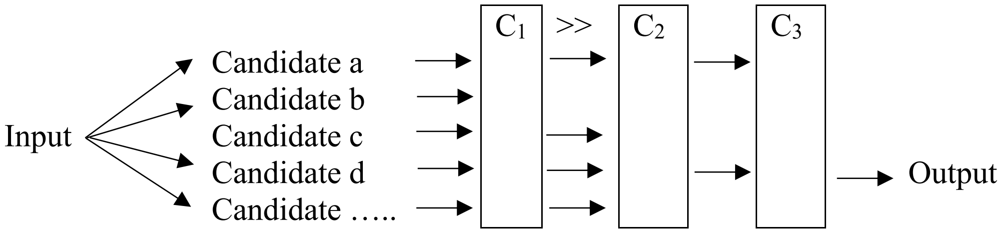


First of all, ICs, *ping-guo* and *mao-mao* are strung into feet. Then *you* and *da* are strung together since their branching direction is the same. Finally, *da* and *li* are both adjoined to the foot *ping-guo* while *chung* is adjoined to the foot *mao-mao*.

2.4 Optimality Theory

The Optimality Theory (OT) is proposed by Prince and Smolensky (1993/2004). OT regards grammars as a set of ranked constraints which are violable. The operation of OT consists mainly of Generator (GEN) and Evaluator (EVAL).

(25) Mapping of input to output in OT grammar



(Kager 1999:22)

DOI:10.6814/NCCU202001428

The figure in (25) is illustrated by Kager (1999). In (25), GEN generates infinite sets of output candidates, which are evaluated by a set of hierarchically ranked constraints ($C_1 >> C_2 >> C_3 \ldots$). Each constraint may eliminate some output candidates until only one output candidate survives.

2.4.1 Faithfulness Constraints and Markedness Constraints

The constraints in Optimality Theory (Prince and Smolensky 1993/2004) are universal. What makes language different from each other is the different rankings of the constraints. OT mainly contains two kinds of constraints, which are faithfulness constraints and markedness constraints.

The correspondence theory provides the framework for defining faithfulness constraints (McCarthy and Prince 1995, 1999). The concept is that each candidate generated by GEN includes an output representation and a relation between the input and the output. The faithfulness constraints of correspondence relation are shown in (26-28).

(26) MAX (No deletion):

Let $input = i_1 i_2 i_3 \ldots i_n$ and $output = o_1 o_2 o_3 \ldots o_m$
Assign one violation mark for every i_x
if there is no o_y where $i_x \mathfrak{R} o_y$

(27) DEP (No epenthesis):

Let $input = i_1 i_2 i_3 \ldots i_n$ and $output = o_1 o_2 o_3 \ldots o_m$
Assign one violation mark for every o_y
if there is no i_x where $i_x \mathfrak{R} o_y$

(28) IDENT (No change):

Let $input = i_1 i_2 i_3 \ldots i_n$ and $output = o_1 o_2 o_3 \ldots o_m$
Assign one violation mark for every pair (i_x, o_y), where $i_x \mathfrak{R} o_y$
and i_x and o_y have different values of feature.

DOI:10.6814/NCCU202001428

As shown in (26), MAX requires no deletion in the output. DEP in (27) prohibits epenthesis in the output while IDENT in (28) demands that output and input have the same values of feature.

Markedness constraints require well-formedness on the output. For example, this research proposes the constraint, *ProsSt, which resquires no prosodic structure in the output. When *ProsSt ranks higher than the faithfulness constraint, MAXProsSt, which requires input prosodic structure preserved in the output, the prosodic structure will not remain in the output.

2.4.2 Generalized Alignment

McCarthy and Prince (1993) propose a schema for defining alignment constraints, as in (29). The alignment schema matches edges of prosodic and/or grammatical constitutes.

(29) Alignment constraint schema (McCarthy and Prince 1993: 80)
ALIGN (Cat1, Edge1, Cat2, Edge2)$=_{def}$
∀ Cat1 ∃ Cat2 such that Edge1 of Cat1 and Edge2 of Cat2 coincide,
where Cat1, Cat2 ∈ PCat∪GCat and Edge1, Edge2 ∈ {Right, Left}.

As shown in (29), the grammatical category, GCat, refers to root, stem, syntactic word, phrase, etc. The prosodic category, PCat, refers to syllable, foot, phonological word, phonological phrase, intonational phrase, etc.

For example, constraint (30a) demands that the left edge of a foot coincide with the left edge of a word. (30b) requires the alignment between the right edge of a word and the right edge of a foot.

(30)

a. Align-Left (foot, word)

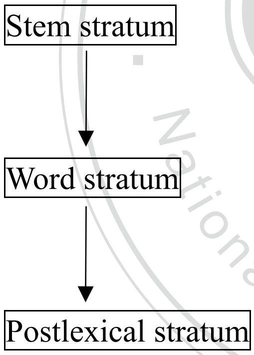
b. Align-Right (word, foot)

The present study examines the alignment between prosodic words and musical beats as well as the alignment between intonational phrases and musical structures.

2.4.3 Stratal OT

Stratal OT has been used to account for opacity and paradigmatic transfer phenomena (Kiparsky 2000, Bermúdez-Otero 1999, 2007). It has its original in derivation lexical phonology (Kiparsky 1982, Mohanan 1982, Booij and Rubach 1987), which examines the intermediate levels between the input and the output.

(31)



As shown in (31), except for the first level, the input of each stratum is defined by the previous one. Each stratum has its own constraint ranking. In other words, stem must satisfy the stem phonology, word must satisfy the word phonology, and phrase must satisfy the phrasal phonology.

In the present study, I observe that children's perception of Mandarin songs exhibits a two-step operation of lexical association. When children perceive the musical pitch into their linguistic input, they associate the linguistic input to their mental lexicon at the lexical level. However, if they fail to associate with an actual term, they then search at the postlexical level where they may or may not successfully associate with an actual term. The separate operations at lexical and postlexical levels show the spirit of Stratal OT.

2.4.4 Local Conjunction

Local conjunction, which is proposed by Green (1994) and Smolensky (1995), combines two constraints into one. Smolensky (1995) proposes the formalized description, as in (32).

(32) Local conjunction

The local conjunction of C_1 and C_2 in domain, C_1 and C_2 is violated when there is some domain of type D in which both C_1 and C_2 are violated.

(Smolensky 1995:10)

Local conjunction rules out the worst of the worst output, which is called the WOW effect. The constraint conjunction is violated only when both of its members are violated. For example, the conjoined markedness constraints prohibit the worst of the worst marked output. Morris (2002) and Łubowicz (2002, 2005) indicate the need to conjoin markedness and faithfulness constraints. The concept is that the conjoined markedness member is activated only when the faithfulness member is violated. Wee (2002) proposes the faithfulness-faithfulness conjunction for tone. The constraint ID-R and ID-C is violated when both the tonal register and the tonal contour change. Hsiao (2015) proposes that both faithfulness-faithfulness and markedness-faithfulness conjuncts are required for the complex tonal chain shifts in Taiwanese.

DOI:10.6814/NCCU202001428

This study proposes a conjoined constraint, which is comprised of two markedness constraints, namely NoShare-σ and $*CC_{CODA}$. NoShare-σ forbids syllables from being shared by two beats, whereas $*CC_{CODA}$ eliminates coda clusters. The conjoined constraint, NoShare-σ & $*CC_{CODA}$ is violated only when a syllable with a coda cluster is shared by two beats.

DOI:10.6814/NCCU202001428

Chapter 3

Language-to-Music Mapping: Onset Cluster

3.1 Introduction

This chapter compares the linguistic mapping and the language-to-music mapping in onset clusters. Segmental changes are observed from Mandarin-accented English in reading and singing. I posit the mapping schema in (33).

(33) Language-to-music mapping: segment

Linguistic input
Production grammar
Linguistic output
Perception grammar
Musical input
Production grammar
Musical output

As proposed in (33), the linguistic output is the English word pronounced by Mandarin speakers. This linguistic output is then perceived as the musical segmental inputs. The perceived musical input is then produced as the musical output, which is sung by the speakers. Segmental changes are observed from the linguistic mapping and language-to-music mapping.

There are two linguistic output variants of monosyllabic syllables with onset clusters. Linguistic output$_1$ preserves the onset cluster and the syllable number

DOI:10.6814/NCCU202001428

remains one. Linguistic output$_2$ involves vowel insertion so that the syllable number becomes two. These linguistic outputs are respectively assigned to one and two musical beats. This chapter investigates how musical beat assignment affects segmental changes and how the prosodic word conditions musical beat assignment in the musical input.

3.2 Data Design

The database is designed to compare segmental changes of onset clusters in the linguistic mapping and the language-to-music mapping. The Mandarin informants include two males and three females, aged between 59 and 72, and are of senior secondary to higher education in Taiwan. All of them have learnt English for at least six years. In order to examine segmental changes between the linguistic output and the musical output, the informants are asked to read and sing the assigned target words. The procedures of reading in step1 and singing in step 2 are introduced in 3.2.1 and 3.2.2 respectively.

3.2.1 Step 1: Reading

The informants read the target words on a piece of paper in step 1. They can see and read the words to ensure that their linguistic output comes from their own input instead of from other peoples' linguistic output.

The target words include onset clusters with different combination of consonants on the sonority scale.[1] The sonority sequencing principle (SSP) is introduced by Sievers (1881) and Jespersen (1904). They propose that in the onset cluster a more

---

[1] The collected onset clusters do not reveal clear correlation to the SSP, which, however, effective on coda clusters, as will be discussed in Chapter 4.

DOI:10.6814/NCCU202001428

sonorous consonant stands closer to the syllable peak than one that is less sonorous. Carr (1993), and Broselow and Finer (1991) list stop and fricative as separate classes. As in (34), stop is the least sonorous, while glide is the most sonorous.

(34) Sonority scale

Class	**Scale**
Glide	5
Liquid	4
Nasal	3
Fricative	2
Stop	1

The data in (35) are examples of target words with onset clusters. The onsets in (35) conform to SSP. In other words, the sonority of the first consonant in the onset cluster should be lower than that of the second consonant, which is closer to the syllable peak. For instance, the onset cluster [bl] is the combination of a less sonorous stop followed by a more sonorous liquid.

(35) Target words with onset clusters

		SSP	Example target words	
a.	stop+liquid	1 4	/**bl**u/ ‘blue’	/**kr**aɪ/ ‘cry’
b.	stop+glide	1 5	/**bj**utɪ/ ‘beauty’	/**kw**in/ ‘queen’
c.	fricative+nasal	2 3	/**sn**ou/ ‘snow’	/**sm**ɔl/ ‘small’
d.	fricative+glide	2 5	/**fl**aɪ/ ‘fly’	/**fr**i/ ‘free’

In step 1, the informants read the target words, such as in (36-39).

(36) Target word: **blue**
Output: [blu], [bulu]

(37) Target word: **play**
Output: [pʰleɪ], [pʰuleɪ]

(38) Target word: **green**
Output: [grin], [gurin]

(39) Target word: **class**
Output: [klæs], [kəlæs]

There are two kinds of linguistic outputs. One is without vowel insertion and the other is with vowel insertion. The variation in pronouncing the onset clusters is schematized as (40).

(40) Linguistic input Linguistic output
/blu/ ‘blue’ → (blu)$_\omega$ Linguistic output$_1$ (σ)
/blu/ ‘blue’ → (bulu)$_\omega$ Linguistic output$_2$ (σσ)

As in (40), the prosodic words are formed in the output. For example, the linguistic input /blu/ yields either (blu)$_\omega$ or (bulu)$_\omega$ in the output. Linguistic output$_1$ preserves the onset cluster and remains monosyllabic, whereas linguistic output$_2$ inserts a vowel into the cluster and becomes disyllabic.

(41) Statistics of the linguistic outputs

Linguistic input	Linguistic output$_1$	Linguistic output$_2$	Total
CCV[2]	(CCV)$_\omega$ (51/86%)	(C**V**CV)$_\omega$ (8/14%)	59 (100%)

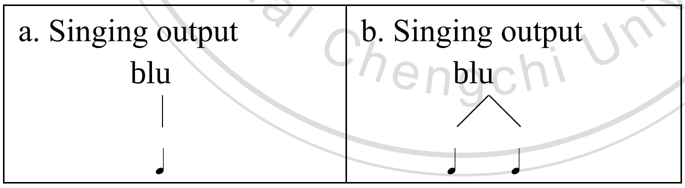
Table (41) shows the linguistic input to linguistic output mapping. There are totally 59 syllables with onset clusters. Fifty-one, or 86%, of them surface as one syllable; eight, or 14%, of them insert a vowel, changing the syllable number to two. The percentage shows that the informants tend to correctly pronounce syllables with onset clusters.

3.2.2 Step 2: Singing

In step 2, the informants sing the linguistic outputs they produce in step 1. As mentioned in 3.2.1, there are two kinds of linguistic outputs, namely, linguistic output$_1$ and linguistic output$_2$. Each of the linguistic outputs is respectively assigned with one and two beats in the language-to-music mapping. Take (blu)$_\omega$ for example.

(42) Linguistic output$_1$ (σ) to singing mapping

Linguistic output$_1$:(blu)$_\omega$



As shown in (42), whether linguistic output$_1$, (blu)$_\omega$, is assigned with one musical beat, as in (42a), or with two musical beats, as in (42b), the singing outputs surface as [blu], where no vowel insertion occurs. The prosodic word structure is removed in the singing output.

---

[2] Some of the target words are with coda, for example, *green*. However, since coda is not the focus in this chapter, only onsets and nuclei are listed.

(43) Statistics of the linguistic output$_1$ (σ) to singing mapping

Linguistic output$_1$	Singing output (♩)	Singing output (♩ ♩)
(CCV)$_\omega$	CCV (50/98%) | ♩	CCV (51/100%) ∧ ♩ ♩
	CVCV (1/2%) ∨ ♩	CVCV (0/0%) | | ♩ ♩
Total	51 (100%)	51 (100%)

The table in (43) shows that there are totally 51 onset clusters that are produced as linguistic output$_1$. These onset clusters are respectively assigned with one and two musical beats. When they are associated with one beat, 50, or 98%, of them are sung as monosyllabic CCV. When they are associated with two beats, still 100% of them are sung as monosyllabic syllables.

Consider the output$_2$ mapping, as in (44).

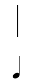
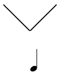
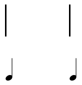
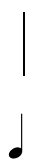
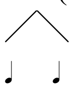
(44) Linguistic output$_2$ (σσ) to singing mapping

Linguistic output$_2$: (bulu)$_\omega$

a. Singing output blu | ♩	b. Singing output blu ∧ ♩ ♩

When (bulu)$_\omega$ is assigned with one musical beat, as in (44a), the singing output is truncated as [blu]. When the same linguistic output is assigned with two musical beats, as in (44b), the singing output is still (blu)$_\omega$. The structure of the prosodic word is removed in the singing output.

(45) Statistics of the linguistic output₂ (σσ) to singing mapping

Linguistic output₂	Singing output (♩)	Singing output (♩ ♩)
(CVCV)ω	CCV (7/87.5%) 	CCV (7/87.5%) 
	CVCV (0/0%) 	CVCV (1/12.5%) 
	CV (1/12.5%) 	CV (0/0%) 
Total	8 (100%)	8 (100%)

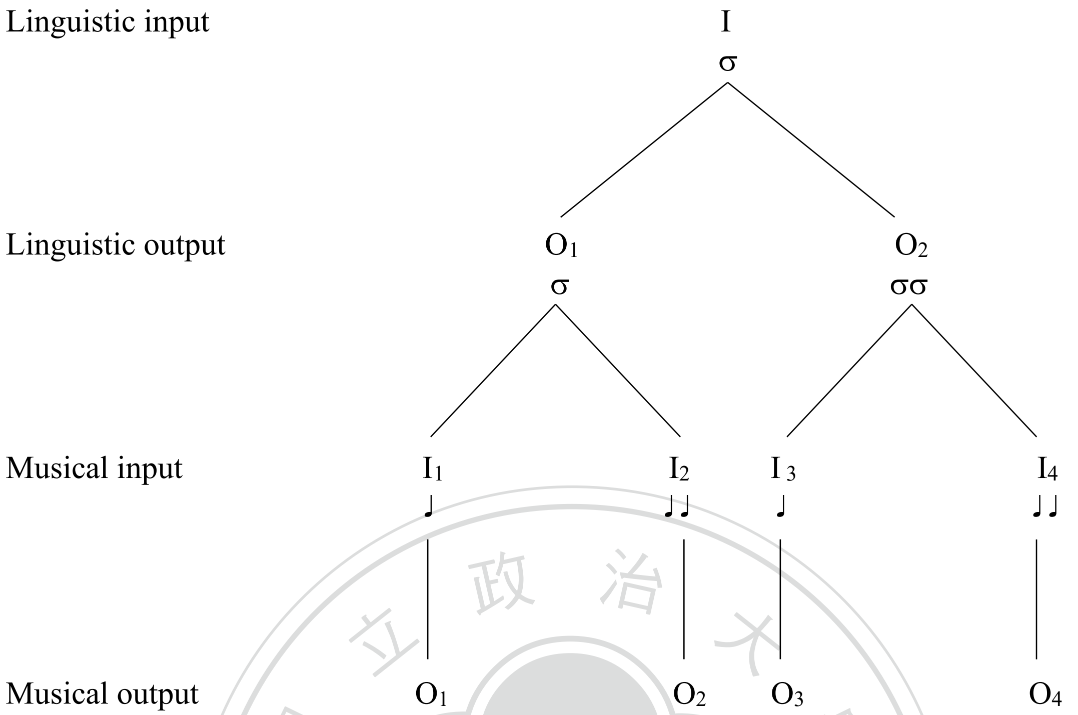
There are totally eight onset clusters that are produced as linguistic output₂ in the database. Each of them is assigned with one and two musical beats respectively. As in table (45), when they are associated with one beat, seven, or 87.5%, of them are sung as CCV. Only one, or 12.5%, of them is sung as CV. None of them are sung as disyllabic syllables. When they are associated with two beats, still seven, or 87.5% of them are sung are sung as monosyllabic syllables.

In brief, two patterns are in order. First, whether the monosyllabic output₁ is associated with one or two musical beats, there is no vowel inserted to resolve the onset cluster. Second, whether the disyllabic linguistic output₂ is assigned with one or two musical beats, the inserted vowel in reading is deleted in singing.

3.3 Language-to-Music Mapping

I propose a model for the language-to-music mapping, as illustrated in (46). Production and perception grammars are shown in the linguistic mapping and language-to-music mapping.

(46) Segmental Model for the Language-to-Music Mapping:



The input on the top is a monosyllable with an onset cluster. The production grammar is shown in the linguistic input to output mapping. There are two linguistics outputs, namely, linguistic output$_1$ and linguistic output$_2$. Linguistic output$_1$ is with monosyllable where no vowel is inserted. Linguistic output$_2$ is pronounced with two syllables where a vowel is inserted to prevent complex consonants. Both of the linguistic outputs are parsed into prosodic words.

The mapping from the linguistic output to the musical input demonstrates a need for the perception grammar. The linguistic outputs are perceived as the musical inputs, and each of them are assigned with one and two musical beats respectively. The prosodic word structure, which may affect beat assignment, is formed in the musical input. Whether linguistic output$_1$ is assigned with one or two beats, the musical segmental input remains monosyllabic. As for linguistic output$_2$, its inserted vowels are deleted whether they are assigned with one two beats.

Finally, the musical inputs are mapped to the musical outputs, where the prosodic word structure is removed.

Take the linguistic input, /blu/, for example. /blu/ is produced as $(\text{blu})_\omega$ or $(\text{bulu})_\omega$ in the linguistic outputs. The monosyllabic $(\text{blu})_\omega$ is referred to as linguistic output$_1$, whereas $(\text{bulu})_\omega$ is referred to as linguistic output$_2$. The linguistic outputs are perceived as the musical inputs, where both the monosyllabic $(\text{blu})_\omega$ and the disyllabic $(\text{bulu})_\omega$ are assigned with one and two beats respectively. Finally, the musical beat association yields the musical output, where the prosodic word structure is removed.

3.4 L_I-to-L_O Mapping: Production Grammar

This section analyzes the linguistic input (L_I) to linguistic output (L_O) mapping under the framework of Optimality Theory (Prince and Smolensky 1993/2004). There are two linguistic outputs. One is with vowel insertion and the other is without vowel insertion. Take /blu/ 'blue' for example. The output of /blu/ can be either $(\text{blu})_\omega$ or $(\text{bulu})_\omega$. The relevant constraints are given in (47-50).

(47) MAX-C

Assign one violation mark for every input consonant that does not show in the output.

(48) DEP-V

Assign one violation mark for every output vowel that does not show in the input.

(49) *CC

Assign one violation mark for every syllable that has a consonant cluster.

(50) ALIGN-E (LEX, ω)

Assign one violation mark for every lexical word whose edges are not aligned with the edges of a prosodic word.

The database shows that 86% of the onset clusters are pronounced faithfully to their input. Therefore, we need faithfulness constraints, MAX-C and DEP-V that demand faithful relation between the input and the output. These constraints forbid consonant deletion and vowel insertion so that the output segments remain the same as the input. *CC is a markedness constraint that forbids complex consonants in a syllable. *CC competes with the faithfulness constraints and select the outputs that do not have consonant clusters. Therefore, in order to surface [blu], MAX-C and DEP-V should be ranked higher than *CC so that the consonant cluster can be preserved. Syntactic lexical words are aligned with the prosodic words in the output, which is governed by ALIGN-E(word, ω).

Tableau (51) shows the competition of these constraints.

(51) L_{I}-to-L_{O} mapping: output_{1}, $(\text{blu})_{\omega}$

/blu/	ALIGN-E (word, ω)	MAX-C	DEP-V	*CC
☞ a. $(\text{blu})_{\omega}$				*
b. $(\text{bulu})_{\omega}$			*!	
c. $(\text{bu})_{\omega}$		*!		
d. blu	*!			*

In tableau (51), MAX-C and DEP-V are ranked higher than *CC. (51b) inserts a vowel and incurs a fatal violation of DEP-V. (51c) deletes a consonant, and thus is ruled out by MAX-C. The output in (51d) is not a prosodic word, so it is ruled out by ALIGN-E(word, ω). (51a) that is faithful to the input is selected, in spite of a violation of *CC.

The database shows that 14% of the coda cluster are produced with an inserted vowel like (bulu)$_\omega$. To obtain (bulu)$_\omega$, *CC should be ranked higher than DEP-V. In this case, consonant clusters are avoided by inserting a vowel instead of deleting any consonant. Therefore, both *CC and MAX-C ranks higher than DEP-V. Tableau (52) examines the selection of the optimal output.

(52) L$_I$-to-L$_O$ mapping: output$_2$, (bulu)$_\omega$

/blu/	ALIGN-E (word, ω)	*CC	MAX-C	DEP-V
☞ a. (bulu)$_\omega$				*
b. (blu)$_\omega$		*!		
c. (bu)$_\omega$			*!	
d. blu	*!	*		

Consider the output$_2$ in (52). (52b) preserves the onset cluster [bl], so it is ruled out by *CC. (52c) conforms to *CC by deleting [l], and thus is ruled out by MAX-C. (52d) is not a prosodic word, so it is ruled out by Align-E(word, ω).

The constraint rankings for the linguistic input-to-output production grammars are summarized in (53).

(53) Linguistic input to linguistic output production grammars

a. Output$_1$: [blu]→(blu)$_\omega$ ALIGN-E(LEX, ω), MAX-C, DEP-V >> *CC

b. Output$_2$: [blu]→(bulu)$_\omega$ ALIGN-E(LEX, ω), *CC, MAX-C >> DEP-V

The target word surfaces with two syllables when *CC is ranked higher than MAX-C and DEP-V. When *CC is ranked at the bottom, the target word is faithful to the input and remains monosyllabic. ALIGN-E(LEX, ω) is the dominant constraint governing both outputs.

3.5 L_O-to-M_I Mapping: Perception Grammar

3.5.1 L_{O1}-to-M_I Mapping

As mentioned in section 3.4, there are two kinds of linguistic outputs. Output$_1$ (L_{O1}) is monosyllabic, whereas output$_2$ (L_{O1}) is disyllabic. This section examines the linguistic output$_1$ (L_{O1}) to the musical input (M_I) mapping.

When L_{O1} is mapped to M_I, it is respectively assigned with one and two musical beats. I first discuss the case where L_{O1} is assigned with one musical beat. Three relevant constraints are proposed in (54-56).

(54) NoStray

Assign one violation mark for every stray element, $\boxed{e}$, in the output.

(55) NoShare-B

Assign one violation mark for every musical beat that is shared by two or more syllables.

(56) NoShare-σ

Assign one violation mark for every syllable that is shared by two or more musical beats.

The constraint NoStray is an undominated constraint, which ensures that every syllable or musical beat is associated. Max-C is ranked above Dep-V, such that consonant deletion is avoided while vowel insertion is possible. NoShare-σ and NoShare-B are ranked at the bottom, as a syllable is often linked to multiple musical beats and vice versa. A partial constraint ranking is proposed in (57).

(57) L_{O1}-to-M_I partial constraint ranking:

NoStray, Max-C >> Dep-V >> NoShare-B, NoShare-σ

DOI:10.6814/NCCU202001428

The tableau in (58) shows how the constraint ranking works.

