# 自然话语的韵律形式和韵律功能探讨

——以汉语普通话的声学特征、韵律分析和语调模拟结果为例

A Study on Form and Function of Prosody Based on Acoustics, Interpretation and Modelling
—with Evidence from the Analysis by Synthesis of Mandarin Speech Prosody

智娜 著

世界图书出版公司

# 自然话语的韵律形式和韵律功能探讨
## ——以汉语普通话的声学特征、韵律分析和语调模拟结果为例

# A Study on Form and Function of Prosody Based on Acoustics, Interpretation and Modelling
## —with Evidence from the Analysis by Synthesis of Mandarin Speech Prosody

智 娜 著

世界图书出版公司
北京·广州·上海·西安

**图书在版编目（CIP）数据**

自然话语的韵律形式和韵律功能探讨 ：以汉语普通话的声学特征、韵律分析和语调模拟结果为例 = A study on form and function of prosody based on acoustics，interpretation and modelling：with evidence from the analysis by synthesis of mandarin speech prosody：英文/智娜著. —北京：世界图书出版公司北京公司，2014. 11

ISBN 978-7-5100-8929-9

Ⅰ. ①自… Ⅱ. ①智… Ⅲ. ①普通话—韵律（语言）—研究—英文 Ⅳ. ①H11

中国版本图书馆CIP数据核字（2014）第 269467 号

**自然话语的韵律形式和韵律功能探讨**
**——以汉语普通话的声学特征、韵律分析和语调模拟结果为例**

**A Study on Form and Function of Prosody Based on Acoustics, Interpretation and Modelling**
**—with Evidence from the Analysis by Synthesis of Mandarin Speech Prosody**

**著　　者**：智　娜
**责任编辑**：夏　丹　李培肖

**出　　版**：世界图书出版公司北京公司
**发　　行**：世界图书出版公司北京公司
（地址：北京市朝内大街137号　邮编：100010　电话：010-64038355）
**销　　售**：各地新华书店
**印　　刷**：虎彩印艺股份有限公司

**开　　本**：787 mm × 1092 mm　1/16
**印　　张**：14
**字　　数**：200千
**版　　次**：2015年5月第1版　2015年5月第1次印刷

ISBN 978-7-5100-8929-9　　**定价**：39.00元

# Abstract

An analysis-by-synthesis study on Mandarin speech prosody is conducted in the present book. The features of Mandarin speech prosody are discussed by focusing on two salient aspects: the function of prosody and the form of prosody. The study attempts to find a plausible way in which the two aspects can be mapped onto each other through the functional analysis of prosody and the multi-level formal representation. The form of Mandarin speech prosody is a complex F0 picture due to the simultaneous uses of pitch contours by both lexical tones and sentential intonation. The phenomenon of tone sandhi in speech context triggers more puzzling issues when researchers are confronted with the acoustic form of Mandarin prosody. The functional use of prosody in Mandarin speech concerns: at the lexical level for word identity (Tone 1, Tone 2, Tone 3, Tone 4, and Tone 0); at the sentential level for prominence marking (sentence accents) and the indication of prosodic boundaries (intonation boundary tones). In the present study, the analysis of prosodic function at the two levels provides a basic framework in coding the surface melodic form of Mandarin prosody, which consists of pitch contours in tonal units and boundary tones at the beginning and end of intonation unit. For the formal representation of Mandarin speech prosody, the surface F0 contour of each utterance is coded into a sequence of INTSINT symbols, and subject to the Prozed tool for speech synthesis. It is shown

that the synthesized stimuli derived from the symbolic coding can closely follow the melodic features and correctly express the prosodic function of the original Mandarin utterances. The present study employs acoustic data, symbolic coding, and speech synthesis for the derivative mapping between prosodic function and form, which aims to interpret the complex prosodic phenomenon, and provide an insight for the annotation and analysis of Mandarin speech prosody.

# Contents

CHAPTER 1 Introduction ..... 1

1.1 Background of prosodic study of natural languages ..... 1

1.1.1 Function of prosody ..... 5

1.1.2 Form of prosody ..... 9

1.2 Mandarin prosody: issues and previous approaches ..... 19

1.2.1 Introduction of Mandarin ..... 19

1.2.2 Earlier studies on the interaction between tone and intonation ..... 21

1.2.3 Previous studies on rhythmic accent ..... 26

1.2.4 Recent studies: from data to modelling ..... 31

1.3 Objective of the present study ..... 44

1.3.1 The data of the study ..... 44

1.3.2 A plausible mapping between form and function ..... 45

CHAPTER 2 Literature Review of Prosodic Study ..... 53

2.1 Introduction ..... 53

2.2 The British School ..... 53

2.3 The Dutch School ..... 57

2.4 The American School ..... 60

2.4.1 The level approach ..... 60
2.4.2 The Autosegmental-metrical approach ..... 61
2.4.3 The ToBI labeling system ..... 65
2.4.4 Pan-Mandarin system ..... 68
2.5 Summary ..... 70
CHAPTER 3 Speech Corpora and Labeling Methodology ..... 73
3.1 Introduction ..... 73
3.1.1 Daily-conversation corpus ..... 73
3.1.2 Read-speech corpus ..... 74
3.2 Data selection and labeling ..... 75
3.3 Summary ..... 81
CHAPTER 4 Three Functional Components of Mandarin Prosody ..... 83
4.1 Introduction ..... 83
4.2 Lexical tone ..... 83
4.2.1 The functional role of lexical tone ..... 83
4.2.2 The paradigmatic annotation of lexical tone ..... 84
4.2.3 The autosegmental property of tone ..... 86
4.2.4 The phonological representation of tone: level vs. contour systems ..... 100
4.2.5 Tone sandhi in context ..... 104
4.3 Rhythmic accent ..... 123
4.3.1 Accentual hierarchy ..... 123
4.3.2 The prosodic structure of Mandarin speech ..... 138