(58) L_{O1}-to-M_I mapping: onset cluster (♩)

L_{O1}: (blu)$_\omega$ M_I: (blu)$_\omega$
⋮
♩

(blu)$_\omega$ ♩	NoStray	Max-C	Dep-V	NoShare-B	NoShare-σ
☞a.(blu)$_\omega$ ⋮ ♩					
b. (bl$\boxed{\text{u}}$)$_\omega$ $\boxed{♩}$	*!*				
c. (bu)$_\omega$ ⋮ ♩		*!			
d. (bulu)$_\omega$ ⋱⋮ ♩			*!	*	

Tableau (58) examines the L_{O1}-to-M_I mapping where one musical beat is assigned. Max-C rules out (58c), where [l] is deleted, while Dep-V rules out (58d), where [u] is inserted. NoStray is fatally violated by (58b), in which there is a stray syllable and a stray musical beat. (58a) incurs no constraint violation and is selected as the optimal output.

When L_{O1} is assigned with two beats, the mapped M_I is still (blu)$_\omega$, as shown in tableau (59).

(59) L_{O1}-to-M_I mapping: onset cluster (♩♩)

L_{O1}: $(blu)_\omega$ M_I: $(blu)_\omega$

⋰⋱

♩ ♩

$(blu)_\omega$ ♩ ♩	NoStray	Max-C	Dep-V	NoShare-B	NoShare-σ
☞a.$(blu)_\omega$ ⁝⋱ ♩ ♩					*
b. $(blu)_\omega$ ⋮ ♩$\boxed{\text{♩}}$	*!				
c. $(bulu)_\omega$ ⋮ ⋮ ♩ ♩			*!		
d. $(bu)_\omega$ ⁝⋱ ♩ ♩		*!			*

The syllable in (59a) is shared by two musical beats, violating the bottom-ranked NoShare-σ, but it is still selected as the optimal output. In (59b), there is one unlinked musical beat, and thus is eliminated by NoStray. Dep-V and Max-C then rule out (59c) and (59d) respectively.

3.5.2 L_{O2}-to-M_I Mapping

This section discusses the mapping from linguistic output₂ (L_{O2}) to the musical input (M_I). There are two syllables in L_{O2}, which are linked to either one or two musical beats in M_I.

When the L_{O2}, $(bulu)_\omega$, is assigned with one musical beat, it is mapped to monosyllabic $(blu)_\omega$, as shown in tableau (60).

DOI:10.6814/NCCU202001428

(60) L_{O2}-to-M_I mapping: onset cluster (♩)

L_{O2}: (bulu)$_\omega$ M_I: (blu)$_\omega$
⋮
♩

(bulu)$_\omega$ ♩	NoStray	Max-C	Dep-V	NoShare-B	NoShare-σ
☞a. (blu)$_\omega$ ⋮ ♩					
b. (bl$\boxed{\text{u}}$)$_\omega$ $\boxed{\text{♩}}$	*!*				
c. (bu)$_\omega$ ⋮ ♩		*!			
d. (bulu)$_\omega$ ⋱⋮ ♩				*	

In the previous tableau in (58), the output in (58d), (bulu)$_\omega$, is ruled out by Dep-V due to the insertion of [u]. However, the same linguistic output in (60d) involves no vowel insertion, and thus Dep-V is inactive. NoStray and Max-C rule out (60b) and (60c) respectively. NoShare-B favors (60a) over (60d), where the musical beat is shared by two syllables. As a result, (60a) emerges.

When the L_{O2}, (bulu)$_\omega$, is assigned with two musical beats, it is still mapped to monosyllabic (blu)$_\omega$, which is linked to two musical beats. However, the tableau evaluation in (61) makes an incorrect prediction.

DOI:10.6814/NCCU202001428

(61) L_{O2}-to-M_{I} mapping: onset cluster (♩♩)

L_{O2}: (bulu)$_\omega$ M_{I}: (blu)$_\omega$
⁚˙.
♩ ♩

(bulu)$_\omega$ ♩ ♩	NoStray	Max-C	Dep-V	NoShare-B	NoShare-σ
(☞)a. (blu)$_\omega$ ⁚˙. ♩ ♩					*!
b. (blu)$_\omega$ ⁝ ♩$\boxed{♩}$	*!				
c. (bu)$_\omega$ ⁚˙. ♩ ♩		*!			*
☛d. (bulu)$_\omega$ ⁝ ⁝ ♩ ♩					

NoShare-σ undesirably rules out (61a), the real optimal output, as indicated by the parenthesized white right-headed hand symbol, (☞), and (61d) is wrongly selected, as indicated by the parenthesized black right-headed hand symbol, ☛. To exclude the unwanted candidate in (61d), I propose the constraint in (62).

(62) Align-R(♩, ω):

Assign one violation mark for every musical beat, ♩, that is not linked to the rightmost syllable in a prosodic word, ω.

This constraint must be ranked higher than NoShare-σ so that all the musical beats are associated with the final syllable. The enriched constraint ranking is provided in (63).

(63) L_{O2}-to-M_I constraint ranking (enriched)

NOSTRAY, MAX-C >> DEP-V >> ALIGN-R(♩, ω) >> NOSHARE-B, NOSHARE-σ

Consider now the tableau in (64).

(64) L_{O2}-to-M_I mapping: onset cluster (♩♩)

L_{O2}: $(bulu)_ω$ M_I: $(blu)_ω$
⁚˙.
♩ ♩

$(bulu)_ω$ ♩ ♩	NOSTRAY	MAX-C	DEP-V	ALIGN-R(♩, ω)	NOSHARE-B	NOSHARE-σ
☞a. $(blu)_ω$ ⁚˙. ♩ ♩						*
b. $(blu)_ω$ ⁚ ♩ $\boxed{♩}$	*!					
c. $(bu)_ω$ ⁚˙. ♩ ♩		*!				*
d. $(bulu)_ω$ ⁚ ⁚ ♩ ♩				*!		

In (64d), each syllable is associated with a musical beat, and thus incurs a fatal violation of ALIGN-R(♩, ω). Eventually, (64a) is selected as the optimal output, where the leftmost [u] is deleted while the rightmost [u] is shared by two musical beats.

3.6 M_I-to-M_O Mapping: Production Grammar

The mapping between the musical input and the musical output is subject to the production grammar that preserves syllable-beat association and removes prosodic structure in the musical output. Three constraints are posited in (65-67), and the constraint ranking is summarized in (68).

(65) ID-ASSOC:

Assign one violation mark for every output association line that is not identical to that in the musical input.

(66) *PROSST:

Assign one violation mark for every prosodic structure in the output.

(67) MAX-PROSST:

Assign one violation mark for every prosodic structure in the musical input that does not have a correspondent in the musical output.

(68) M_I-to-M_O constraint ranking:

ID-ASSOC, *PROSST >> MAX-PROSST

ID-ASSOC must be undominated so that the syllable-beat association in the musical input is retained in the musical output. *PROSST must outrank MAX-PROSST to ensure that there is no prosodic structure in the musical output. The tableaux in (69-70) show how the constraint ranking works.

(69) M_I-to-M_O mapping: onset cluster (σ/♩)

(blu)$_\omega$ ⋮ ♩	ID-ASSOC	*PROSST	MAX-PROSST
☞a. blu ⋮ ♩			*
b. (blu)$_\omega$ ⋮ ♩		*!	
c. bl$\boxed{u}$ $\boxed{♩}$	*!		

(70) M_I-to-M_O mapping: onset cluster (σ/♩♩)

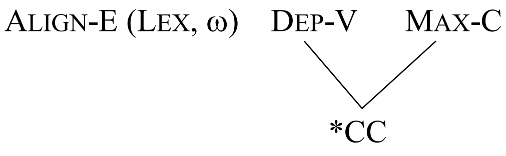
(blu)$_\omega$ ⁚⋱ ♩ ♩	ID-ASSOC	*PROSST	MAX-PROSST
☞a. blu ⁚⋱ ♩ ♩			*
b. (blu)$_\omega$ ⁚⋱ ♩ ♩		*!	
c. blu ⁚ ♩ $\boxed{♩}$	*!		

Both (69b) and (70b) are ruled out by *PROSST, as they preserve the prosodic words in the output. On the other hand, (69c) and (70c) are eliminated by ID-ASSOC, since the association lines are changed.

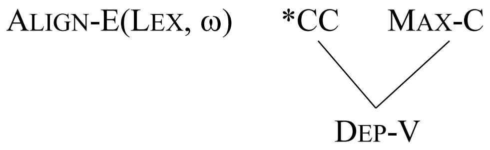
3.7 Summary

The mapping of an onset cluster from the linguistic input to the linguistic output involves interactions between segmental faithfulness and markedness, as well as prosodic alignment. Two output variants are observed: L_{O1} is monosyllabic and faithful, while L_{O2} is disyllabic, with a vowel inserted to resolve the onset cluster. I have proposed a set of relevant constraints, which are subject to flexible rankings, as illustrated by the Hasse diagrams in (71-72).

(71) L_I-to-L_{O1} mapping: (σ)



(72) L_I-to-L_{O2} mapping: (σσ)



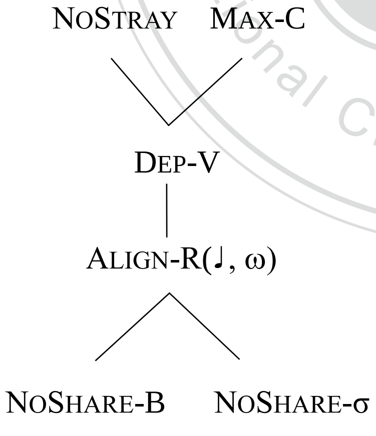
In the mapping to the musical input, the two linguistic outputs (L_{O1} and L_{O2}) are assigned with one or two musical beats, as in (73).

(73) L_{O1}-to-M_I Mapping: σ (♩, ♩♩)

L_{O2}-to-M_I Mapping: σσ (♩, ♩♩)

In spite of the fact that there are two constraint rankings in the L_I-to-L_O mapping, the perception grammar in the L_O-to-M_I mapping lies in a single constraint ranking, as in (74).

(74) L_O-to-M_I mapping



The mapping between the musical input (M_I) and the musical output (M_O) is governed by three constraints, ID-ASSOC, *PROSST, and MAX-PROSST, as illustrated in tableaux (69-70), where the optimal outputs preserve the musical beat association but remove the constructions of the prosodic word.

DOI:10.6814/NCCU202001428

Chapter 4

Language-to-Music Mapping: Coda Cluster

4.1 Introduction

This chapter continues to discuss the Mandarin-accented English in reading and singing, with a focus on coda clusters. Again, the segmental changes are examined in both the linguistic mapping and the language-to-music mapping. Unlike the onset clusters, the English coda clusters collected encounter the sonority sequencing principle (SSP). The segmental changes in relation to the coda clusters further support the argument for the separation of the perception grammar and the production grammar. Again, the linguistic output, the English words read by the Mandarin speakers, is perceived as the musical segmental input for singing. There are two linguistic output variations of the monosyllabic syllables with coda clusters. Linguistic output$_1$ preserves the coda cluster and remains monosyllabic. Linguistic output$_2$ involves vowel insertion so that the syllable surfaces with two syllables. The linguistic outputs are formed into prosodic words. Both of the linguistic outputs are respectively assigned with one and two musical beats. This chapter looks closely into how musical beat assignment affects the segmental changes and how the SSP and prosodic wordhood restrict the musical beat assignment in the musical input.

4.2 Data Design

The target monosyllabic words are designed to compare segmental changes of coda clusters in linguistic mappings and language-to-music mappings. The informants are the same as those mentioned in Chapter 3. They are asked to read and sing the

DOI:10.6814/NCCU202001428

assigned target words that contain coda clusters. Section 4.2.1 introduces the reading forms of the informants while 4.2.2 examines their singing forms.

4.2.1 Step 1: Reading

The target words contain coda clusters selected with different combinations of consonants on the sonority scale so that the influence of the Sonority Sequencing Principle (SSP) can be observed. The SSP is introduced by Sievers (1881) and Jespersen (1904). They propose that more sonorous segments stand closer to the syllable peak than syllables that are less sonorous. Carr (1993), and Broselow and Finer (1991) list the stop and fricative as separate classes. As shown in (75), *stop* is the least sonorous, while *glide* is the most sonorous.

(75) Sonority scale

Class	**Scale**
Glide	5
Liquid	4
Nasal	3
Fricative	2
Stop	1

The data in (76) are examples of the target words.

(76) Target words with coda clusters

		SSP	Example target words	
a.	liquid+stop	4 1	/park/ ‘park’	/maɪld/ ‘mild’
b.	fricative+stop	2 1	/dɛsk/ ‘desk’	/sɔft/ ‘soft’
c.	nasal+stop	3 1	/mind/ ‘maɪnd’	/ɪnk/ ‘ink’

d.	stop+stop	1 1	/ækt/ ‘act’	/kɛpt/ ‘kept’

The coda segment that is closer to the peak is more sonorous than the one that is further away from the peak. The codas in (76a, b) conform to the SSP, while those in (76c, d) violate the SSP; both /kt/ and /pt/ are stop-stop strings.

In the first step, the informants are asked to read the target words, as initiated in (77-80). There are two kinds of linguistic outputs. One preserves the coda cluster, while the other resolves the coda cluster by inserting a vowel.

(77) Target word: **soft**
Output: [sɔft], [sɔftə]

(78) Target word: **mind**
Output: [maɪnd], [maɪndə]

(79) Target word: **act**
Output: [ækt], [æktə]

(80) Target word: **kept**
Output: [kɛpt], [kɛptə]

(77-80) show the variation in reading the coda clusters, which is schematized in (81).

(81) Linguistic input Linguistic output

$$\text{/ækt/ ‘act’} \begin{cases} \nearrow (\text{ækt})_{\omega} & \text{Linguistic output}_1\ (\sigma) \\ \searrow (\text{æktə})_{\omega} & \text{Linguistic output}_2\ (\sigma\sigma) \end{cases}$$

As in (81), the linguistic outputs are in form of prosodic words (ω). The linguistic input, /ækt/, yields either (ækt)$_{\omega}$ or (æktə)$_{\omega}$ in the output. Linguistic output$_1$ preserves the coda cluster and remains monosyllabic, whereas linguistic output$_2$ inserts a vowel and becomes disyllabic.

(82) Statistics of the linguistic outputs

Linguistic input	Linguistic output$_1$	Linguistic output$_2$	Total
VCC[3]	$(VCC)_\omega$ (129/86%)	(VCC**V**)$_\omega$ (21/14%)	150(100%)

The table in (82) shows the mapping between the linguistic input and the linguistic output. There are 150 syllables with coda clusters. 129, or 86%, of them surface with coda clusters and the syllable number remains one. 21, or 14%, of them insert a vowel and the syllable number is changed into two. This shows that most informants tend to preserve the coda clusters.

4.2.2 Step 2: Singing

In step 2, the informants sing the linguistic outputs they produce in step 1. As mentioned in 4.2.1, there are two kinds of linguistic outputs, namely, linguistic output$_1$ and linguistic output$_2$. Each of the linguistic outputs is respectively assigned with one and two with beats in the language-to-music mapping. Take **(**ækt**)**$_\omega$ for example.

(83) Linguistic output$_1$ (σ) to singing mapping

Linguistic output$_1$ (ækt**)**$_\omega$

a. Singing output	b. Singing output
ækt │ ♩	æktə │ │ ♩ ♩

---

[3] The target words may have onset, for example, mind [maɪnd]. However, since onset is not the focus in this chapter, nuclei and codas are shown only.

As in (83), when linguistic output$_1$, (ækt)$_\omega$, is assigned with one musical beat, as in (83a), the singing output surfaces as [ækt], where no vowel insertion occurs. On the other hand, when linguistic output$_1$, (ækt)$_\omega$, is assigned with two musical beats, as in (83b), the singing output surfaces as [æktə], where a vowel is inserted. The prosodic word structures are removed in both of the singing outputs.

(84) Statistics of the linguistic output$_1$ (σ) to singing mapping

Linguistic output$_1$	Singing output (♩)	Singing output (♩ ♩)
(VCC)$_\omega$	VCC (121/93.8%) │ 𝄽	VCCV (86/66.7%) │ │ ♩ ♩
	VCCV (8/6.2%) ╲╱ ♩	VCC (43/33.3%) ╱╲ ♩ ♩
Total	129 (100%)	129 (100%)

As in (84), totally 129 coda clusters belong to linguistic output$_1$. Each output$_1$ is respectively assigned with one and two musical beats. When they are associated with one musical beat, 121, or 93.8%, of them surface as VCC (σ). When they are associated with two musical beats, 87, or 66.7%, of them still surface as VCC (σ).

Consider the output$_2$ mapping, as in (85).

(85) Linguistic output$_2$ (σσ) to singing mapping

Linguistic output$_2$ (æktə)$_\omega$

a. Singing output	b. Singing output
æktə ╲╱ ♩	æktə │ │ ♩ ♩

When (æktə)$_\omega$ is assigned with one musical beat, as in (85a), the singing output is [æktə]. When the same linguistic output is assigned with two musical beats, as in (85b), the singing output is still [æktə], where each syllable is linked to one musical beat respectively. The structure of the prosodic word is removed in the singing output.

(86) Statistics of the linguistic output$_2$ (σσ) to singing mapping

Linguistic output$_2$	Singing output (♩)	Singing output (♩ ♩)
(VCCV)$_\omega$	VCCV (17/81%) ♩	VCCV (18/85.7%) ♩ ♩
	VCC (4/19%) ♩	VCC (3/14.3%) ♩ ♩
Total	21 (100%)	21 (100%)

The table in (86) shows that there are totally 21 linguistic output$_2$ in the database. When they are associated with one musical beat, 17, or 81%, of them are sung as two syllables. When they are associated with two beats, 18, or 85.7%, of them are still sung as two syllables.

4.3 Language-to-music Mapping

The model I proposed in (46) of Chapter 3 is applicable here. The monosyllabic input yields two linguistic outputs, which are parsed into prosodic words. Each of the outputs is assigned with one and two musical beats. Linguistic output$_1$ is read with one syllable, where no vowel is inserted. Linguistic output$_2$ is read with two syllables, where a vowel is inserted to resolve the coda cluster.

The mapping from the linguistic output to the musical input demonstrates a need for the perception grammar. The linguistic outputs are perceived as the musical inputs, and assigned with one and two musical beats. The prosodic word structure,

DOI:10.6814/NCCU202001428

which may affect beat assignment, is shown in the musical input. When linguistic output$_1$ is associated with one musical beat, the coda cluster is preserved and the syllable remains monosyllabic. When linguistic output$_1$ is assigned with two musical beats, a vowel is inserted and the syllable becomes disyllabic. On the other hand, linguistic output$_2$ remains disyllabic, regardless of the number of the musical beat assigned. Finally, the musical inputs are mapped to the musical outputs, where the prosodic word structure is removed.

4.4 L_I-to-L_O Mapping: Production Grammar

This section analyzes the linguistic input (L_I) to the linguistic output (L_O) mapping under the framework of Optimality Theory (Prince and Smolensky 1993/2004). There are two linguistic outputs. One is with vowel insertion and the other is without vowel insertion. Take /ækt/ 'act,' for example. The output of /ækt/ can be either (ækt)$_\omega$ or (æktə)$_\omega$. Since a word like (ækt)$_\omega$ violates the SSP, the constraint in (87) is thus relevant.

(87) SSP:

> Assign one violation mark for every syllable whose coda does not rise in sonority toward the nucleus, or whose coda does not decrease in sonority from the nucleus.

This constraint is ranked relatively low so that (ækt)$_\omega$ can surface. Adding this constraint to the constraint rankings proposed in Chapter 3, the two output variants can be evaluated through the tableaux in (88-89).

(88) L_I-to-L_O mapping: output$_1$, (ækt)

/ækt/ 'act'	ALIGN-E (LEX, ω)	MAX-C	DEP-V	SSP	*CC
☞ a. (ækt)$_\omega$				*	*
b. (æktə)$_\omega$			*!		
c. (æk)$_\omega$		*!			
d. ækt	*!			*	*

In tableau (88), MAX-C and DEP-V are ranked higher than *CC so that (ækt)$_\omega$ will not be eliminated. (88b) inserts a vowel and incurs a fatal violation of DEP-V, so it is ruled out. (88c) deletes a consonant, and thus is ruled out by MAX-C. The candidate in (88d) is not a prosodic word, fatally violating ALIGN-E (LEX, ω). (88a) is thus selected as the optimal output, in sacrifice of the SSP.

(89) L_I-to-L_O mapping: output$_2$, (æktə)

/ækt/ 'act'	ALIGN-E (LEX, ω)	MAX-C	*CC	SSP	DEP-V
a. (ækt)$_\omega$			*!	*	
☞ b. (æktə)$_\omega$					*
c. (æk)$_\omega$		*!			
d. ækt	*!		*	*	

As in (89), *CC is ranked higher than DEP-V. (89a) preserves the coda cluster [kt], which is ruled out by *CC. (89c) deletes [t], and thus is ruled out by MAX-C. (89d) is not a prosodic word, and thus is ruled out by ALIGN-E (LEX, ω). (89b) thus emerges.

The constraint rankings for the linguistic input-to-output production grammar are summarized in (90).

(90) Linguistic input to linguistic output production grammar

a. Output$_1$: [ækt]→ (ækt)$_\omega$ ALIGN-E(LEX, ω), MAX-C, DEP-V >> SSP >> *CC

b. Output$_2$: [ækt]→ (æktə)$_\omega$ ALIGN-E(LEX, ω), MAX-C, *CC >> SSP >> DEP-V

In either ranking above, the SSP is not crucial.

4.5 L_O-to-M_I Mapping: Perception Grammar

4.5.1 L_{O1}-to-M_I Mapping

This section examines the monosyllabic linguistic output$_1$ (L_{O1}) to the musical input (M_I) mapping. L_{O1} is perceived as the musical segmental input and is respectively assigned with one and two musical beats. I first discuss the case that L_{O1} is assigned with one musical beat. Given the constraints posited in Chapter 3 with the addition of the SSP, the coda cluster mapping is illustrated in tableau (91), where the bottom-ranked constraints are omitted due to the limited space.

(91) L_{O1}-to-M_I mapping: coda cluster (♩)

L_{O1}: (ækt)$_\omega$ M_I: (ækt)$_\omega$
⋮
♩

(ækt)$_\omega$ ♩	NOSTRAY	MAX-C	DEP-V	SSP	ALIGN-R (♩, ω)
☞a. (ækt)$_\omega$ ⋮ ♩				*	
b.($\boxed{\text{æ}}$kt)$_\omega$ $\boxed{♩}$	*!*			*	
c. (æk)$_\omega$ ⋮ ♩		*!			
d. (æktə)$_\omega$ ⋮⋰ ♩			!*		

DOI:10.6814/NCCU202001428

There are a stray syllable and a stray musical beat in (91b), which incurs a fatal violation of NoStray. (91c) deletes a consonant to satisfy the SSP, but violates the higher-ranked MAX-C. (91d) is ruled out by Dep-V with an insertion of [ə]. Align-R(♩, ω) is inactive here because there is only one musical beat and one syllable. Eventually, (91a) emerges.

However, this constraint ranking selects the wrong output in tableau (92).

(92) L_{OI}-to-M_I mapping: coda cluster (♩♩)

L_{OI}: (ækt)$_{\omega}$　　M_I: (æktə)$_{\omega}$
⋮ ⋮
♩ ♩

(ækt)$_{\omega}$ ♩ ♩	NoStray	Max-C	Dep-V	SSP	Align-R (♩, ω)
(☞)a. (æktə)$_{\omega}$ ⋮ ⋮ ♩ ♩			*!		*
b. (ækt)$_{\omega}$ ⋮ ♩ $\boxed{♩}$	*!			*	
☛c. (ækt)$_{\omega}$ ⁚⋱ ♩ ♩				*	
d. (æk)$_{\omega}$ ⁚⋱ ♩ ♩		*!			

The real optimal output is (92a), but it is undesirably ruled out by NoStray, as indicated by the parenthesized right-head (☞) symbol. (92c) is incorrectly selected as the optimal output, as indicated by the black right-head ☛ symbol. In order to select the correct output, I posit the conjoined constraint in (93).

(93) NOSHARE-σ & *CC$_{\text{CODA}}$:

Assign one violation mark for every syllable that contains complex coda and is shared by two musical beats, ♩♩.

The unconjoined NOSHARE-σ and *CC$_{\text{CODA}}$ are ranked at the bottom. When NOSHARE-σ and *CC$_{\text{CODA}}$ are conjoined, the constraint ranking is substantially promoted. The purpose of the local conjunction is to rules out the worst of the worst output, which is called the WOW effect (Green 1993; Smolemsky 1993, 1995; Moreton & Smolensky 2002, among others). The conjoined NOSHARE-σ & *CC$_{\text{CODA}}$ is violated only when both NOSHARE-σ and *CC$_{\text{CODA}}$ are violated. This constraint indicates the dispreferrence of musical lengthening on a closed syllable with a complex coda, as in tableau (94).

(94) L_{O1}-to-M_I mapping: coda cluster (♩♩)

L_{O1}: (ækt)$_{\omega}$ M_I: (æktə)$_{\omega}$
⋮ ⋮
♩ ♩

(ækt)$_{\omega}$ ♩ ♩	NOSTRAY	NOSHARE-σ & *CC$_{\text{CODA}}$	MAX-C	DEP-V	SSP	ALIGN-R (♩, ω)
☞a. (æktə)$_{\omega}$ ⋮ ⋮ ♩ ♩				*		*
b. (ækt)$_{\omega}$ ⋮ ♩ ♩	*!				*	
c. (ækt)$_{\omega}$ ⋮⋱ ♩ ♩		*!			*	
d. (æk)$_{\omega}$ ⋮⋱ ♩ ♩			*!			

(94c) is now ruled out by NoShare-σ & *CC$_{\text{CODA}}$ as expected, whereby (94a) is enabled to surface. It should be noted that when a monosyllabic syllable with an onset cluster like (blu)$_{\omega}$ is assigned with two musical beats, its surface form is still (blu)$_{\omega}$; since it does not have a complex coda, the conjoined constraint NoShare-σ & *CC$_{\text{CODA}}$ would be irrelevant.

4.5.2 L$_{O2}$-to-M$_I$ Mapping

This section discusses the mapping from linguistic output$_2$ (L$_{O2}$) to the musical input (M$_I$). The disyllabic L$_{O2}$ is linked to either one or two musical beats in M$_I$. When the L$_{O2}$, (æktə)$_{\omega}$, is assigned with one musical beat, it is mapped to disyllabic (æktə)$_{\omega}$, as shown in tableau (95).

(95) L$_{O2}$-to-M$_I$ mapping: coda cluster (♩)

L$_{O2}$: (æktə)$_{\omega}$ M$_I$: (æktə)$_{\omega}$
˙.. ˙
♩

(æktə)$_{\omega}$ ♩	NoStray	NoShare-σ & *CC$_{\text{CODA}}$	Max-C	Dep-V	SSP	Align-R (♩, ω)
☞a. (æktə)$_{\omega}$ ⁚.˙ ♩						
b. ($\boxed{\text{æ}}$kt)$_{\omega}$ $\boxed{♩}$	*!*				*	
c. (æk)$_{\omega}$ ⋮ ♩			*!			
d. (ækt)$_{\omega}$ ⋮ ♩					*!	
e. (ktə)$_{\omega}$ ⋮ ♩					*!	

As previously shown in (91d), the candidate (æktə)$_{\omega}$ is not selected because of the violation of Dep-V. However, (æktə)$_{\omega}$ in (95a) is selected as the optimal output in that there is already an inserted vowel, and it does not violate Dep-V. Both (95d) and (95e) are ruled out by SSP. (95b) has an unassociated syllable and an unassociated musical beat, and is ruled out by NoStray. In (95c), [t] is deleted and fatally violates Max-C.

When the L_{O2}, (æktə)$_{\omega}$, is assigned with two musical beats, it is still mapped to the disyllabic (æktə)$_{\omega}$, where each syllable is linked to one musical beat, as in (96).

(96) L_{O2}-to-M_I mapping: coda cluster (♩♩)

L_{O2}: (æktə)$_{\omega}$ M_I: (æktə)$_{\omega}$
⋮ ⋮
♩ ♩

(æktə)$_{\omega}$ ♩ ♩	NoStray	NoShare-σ & *CC$_{Coda}$	Max-C	DEP-V	SSP	Align-R (♩, ω)
☞a. (æktə)$_{\omega}$ ⋮ ⋮ ♩ ♩						*
b. (ækt)$_{\omega}$ ⋮ ♩$\boxed{♩}$	*!				*	
c. (ækt)$_{\omega}$ ⁝˙. ♩ ♩		*!			*	
d. (æk)$_{\omega}$ ⁝˙. ♩ ♩			*!			*

In (96c), two musical beats are linked to a closed syllable with a complex coda, which is ruled out by the conjoined NoShare-σ & *CC$_{Coda}$. (96b) is ruled out by NoStray, and (96d) by Max-C. Eventually, (96a) emerges as the optimal output.

4.6 M_I-to-M_O Mapping: Production Grammar

Like the mapping of the onset cluster, the production grammar that maps the coda cluster from the musical input (M_I) to the musical output (M_O), preserves the musical beat association but removes the prosodic structure such as a prosodic word. Given the constraints posited in Chapter 3, tableau (97-99) show how these constraints compete with each other.

(97) M_I-to-M_O mapping: coda cluster (σ/♩)

(ækt)$_\omega$ ⋮ ♩	ID-ASSOC	*ProsSt	MAX-PROSST
☞a. ækt ⋮ ♩			*
b. (ækt)$_\omega$ ⋮ ♩		*!	
c. $\boxed{\text{æ}}$kt $\boxed{♩}$	*!		*

(98) M_I-to-M_O mapping: coda cluster (σσ/♩)

(æktə)$_\omega$ ⋱⋰ ♩	ID-ASSOC	*ProsSt	MAX-PROSST
☞a. æktə ⋱⋰ ♩			*
b. (æktə)$_\omega$ ⋱⋰ ♩		*!	
c. ækt$\boxed{\text{ə}}$ ⋮ ♩	*!		*

DOI:10.6814/NCCU202001428

(99) M_I-to-M_O mapping: coda cluster (σσ/♩♩)

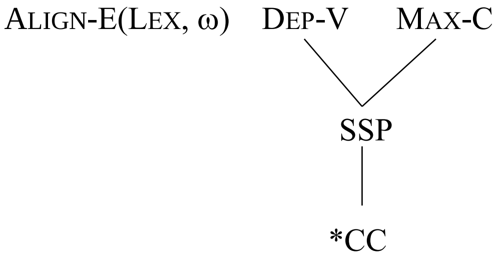
(æktə)$_\omega$ ⋮ ⋮ ♩ ♩	ID-ASSOC	*ProsSt	MAX-PROSST
☞a. æktə ⋮ ⋮ ♩ ♩			*
b. (æktə)$_\omega$ ⋮ ⋮ ♩ ♩		*!	
c. æktə ⋮ ♩ ♩	*!		*

In (97-99), candidates (b) are ruled out by *ProsSt, as they preserve the prosodic word in the output, and candidates (c) by ID-ASSOC, since the association lines are changed. It should be noted that when (ækt)$_\omega$ of L_{O1} and (æktə)$_\omega$ of L_{O2} are assigned with two musical beats, both emerge as (æktə)$_\omega$ in M_I, i.e., as the input in (99).

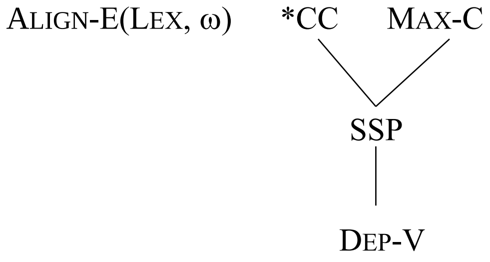
4.7 Summary

The mapping of a coda cluster also yields two linguistic outputs: L_{O1} is monosyllabic and faithful, while L_{O2} is disyllabic, with a vowel inserted to resolve the coda cluster. I have added the SSP to the set of constraints posited in Chapter 3, and the constraint rankings are enriched as in (100-101).

(100) L_I-to-L_{O1} mapping: (σ)

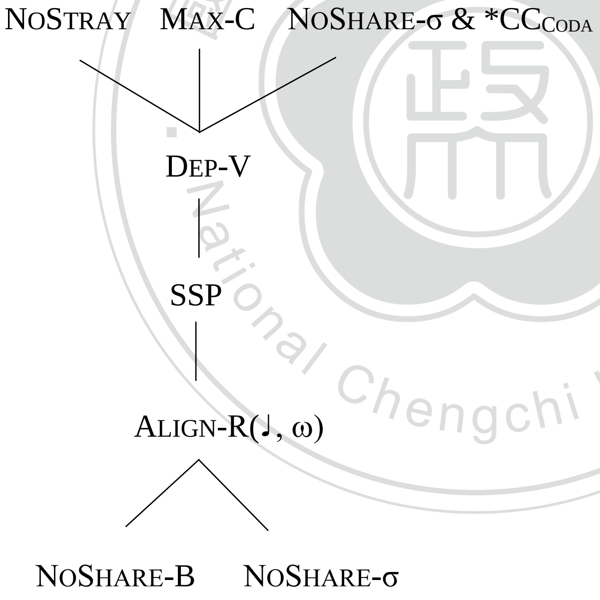


(101) L_I-to-L_{O2} mapping: (σσ)



In the L_I-to-L_O mapping, there are two constraint rankings yielding two linguistic outputs, but in the L_O-to-M_I perception grammar, there is only a single constraint ranking, as in (102), where I have observed the influence of SSP on the insertion of vowels in singing.

(102) L_O-to-M_I Mapping



As posited in Chapter 3, the constraints Id-Assoc, *ProsSt, and Max-ProsSt serve to map between the musical input (M_I) to the musical output (M_O). The musical output preserves the musical beat association but removes the structure of the prosodic word.

DOI:10.6814/NCCU202001428

Chapter 5

Language-to-Music Mapping: Foot and Musical Beat Assignment

5.1 Introduction

Previous studies (Fang and Su 2005, Li 2009) have found that Mandarin children's song values the rhythmic correspondence between lyric and song, which facilitates Mandarin learning of children. In terms of the connection between lyric and song, this chapter examines the language-to-music mapping of rhythmic structures in Mandarin children's songs with a focus on the foot.

The mapping schema provided in Chapter 3 is also applicable here, which is reproduced below.

(103) Language-to-music mapping: rhythm

Linguistic input
Production grammar
Linguistic output
Perception grammar
Musical input
Production grammar
Musical output

The syntactic structure of the lyrics in the linguistic input yields the prosodic structure in the linguistic output through the production grammar. The linguistic output is then mapped to the musical input through the perception grammar. The prosodic structure in the musical input serves to condition the musical beat assignment, but is removed in the musical output through the musical production grammar.

This study constructs a database of Mandarin children's songs based on five lyric books that contain thirty-nine songs.[4] These books are for children from zero to eight years old, and the lyrics are simple and close to children's daily lives. The songs in the abovementioned books are new songs, but not traditional songs. The composing of the musical melodies is based on existing lyrics.

This chapter discusses three questions of the language-to-music mapping. First, how is a foot formed in the linguistic output? Second, how does the structure of the foot affect the assignment of musical beats? Finally, how does the musical structure emerge in the musical output?

5.2 L_I-to-L_O Mapping: Foot Formation

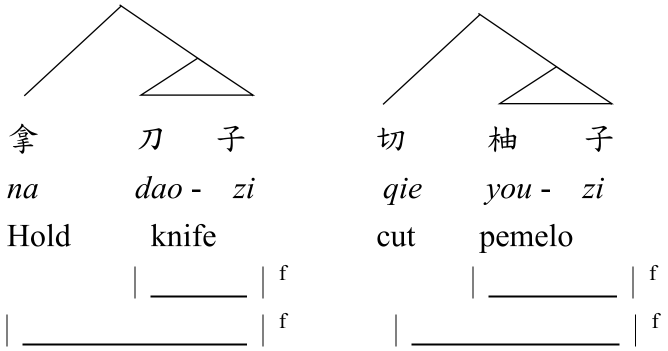
Along the lines of Shih (1986) and Hsiao (1991), I consider the structure of the prosodic foot is formed in the linguistic output on four conditions. First, a pair of syllables that are ICs constitutes a binary foot. Second, adjacent syllables are paired into a disyllabic foot. Third, unparsed monosyllable joins its adjacent foot to form a bigger foot. Finally, two adjacent syllables that syntactically branch in the opposite direction cannot form a foot.

A couple of trisyllabic examples are given in (104).

---

[4] The five books are *zao an wan an* 早安晚安, *xiao yu di* 小雨滴, *xiao hou zi* 小猴子, and *1234 dong dong ti cao* 1234 動動體操. One traditional song is included in one of the books, and is excluded from the database because it is not clear whether the lyrics or the music is composed first.

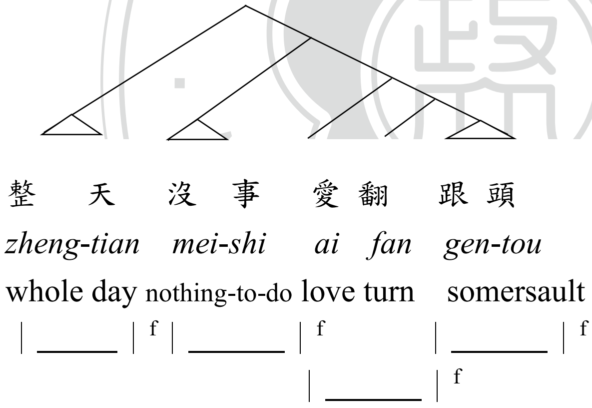
(104) (a) ‘Hold a knife.’ (b) ‘Cut a pomelo.’



In (104), *dao-zi* and *you-zi* form two binary feet first. The leftover monosyllables, *na* and *qie*, then join the existing binary feet to form trisyllabic feet respectively.

Consider the first two syllables in (105).

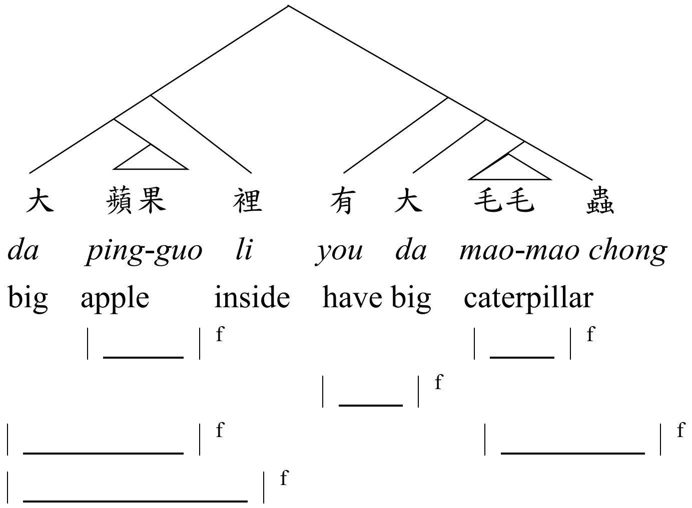
(105) ‘(He) loves to somersault when (he) has nothing to do all day long.’



As in (105), after the three pairs of ICs, *zheng-tian*, *mei-shi*, and *gen-tou*, form binary feet, the adjacent unparsed syllables, *ai* and *fan*, are paired into a disyllabic foot.

Unlike *ai* and *fan*, the adjacent syllables, *li* and *you*, in (106) cannot form a foot, since they branch in the opposite direction.

(106) ‘There is a big caterpillar in a big apple.’



In (106), *ping-guo* and *mao-mao* are ICs, and thus they are grouped into disyllabic feet respectively. At this point, *li* and *you* syntactically branch in the opposite direction so that they cannot form a foot; the unparsed *you* and *da* are then paired into a disyllabic foot. The leftover monosyllables, *da, li* and *chong*, subsequently join the adjacent existing feet to trisyllabic and tetrasyllabic feet.

Foot constraints governing the foot formation are proposed in (107-110), based on the concepts developed in work by Shih (1986), Hsiao (1991), and Lin (2001).

(107) Align-E(IC, Ft):

Assign one violation mark for every pair of syllables that are ICs whose edges do not coincide with the edges of a foot.

(108) FtBin:

Assign one violation mark for every foot that contains more than two syllables.

(109) *MONO-FT:

Assign one violation mark for every monosyllabic foot in the output.

DOI:10.6814/NCCU202001428

(110) Parse-σ:

Assign one violation mark for every syllable that is not parsed into any foot in the output.

The constraints Parse-σ and *MONO-FT must be undominated to ensure that every syllable is parsed into a foot but not a monosyllabic foot. Align-E(IC, Ft) is also ranked at the top to match IC edges with foot edges. FtBin is ranked at the bottom, such that a trisyllabic or larger foot can occur. A partial constraint ranking is posited in (111).

(111) L_I-to-L_O partial constraint ranking:

Parse-σ, Align-E(IC, Ft), *MONO-FT >> FtBin

Tableau (112) shows how this constraint ranking works.

(112) = (104)

Input: [*na* [*dao-zi*]$_{NP}$]$_{VP}$ [*qie* [*you-zi*]$_{NP}$]$_{VP}$
拿 刀 子 切 柚 子
Hold knife cut pomelo
'Hold a knife and cut a pomelo.'

Candidates:

a. (*na* (*dao-zi*)$_{Ft}$)$_{Ft}$ (*qie* (*you-zi*)$_{Ft}$)$_{Ft}$
b. *na* (*dao-zi*)$_{Ft}$)$_{Ft}$ (*qie* (*you-zi*)$_{Ft}$)$_{Ft}$
c. (*na dao*)$_{Ft}$ (*-zi qie*)$_{Ft}$ (*you-zi*)$_{Ft}$
d. (*na*)$_{Ft}$ (*dao-zi*)$_{Ft}$ (*qie* (*you-zi*)$_{Ft}$)$_{Ft}$

L$_{\text{I}}$-to-L$_{\text{O}}$: Parse-σ, Align-E(IC, Ft), *MONO-FT >> FtBin

	Parse-σ	Align-E(IC, Ft)	*MONO-FT	FtBin
☞ a.				**
b.	*!			*
c.		*!		
d.			*!	*

The footing pattern in (112b) is ruled out by Parse-σ, as *na* is left unfooted. The pair of the ICs, *dao* and *zi*, in (112c) is not aligned with any foot, incurring a fatal violation of Align-E(IC, Ft). The monosyllabic foot, *na*, in (112d) violates *MONO-FT, and thus is ruled out. As a consequence, (112a) is chosen as the optimal output.

The constraint ranking in (111) also correctly predicts that the fifth and sixth syllables in (113) form a disyllabic foot.

(113) = (105)

Input: [[*zheng-tian*]$_{\text{ADV}}$[*mei-shi*]$_{\text{VP}}$ [*ai* [*fan* [*gen-tou*]$_{\text{NP}}$]$_{\text{VP}}$]$_{\text{VP}}$

整 天 沒 事 愛 翻 跟 頭

whole day nothing-to-do love turn somersault

'(He) loves to somersault when (he) has nothing to do all day long.'

Candidates:

a. (*zheng-tian*)$_{\text{Ft}}$ (*mei-shi*)$_{\text{Ft}}$ (*ai fan*)$_{\text{Ft}}$ (*gen-tou*)$_{\text{Ft}}$

b. (*zheng-tian*)$_{\text{Ft}}$ (*mei-shi*)$_{\text{Ft}}$ *ai* (*fan* (*gen-tou*)$_{\text{Ft}}$)$_{\text{Ft}}$

c. (*zheng-tian*)$_{\text{Ft}}$ (*mei-shi*)$_{\text{Ft}}$ (*ai*)$_{\text{Ft}}$ (*fan* (*gen-tou*)$_{\text{Ft}}$)$_{\text{Ft}}$

d. (*zheng-tian*)$_{\text{Ft}}$ (*mei-shi*)$_{\text{Ft}}$ (*ai*)$_{\text{Ft}}$ (*fan gen-tou*)$_{\text{Ft}}$

DOI:10.6814/NCCU202001428

L_I-to-L_O: Parse-σ, Align-E(IC, Ft), *MONO-FT >> FtBin

	Parse-σ	Align-E(IC, Ft)	*MONO-FT	FtBin
☞ a.				
b.	*!			*
c.			*!	*
d.		* (!)	* (!)	*

Again, (113b) violates Parse-σ, as *ai* is not parsed into any foot, whereas *ai* in (113c-d) constitutes a foot alone and violates *MONO-FT. The left edge of the pair of the ICs, *gen* and *tou* in (113d) is not aligned with the left edge of a foot, and thus is ruled out by Align-E(IC, Ft). Finally, (113a) emerges.

The same constraint ranking, however, renders an incorrect prediction in tableau (114).

(114) = (106)
Input: [[*da* [*ping-guo*]$_{\text{NP}}$]$_{\text{NP}}$ *li*]$_{\text{PP}}$ [*you* [*da* [[*mao-mao*] *chong*]$_{\text{NP}}$]$_{\text{NP}}$]$_{\text{VP}}$,
大 蘋 果 裡 有 大 毛 毛 蟲
big apple inside have big caterpillar
‘There is a big caterpillar in a big apple.’

Candidates:

a. (((*da* (*ping-guo*)$_{\text{Ft}}$)$_{\text{Ft}}$ *li*)$_{\text{Ft}}$ (*you da*)$_{\text{Ft}}$ ((*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$
b. (*da* (*ping-guo*)$_{\text{Ft}}$)$_{\text{Ft}}$ (*li you*)$_{\text{Ft}}$ (*da* (*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$
c. (*da* (*ping-guo*)$_{\text{Ft}}$ *li*)$_{\text{Ft}}$ *you* (*da* ((*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$)$_{\text{Ft}}$
d. (*da ping-*)$_{\text{Ft}}$ (guo *li*) $_{\text{Ft}}$ (*you* (*da* ((*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$)$_{\text{Ft}}$
e. (*da* (*ping-guo*)$_{\text{Ft}}$ *li*)$_{\text{Ft}}$ (*you* (*da* ((*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$)$_{\text{Ft}}$)$_{\text{Ft}}$
f. ((*da* (*ping-guo*)$_{\text{Ft}}$)$_{\text{Ft}}$ *li*)$_{\text{Ft}}$ (*you*)$_{\text{Ft}}$ (*da* ((*mao-mao*)$_{\text{Ft}}$ *chong*)$_{\text{Ft}}$)$_{\text{Ft}}$

L_I-to-L_O: Parse-σ, Align-E(IC, Ft), *MONO-FT >> FtBin

	Parse-σ	Align-E(IC, Ft)	*MONO-FT	FtBin
(☞) a.				***!
☛ b.				**
c.	*!			***
d.		*!		***
e.				***!*
f.			*!	***!*

The foot structure in (114b) erroneously wins over (114a) by one less violation of FtBin, as indicated by the black right-headed hand symbol ☛. The real optimal output is (114a), as indicated by the parenthesized white right-headed hand symbol (☞). In order to obtain (114a), the constraint in (115) must be added to the partial constraint ranking provided in (111).

(115) NOSTRADDLE-Ft:

Assign a violation mark for every foot consisting of syllables that syntactically branch in the opposite direction.

NOSTRADDLE-Ft forbids syllables that syntactically branch in opposite direction to form a foot, which I refer to as a straddle foot. The enriched constraint ranking is provided in (116) and the top-ranking of NOSTRADDLE-Ft now successfully removes the straddling foot in (117b).

(116) L_I-to-L_O constraint ranking (enriched):

NOSTRADDLE-Ft, Parse-σ, Align-E(IC, Ft), *MONO-FT >> FtBin

DOI:10.6814/NCCU202001428

(117) = (114)

L_I-to-L_O: NoStraddle-Ft, Parse-σ, Align-E(IC, Ft), *Mono-Ft >> FtBin

	NoStraddle-Ft	Parse-σ	Align-E(IC, Ft)	*Mono-Ft	FtBin
☞ (114a).					***
(114b).	*!				**
(114c).		*!			***
(114d).			*!		***
(114e).					***!*
(114f).				*!	***!*

When the linguistic output is perceived into the musical input, the foot structure plays a role in assigning musical beats, as will be discussed next.

5.3 L_O-to-M_I Mapping: Footing and Musical Beat Assignment

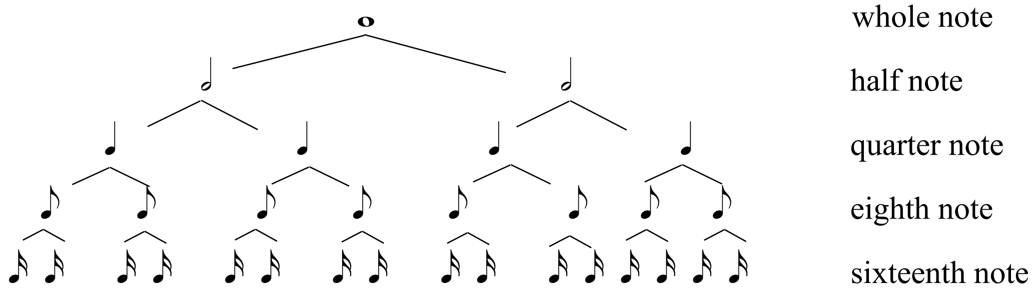
The perception grammar that maps the linguistic footing output to the musical input rhythm involves three questions. First, in what way can a musical beat be shared by syllables? Second, in what way can a syllable be shared by musical beats? Finally, how are musical beats confined by the domain of the foot? Before I enter the discussions, some musical basics are introduced in the following section.

5.3.1 Some Musical Basics

The names and durations of the musical notes are given in (118).

DOI:10.6814/NCCU202001428

(118)



The whole note (𝅝) is the top note, which is equal to two half notes (𝅗𝅥). A half note is equal to two quarter notes (♩). The duration of one quarter note is equal to two eighth notes (♪). One eighth note is equal to two sixteenth notes (𝅘𝅥𝅯). When there is a dot besides a note, the duration of that note becomes 1.5 times of itself. For example, if the length of ♩ represents one beat, ♩. indicates one and a half beat.

There are two kinds of time signatures in the database, namely 4/4 and 6/8. The time signature indicates the beat number in each measure and the note value that is equal to a musical beat.

(119) Time signature: 4/4

Music metric accent (primary accent: > secondary accent: >)

\> >

| ♩ ♩ ♩ ♩ |

As in (119), the time signature is 4/4. The first digit is the beat number in each measure, whose edges are signaled by bar lines, |. The second digit indicates the note value that is equal to a beat. Therefore, 4/4 means that there are 4 beats in each measure and a quarter note is equal to a beat. In the 4/4 time signature, the default primary accent (>) is assigned to the first beat while the default secondary accent (>) is assigned to the third beat.

DOI:10.6814/NCCU202001428

The time signature in (120) is 6/8, which means that there are six beats in each measure and an eighth note (♪) is equals to a beat. In the 6/8 time signature, the first beat has the primary accent while the fourth beat has the secondary accent.

(120) Time signature: 6/8
Music metric accent (primary accent: > secondary accent: >)
\> >
|♪ ♪ ♪ ♪ ♪ ♪|

When two notes are linked by a concave arc line (‿), the duration is the combination of the two notes.

(121) Two musical beats
♩‿♩

As in (121), the linked quarter-notes signal two musical beats when a quarter note is equal to a beat.

5.3.2 Musical Beat Assignment

The mapping of the lyric output rhythm to the musical input is keyed to the assignment of musical beats. Two relevant constraints are posited in (122-123).

(122) NoShare-σ(B≥5):
Assign one violation mark for every syllable that is shared by five or more musical beats.

(123) NoSplit-IC(B):

> Assign a violation mark for every pair of syllables that are syntactic ICs but are not associated with the same musical beat.

As previously proposed in Chapter 3, NoStray stems from the notion of stray erasure, and bans any unassociated musical beat or syllable. NoShare-σ(B≥5) indicates the musical limit on syllable lengthening, and disallows a syllable to be shared by five or more than five musical beats. NoSplit-IC(B) suggests the tight connection between ICs, requiring a pair of ICs to share a musical beat. NoStray and NoShare-σ(B≥5) are ranked above NoSplit-IC(B), as ICs do not always share a musical beat. NoShare-σ and NoShare-B, which are posited in Chapter 3, are ranked at the bottom, as a syllable is often linked to multiple musical beats to prevent stray elements and vice versa.

The partial constraint ranking is given in (124), and the tableau in (125) shows how this constraint ranking works.

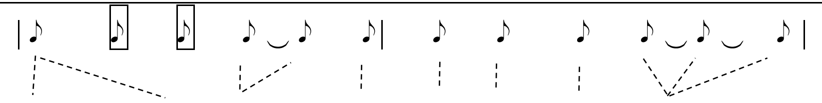
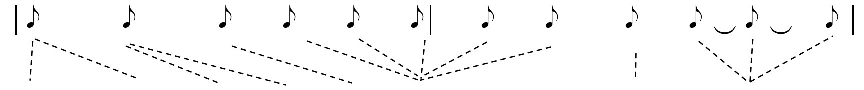
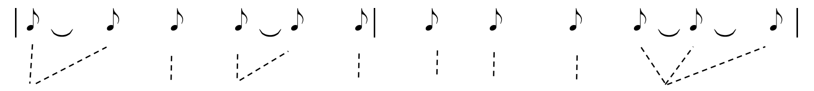
(124) L_O-to-M_I partial constraint ranking

NoStray, NoShare-σ(B≥5) >> NoSplit-IC(B), NoShare-B, NoShare-σ

(125)
L_O: (*zheng-tian*)$_{Ft}$ (*mei-shi*)$_{Ft}$ (*ai fan*)$_{Ft}$ (*gen-tou*)$_{Ft}$
整 天 沒 事 愛 翻 跟 頭
whole day nothing-to-do love turn somersault
‘(He) loves to somersault when (he) has nothing to do all day long.’

Music time signature: 6/8 (There are six beats in each measure.)
| ♪ ♪ ♪ ♪ ♪ ♪ |♪ ♪ ♪ ♪ ♪ ♪ |

Candidates:

a.	 (*zheng*__- *tian*)$_{Ft}$ (*mei*__- *shi*)$_{Ft}$ (*ai fan*)$_{Ft}$ (*gen -tou*____)$_{Ft}$
b.	 (*zheng* - *tian*)$_{Ft}$ (*mei*__- *shi*)$_{Ft}$ (*ai fan*)$_{Ft}$ (*gen* - *tou*______)$_{Ft}$
c.	 (*zheng* - *tian*)$_{Ft}$ (*mei*- *shi*)$_{Ft}$ (*ai fan*)$_{Ft}$ (*gen* - *tou*____)$_{Ft}$

L$_O$-to-M$_I$: NOSTRAY, NOSHARE-σ(B≥5) >> NOSPLIT-IC(B), NOSHARE-B, NOSHARE-σ

	NOSTRAY	NOSHARE-σ(B≥5)	NOSPLIT-IC(B)	NOSHARE-B	NOSHARE-σ
☞ a.			***		***
b.	*!*		**	*	**
c.		*!	**	**	**

The linguistic output of (125) emerges from tableau (113). The floating musical beats in (125b) incur fatal violations of NOSTRAY. (125c) is ruled out by NOSHARE-σ(B≥5), as *fan* is shared by five musical beats. Consequently, (125a) is selected as the optimal output, in spite of three violations of NOSPLIT-IC(B) and NOSHARE-σ.

The linguistic output selected in tableau (117) is mapped into music in (126).

(126)

L$_O$: ((*da* (*ping- guo*)$_{Ft}$)$_{Ft}$ *li*)$_{Ft}$ (*you da*)$_{Ft}$ ((*mao-mao*)$_{Ft}$ *chong*)$_{Ft}$

大 蘋 果 裡 有 大 毛 毛 蟲

big apple inside have big caterpillar

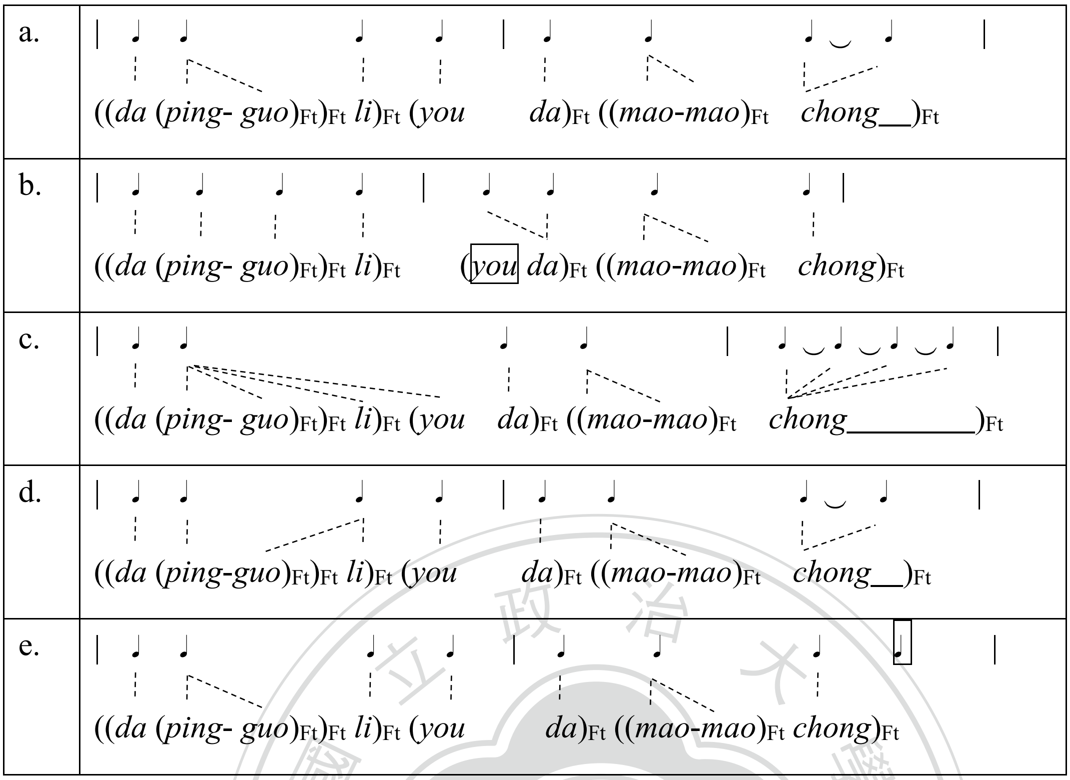
‘There is a big caterpillar in a big apple.’

Music time signature: 4/4 (There are four beats in each measure.)

| ♩ ♩ ♩ ♩ | ♩ ♩ ♩ ♩ |

DOI:10.6814/NCCU202001428

Candidates:



L_O-to-M_I: NoStray, NoShare-σ(B≥5) >> NoSplit-IC(B), NoShare-B, NoShare-σ

	NoStray	NoShare-σ(B≥5)	NoSplit-IC(B)	NoShare-B	NoShare-σ
☞ a.				**	*
b.	*!		*	*	*
☛ c.				**	*
d.			*!	**	*
e.	*!			**	

There is a floating syllable in (126b) and a floating musical beat in (126e); hence both are ruled out by NoStray. The pair of ICs, *ping-guo*, in (126d) is associated by two different musical beats, and thus is ruled by by NoSplit-IC(B). The real optimal output is (126a), but (126c) wrongly emerges as well, as indicated by the ☛ symbol. To remove (126c), we need a constraint to prevent syllables that belong to different feet to share the same musical beat, as in (127).

(127) NoStraddle-B(Ft):

Assign one violation mark for every beat that is shared by syllables from different feet.

The constraint NoStraddle-B(Ft) is also top-ranked as shown in the enriched constraint ranking in (128). Tableau (126) can now be reanalyzed as tableau (129).

(128) L_O-to-M_I constraint ranking (enriched)

NoStraddle-B(Ft), NoStray, NoShare-σ(B≥5) >> NoSplit-IC(B), NoShare-B, NoShare-σ

(129)

L_O-to-M_I: NoStraddle-B(Ft), NoStray, NoShare-σ(B≥5) >> NoSplit-IC(B), NoShare-B, NoShare-σ

	NoStraddle-B(Ft)	NoStray	NoShare-σ (B≥5)	NoSplit-IC(B)	NoShare -B	NoShare -σ
☞ (126a)					**	*
(126b)		*!		*	*	*
(126c)	*!				**	*
(126d)				*!	**	*
(126e)		*!			**	

In (126c), *li* and *you* branch in the opposite direction syntactically, but share the same musical beat; hence it is ruled out by NoStraddle-B(Ft).

5.4 M_I-to-M_O Mapping: Prosody Removal and Association Faithfulness

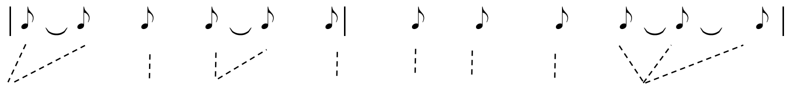
The production grammar that maps the musical input (M_I) to the musical output (M_O) preserves the musical beat association but removes the prosodic structure such as prosodic

DOI:10.6814/NCCU202001428

word. Consider the tableau in (130).

(130)

Musical input:



(*zheng__- tian*)$_{\text{Ft}}$ (*mei__- shi*)$_{\text{Ft}}$ (*ai fan*)$_{\text{Ft}}$ (*gen- tou______*)$_{\text{Ft}}$

whole day nothing-to-do love turn somersault

整 天 沒 事 愛 翻 跟 頭

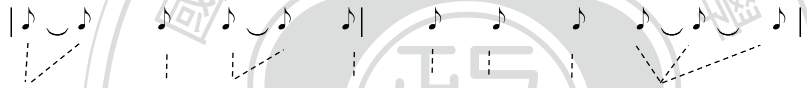
'(He) loves to somersault when (he) has nothing to do all day long.'

Music time signature: 6/8 (There are six beats in each measure.)

| ♪ ♪ ♪ ♪ ♪ ♪ |♪ ♪ ♪ ♪ ♪ ♪ |

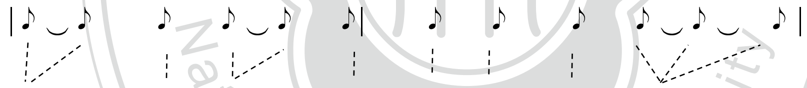
Candidates:

a.



zheng__- tian mei__ - shi ai fan gen- tou_____

b.



(*zheng__- tian*)$_{\text{Ft}}$ (*mei__- shi*)$_{\text{Ft}}$ (*ai fan*)$_{\text{Ft}}$ (*gen- tou______*)$_{\text{Ft}}$

c.

zheng__- tian mei__- shi ai fan gen- tou______

M$_{\text{I}}$-to-M$_{\text{O}}$: ID-ASSOC, *PROSST >> MAX-PROSST

	ID-ASSOC	*PROSST	MAX-PROSST
☞ a.			*
b.		*!	
c.	*!		*

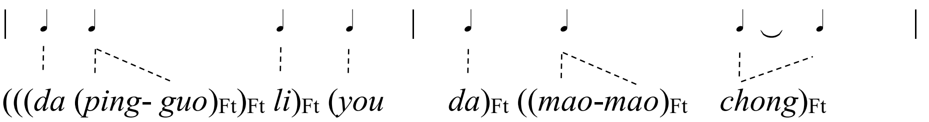
The musical input of (130) emerges from tableau (125). The associated line in (130c) is different from that in the musical input, and (130c) thus is ruled out by ID-ASSOC. In

(130b), the prosodic structure is preserved in the musical output, which fatally violates *PROSST. As a consequence, (130a) emerges.

The M_I-to-M_O mapping in (131) is evaluated by the same constraint ranking.

(131)

M_I:

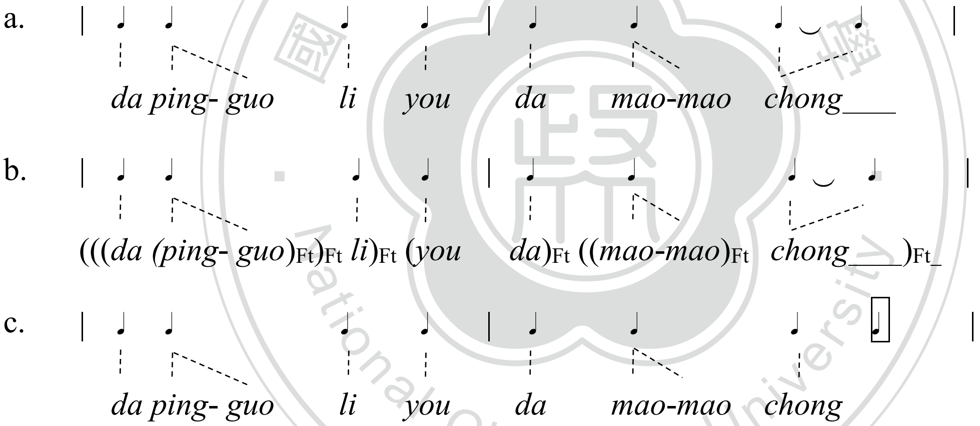


大 蘋 果 裡 有 大 毛 毛 蟲

big apple inside have big caterpillar

'There is a big caterpillar in a big apple.'

Candidates:



M_I-to-M_O: ID-ASSOC, *PROSST >> MAX-PROSST

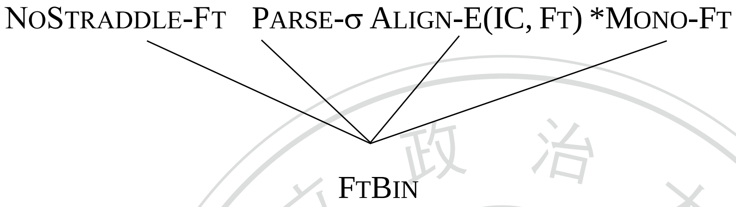
	ID-ASSOC	*PROSST	MAX-PROSST
☞ a.			*
b.		*!	
c.	*!		*

The music input emerges from tableau (129). Again, (131b) and (131c) are ruled by *PROSST and ID-ASSOC respectively. The optimal output is (131a), which surfaces in sacrifice of MAX-PROSST.

5.5 Summary

The mapping of foot structure between the linguistic input and the linguistic output has an access to syntactic structures, in particular, immediate constituency and branching direction. I have proposed a set of constraints to govern the L_I-to-L_O production grammar, as illustrated by the Hasse diagram in (132).

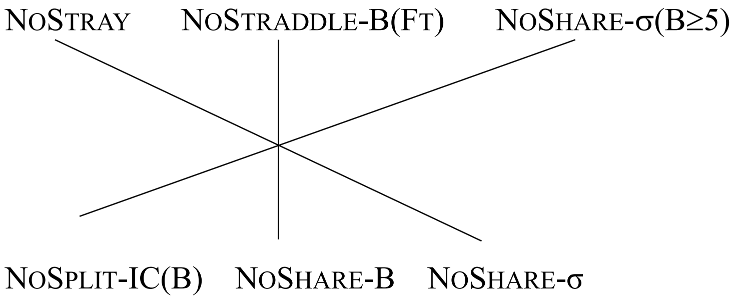
(132) L_I-to-L_O production grammar



The top-ranking of NoStraddle-Ft and Align-E(IC, Ft) prevents syllables that syntactically branch in the opposite direction from forming a foot, but requires syllables that are syntactic ICs to form a foot. The top-ranking of Parse-σ and *Mono-Ft ensures that every syllable is parsed into a foot, but not a monosyllabic foot. The bottom-ranking of FtBin allows trisyllabic or larger foot to occur.

The mapping of the linguistic output prosody to the musical input rhythm lies in the foot structure and the syllable-to-beat assignment. I have posited a set of constraints, as shown by the Hasse diagram in (133), to govern the L_O-to-M_I perception grammar.

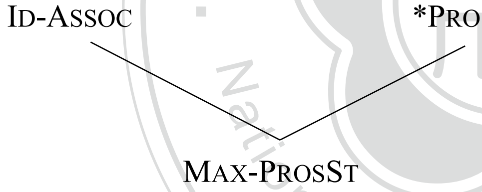
(133) L_O-to-M_I perception grammar



The top-ranking of NoStray bans any unassociated musical beat or syllable in the musical output, that of NoShare-σ(B≥5) indicates musical limit on syllable lengthening, and that of NoStraddle-B(Ft) prohibits a musical beat to straddle two feet. The bottom-ranking of NoSplit-IC(B) then predicts the fact that syllables that are ICs do not always share a musical beat. NoShare-B and NoShare-σ are also ranked at the bottom, as a syllable is often linked to multiple musical beats to prevent stray elements and vice versa.

The mapping between the musical input and the musical output lies in the faithfulness of syllable-beat association and the removal of prosodic structure in the musical output. Three constraints are proposed to govern the M_I-to-M_O production grammar, as indicated by the Hasse diagram in (134).

(134) M_I-to-M_O production grammar



The top-ranking of Id-Assoc retains the musical beat assignment in the musical output. The ranking of *ProsSt above Max-ProsSt removes all prosodic structures in the musical output.

Chapter 6

Language-to-Music Mapping: IP and Musical Beat Assignment

6.1 Introduction

This chapter continues to examine the language-to-music mapping of rhythmic structures in the composing of Mandarin children's songs. In particular, this chapter examines the mapping of intonational phrase (IP) between language and music. The mapping schema in Chapter 5 is reproduced below.

(135) Language-to-Music mapping: rhythm

Linguistic input
Production grammar
Linguistic output
Perception grammar
Musical input
Production grammar
Musical output

This chapter discusses three questions of the language-to-music mapping. First, how is an IP formed in the linguistic output? Second, how does the structure of the IP affect the assignment of musical beats? Finally, how does the musical structure emerge in the musical output?

6.2 L_I-to-L_O Mapping: intonational phrasing

The composing of musical melody for the Mandarin children's songs is based on the given lyrics. The punctuation marks of the lyric lines, such as commas and periods, denote IP boundaries. The following are some examples.

(136) (拿 刀 子)**IP** ，(切 柚 子)**IP** ，(切 了 柚 子)**IP**， (剝 柚子)**IP**。
na dao-zi *qie you-zi* *qie-le you-zi* *bo you zi*
hold knife cut pomelo cut-ASP pomelo peel pomelo
'Hold a knife to cut a pomelo. After cutting the pomelo, peel the pomelo.'

(137) (大 蘋 果 裡 有 大 毛 毛 蟲)**IP** ，
da ping-guo li you da mao-mao chong
big apple inside have big caterpillar
'There is a big caterpillar in a big apple.'

(138) (扭 扭 屁 股 跳 一 跳)**IP**。
niu niu pi- gu tiao yi tiao
shake shake bottom jump one jump
'Shake the bottom and jump.'

There are three commas in (136) and one in (137). Each denotes a right edge of an IP. In (136) and (138), a period also signals the end of an IP. As a result, (136) consists of four IPs, while in either (137) or (138), the entire line constitutes an IP respectively. The mapping of IP between the linguistic input and the linguistic output can be governed by the constraint in (139).

(139) ALIGN-R(PM, IP):
Assign one violation mark for every punctuation mark that does not coincide with the right edge of an IP.

When the linguistic output is perceived into the musical input, the IP structure plays a role in aligning the musical measures and assigning musical beats, as will be discussed next.

6.3 L_O-to-M_I Mapping: IP and Musical Beat Assignment

The perception grammar that maps the linguistic IP output to the musical input rhythm involves three questions. First, how are IP's and musical measures aligned? Second, how does the edge of an IP affect the assignment of musical beats to IC syllables?

6.3.1 Edge alignment between IP and Musical Measure

The database shows that the left edge of an odd-number measure tends to align with the left edge of an IP, found in 100% of the data, as in (140). The right edge of an even-number measure tends to align with right edge of an IP, found in 98% of the data, as in (141).

(140) Odd-number measure alignment

Aligned	113	100%
Unaligned	0	0%
Total	113	100%

(141) Even-number measure alignment

Aligned	111	98.2%
Unaligned	2	1.8%
Total	113	100%

The examples in (142) and (143) illustrate this tendency.

DOI:10.6814/NCCU202001428

(142) |♩ ♪ ♪ ♩ ♩ |1 ♩ ♪ ♪ ♩ ‿ ♩ |2

((*da* (*ping-guo*)$_{\text{FT}}$ *li*)$_{\text{FT}}$ (*you* *da*)$_{\text{FT}}$ (*mao-mao*)$_{\text{FT}}$ *chong*___)$_{\text{FT}}$**)IP**

大 蘋 果 裡 有 大 毛 毛 蟲

big apple inside have big caterpillar

'There is a big caterpillar in a big apple.'

(143) | ♪ ♪ ♩ ♪ ♪ ♩ |1 ♪. 𝅘𝅥𝅯 ♪ ♪ ♪ ♪ ♩ |2

((*na* (*dao-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)IP** **(**(*qie* (*you-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)IP** **(**(*qie-le*)$_{\text{FT}}$ (*you-zi*)$_{\text{FT}}$**)IP** **(**(*bo* (*you-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)IP**

拿 刀 子 切 柚 子 切 了 柚 子 剝 柚 子

hold knife cut pomelo cut-ASP pomelo peel pomelo

'Hold a knife to cut a pomelo. After cutting the pomelo, peel the pomelo.'

The pattern in (142) and (143) shows that an odd-number measure is left-aligned with an IP but not vice versa, and that an even-number measure is right-aligned with an IP but not vice versa.

The right edge of an IP may go across the even-number measure only in the last line of a song, where the final syllable is lengthened to signal the end of the song, as in (144).

(144) | ♩ ♩ ♩ ♩ |11 ♩ ‿ ♩ ♩ ‿ ♩ |12 ♩ ‿ ♩ ‿ ♩ ‿ ♩

((*niu niu*)$_{\text{FT}}$ (*pi - gu*)$_{\text{FT}}$ (*tiao*_______ (*yi*_______ *tiao*_________)$_{\text{FT}}$)$_{\text{FT}}$**)IP**

扭 扭 屁 股 跳 一 跳

Shake shake bottom jump one jump

'Shake bottom and jump.'

The measure alignment can be governed by the constraints in (145-148).

(145) ALIGN-L(M_N, IP):

Let N ∈ odd number.

Assign one violation mark for every odd number measure, M_N, whose left edge does not coincide with the left edge of an IP.

(146) ALIGN-R(M_{2N}, IP):

Let 2N ∈ even number.

Assign one violation mark for every even number measure, M_{2N}, whose right edge does not coincide with the right edge of an IP.

(147) ALIGN-L(IP, M_N):

Let N ∈ odd number.

Assign one violation mark for every IP, whose left edge does not coincide with the left edge of the odd number measure, M_N.

(148) ALIGN-R(IP, M_{2N}):

Let 2N ∈ even number.

Assign one violation mark for every IP, whose right edge does not coincide with the right edge of the even number measure, M_{2N}.

The constraints ALIGN-L(M_N, IP) and ALIGN-R(M_{2N}, IP) must be ranked above ALIGN-L(IP, M_N) and ALIGN-R(IP, M_{2N}), as in (149).

(149) L_O-to-M_I partial constraint ranking:

ALIGN-L(M_N, IP), ALIGN-R(M_{2N}, IP) >> ALIGN-L(IP, M_N), ALIGN-R(IP, M_{2N})

This constraint ranking correctly predicts the measure alignment in (150).

DOI:10.6814/NCCU202001428

(150)

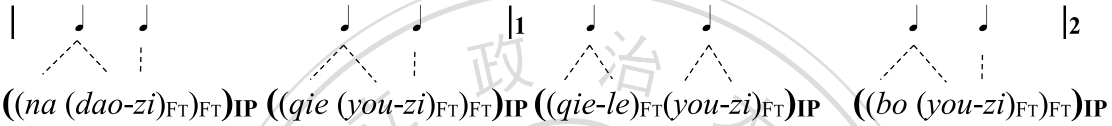
L$_{\text{O}}$: **(**(*na* (*dao-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)**$_{\textbf{IP}}$ **(**(*qie* (*you-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)**$_{\textbf{IP}}$ **(**(*qie-le*)$_{\text{FT}}$(*you-zi*)$_{\text{FT}}$**)**$_{\textbf{IP}}$ **(**(*bo* (*you-zi*)$_{\text{FT}}$)$_{\text{FT}}$**)**$_{\textbf{IP}}$

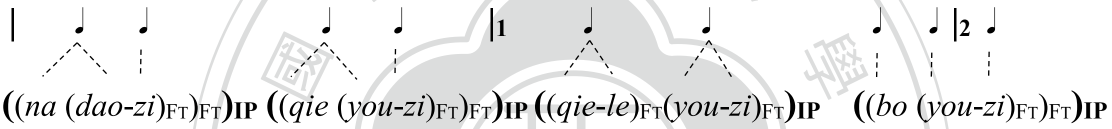
拿刀子	切柚子	切了柚子	剝柚子。
hold knife	cut pomelo	cut-ASP pomelo	peel pomelo

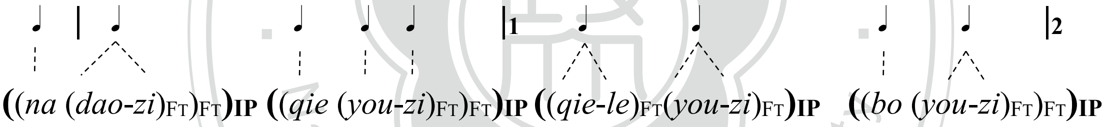
‘Hold a knife to cut a pomelo. After cutting the pomelo, peel the pomelo.’

Music time signature: 4/4 (There are four beats in each measure.)

Candidates:

a.


b.


c.


L$_{\text{O}}$-to-M$_{\text{I}}$: Align-L(M$_{\text{N}}$, IP), Align-R(M$_{\text{2N}}$, IP) >> Align-L(IP, M$_{\text{N}}$), Align-R(IP, M$_{\text{2N}}$)

	Align-L(M$_{\text{N}}$, IP)	Align-R(M$_{\text{2N}}$, IP)	Align-L(IP, M$_{\text{N}}$)	Align-R(IP, M$_{\text{2N}}$)
☞ a.			***	***
b.		*!	***	****
c.	*!		****	***

The linguistic output is obtained from (136). The right edge of measure 2 in (150b) is unaligned with that of an IP, and thus is ruled out by Align-R(M$_{\text{2N}}$, IP). In (150c), measure 1 is not left-aligned with an IP, and thus is ruled out by Align-L(M$_{\text{N}}$, IP). Hence (150a) is chosen as the optimal output.

The constraint ranking in (149), however, is unable to account for the measure alignment in (144), where measure 12 is not right-aligned with an IP. A constraint like (151) is necessary.

(151) ALIGN-R (SONG, IP):

Assign one violation mark for every song whose right edge does not coincide with the right edge of an IP.

This constraint requires that the final measure of a song be right-aligned with an IP, and it must outrank ALIGN-R(M_{2N}, IP). The partial constraint ranking in (149) is now modified and enriched as that in (152).

(152) L_O-to-M_I constraint ranking (enriched)

ALIGN-R(SONG, IP), ALIGN-L(M_N, IP) >> ALIGN-R(M_{2N}, IP) >>
ALIGN-L(IP, M_N), ALIGN-R(IP, M_{2N})

This constraint ranking correctly choses (153a) as the optimal output.

(153)

L_O: ((*niu* *niu*)$_{FT}$ (*pi gu*)$_{FT}$ (*tiao* (*yi* *tiao*)$_{FT}$)$_{FT}$)$_{\mathbf{IP}}$

扭 扭 屁股 跳 一 跳

shake shake bottom jump one jump

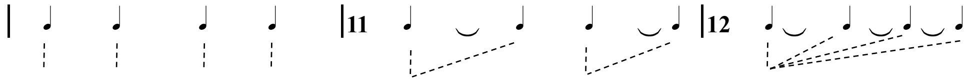
‘Shake the bottom and jump.’

Music time signature: 4/4 (There are four beats in each measure.)

| ♩ ♩ ♩ ♩ $|_N$ ♩ ♩ ♩ ♩ $|_{2N}$

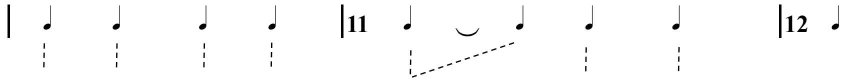
Candidates:

a.



((*niu* *niu*)$_{FT}$ (*pi* - *gu*)$_{FT}$ (*tiao*____ (*yi* ____ *tiao*____________)$_{FT}$)$_{FT}$**)**$_{\mathbf{IP}}$

b.



((*niu niu*)$_{FT}$ (*pi* *gu*)$_{FT}$ (*tiao*____ (*yi* *tiao*)$_{FT}$)$_{FT}$**)**$_{\mathbf{IP}}$

L$_O$-to-M$_I$: ALIGN-L(M$_N$, IP), ALIGN-R(SONG, IP) >> ALIGN-R(M$_{2N}$, IP)

	ALIGN-L(M$_N$, IP)	ALIGN-R(SONG, IP)	ALIGN-R(M$_{2N}$, IP)
☞ a.			*
b.		*!	

The linguistic output is obtained from (138). In (153b), measure 12 is right-aligned with the final IP in conformity of ALIGN-R(M$_{2N}$, IP), but the final measure of the song is unaligned. Hence it is ruled out by the higher-ranked ALIGN-R(SONG, IP).

The constraint ranking in (153) also correctly select (154a) as the optimal output.

(154)

L$_O$: **(**(*beng- beng*)$_{FT}$ (*tiao-tiao*)$_{FT}$ (*wo* *shi*)$_{FT}$ (*xiao* (*tu-zi*)$_{FT}$)$_{FT}$**)**$_{\mathbf{IP}}$

蹦 蹦	跳 跳	我	是	小	兔 子	
leap leap	jump jump	1SG	be	little bunny		

'Caper. I am a bunny.'

Music time signature: 4/4 (There are four beats in each measure.)

DOI:10.6814/NCCU202001428

Candidates:

a.

((*beng-beng*)$_{\text{FT}}$ (*tiao-tiao*)$_{\text{FT}}$ (*wo shi*) (*xiao tu - zi*__________)$_{\text{FT}}$)$_{\text{FT}}$**)**$_{\textbf{IP}}$

b.

((*beng-beng*)$_{\text{FT}}$ (*tiao-tiao*)$_{\text{FT}}$ (*wo shi*) (*xiao tu- zi*)$_{\text{FT}}$)$_{\text{FT}}$**)**$_{\textbf{IP}}$

L_O-to-M_I: ALIGN-L(M_N, IP), ALIGN-R(SONG, IP) >> ALIGN-R(M_{2N}, IP)

	ALIGN-L(M_N, IP)	ALIGN-R (SONG, IP)	ALIGN-R(M_{2N}, IP)
☞ a.			*
b.		*!	

In (154b), the final measure of the song is unaligned so it is eliminated by ALIGN-R(SONG, IP).

The intonational phrasing in (150), (153), and (154) also indicates another phenomenon, namely, an IP-final syllable is musically longer than its preceding syllable. The following section will discuss musical lengthening at the right edge of an IP

6.3.2 IP-final Lengthening

In Mandarin Children's songs, an IP-final syllable is usually associated with a longer musical beat, i.e., its musical duration is longer than the preceding syllable. In the database, there are 258 IPs, and 218 of their final syllables are aligned with longer musical beats, found in 84.5% of the data. The example in (155) shows the final lengthening.

(155)

((*yu-dao*)$_{\text{FT}}$ (*lu-deng*)$_{\text{FT}}$ ((*wang qian*)$_{\text{FT}}$ *kai*)$_{\text{FT}}$**)**$_{\textbf{IP}}$

遇 到 綠 燈 往 前 開

meet green light toward front drive

'Drive ahead when the traffic light turns green.'

The time signature of (155) is 4/4, which means that there are four beats in each measure and a quarter note (♩) is equal to one beat. The IP final syllable, *kai*, has one musical beat, which is longer than the preceding syllable, *qian*, which is set to a demibeat.

(156) | ♪ ‿ ♪ ♪ ♪ ‿ ♪ ♪ | ♪ ‿ ♪ ♪ ♪ ‿ ♪ ‿ ♪ |
((*lan*____*tian*)$_{\text{Ft}}$ (*lan*____ *ya*)$_{\text{Ft}}$ ((*bai* *yun*)$_{\text{Ft}}$ *bai*________)$_{\text{Ft}}$)$_{\textbf{IP}}$
藍 天 藍 呀 白 雲 白
blue sky blue FP white cloud white
‘The sky is blue and the cloud is white.’

(157) | ♪ ‿ ♪ ♪ ♪ ‿ ♪ ♪ | ♪ ♪ ♪ ♪ ‿ ♪ ‿ ♪ |
((*zheng*___*-tian*)$_{\text{Ft}}$ (*mei*___ *- shi*)$_{\text{Ft}}$ (*ai* *fan*)$_{\text{Ft}}$ (*gen - tou*________)$_{\text{Ft}}$)$_{\textbf{IP}}$
整 天 沒 事 愛 翻 跟 頭
whole day nothing-to-do love turn somersault
‘(He) loves to somersault when (he) has nothing to do all day long.’

The time signature of (156) and (157) are both 6/8, which means that there are six beats in each measure and an eighth-note (♪) is equal to one beat. Musical lengthening occurs here. In (156) and (157), the IP-final syllables, *bai* and *tou* are associated with three musical beats respectively, while their preceding syllables, *yun* and *gen*, are associated with one musical beat each. A constraint of sequential markedness is needed to govern the IP-final musical lengthening, as posited in (158).

(158) $*\sigma_{n-1} \geq \sigma_n]_{IP}$

Assign one violation mark for every IP-final syllable whose musical duration is equal to or shorter than its preceding syllable.

DOI:10.6814/NCCU202001428

This constraint serves to rule out the illegal candidates in (159).

(159)

L_O: **(**(*yu-dao*)$_{\text{Ft}}$ (*lu-deng*)$_{\text{Ft}}$ ((*wang* *qian*)$_{\text{Ft}}$ *kai*)$_{\text{Ft}}$**)**$_{\textbf{IP}}$

遇 到 綠 燈 往 前 開。

meet green light toward front drive

‘Drive ahead when the traffic light turns green.’

Music time signature: 4/4 (There are four beats in each measure.)

Candidates:

a.


((*yu-dao*)$_{\text{Ft}}$ (*lu-deng*)$_{\text{Ft}}$ ((*wang qian*)$_{\text{Ft}}$ *kai*)$_{\text{Ft}}$**)**$_{\textbf{IP}}$

b.


((*yu-dao*)$_{\text{Ft}}$ (*lu-deng*)$_{\text{Ft}}$ ((*wang qian*)$_{\text{Ft}}$ *kai*)$_{\text{Ft}}$**)**$_{\textbf{IP}}$

c.


((*yu-dao*)$_{\text{Ft}}$ (*lu-deng*)$_{\text{Ft}}$ ((*wang qian*___)$_{\text{Ft}}$ *kai*)$_{\text{Ft}}$**)**$_{\textbf{IP}}$

	*$\sigma_{n-1} \geq \sigma_n]_{IP}$
☞ a.	
b.	*!
c.	*!

In (159b), the duration of the IP-final syllable, *kai*, is equal to the preceding syllable, *qian*, and in (159c), *kai* is shorter than *qian*. Both violate *$\sigma_{n-1} \geq \sigma_n]_{IP}$ fatally. This constraint may interact with the constraints of musical beat assignment, as shown in (160).

(160)

L$_{O}$: ((*zheng* - *tian*)$_{\text{Ft}}$ (*mei* - *shi*)$_{\text{Ft}}$ (*ai* *fan*)$_{\text{Ft}}$ (*gen* - *tou*)$_{\text{Ft}}$)$_{\text{IP}}$

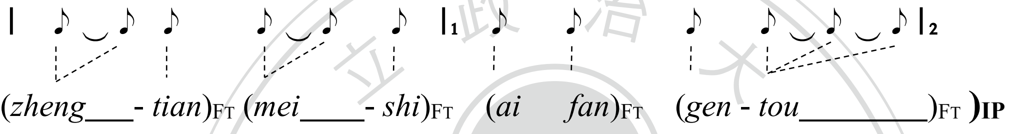
整 天 沒 事 愛 翻 跟 頭

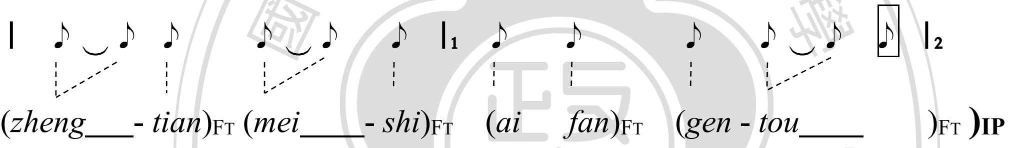
whole day nothing-to-do love turn somersault

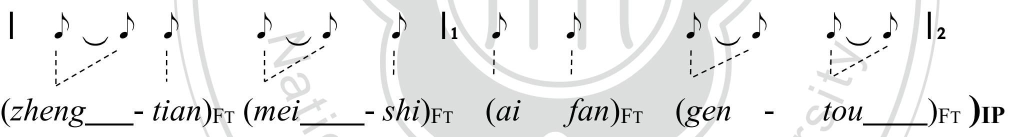
'(He) loves to somersault when (he) has nothing to do all day long.'

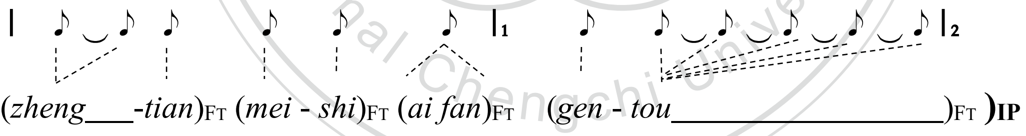
Music time signature: 6/8 (There are six beats in each measure.)

Candidates:

a.


b.


c.


d.


L$_{O}$-to-M$_{I}$: NOSTRAY, $*\sigma_{n-1} \geq \sigma_{n}]_{\text{IP}}$, NOSHARE-σ(B≥5) >> NOSPLIT-IC(B)

	NOSTRAY	$*\sigma_{n-1} \geq \sigma_{n}]_{\text{IP}}$	NOSHARE-σ(B≥5)	NOSPLIT-IC(B)
☞ a.				***
b.	*!			***
c.		*!		***
d.			*!	***

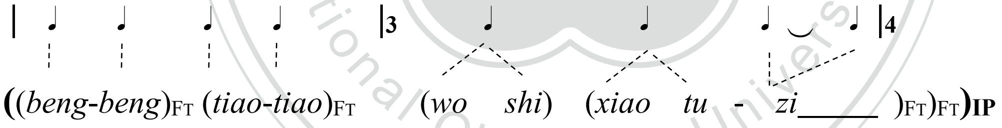
The time signature here is 6/8, where an eighth-note (♪) is equal to one beat. There is a floating beat in (160b), which is then ruled out by NoStray. In (160c), the duration of the IP-final syllable, *tou*, is equal to its preceding syllable *gen*, and thus is ruled out by $*\sigma_{n-1} \geq \sigma_n]_{IP}$. In (160d), the final syllable is linked to five musical beats, incurring a fatal violation of NoShare-σ(B≥5). Consequently, (160a) is selected as the optimal output.

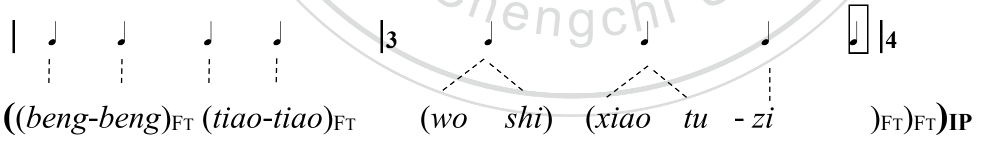
The constraint ranking in (160) also selects (161a) as the optimal output.

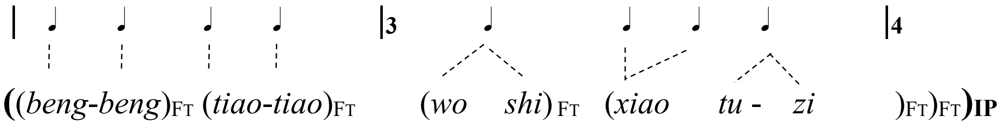
(161)
L_O: **(**(*beng-beng*)$_{FT}$ (*tiao-tiao*)$_{FT}$ (*wo shi*)$_{FT}$ (*xiao* (*tu-zi*)$_{FT}$)$_{FT}$**)$_{IP}$**
蹦 蹦 跳 跳 我 是 小 兔 子
leap leap jump jump 1SG be little bunny
'Caper. I am a bunny.'

Music time signature: 4/4 (There are four beats in each measure.)

Candidates:

a.


b.


c.


L_O-to-M_I: NoStray, $*\sigma_{n-1} \geq \sigma_n]_{IP}$, NoShare-σ(B≥5) >> NoSplit-IC(B)

	NoStray	$*\sigma_{n-1} \geq \sigma_n]_{IP}$	NoShare-σ(B≥5)	NoSplit-IC(B)
☞ a.				***
b.	*!			***
c.		*!		**

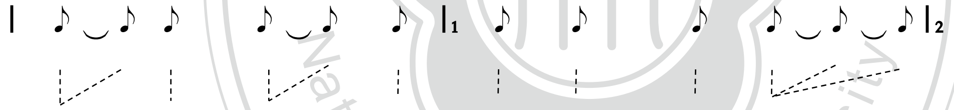
The time signature here is 4/4, where a quarter-note (♩) is equal to one beat. NoStray rules out (161b,) which has a floating beat. In (161c), the duration of the IP-final syllable, *zi*, is equal to its preceding syllable, *tu*, and hence is ruled out by $*\sigma_{n-1} \geq \sigma_n]_{IP}$. As a result, (161a) emerges. The lyric line of (161) is identical to that of (154), but their mappings to music are different. However, their mappings to music are both governed by the proposed constraint rankings.

6.4 M_I-to-M_O Mapping: Prosody Removal and Association Faithfulness

In the mapping from the musical input to the musical output, again, the prosodic structure is removed, while syllable-beat association is faithful to the musical input.

(162)

M_I:



((*zheng* - *tian*)$_{FT}$ (*mei* - *shi*)$_{FT}$ (*ai* *fan*)$_{FT}$ (*gen* - *tou*________)$_{FT}$ **)**$_{\mathbf{IP}}$

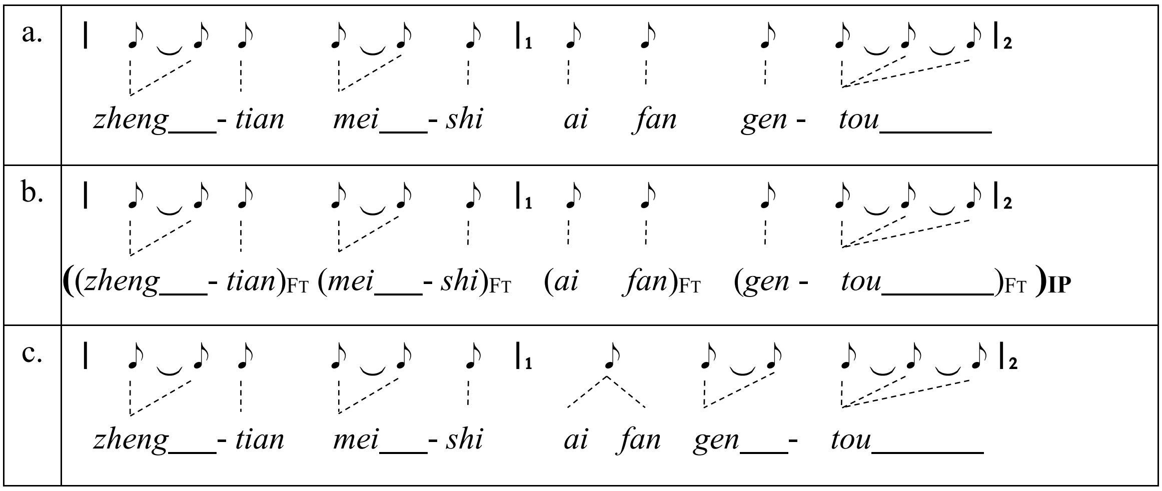
整 天 沒 事 愛 翻 跟 頭

whole day nothing-to-do love turn somersault

‘(He) loves to somersault when (he) has nothing to do all day long.’

Music time signature: 6/8 (There are six beats in each measure.)

Candidates:



$\mathrm{M_I}$-to-$\mathrm{M_O}$: Id-Assoc, *ProsSt >> Max-ProsSt

	Id-Assoc	*ProsSt	Max-ProsSt
☞ a.			*
b.		*!	
c.	*!		*

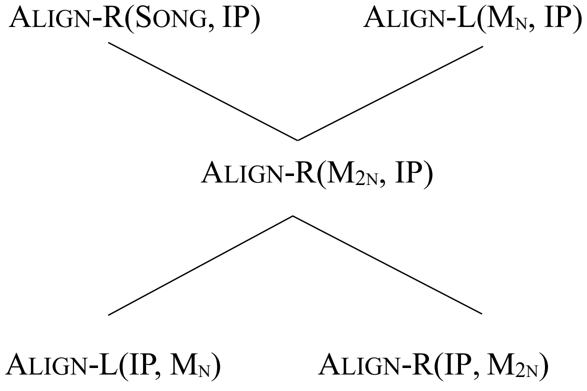
The musical input is selected from tableau (160). The prosodic structure in (162b) is retained in the musical output, and thus is ruled out by *ProsSt. The associated lines in (162c) are different from those in the musical input, incurring a fatal violation of Id-Assoc. Eventually, (162a) emerges as the optimal output.

6.5 Summary

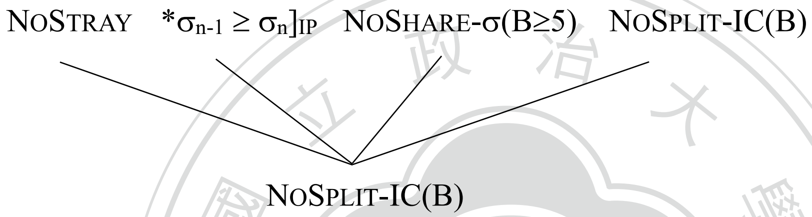
The intonational phrases in the lyric output are indicated by the punctuation marks. I have posited the constraint Align-R(Pm, IP) to govern this pattern. The mapping of the intonational phrases to the musical input is keyed to the measure-to-IP alignment and the IP-final musical lengthening. I have posited two sets of constraints, as shown by the Hasse diagrams in (163-164), to govern the $\mathrm{L_O}$-to-$\mathrm{M_I}$ perception grammar.

DOI:10.6814/NCCU202001428

(163) L_O-to-M_I perception grammar: measure-to-IP alignment



(164) L_O-to-M_I perception grammar: musical beat assignment

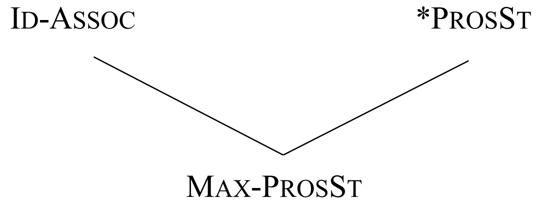


In (163), the top-ranking of ALIGN-R(SONG, IP) and ALIGN-L(M_N, IP) ensures that the odd-number measure is left-aligned with an IP, and the final measure of a song is right-aligned with an IP. The medial ranking of ALIGN-R(M_{2N}, IP) requires the even-number measure to be right-aligned with an IP, unless it is followed by the final measure of a song. On the other hand, in (164), the top-ranking of NOSTRAY, *$\sigma_{n-1} \geq \sigma_n]_{IP}$, NOSHARE-$\sigma$(B≥5) and NOSPLIT-IC(B) predicts four patterns. First, no stray syllable or musical beat is allowed. Second, the IP-final syllable must be musically longer than its preceding syllable. Third, a musical beat cannot be linked to five or more musical beats. Finally, a pair of ICs cannot be linked to different musical beats.

As discussed in Chapter 3-5, the mapping between the musical input and the musical output is governed by three constraints, ID-ASSOC, *PROSST and MAX-PROSST. The Hasse diagram is reproduced in (165).

DOI:10.6814/NCCU202001428

(165) M_I-to-M_O production grammar



The ranking of *PROSST above MAX-PROSST avoids the presence of foot structure and IP structure in the musical output. The top-ranking of ID-ASSOC requires the preservation of the musical beat assignment.

Chapter 7

Music-to-Language Mapping: Tone and Term Association

7.1 Introduction

This chapter examines children's perception of the musical melody in the singing of Mandarin songs. I observe how the children map the musical pitches to linguistic tones, and how they associate the tones to the words or phrases in their mental lexicon. The music-to-language mapping is generalized by the schema in (166).

(166) Music-to-language mapping: tone

Musical output
Perception grammar
Linguistic input
Production grammar
Linguistic output

Singing words or phrases are mapped to the linguistic input forms with similar tonal values through the perception grammar. The children associate what they have heard to their mental lexicon through the production grammar. In particular, I examine how disyllabic musical pitches (rising, falling, and level) are transformed into linguistic tones, and how the toned disyllabic strings are interpreted and produced in the output.

7.2 Some Basics and the Proposed Corresponding Principles

In Chao's (1930) five-scale notation system, Mandarin has four lexical tones, as

DOI:10.6814/NCCU202001428

listed in (167), including 55, which represents high level, 35, which represents rising, 21, which represents low, and 53, which represents falling.

(167) Tone notations: numeric values

T1	T2	T3	T4
55	35	21	53

Musical pitches show the highness or lowness of a sound. The database provides a variety of musical pitch combinations. For the convenience of discussion, the pitch names are converted into numeric musical notation. The pitch name of a G major scale is given in (168). For instance, the numeric musical notation for G is 1, and C refers to the middle C on the piano keyboard.

(168) G major scale and numeric musical notation conversion

Pitch name	G	A	B	C	D	E	F#	G
Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	1̇

In this study, I propose a corresponding scale between musical pitch and language tone, as in (169).

(169) Musical pitch to language tone corresponding scale

<table>
<tr><td rowspan="2">Music</td><td>Pitch</td><td>7̣, 1, #1</td><td>2, #2, 3</td><td>4, #4, 5</td><td>#5, 6, #6</td><td>7, 1, #1̇</td></tr>
<tr><td>Register</td><td>[-upper]</td><td>Unspecified</td><td colspan="3">[+upper]</td></tr>
<tr><td rowspan="2">Language</td><td>Tone</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Register</td><td colspan="2">[-upper]</td><td>Unspecified</td><td colspan="2">[+upper]</td></tr>
</table>

DOI:10.6814/NCCU202001428

There are totally fifteen musical pitches in the database. The distance of each pitch is a half step (半音), which is the distance from one key to the next adjacent key on the keyboard. For example, the musical pitches,**7̣**, **1**, and **#1**, correspond to 1 in linguistic tone, the musical pitches, **2**, **#2**, and **3**, correspond to 2 in linguistic tone, and so forth.

The music register, as proposed in (169), may affect the tonal perception. Yip (1989, 2007) proposes two binary features, [-upper] and [+upper], to characterize the tonal registers; the tone ranges from 1 to 2 is of the [-upper] register, that from 4 to 5 is of the [+upper] register, whereas 3 is unspecified. The term "register" in fact comes from music. In terms of the pitch-tone correspondence, I propose here that the musical pitch ranges from **7̣** to **#1** pertains to the [-upper] register, that from **4** to **#i** pertains to the [+upper] register, but that from **2** to **3** is unspecified. The [-upper] and [+upper] registers in music correspond to the level tones. The unspecified pitch range can be mapped to either the [+upper] register or the [-upper] register.

The distance between two musical pitches is also a factor that decides the perceived tones. The larger distance between the musical pitches, the larger distance between the perceived tones. The scale in (170) shows the pitch distance that is calculated by the number of half step, namely, *0.5*. Two half steps add up to one whole step, which is counted as *1* in distance. For example, the distance from pitch **1** to **5** is *3.5*. The distance from a pitch to itself is *0*.

(170) Distance of musical pitches

Pitch	**7̣**	**1**	**#1**	**2**	**#2**	**3**	**4**	**#4**	**5**	**#5**	**6**	**#6**	**7**	**1**	**#i**
Distance	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	

The distance between two language tones is shown in (171). For example, the tonal distance from 1 to 5 is *4*. The distance from any tone to itself is *0*.

(171) Distance of tones

Lg tone	1		2		3		4		5
Distance		1		1		1		1	

The correspondence between the musical pitch distance and the tone distance is proposed in (172).

(172) Pitch distance mapping principle

Musical dist. = 0 ℜ Tonal dist. = 0
0.5 ≤ Musical dist. ≤ 3 ℜ Tonal dist. = 2
Musical dist. ≥ 3.5 ℜ Tonal dist. = 4

The *0* music distance corresponds to *0* tone distance. When the musical pitch distance is equal to or more than *0.5* but less than or equal to *3*, the corresponding tone distance is *2*. In Mandarin, 11 and 22, whose distance is *1*, are usually perceived as low both. However, 11 and 33, whose distance is *2*, are not regarded the same. When the musical pitch distance is equal to or more than *3.5*, the corresponding tone distance is *4*. This is because there is no limit for musical pitch distance, but the tone distance is at most *4* in Mandarin (and other Chinese dialects).

7.3 Data Design

The data are designed to examine two questions of the music-to-language mapping. First, how is musical pitch interpreted as linguistic tone by children? What

DOI:10.6814/NCCU202001428

is the relationship between pitch-tone faithfulness and term association?

The data are collected from normal-hearing preschool children, aged from four to six. The children live in Taipei and their parents are all native Mandarin speakers. There are totally six children participating in this project. Three of them are female and three are male.[5] All of the children can precisely repeat Mandarin tones and musical melodies, which indicates that they can correctly perceive Mandarin tones and musical melodies.

The children are asked to listen to the singing of disyllabic terms. The words that children are familiar to are animals, body parts, food and drink, natural vocabularies, common verbs, etc., mainly taken from Liu & Chen's (2015) study that evaluates children from sixteen months to thirty-six months. The evaluation inventory of Liu & Chen (2015) is constructed on the basis of MCDI (Mandarin-Chinese Communicative Development Inventories, Taiwan) proposed by Liu & Tsao (2010). Some examples of the disyllabic terms in the present study are given in in (173).

(173) Examples of the familiar disyllabic words/phrases

Familiar words/phrases	
*bing*21 *gan*55 'cookie' 餅乾	*yan*21 *jing*55 'eye' 眼睛
*tai*53 *yang*35 'sun' 太陽	*yi*35 *bian*53 'one time' 一遍
*shu*55 *bao*55 'bookbag' 書包	*zao*35 *qi*21 'get up early' 早起

The data contains 359 tokens of disyllabic words or phrases that children are familiar or unfamiliar with. The words or phrases are set to different combination of musical pitches from (173). Each syllable is set to one musical pitch, whose duration is one beat. The syllables are sung by a female native Mandarin speaker with an

---

[5] In spite of the fact that Lin (1968) finds that there are no significant sex differences in language development of preschool children, both sexes of children are included in this study.

DOI:10.6814/NCCU202001428

electric keyboard. The singer is well-trained in music and singing so that the musical pitches are produced accurately. The children are asked to tell the possible lyrics that are sung by the singer. The following are some examples.

(174) Examples of the singing words/phrases

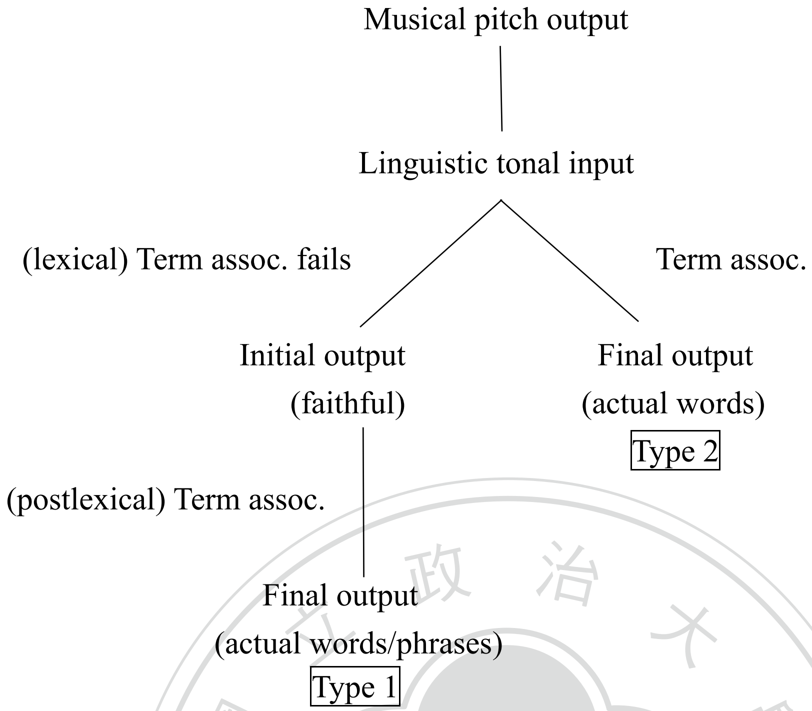
	Pitch contours	Singing outputs
a.	Falling	*bing*$^{\#}$**5** *gan*$^{\#}$**1** ‘cookie’ 餅乾 ↘
b.	Rising	*tai*$^{\#}$**1** *yang*$^{\#}$**5** ‘sun’ 太陽 ↗
c.	Level	*shi*$^{\#}$**5** *jie*$^{\#}$**5** ‘world’ 世界 →

The singing output in (174a) display a falling pitch, whereas that in (174b) shows rising pitch. The pitch contour in (174c) is level.

7.4 Musical Pitch to Linguistic Tone Mapping

This section discusses the mapping from the musical pitch to the linguistic tone, and children’s association of the words or phrases to their mental lexicon. I propose a mapping model in (175).

(175) Musical pitch to linguistic tone mapping: associated terms



Two types of operations in term association are observed. Some of the children unexpectedly produce an initial output and then correct that to an actual word, which is referred to the type 1 operation. In the type 1 operation, term association fails at the lexical level, where unknown initial outputs are produced, but succeeds at the postlexical level, where actual words or phrases are produced. In the type 2 operation, term association succeeds at the lexical level, and actual words or phrases are directly produced. The children can mostly produce the actual terms directly, but there are some cases where they produce initial outputs first. The table in (176) presents some statistics.

(176) Statistics of the associated terms

	Tokens	Percentage
Type 1	8	3.7%
Type 2	209	96.3%
Total	217	100%

There are 217 tokens of associated terms in the database. Eight (3.7%) of them are produced through certain initial outputs to the final outputs, which pertains to the type 1 operation. 209 (96.3%) of them are directly produced as the final outputs, which pertains to the type 2 operation. In spite of the relative small number of the tokens in the type 1 operation, the fact that the tonal contour of the initial output is faithful to the singing pitch contour makes possible the hypothesis that the initial output is identical to the linguistic input. The existence of the initial output also supports the fact that there is perception in the music-to-language mapping.

However, some of the familiar words are not associated by the children. The unassociated terms will be discussed in Chapter 8.

7.4.1 Type 1 Operation

Based on the singing, some of the children may produce an initial output and then modify it into the final output. The following are some examples

(177) Examples of the type 1 operation

Musical output	**Initial tonal output**	**Final tonal output**
a. *bing*$^{\#}$**5** *gan*$^{\#}$**1** 'cookie'　餅乾	*bing*55 *gan*21	*bing*21 *gan*55 'cookie'　餅乾
b. *tai*$^{\#}$**1** *yang*$^{\#}$**5** 'sun'　太陽	*tai*21 *yang*55	*tai*53 *yang*35 'sun'　太陽

DOI:10.6814/NCCU202001428

c. *shu*$^{\#}$**1** *bao*$^{\#}$**5** 'bookbag' 書包	*shu*21 *bao*55	*shu*55 *bao*55 'bookbag' 書包
d. *shi*$^{\#}$**5** *jie*$^{\#}$**5** 'world' 世界	*shi*55 *je*55	*shi*53 *jie*53 'world' 世界

The initial outputs are faithful to the musical output, whereas the final outputs are the words or phrases that the children associate to their mental lexicon. In (177a), the musical output is *bing*$^{\#}$**5** *gan*$^{\#}$**1**, which is identical to the initial tonal output *bing*55 *gan*21. However, the tone of the final output, *bing*21 *gan*55 is totally different from *bing*$^{\#}$**5** *gan*$^{\#}$**1**. The fact that the initial tonal output is totally faithful to the musical pitch output supports the argument that the initial tonal output is the same as the perceived linguistic tonal input. Consider the pitch distance in (178).

(178) Musical pitch distance and tonal distance

Musical output	**Linguistic tonal input**
a. *bing*$^{\#}$**5** *gan*$^{\#}$**1** 'cookie' 餅乾 Dist: 3.5	*bing*55 *gan*11 Dist: 4
b. *tai*$^{\#}$**1** *yang*$^{\#}$**5** 'sun' 太陽 Dist: 3.5	*tai*11 *yang*55 Dist: 4
c. *shu*$^{\#}$**1** *bao*$^{\#}$**5** 'bookbag' 書包 Dist: 3.5	*shu*11 *bao*55 Dist: 4
d. *shi*$^{\#}$**5** *jie*$^{\#}$**5** 'world' 世界 Dist: 0 Reg: [+upper]	*shi*55 *je*55 Dist: 0 Reg: [+upper]

Given the pitch distance mapping principle in (172), the musical pitch distance of *bing*$^{\#}$**5** *gan*$^{\#}$**1** in (178a), is *3.5* while the corresponding tonal distance of the linguistic tonal input, *bing*55 *gan*11, is *4*, where the musical pitch distance is preserved. The disyllabic tonal contour of the tonal input, *bing*55 *gan*11 is falling, which is identical to the pitch contour of the musical output. Since the children fail to associate the linguistic input to their mental lexicon at the lexical level, the linguistic tonal input is produced as the initial output, where 11 is adjusted as 21. The children then successfully associate the term to their mental lexicon at the postlexical level, where they produce the actual NP, *bing*35 *gan*55, ‘cookie’. The pitch distance in (178b-c) is preserved in the same way.

In (178d), the musical pitch is level. Mandarin has a high tone, 55, and a low tone, 21, which is like 11. The music register may affect the mapping of level tones. The music register of *shi*$^{\#}$**5** *je*$^{\#}$**5** is [+upper], which corresponds to the [+upper] tone register of *shi*55 *jei*55 according to the corresponding scale in (169). Since the children fail to associate the linguistic input *shi*55 *jei*55 to their mental lexicon at the lexical level, *shi*55 *jei*55 is produced as the initial output. At the postlexical level, the children successfully associate the initial output with a familiar term, *shi*53 *jei*53 ‘world’, and the final output is produced.

7.4.2 Type 2 Operation

In most cases, the children may directly produce the actual words as their final outputs, without producing the initial tonal outputs. Given the fact that the initial tonal output is faithful to the musical pitch output in the type 1 operation, we can posit a linguistic tonal input, which is faithful to the musical pitch output as well. I also posit the faithful linguistic tonal input for the type 2 operation. The following are some

DOI:10.6814/NCCU202001428

examples.

(179) Examples of the type 2 operation

Musical output	**Linguistic tonal input**	**Final tonal output**
a. *bing*$^{\#}$**5** *gan*$^{\#}$**1** ‘cookie’ 餅乾	*bing*55 *gan*11.	*bing*35 *gan*55 ‘cookie’ 餅乾
b. *tai***1** *yang***5** ‘sun’ 太陽	*tai*11 *yang*55.	*tai*53 *yang*35 ‘sun’ 太陽
c. *shu***1** *bao***5** ‘bookbag’書包	*shu*11 *bao*55.	*shu*55 *bao*55 ‘bookbag’ 書包
d. *shi***5** *jie***5** ‘world’ 世界	*shi*55 *jie*55.	*shi*53 *jie*53 ‘world’ 世界
e. *la***5** *ba***5** ‘raise’ 拉拔	*la*55 *ba*55.	*la*21 *ba*55 ‘trumpet’ 喇叭

At the lexical level, the linguistic input is successfully associated to an actual word, whose tone is often not identical to the linguistic input. The final outputs are words that the children are familiar with. The associated words are sometimes not the same as the original singing output. For example, the singing output in (179e) is *la***5** *ba***5** ‘raise’ (拉拔), but the associated word is *la*21 *ba*55 ‘trumpet’ (喇叭).

7.5 M_O-to-L_I Mapping: Pitch Perception

In both types of operations discussed above, I posit a linguistic input that corresponds to the musical pitch output. This section discusses the perception grammar in the mapping from the musical output (M_O) to the linguistic input (L_I) through three kinds of disyllabic pitch contours, falling, rising and level.

7.5.1 Falling Contour

The disyllabic tonal contour of the linguistic input is basically faithful to the musical pitch output. For example, the falling musical pitch **$^{\#}$5-$^{\#}$1** may be mapped as the tonal strings 55-11 or 55-33. Four constraints are given in (180-183).

(180) ID-CTR:

Assign one violation mark for every pitch contour in S_2 that is not identical to its correspondent in S_1.

(181) MAX-DIST:

Assign one violation mark for every pitch distance in S_2 that does not have a correspondent in S_1.

(182) No-DIST=1:

Assign one violation mark for every pair of linguistic tones whose distance is equal to 1.

(183) No-DIST=4:

Assign one violation mark for every pair of linguistic tones whose distance is equal to 4.

In (180), S_2 can refer to the linguistic input, while S_1 to the musical output. The constraint, ID-CTR, in (180) is proposed to ensure the faithful correspondence of the pitch contour. The mapping to the linguistic input is also affected by the pitch distance of the musical output; specifically, the pitch distance in the musical output is preserved in the linguistic input, a fact that can be governed by the constraint MAX-DIST in (181). The constraints No-DIST=1 in (182) and No-DIST=4 in (183) indicate markedness on the tonal distance of the disyllabic output tones. ID-CTR, No-DIST=1, and MAX-DIST must be ranked above No-DIST=4 to preserve the contour shape and the pitch distance but prevent mappings like [**$^{\#}$5-$^{\#}$1** → 55 44], etc. The partial constraint ranking is given in (184).

(184) M_O-to-L_I partial constraint ranking:

ID-CTR, No-DIST=1, MAX-DIST >> No-DIST=4

DOI:10.6814/NCCU202001428

Tableau (185) shows the candidate evaluation.

(185) M_O-to-L_I mapping

Musical output: *tian*$^{\#}$**2**-*jia*$^{\#}$**1** ‘heaven’ 天家[6]

Musical pitch distance: 1

	ID-CTR	NO-DIST=1	MAX-DIST	NO-DIST=4
☞ a. *tian*55 *jia*33 Tonal dist. 2				
☞ b. *tian*33 *jia*11 Tonal dist. 2				
c. *tian*33 *jia*55 Tonal dist. 2	*!			
d. *tian*55 *jia*44 Tonal dist. 1		* (!)	* (!)	
e. *tian*55 *jia*11 Tonal dist. 4			*!	*

Tableau (185) shows that the musical output pitch *tian*$^{\#}$**2** *jia*$^{\#}$**1** can be mapped into two possible linguistic tonal inputs, *tian*55 *jia*33 and *tian*33 *jia*11. (185c) is ruled out by ID-CTR because its musical pitch contour is rising whereas the linguistic input contour is falling. The tone distance in (185d) is ruled out by NO-DIST=1 and MAX-DIST. According to the pitch distance mapping principle in (172), if the pitch distance in music is between 0.5 and 3, its corresponding tonal distance is 2. In this case, the tonal distance in (185e) is not preserved, and thus is ruled out by MAX-DIST.

---

[6] All of the children informants are attending Sunday school at church. Therefore, *tian*55 *jia*55, ‘heaven,’ is a familiar word to them.

DOI:10.6814/NCCU202001428

7.5.2 Rising contour

The mapping of the rising pitch contour in music is also preserved in the linguistic tonal input. The tableau in (186) serves as an example.

(186) M_O-to-L_I mapping

Musical output: *shi*$^{\#}$**1** *tou*$^{\#}$**2** ‘stone’ 石頭

Musical pitch distance: 1

	ID-CTR	NO-DIST=1	MAX-DIST	NO-DIST=4
☞ a. *shi*11 *tou*33 Tonal dist. 2				
☞ b. *shi*33 *tou*55 Tonal dist. 2				
c. *shi*55 *tou*44 Tonal dist. 1	*(!)	*(!)	*(!)	
d. *shi*33 *tou*44 Tonal dist. 1		*(!)	*(!)	
e. *shi*11 *tou*55 Tonal dist. 4			*!	*

Tableau (186) shows the rising musical pitch contour and its corresponding linguistic input. The linguistic tonal contour of (186c) is falling, and thus is ruled out by the dominant constraint ID-CTR. (186d) is ruled out by NO-DIST=1 and MAX-DIST. Again, based on the pitch distance mapping principle in (172), if the pitch distance in music is between 0.5 and 3, its corresponding tonal distance is 2. The pitch distance in (186e) is not preserved, fatally violating MAX-DIST. As a result, both (186a) and (186b) emerge.

Tableau (187) shows another example of the preservation of the rising pitch contour and the musical pitch distance in the linguistic tonal input.

(187) $\mathrm{M_O}$-to-$\mathrm{L_I}$ mapping

Musical output: *tai***1** *yang***5** ‘sun’ 太陽

Musical pitch distance: 3.5

	ID-CTR	NO-DIST=1	MAX-DIST	NO-DIST=4
☞ a. *tai*11 *yang*55 Tonal dist. 4				*
b. *tai*55 *yang*11 Tonal dist. 4	*!			*
c. *tai*11 *yang*22 Tonal dist. 1		*(!)	*(!)	
d. *tai*11 *yang*33 Tonal dist. 2			*!	

ID-CTR rules out (187b), whose tonal contour is falling. (187c) is ruled out by NO-DIST=1 and MAX-DIST since the tonal distance is 1 and the pitch distance is not preserved. The pitch distance in (187d) is also not preserved, fatally violating MAX-DIST. Eventually, (187a) is selected as the optimal output.

7.5.3 Level contour

The register of the musical pitch may affect the perception of level contour. As posited in (169), the musical pitches range from **7** to $^{\#}$**1** pertains to the [-upper] register, that from **4** to $^{\#}$**i** pertains to the [+upper] register, but that from **2** to **3** is unspecified. The [+upper] and [-upper] registers in music correspond to the level tones. The unspecified pitch range can be mapped to either the [+upper] register or the [-upper] register in language. In terms of Yip (1989, 2007), the tonal range from 1 to 2 is of the [-upper] register, that from 4 to 5 is of the [+upper] register, while 3 is

DOI:10.6814/NCCU202001428

unspecified. In the M_O to L_I mapping, correspondence of register is observed, which can be governed by the constraint in (188).

(188) ID-REG:

Assign one violation mark for every register in S_2 that is not identical to its correspondent in S_1.

This constraint must be undominated as well, such that the register can be retained. The enriched constraint ranking is given in (189).

(189) M_O-to-L_I constraint ranking (enriched):

ID-REG, ID-CTR, NO-DIST=1, MAX-DIST >> NO-DIST=4

Take the familiar word, *shi*$^{\#}$**5** *jie*$^{\#}$**5**, 'world', for example. It is perceived as *shi*55 *jie*55. The musical output is of [+upper] register and the linguistic tonal input preserves the [+upper] register. The tableau in (190) illustrates the function of ID-REG.

(190) M_O-to-L_I perception grammar

Musical output: *shi*$^{\#}$**5** *jie*$^{\#}$**5** 世界 'world'

Register: [+upper][+upper]

Musical pitch distance: 0

	ID-REG	NODIST=1	MAX-DIST	ID-CTR	NODIST=4
☞ a. *shi*55 *jie*55 +U +U Tonal dist. 0					
☞ b. *shi*33 *jie*33 +U +U Tonal dist. 0					
c. *shi*11 *jie*11 -U -U	*!*				

DOI:10.6814/NCCU202001428

Tonal dist. 0					
d. *shi*33 *jie*55 +U +U Tonal dist. 2				*!	

Both (190a) and (190b) are selected as the optimal outputs since they preserve the [+upper] register, in conformity of ID-REG. (190c) is ruled out by ID-REG because the register is changed to [-upper]. The contour of (190d) is rising, which is different from the musical output, and thus is ruled out by ID-CTR.

7.6 L$_I$-to-L$_O$ Mapping: Type 1 Operation

As proposed in (175), the model of the musical pitch to linguistic tone mapping suggests two types of production operation. This section discusses the type 1 operation, in which term association fails at the lexical level and an initial output is produced. The initial output in turn is taken as the postlexical input, which then is successfully associated and yields the final output. Three constraints are relevant, as in (191-193)

(191) ASSOC-TERM:

Assign one violation mark for every term that is not associated to the mental lexicon.

(192) TONOTACTICS:

Assign one violation mark for every tone such as 33 and 44 that does not exist in Mandarin.

DOI:10.6814/NCCU202001428

(193) ID-T:

Assign one violation mark for every output tone that is not identical to its correspondent in the input.

AS in (191), ASSOC-TERM is developed from the notion of Boersma's (2001) constraint, *LEX, which is defined as not recognizing an utterance as a lexical item in the comprehension grammar. In this research, I consider that ASSOC-TERM operates in the production grammar. The tonal output of an associated term is often unfaithful to the input, and thus ASSOC-TERM is top-ranked. As in (192), TONOTACTICS indicates that there are no tones like 33 and 44 in Mandarin, and is ranked above ID-T, which requires tonal faithfulness between the linguistic input and the linguistic output. The constraint ranking is posited in (194).

(194) L_I-to-L_O constraint ranking:

ASSOC-TERM, TONOTACTICS >> ID-T

Tableau (195) shows how the constraint ranking works.

(195) L_I-to-L_O mapping: initial output

Linguistic input: *tai*11 *yang*55 'sun' 太陽

Initial output: *tai*11[7] *yang*55

	ASSOC-TERM	TONOTACTICS	ID-T
☞ a. *tai*11 *yang*55	*		*
b. *tai*33 *yang*55	*	*!	

---

[7] The third tone in Mandarin is often perceived as a low tone. Native Mandarin speakers do not differentiate between 11 and 21.

In tableau (195), the children fail to associate the linguistic input with the actual term at the lexical level such both candidates violate ASSOC-TERM. TONOTACTICS favors (195a) over (195b), so *tai*11 *yang*55 emerges as the initial output, which then serves as the postlexical input and activates ASSOC-TERM again, as in tableau (196).

(196) L_I-to-L_O mapping: final output

Initial output/postlexical input: *tai*11 *yang*55

Final output: *tai*53 *yang*35

	ASSOC-TERM	TONOTACTICS	ID-T
☞ a. *tai*53 *yang*35 太陽 'sun'			**
b. *tai*11 *yang*55	*!		
c. *tai*33 *yang*55	* (!)	* (!)	*

At the postlexical level, ASSOC-TERM successfully selects (196a) as the optimal output, where the NP, *tai*53 *yang*35, emerges. The two-step operation of ASSOC-TERM is developed from Kiparsky's (2010, 2015) Stratal OT, which considers that each linguistic level has an individual grammar. In this spirit, I observe that ASSOC-TERM operate separately at lexical and postlexical levels.

7.7 L_I-to-L_O Mapping: Type 2 Operation

In the type 2 operation, the linguistic input is successfully associated to a familiar term at the lexical level, and thus an actual word is directly produced. Tableau (197) demonstrates the production grammar from the linguistic input to the linguistic final output.

(197) L_I-to-L_O mapping: final output

Linguistic input: *tai*11 *yang*55

Final input: *tai*53 *yang*35

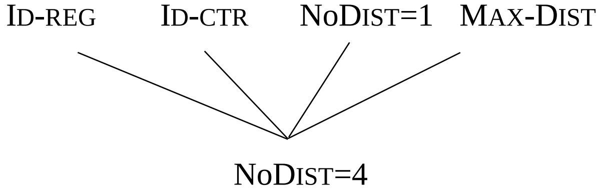
	Assoc-term	Tonotactics	Id-T
☞ a. *tai*53 *yang*35 太陽 'sun'			**
b. *tai*11 *yang*55	*!		
c. *tai*33 *yang*55	* (!)	* (!)	*

(197a) is the only term that is associated and thus is selected as the optimal output, regardless that fact that it violates Id-T. (197b-c) are unassociated terms, and thus are ruled out by Assoc-term.

7.8 Summary

This chapter has examined closely how the children map the singing outputs to their linguistic outputs. I have proposed a model, as in (175) that distinguishes the perception grammar from the production grammar. In the perception grammar, I have posited principles to account for the correspondence of musical pitch, register, and distance in the M_O-to-L_I mapping, which is governed by a set of constraints, as summarized by the Hasse diagram in (198).

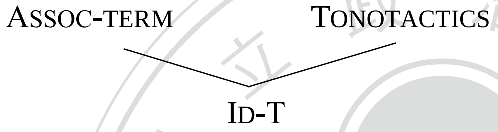
(198) M_O-to-L_I perception grammar: pitch-tone correspondence



DOI:10.6814/NCCU202001428

In the production grammar, two types of operations are observed. In the type 1 operation, term association fails at the lexical level and an initial output is produced. The initial output then serves as the postlexical input, which is successfully associated and yields the final output. In the type 2 operation, the linguistic input is successfully associated to a familiar term at the lexical level, and thus an actual word is directly produced. A set of constraints are posited to govern the production grammar, which are ranked as in (199).

(199) L_I-to-L_O production grammar: term association



To summarize, the M_O-to-L_I perception grammar values the faithfulness relation between the musical pitch and the linguistic tone. The L_I-to-L_O production grammar requires term association at the lexical and postlexical levels.

Chapter 8

Music-to-Language Mapping: Tone and Markedness

8.1 Introduction

This chapter continues to examine how children map the musical pitches to linguistic tones in the singing of Mandarin songs, with a focus on the failure of term association. Precisely, I raise the question that when the children fail to associate the tonal strings to the words or phrases in their mental lexicon, how the final linguistic output is produced. In particular, I look at the role of markedness constraints in the tonal production.

8.2 Data Design

The data collected from the preschool children also contain words or phrases that the children are not familiar to, such as technical terms, idioms, and other less frequent words. Some examples are given in (200).

(200) Examples of the unfamiliar disyllabic words/phrases

Unfamiliar words/phrases
*shu*35 *jia*53 'ransom' 贖價
*ju*53 *fa21* 'syntax' 句法
*li*35 *li*35 'flourish' 離離
*you*53 *you*21 'and have' 又有

The syllables are sung by a female native Mandarin speaker with electric keyboard. Each syllable is set to one musical pitch, whose duration is one beat. The singer is

DOI:10.6814/NCCU202001428

well-trained in music and singing so that the musical pitches are produced accurately. The children are asked to tell the possible lyrics that are sung by the singer. The unfamiliar words and phrases are sung with different musical pitch contours. The following are some examples.

(201) Examples of the singing words/phrases

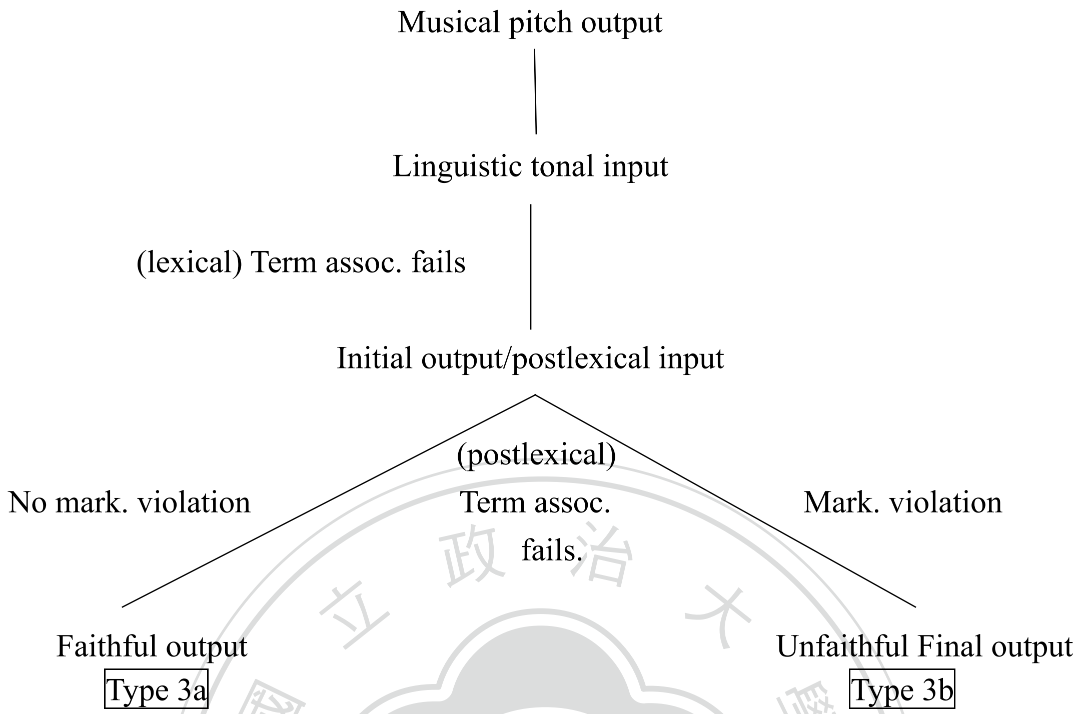
	Music contour	Singing output
a.	Falling	*yin*6 *shi*$^{\#}$**4** ‘diet’ 飲食 ↘
b.	Rising	ji7 jiao$^{\#}$**1** ‘haggle’ 計較 ↗
c.	Level	*li*5 *li* 5 ‘flourish’ 離離 →

The musical output in (201a) displays a falling contour, while that in (201b) shows rising contour. The pitch contour in (201c) is level.

8.3 Musical Pitch to Linguistic Tone Mapping

This section discusses the mapping from the musical pitch to the linguistic tone in unfamiliar terms, without association of the perceived tonal strings to their mental lexicon. In this case, an initial output is produced. The fact that the initial tonal output is totally faithful to the musical pitch output supports the argument that the initial tonal output is the same as the perceived linguistic tonal input. The emergence of the final linguistic output is then keyed to the interaction of faithfulness and markedness constraints. I have proposed the model in (202) to account for the unassociated term mapping.

(202) Musical pitch to linguistic tone mapping: unassociated terms



In the failure of term association at the lexical level, an initial output is produced and serves as the postlexical input. The model in (202) indicates that when term association also fails at the postlexical level, two types of production operations are in order, here referred to as type 3a and type 3b. In the type 3a operation, the postlexical input is faithfully mapped as the final output, with no crucial violation of any markedness constraint. In the type 3b operation, the postlexical input crucially violates some markedness constraints and is subject to segmental adjustments, such that the emerged final output is not faithful to the postlexical input. The table in (203) presents some statistics.

(203) Statistics of the unassociated terms

	Tokens	Percentage
Type 3a	105	73.9%
Type 3b	37	26.1%
Total	142	100%

There are 142 tokens of the unassociated terms in the database. 105 (73.9%) of them are produced faithfully to the initial output (postlexical input), which pertains to the type 3a operation. 37 (26.1%) of them are unfaithful to the initial output (postlexical input), which pertains to the type 3b operation.

8.3.1 Type 3a Operation

Based on the singing, some of the children may produce an initial output which faithfully yields the final output. The following are some examples.

(204) Examples of the type 3a operation

Musical output	**Initial tonal output**	**Final tonal output**
a. *si***6** *zhou***6** ‘surroundings’ 四周	*si*55 *zhou*55	*si*55 *zhou*55
b. *shi***6** *shi***6** ‘constantly’ 時時	*shi*55 *shi*55	*shi*55 *shi*55
c. *yin***6** *shi*$^{\#}$**4** ‘diet’ 飲食	*yin*55 *shi*33	*yin*55 *shi*35
d. *ji***7** *jiao*$^{\#}$**1** ‘haggle’ 計較	*ji*33 *jiao*55	*ji*35 *jiao*55
e. *yu***1** *san***5** ‘umbrella’ 雨傘	*yu*11 *san*55	*yu*21 *san*55

DOI:10.6814/NCCU202001428

In (204a), the musical output, *si***6** *zhou***6**, which is identical to the initial tonal output *si*55 *zhou*55, is then faithfully mapped as the final output. It should be noted that in (204f), *yu***1** *san***5**, 'umbrella', is supposed to be a word that children recognized. However, since they fail to associate *yu***1** *san***5** to their mental lexicon, it is regarded as an unfamiliar word.

Consider the pitch distance in (205).

(205) Musical pitch distance and tonal distance

Musical output	**Linguistic tonal input**
a. *si***6** *zhou***6** 'surroundings' 四周 Dist: 0 Reg: [+upper]	*si*55 *zhou*55 Dist: 0 Reg: [+upper]
b. *shi***6** *shi***6** 'constantly' 時時 Dist: 0 Reg: [+upper]	*shi*55 *shi*55 Dist: 0 Reg: [+upper]
c. *yin***6** *shi***#4** 'diet' 飲食 Dist: 1.5	*yin*55 *shi*33 Dist: 2
d. *ji***7** *jiao***#1** 'haggle' 計較 Dist: 1	*ji*33 *jiao*55 Dist: 2
e. *yu***1** *san***5** 'umbrella' 雨傘 Dist: 3.5	*yu*11 *san*55 Dist: 4

In (205a-b), the musical pitch contour is level. Mandarin has a high tone, 55, and a low tone, 21, which is like 11. As discussed in Chapter 7, the music register may affect the mapping of level tones. In (205a), the music register of *si***6** *zhou***6** is [+upper], which corresponds to the [+upper] tone register of *si*55 *zhou*55. Since the children fail to associate the linguistic input, *si*55 *zhou*55, to their mental lexicon at the lexical and the postlexical level, *si*55 *zhou*55 is produced as the initial output and eventually as the final output. The mapping in (205b) is operated in the same way.

Given the pitch distance mapping principle in (172) of Chapter 7, the musical pitch distance of *yin***6** *shi*#**4** in (205c), is 1.5, and the corresponding tonal distance is 2, which is preserved in the linguistic tonal input, *yin*55 *shi*33. The disyllabic tonal contour of the tonal input, *yin*55 *shi*33, is falling, which is identical to the pitch contour of the musical output. Since the children fail to associate the linguistic input, *yin*55 *shi*33, to their mental lexicon at the lexical and postlexical level, *yin*55 *shi*33 is produced as the initial output, which is then faithfully mapped as the final output. The pitch distance in (205d, e) is preserved in the same way.

8.3.2 Type 3b Operation

When the children fail to associate the initial tonal outputs to their mental lexicon, the initial tonal outputs may incur crucial violations of some markedness constraints and be subject to surface well-formedness. Consequently, the final outputs may not be faithful to the tonal inputs. The unassociated terms with the unfaithful outputs are shown in table (206).

(206) Examples of the type 3b operation

Musical output	**Linguistic tonal input**	**Final tonal output**
a. *you***1** *you***1** ‘and have’ 又有 Dist: 0 Reg: [-upper]	*you*11 *you*11 Dist: 0 Reg: [-upper]	*you*35 *you*21
b. *li***1** *li***1** ‘flourish’ 離離 Dist: 0 Reg: [-upper]	*li*11 *li*11 Dist: 0 Reg: [-upper]	*li*35 *li*21

c. *mei*$^{\#}$**4** *mao*$^{\#}$**4** ‘eyebrows’ 眉毛 Dist: 0 Reg: [+upper]	*mei*55 *mao*55 Dist: 0 Reg: [+upper]	*mei*35 *mao*55

In (206a), the music register of *you***1** *you***1** is [-upper], which corresponds to the [-upper] tonal register of the linguistic input, *you*11 *you*11. However, it violates the OCP (Leben 1973, Goldsmith 1976), which disallows adjacent low tones in Mandarin. Consequently, the final output, *you*35 *you*21 is not faithful to the tonal input. (206b) is operated in the same way. In (206c), *mei*$^{\#}$**4** *mao*$^{\#}$**4** ‘eyebrows’, is supposed to be a word that children are familiar with. Nevertheless, since *mei*$^{\#}$**4** *mao*$^{\#}$**4** fails to be associated to the children’s mental lexicon, it is regarded as an unfamiliar word. In (206c), the music register of *mei*$^{\#}$**4** mao$^{\#}$**4** is [+upper], which corresponds to the [+upper] tone register of *mei*55 *mao*55. However, the syllable-tone combination of *mei*55 does not exist in Mandarin, and is thus excluded. Consequently, the final output, *mei*35 *mao*55 is not faithful to the tonal input.

8.4 M_O-to-L_I Mapping: Pitch Perception

The perception grammar that maps the musical output (M_O) to the linguistic input (L_I) is the same as that discussed in section 7.5. A falling musical pitch like **6-**$^{\#}$**4** may be mapped as the tonal strings 55-33 or 33-11; when the musical pitch distance is 1.5, its corresponding tonal distance is 2, and the pitch distance is preserved. A rising musical pitch like **7**-$^{\#}$**1** may be mapped as the tonal strings 11-33 or 33-55; when the musical pitch distance is 1, its corresponding tone distance is 2, and this pitch distance is preserved. A level musical pitch like **4-4** may be mapped as the tonal strings 55-55 or 33-33; when the musical pitch distance is 0, its corresponding tone distance is also

DOI:10.6814/NCCU202001428

0, and this pitch distance is preserved. The constraint ranking that governs the M_O-to-L_I mapping is proposed in (189) of Chapter 7, and reproduced in (207).

(207) M_O-to-L_I constraint ranking

ID-REG, ID-CTR, NO-DIST=1, MAX-DIST >> NO-DIST=4

ID-CTR requires faithful correspondence of pitch contour, and ID-REG requires faithful correspondence of register. MAX-DIST requires preservation of pitch distance. NO-DIST=1 and NO-DIST=4 lay down restrictions on the shape of the disyllabic contour. ID-REG, ID-CTR, NO-DIST=1 and MAX-DIST must be ranked above NO-DIST=4 to preserve the contour shape and the register, but to avoid mappings like [**$^{\#}$6-$^{\#}$1** → 55-44], [**1-1** → 55-55],etc.

8.5 L_I-to-L_O Mapping: Type 3a Operation

As proposed in (202), the schema of musical pitch to linguistic tone suggests two types of production operation. This section discusses type 3a operation, in which term association fails at the lexical level and postlexical level while the final output is faithful to the linguistic tonal input. Two constraints are relevant, ASSOC-TERM and ID-T. The partial constraint ranking is posited in (208).

(208) L_I-to-L_O partial constraint ranking:

ASSOC-TERM >> ID-T

Tableau (209) shows how these constraint ranking works.

(209) L_I-to-L_O production grammar: faithful output

Linguistic input: *yin*55 *shi*33 ‘diet’ 飲食

Final output: *yin*55 *shi*35

	ASSOC-TERM	ID-T
☞ a. *yin*55 *shi*35	*	*
b. *yin*33 *shi*55	*	*!*
c. *yin*35 *shi*55	*	*!*

In tableau (209), the terms are unassociated at either lexical or postlexical level, and the constraint ASSOC-TERM is violated by each of the candidates. The candidate in (209a) wins over the others by one less violation of ID-T.

8.6 L_I-to-L_O Mapping: Type 3b Operation

In the type 3b operation, the final output is unfaithful to the linguistic input. The markedness constraints rule out the faithful output and select the output that conforms to surface well-formedness. Three constraints are relevant for this mapping, as in (210-212).

(210) OCP-LOW:

Assign one violation mark for every pairs of low tones.

(211) TONOTACTICS:

Assign one violation mark for every tone such as 33 and 44 that does not exist in Mandarin.

(212) ID-T-R:

Assign one violation mark for every rightmost tone that is not identical to its correspondent in the input.

DOI:10.6814/NCCU202001428

OCP-Low prohibits consecutive low tones, and forces one of the tones to change. Id-T-R preserves the right tone and allows the left tone to change. The constraint Tonotactics bans any tone that does not exist in Mandarin, and must be ranked higher than Id-T. The enriched constraint ranking is posited in (213).

(213) L_I-to-L_O constraint ranking (enriched):
Assoc-Term, Tonotactics, OCP-Low >> Id-T, Id-T-R;

Tableau (214) shows how this constraint works.

(214) L_I-to-L_O production grammar: unfaithful output
Initial output/Postlexical input: *li*11 *li*11 ‘flourish’ 離離
Final output: *li*35 *li*11

	Assoc-term	Tonotactics	OCP-Low	Id-T	Id-T-R
☞ a. *li*35 *li*11	*			*	
b. *li*11 *li*11	*		*!		
c. *li*11 *li*35	*			*	*!
d. *li*33 *li*11	*	*!		*	

The markedness constraints rule out (214b, d) respectively, while (214c) is ruled out by the faithfulness constraint, Id-T-R. As a consequence, (214a) emerges as the optimal output, where *li*35 *li*21, is different from the postlexical input, *li*11 *li*11.

Another example lies in the phonotactics. For instance, the musical output, *mei***5** *mao***5**, perceived as *mei55 mao55* in the linguistic input, which, in the failure of lexical term association, is faithfully mapped as the initial output, which in turns serves as the postlexical input. In the failure of postlexical term association, the postlexical input, *mei55 mao55,* is eventually *mei35 mao55*, which is an unfaithful output. This unfaithful mapping is attributed to the fact that the syllable-tone

DOI:10.6814/NCCU202001428

combination of *mei55* is illegal in the Mandarin output. The constraint PHONOTACTICS, as in (215), is thus needed for the following discussions.

(215) PHONOTACTICS:

Assign one violation mark for every illegal syllable-tone combination in the output.

PHONOTACTICS should be undominated, and the partial constraint ranking is expended as (216).

(216) L_I-to-L_O constraint ranking (enriched):

ASSOC-TERM, TONOTACTICS, PHONOTACTICS, OCP-LOW >> ID-T, ID-T-R

Tableau (217) demonstrates the constraint interaction.

(217) L_I-to-L_O production grammar:

Initial output/Postlexical input: *mei*55 *mao*55 ‘eyebrows’ 眉毛

Final output: *mei*35 *mao*55

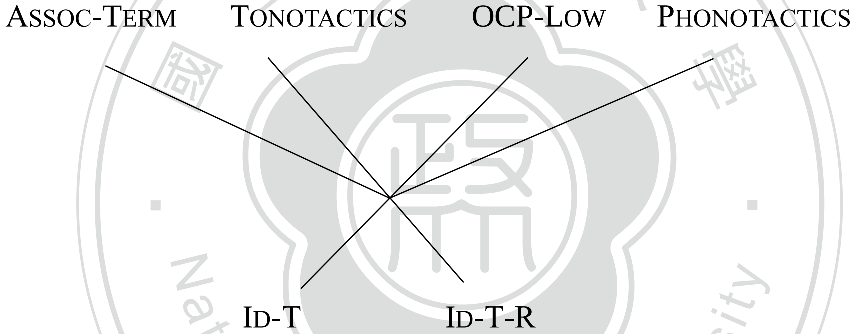
	ASSOC- TERM	TONOTACTICS	PHONOTACTICS	ID-T
☞ a. *mei*35 *mao*55	*			*
b. *mei*55 *mao*55	*		*!	

Since the children fail to associate the term to their mental lexicon, the constraint ASSOC-term is violated by both candidates. The syllable-tone combination of *mei55* in (217b) violates PHONOTACTICS. (217a) thus surfaces and an unfaithful output is selected.

8.7 Summary

This chapter has addressed a third type of operation in the music-to-language mapping. The perception grammar that maps M_O to L_I is the same as that in type 1 and type 2, as discussed in Chapter 7. This third type of operation further consists of two subtypes, type 3a and type 3b, in both of which term association fails at the lexical and postlexical levels. The mapping from the postlexical input to the final output is keyed to the interactions of faithfulness and markedness constraints. I have posited a set of constraints, which are ranked as in (218).

(218) L_I-to-L_O production grammar: term association



The M_O-to-L_I perception grammar values the faithfulness relation between the musical pitch and the linguistic tone. The L_I-to-L_O production grammar requires term association at the lexical and postlexical levels. If term association fails, the final linguistic output will be faithful to the postlexical input; however, when surface markedness constraints are violated, an unfaithful output will emerge.

Chapter 9

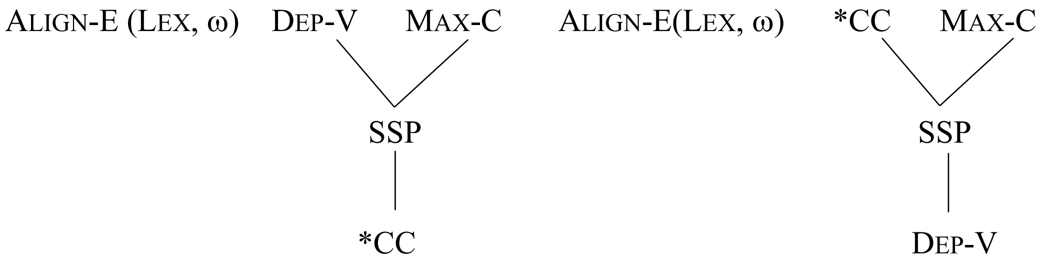
Conclusion and Further Issue

9.1 Conclusion

This study probes into the connection between language and music through the perception and production grammar in English and Mandarin songs. The interaction between language and music is revealed in three aspects, namely, segment, rhythm, and tone.

Chapters 3 and 4 compare the linguistic mapping and the language-to-music mapping in onset and coda clusters through Mandarin accented English. The linguistic mapping in onset clusters and coda clusters involves the interaction between segmental faithfulness and markedness, as well as prosodic alignment, which yields two output variants: the monosyllabic L_{O1} and the disyllabic L_{O2}. At this stage, the SSP is not crucial in either constraint ranking. Consider the Hasse diagrams reproduced in (219).

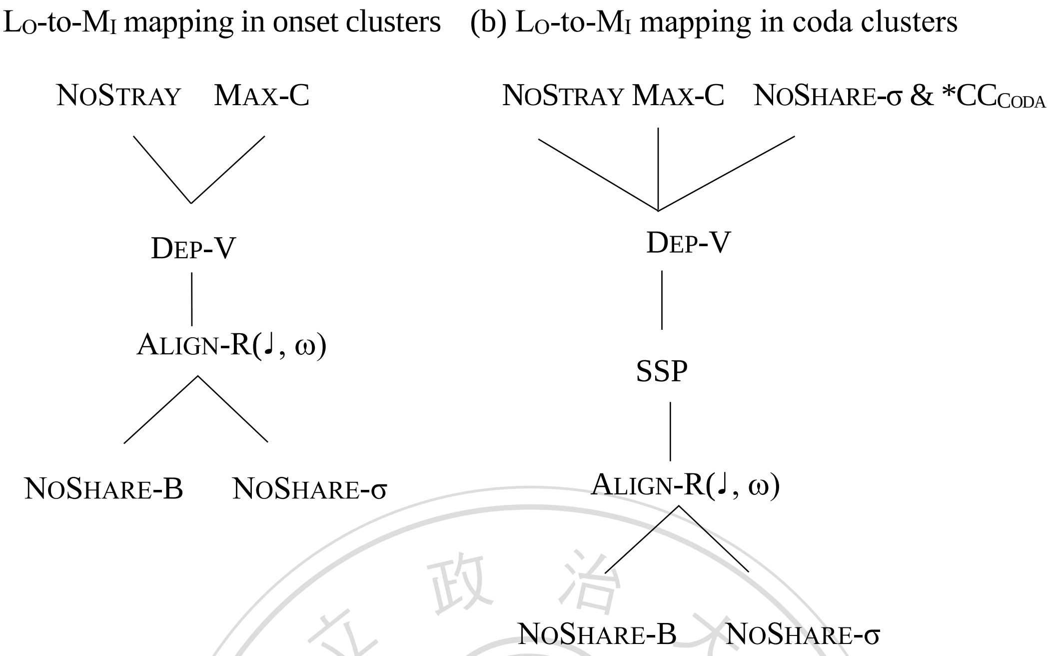
(219) L_I-to-L_O mapping in onset and coda clusters


(a) L_I-to-L_{O1} mapping: (σ)

(b) L_I-to-L_{O2} mapping: (σσ)

(220a) exhibits the L_O-to-M_I perception grammar in onset clusters, while (220b) presents that in coda clusters.

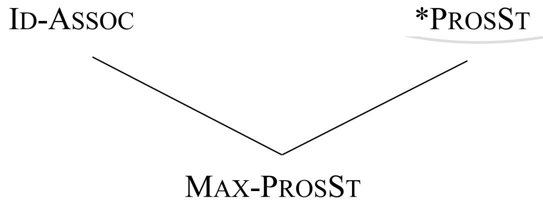
(220)


(a) L_O-to-M_I mapping in onset clusters (b) L_O-to-M_I mapping in coda clusters

In (220b), the conjoined constraint, NoShare-σ & *CC$_{CODA}$, is dominant and SSP is observed to affect the insertion of vowels in singing.

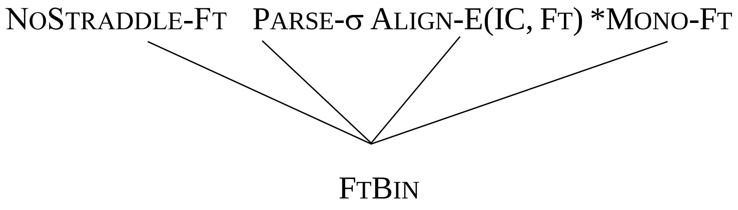
The M_I-to-M_O production grammar in onset and coda clusters requires faithfulness of syllable-beat association and the removal of prosodic structure in the musical output as reproduced in the Hasse diagram in (221).

(221) M_I-to-M_O production grammar



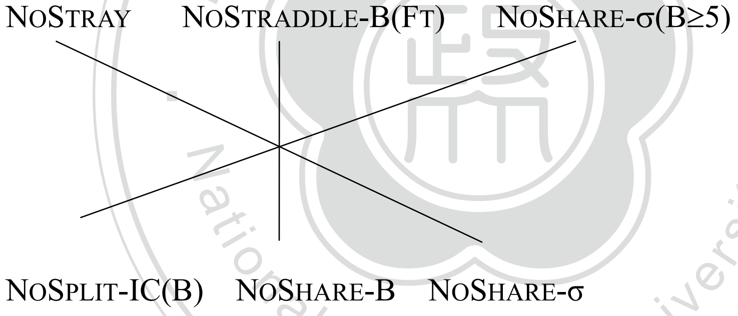
Chapters 5 and 6 examine the language-to-music mapping with the focus on the composing of Mandarin children's song. In Chapter 5, I have examined how the foot and IP structures affect the assignment of musical beats. The foot is formed based on the syntactic structures, in which immediate constituent (IC) and branching direction are keyed to deciding the output. The constraint ranking for foot formation is reproduced in the Hasse diagram in (222).

(222) L_I-to-L_O production grammar



The L_O-to-M_I perception grammar of the syllable-to-beat assignment of foot is exhibited in Hasse diagram reproduced in (223).

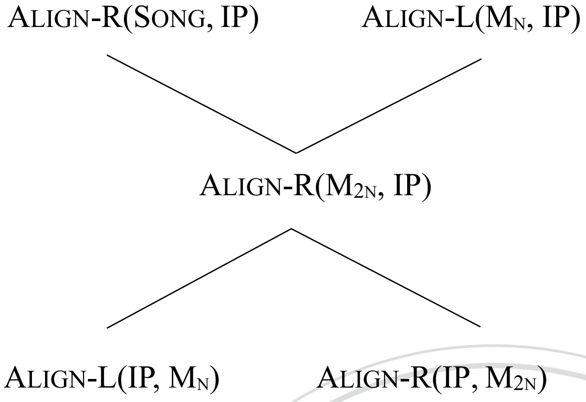
(223) L_O-to-M_I perception grammar



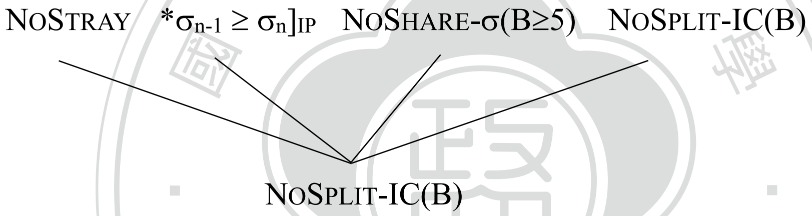
As in (223) and in (220) proposed previously, both NoShare-B and NoShare-σ are ranked at the bottom, as a syllable is often linked to multiple musical beats to prevent stray elements and vice versa.

Chapter 6 continues to examine the mapping of rhythmic structures focalizing on the intonational phrase (IP). The IPs in the lyric output are defined by the punctuation marks, which is captured by Align-R(Pm, IP). The mapping of the IP to the musical input is crucial to the measure-to-IP alignment and the IP-final musical lengthening. The L_O-to-M_I perception grammar is governed by two sets of constraints as reproduced in the Hasse diagrams in (224) and (225).

(224) L_O-to-M_I perception grammar: measure-to-IP alignment



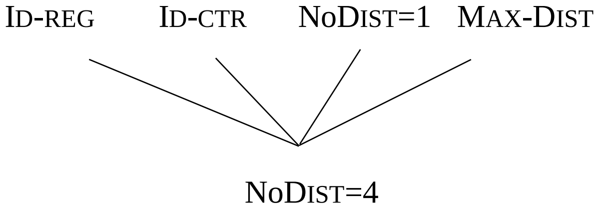
(225) L_O-to-M_I perception grammar: musical beat assignment



The mapping between the musical input (M_I) and the musical output (M_O) in Chapter 5-6 is governed by three constraints, Id-Assoc, *ProsSt and Max-ProsSt. As previously posited in Chapter 3-4, the musical output preserves the musical beat association but removes the structure of the prosodic structures.

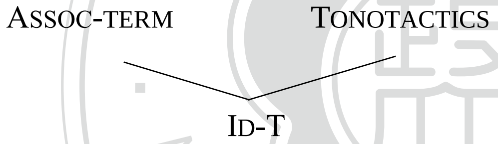
In Chapter 7-8, the music-to-language mapping is examined through children's perception of the singing of Mandarin songs. Singing words or phrases are mapped to the linguistic input forms with similar tonal values through the perception grammar. The perception grammar is accounted by the correspondence of musical pitch, register, and distance in the M_O-to-L_I mapping, which is governed by a set of constraints reproduced in the Hasse diagram in (226).

(226) M_O-to-L_I perception grammar: pitch-tone correspondence



The children associate what they have heard to their mental lexicon through the L_I-to-L_O production grammar. Four types of operations are observed. In the type 1 and type 2 operations, term association is achieved at some points and the final outputs are actual words or phrases as governed by the constraint ranking reproduced in the Hasse diagram in (227).

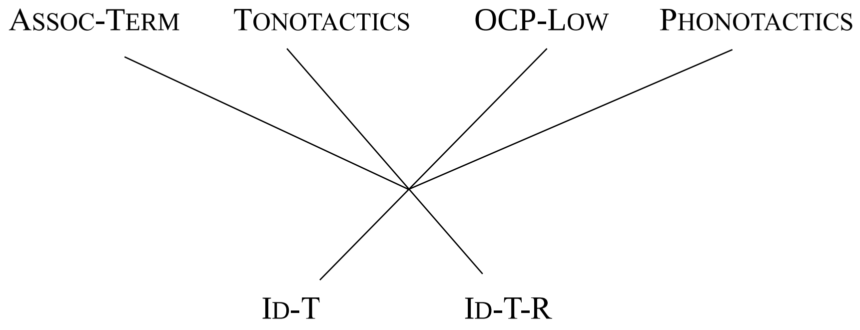
(227) L_I-to-L_O production grammar: term association



The top-ranked ASSOC-TERM selects the associated term whose tone may be different from the linguistic input.

In the type 3a and 3b operations, term association fails at both lexical and postlexical level. If term association fails, the final linguistic output will be faithful to the postlexical input; however, when surface markedness constraints are violated, an unfaithful output will emerge. The relevant Hasse diagram is reproduced in (228).

(228) L_I-to-L_O production grammar: term association



Taken together, Chapters 3 to 8 demonstrate segmental changes, rhythmic alignment, beat assignment, and linguistic tonal changes in the mapping between language and music. From the perspective of segment and rhythm, the language-to-music mapping shows that the perception grammar plays an important role. Tonal changes in the music-to-language mapping also support that there is an independent perception grammar.

9.2 Further issue

In Chapters 7 and 8, I have examined how disyllabic musical pitches are transformed into linguistic tones. Each syllable is associated with one musical beat with equal musical prominence. This study has also observed the musical pitch to linguistic tone mapping in multisyllabic phrases and sentences, where syllables are mapped to accented (>) or unaccented musical beats as exemplified in (229).

(229) 'His name is Jesus'

> (musical prominence)
*ta***5** *ming***7** *jiao***6** *ye***5**-*su***3** (musical pitch)
*ta*55 *ming*53 *jiao*53 *ye*55 *su*55 (linguistic tone)
he name call Jesus

In (229), the linguistic final output, *ta*55, *jiao*53, and *ye*55-*su*55 are terms that can be associated by the children, whereas, *ming*53 is an unassociated term. Since *ming*53 is the unassociated term, its linguistic tonal output is not influenced by lexicon association. However, why is *ming***7** perceived as *ming*53 instead of *ming*55, which is faithful to the singing output? It is observed that some of the children perceive the accented (>) singing output as the linguistic tone, 53. Future studies can examine how the pitches of the accented musical beats are perceived into linguistic tones.

DOI:10.6814/NCCU202001428

Bibliography

English References

Bermúdez-Otero, R. (1999). *Constraint interaction in language change: quantity in English and Germanic* (doctoral dissertation). University of Manchester, UK.

Bermúdez-Otero, R. (2007). Diachronic phonology. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 497-517). Cambridge, Mass.: Cambridge University Press.

Boersma, P. (2001). Phonology-semantics interaction in OT, and its acquisition. In R. Kirchner, W. Wikeley and J. Pater (Ed.), *Papers in experimental and theoretical linguistics*, *6*, 24-35. Edmonton: University of Alberta.

Booij, G. E. & Rubach, J. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry*, *18 (1),* 1-44.

Broselow, E., and Finer, D. (1991). Parameter setting gin second language phonology and syntax. *Second Language Research*, *7*, 35-59.

Carr, P. (1993). *Phonology*. New York: St. Martins Press.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24-27.

Chen, M. (1984). Unfolding Latent Principles of Literary Taste: Poetry as a Window onto Language. *Tsinghua Journal of Chinese Studies*, *16*, 203-240.

Downing, B. (1970). Syntactic Structure and Phonological Phrasing in English (doctoral dissertation). University of Texas at Austin, Austin.

Downing, B. (1973). Parenthesization Rules and Obligatory Phrasing. *Papers in Linguistic*, *6*, 108-128.

Goldsmith, J. A. (1976). *Autosegmental Phonology* (doctoral dissertation). MIT, Cambridge, Mass.

DOI:10.6814/NCCU202001428

Green, T. (1993). The conspiracy of completeness. Paper presented at Rutgers Optimality Workshop1.

Halle, J. & Lerdahl F. (1993). A generative textsetting model. *Current Musicology*, *55*, 3-23.

Hayes, B. (2005). Textsetting as constraint conflict. In J.-L. Aroui and A. Arleo (Eds.), *Towards a typology of poetic form* (pp. 43-61). Amsterdam: Elsevier.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Hsiao, Y. E. (1991). *Syntax, rhythm and tone: a triangular relationship*. Taipei: The Crane Publishing.

Hsiao, Y. E. (2006). The metrical structure of Changhua folk verse. *Concentric: Studies in Linguistics*, *32 (1)*, 1-29.

Hsiao, Y. E. (2007). The metrical structure of Taiwanese nursery rhymes: a corpus study. In D. L. Hoover and S. Lattig (Eds.), *Stylistics: prospect & retrospect* (pp. 85-106). Amsterdam: Rodopi.

Hsiao, Y. E. (2015). Tonal chain shifts in Taiwanese: a comparative markedness approach. In Y. E. Hsiao, L.-H. Wee (Eds.), *Capturing phonological shades within and across languages* (pp. 142-165). Newcastle upon Tyne: Cambridge Scholars Publishing.

Inkelas, S. (1989). *Prosodic constituency in the lexicon* (doctoral dissertation). Stanford University, Stanford.

Ito, J., Kubozono, H., Mester, A. & Tanaka, S. (2019). Kattobase: the linguistic structure of Japanese baseball chants. In K. Hout, A. Mai, A. McCollum, S. Rose, and M. Zaslansky (Eds.), *Proceedings of the 2018 Annual Meetings on Phonology.* Tokyo, Japan: Kurosio Publishing.

DOI:10.6814/NCCU202001428

Jespersen, O. (1904). *Lehrbuch der phonetik.* Leipzig: Teubner.

Kager, R. (1999). *Optimality theory*. New York: Cambridge University Press.

Kenstowicz, M. (2003). The role of perception in loanword phonology. A review of Les emprunts linguistiques d'origine européenne en Fon by Flavien Gbéto Köln: Rüdiger Köppe Verlag. 2000. *Studies in African Linguistics*, *32*, 95-112.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst, N. Smith (Eds.), *The structure of phonological representations (PartI)* (pp. 131-175). Dordrecht: Foris Publications.

Kiparsky, Paul. (2010). Reduplication in stratal OT. In L. Uyechi and L.-H. Wee (Eds.), *Reality exploration and discovery: pattern interaction in language & Life* (pp. 125-142.).

Leben, W. R. (1973). *Suprasegmental phonology* (doctoral dissertation). MIT, Cambridge, Mass.

Lerdahl, F. & Jackendoff, R. (1983). *A generative rheory of tonal music.* Cambridge, Mass.: MIT Press.

Lindblom, B. (1978). Final lengthening in speech and music. In E.Gårding, G. Bruce, and R. Bannert (Eds.), *Nordic prosody: papers from a symposium* (pp.85-102). Lund: Department of Linguistics, Lund University

Łubowicz, A. (2002). Derived environment effects in optimality theory. *Lingua*, *112 (4)*, 243-280

Łubowicz, A. (2005). Locality of conjunction. In J. D. Alderete, C. Hang and A. Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics* (pp. 254-262). Somerville: Cascadilla Proceedings Project.

McCarthy, J. J. and Prince, A. (1995). Faithfulness and reduplicative identity. In: B. Jill, L. Walsh Dickey, and S. Urbanczyk (Eds.), *University of Massachusetts*

DOI:10.6814/NCCU202001428

Occasional Papers in Linguistics 18 (pp.249-384). Amherst, MA: GLSA Publications.

McCarthy, J. J. and Prince, A. (1999). Faithfulness and identity in prosodic morphology. In R. Kager, H. van der Hulst, and W. Zonneveld (Eds.), The *Prosody–Morphology Interface* (pp. 218-309). Cambridge, Mass.: Cambridge University Press. [Excerpted in John J. McCarthy (Ed.) *Optimality Theory in Phonology: A Reader*, Malden, MA & Oxford: Blackwell (2004).]

Mohanan, K. P. (1982). Lexical Phonology (doctoral dissertation). MIT, Cambridge, Mass.

Morris, R. E. (2002). Coda obstruents and local constraint conjunction in North-Central Peninsular Spanish. In T. Satterfield, C. Tortora and D. Cresti (Eds.), *Current issues in Romance Languages* (pp. 207-224). The Netherlands: John Benjamins B.V.

Nespor, M., and Irene Vo. (1986). *Prosodic phonology* (doctoral dissertation). Dordrecht: Foris.

Prince, A., and Smolensky, P. (1993). Optimality theory: constraint interaction in generative grammar. Manuscript. Rutgers University.

Prince, A. and Smolensky, P. (2004). *Optimality theory: constraint interaction and generative grammar*. Blackwell Publishing.

Schellenberg, M. (2009). Singing in a tone language: Shona. In A. Ojo and L. Moshi (Eds.), *Proceedings of the 39th Annual Conference on African Linguistics* (pp. 137-144). Somerville, MA: Cascadilla Proceedings Project.

Schreuder, M. (2006). Prosodic Processes in Language and Music (doctoral dissertation). Research School of Behavioral and Cognitive Neurosciences (BCN), Groningen.

DOI:10.6814/NCCU202001428

Selkirk, E. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and M.-L. Kean (Eds.), *Juncture* (pp. 107-129). Saratoga, CA: Anma Libri.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Shih, C. L. (1986). *The Prosodic Domain of Tone Sandhi in Chinese*. San Diego: University of California dissertation.

Sievers, E. (1881). *Grundzüge der Phonetik*. Leipzig, Germany: Breitkopf and Hartel.

Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology*, *9*, 289-328.

Smolensky, P. (1995) On the internal structure of the constraint component CON of UG. Handout presented in the University of Arizona.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*, 720-731.

Sundberg, J and Lindblom, B. (1991). Generative theories for describing musical structure. In P. Howell, R. West, and I. Cross (Eds.), *Representing musical structure* (pp. 245-272). London: Academic Press.

Wee, L. H. (2002). Of chains and weights. Handout for seminar series Grammaticality of the Fittest, City University of Hong Kong.

Wee, L. H. (2007). Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics*, *35*, 128-144.

Wong, P. and Diehl, R. (2002). How can the lyrics of a song in a tone language be understood? *Psychology of Music*, *30*, 202-209.

Yip, M. (1989). *Hanyu fangyan gaiyao* [*An introduction to Chinese dialects*]. 2nd edn. Beijing: Wenzi Gaige Chubanshe.

DOI:10.6814/NCCU202001428

Yip, M. (1993). Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics*, *1 (1)*, 261-291.

Yip, M. (2007). *Tone*. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 229-252). Cambridge, Mass.: Cambridge University Press.

Chinese References

方金雅、蘇姿云 (Fang and Su 2005)。童謠教學對幼兒聲韻覺識影響之研究。*高雄師大學報：教育與社會科學類*，*19*，1-19。

李金青 (Li 2009)。*兒歌的韻律研究* (碩士論文)，國立臺東大學，臺東。

李紫蓉、楊麗鈴 (2007)。*手指遊戲動動兒歌-小猴子*。臺北：信誼基金出版社。

李紫蓉、朱雲嵩、林炳興、張硯田、蘇珊 (2011)。*早安晚安：給0~3歲的自理兒歌*。臺北：信誼基金出版社。

李紫蓉、鄭榮珍、張麗雪、蘇珊、張硯田、陳中申 (2014)。*小雨滴*。臺北：信誼基金出版社。

李紫蓉、林世仁、蘇珊 (2017)。*1234動動體操*。臺北：信誼基金出版社。

宋凱琳 (Sung 2012)。從優選理論看語言與音樂的互動：從詩經入樂談起。*華語文教學研究*，*9 (4)*，17-39。

林清山 (Lin 1968)。男女學前兒童語言發展各變數的比較研究。*心理與教育*，*2*，121-137。

林蕙珊 (Lin 2001)。從「優選理論」來談國語的三聲變調。*聲韻論叢*，*10*，497-524。

黃婷 (Huang 2007)。華語手指謠之韻律研究。*華語文教學研究*，*4 (2)*，29-42。

劉惠美、曹峰銘 (Liu and Tsao 2010)。*華語嬰幼兒溝通發展量表*。臺北：心理。

劉惠美、陳昱君(Liu and Chen 2015)。華語嬰幼兒表達性詞彙的語意內容及詞類組成之發展。*教育心理學報*，*47 (2)*，217-242。

DOI:10.6814/NCCU202001428

VITA

Education

2020 Ph.D. in Linguistics, National Chengchi University

2010 M.A. in Linguistics, National Chengchi University

2007 B.A. in Chinese, National Taiwan Normal University

Journal Paper

Ling, Wang-Chen. 2015. Circular tonal chain shifts in *Jiaoxian* compound words. Language and Linguistics (語言暨語言學). 16.6. [SSCI & AHCI].

Conference Papers

Ling, Wang-Chen. 2018. The Output-to-output Tonal Correspondence in Language and Music. International Conference on Tone and Intonation (TIE2018).

Ling, Wang-Chen. 2018. The Competition between Contour and Register Correspondence in Music-to-Language Perception: Evidence from Mandarin Child Songs. The 51st International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL51)

Ling, Wang-Chen. 2018. Child perception of Mandarin Songs: An Optimality Theory Analysis. The 7the International Conference on Phonology and Morphology.

Ling, Wang-Chen. 2017. The Conformity of Language and Melody in Mandarin Children's Song. 37th Annual Conference of the Poetics And Linguistics Association. West Chester University, Pennsylvania, USA.

Ling, Wang-Chen. 2016. Text-setting of Mandarin tone in children's song: a corpus analysis. The 28th annual North American Conference on Chinese Linguistics (NACCL-28). Brigham Young University, Utah, USA.

Ling, Wang-Chen. 2015. Circular Tonal Chain Shifts in Leling Pre-light Syllable Positions, National Conference on Linguistics (NCL 2014). National Chengchi University, Taipei.

Ling, Wang-Chen. 2014. Mandarin clipped words of government committees in Taiwan-An OT analysis. The 22nd Annual Conference of the IACL & the 26th North American Conference on Chinese Linguistics (IACL-22 & NACCL-26). University of Maryland, Maryland, USA.

DOI:10.6814/NCCU202001428

Ling, Wang-Chen. 2013. Disyllabic and trisyllabic tone sandhi in Leling. The 4th International Theoretical Phonology Conference (TPC-4). National Chengchi University, Taipei.

Ling, Wang-Chen. 2012. A corpus-based study on the semantic relatedness of the Mandarin polysemous word xia1, blind. NCL2012. Southern Taiwan University, Tainan.

Ling, Wang-Chen. 2012. The disyllabic tone sandhi in Rizhao. The 20th Annual Conference of the International Association of Chinese Linguistics (IACL-20). The Hong Kong Polytechnic University, Hong Kong.

Ling, Wang-Chen. 2009. The phonological-musical interaction in Chinese hymns: an OT analysis. Mao Kong Forum—The First International Graduate Student Conference on Modern Phonology. National Chengchi University, Taipei.

Ling, Wang-Chen. 2009. Textsetting strategies in Mandarin hymns. The 9th International Conference on Teaching of Chinese Language. Taipei

Ling, Wang-Chen. 2009. The phonological-musical strategies in textsetting of Chinese Bible verses: a corpus-based analysis. The 42nd International Conference on Sino-Tibetan Languages and Linguistics. Payap University, Chiang Mai, Thailand.

Ling, Wang-Chen. 2009. The Mandarin (reflexive) pronoun of SELF (zi4ji3) in bible: a corpus-based study. The Fifth Corpus Linguistics Conference. Liverpool University. Liverpool, England.

Ling, Wang-Chen. 2009. Lyric modifications in Mandarin hymns as a case of music template satisfaction. The 2009 Graduate Student Workshop on Phonology. National Chengchi University, Taipei.

Ling, Wang-Chen & Wang, Li-Ting. 2008. Mandarin-Accented English: A Syllable Template Perspective. 2008 Student Workshop on East Asian Linguistics. National Tsinghua University, Hsinchu, Taiwan.

Grants and Scholarships

2020 Member of the Phi Tau Phi Scholastic Honor Society of the Republic of China

2019 Graduate student scholarship (Ph.D. program)

2018-2020 Research Assistantship of Ministry of Science and Technology (MOST107-2410-H-004-115-MY2)

2018 Teaching Assistantship of Taiwan Southern Min and Taiwan Culture, NCCU

2018 NCCU Travel Grant for attending the 51st International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL51)

2018 College of Foreign Languages & Literature, NCCU Travel Grant for attending the 7th International Conference on Phonology and Morphology (ICPM7)

2017 Teaching Assistantship of Language and World Civilization, NCCU

2017 Ministry of Science and Technology (MOST) Travel Grant for attending the 37th Annual Conference of the Poetics And Linguistics Association

2016-2018 Research Assistantship of MOST (MOST 105-2410-H-004-181-MY2)

2016 MOST Travel Grant for attending the 28th North American Conference on Chinese Linguistics

2014-2016 Research Assistantship of MOST (MOST 103-2410-H-004-086-MY2)

2014 MOST Travel Grant for attending the 22nd Annual Conference of the IACL & the 26th North American Conference on Chinese Linguistics.

2013-2014 Research Assistantship of National Science Council (NSC) (NSC 102-2410-H-004-080)

2012-2013 Research Assistantship of National Science Council (NSC) (NSC101-2410-H-004-170)

2012-2013 Teaching Assistantship of Language and World Civilization, NCCU

2010-2014 Graduate student scholarship (Ph.D. program)

2010 Member of the Phi Tau Phi Scholastic Honor Society of the Republic of China

2010 National Chengchi University Travel Grant for attending the 42nd International Conference on Sino-Tibetan Languages and Linguistics

2009-2010 Graduate student scholarship (M.A. program)

2009 National Science Council (NSC) Travel Grant for attending the Fifth Corpus Linguistics Conference. Liverpool, England: Liverpool University.

2008-2009 Research Assistantship of National Science Council (NSC) (NSC 97-2410-H-004-111-MY3)

2007-2008 Research Assistantship of National Science Council (NSC) (NSC 96-2411-H-004-004)

2007 Research Assistantship of National Science Council (NSC) (NSC 95-2411-H-003-022)

DOI:10.6814/NCCU202001428