4.3.3 Prosodic grouping and tone-sandhi domain ..... 142
4.4 Intonation ..... 155
4.4.1 The melodic events of Mandarin speech prosody ..... 155
4.5 Summary ..... 175
CHAPTER 5 Analysis by Synthesis of Mandarin Prosody ..... 177
5.1 Introduction ..... 177
5.2 Evaluating the predicted data with Prozed ..... 179
5.3 Discussion ..... 185
5.4 Summary ..... 191

CHAPTER 6 Conclusion ..... 193

Bibliography ..... 197

# CHAPTER 1 Introduction

## 1.1 Background of prosodic study of natural languages

Prosody refers to the speech characteristics beyond the text level, thus it is often considered as the "suprasegmental" aspect of spoken language. In everyday speech, prosody together with the lexical-syntactic information of spoken text contributes to the interpretation of speech, which is illustrated by an equation in Hirst (2011a) as, *Speech = Text + Prosody.* In social interactions, prosody performs salient functions for the need of communicative exchange between speech participants.

According to the analyses of a variety of natural discourse data in the book, *The Music of Everyday Speech* (Wennerstrom 2001), the way in which prosody contributes to the processing and understanding of spoken language could be found in various aspects. For instance, prosody indicates the coherence of speech by means of pause, length, and other intonational cues at boundaries of speech units. Such function of indicating the finality/non-finality of speech is widely used in the turn-taking between participants of the conversational interaction; Prosody also contributes to the illocutionary force of speech by means of certain intonation patterns, which can be observed from the speech

productions of lawyers in the court room, where intentional speech acts are achieved with particular intonational cues. Prosody can also provide evidence in understanding the subtle features of human social behaviour with the indication of "tone concord" which refers to the synchronized matching of speakers' pitch range, key level and rhythmic structure in a conversation, in particular, speech participants adopt the similar pitch level in matching with each other at one's completion of the turn and the other's turn-taking. It has been observed that such agreement in "tone concord" can be achieved by conversants in a harmonious context of interaction, whereas in less supportive situations, there occurs the "concord breaking" conveyed by discordant responses from conversants, i.e., using high key level competitively as a power struggle, or breaking down the regularity of rhythmic intervals in the interactive speech exchange; Prosody is also an essential non-verbal cue in expressing the emotions and attitudes of speakers. It is found that story-tellers often make full use of prosody to avoid monotonous speech by means of variating pitch, loudness, length and speech rate, etc. to achieve the purpose of enriching the narrative production with strong emotional colour in attracting the attention of listeners; Due to the importance of prosody in spoken language, it is proposed by Wennerstrom (2001) that the prosodic acquisition should be highlighted in the materials of second-language teaching.

Marotta (2008) highlighted the sociolinguistic status of prosody, as she held that prosodic elements such as intonation serves as an important socio-phonetic cue, with which listeners can perceive and distinguish the different varieties of the same language, as found in Italian (Marotta & Sardelli 2003, 2007; Marotta *et al.* 2004). For the phonological status of prosody, she proposed to draw a boundary between the language core grammar level and the pragmatic level, i.e., the sociolinguistic use.

Cresti (1995) discussed the discourse function of prosody, and

proposed "language as an act", in which the intonation of an utterance serves the function of a "systematic marker of informational unit, whose fundamental principle is establishment of illocution". From her observation of an Italian spoken corpus, she summarized that the (topic)-comment pattern can be regarded as the basic pattern in Italian oral speech. Topic is optional as it mainly carries given information, while comment is privileged and sufficient in performing the main function of speech act, as it usually corresponds to the predicate or a whole sentence, conveying new information.

Prosody, as a compositional term, has a broad definition, and researchers may have different understandings on the subject. For some researchers, prosody reflects speakers' emotion and attitude. According to Pike (1945: 10), the distinctiveness of intonational meaning is not defined by the grammatical sentence type, but rather by the attitude of the speaker at the time when the utterance is given. Bolinger (1989:1) also emphasized the paralinguistic function of intonation and defined intonation for its function in reflecting speakers' inner states, and as a "nonarbitrary, sound symbolic system with intimate ties to facial expression and bodily gesture".

Hirst & di Cristo (1998) held that prosody conveys semantic meanings and interpersonal functions together with the lexical and syntactic components of an utterance. Ladd (1996) summarized the linguistic and paralinguistic facts of intonation, and regarded intonation as "the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way". In his understanding, intonation has three main compulsory features, namely, (a) "suprasegmental"; (b) "'postlexical' or sentence-level pragmatic meanings"; and (c) "linguistically structured".

From the above review on the literature of prosody, one can note that the prosodic study of natural languages can extend to a very broad scale, due to the fact that prosody is the general non-verbal aspect

of oral speech, which could encompass a variety of linguistic and paralinguistic phenomena. Hirst (2001) claimed that a considerable number of factors contribute to a language's prosodic features, and they could be universal, language specific, dialectal, individual, syntactic, phonological, semantic, pragmatic, discursive, attitudinal, emotional, and the list is obviously not complete.

In the present study, I shall focus the discussions of prosody onto its linguistic functional aspect. The book consists of six chapters. Chapter 1 introduces the background of speech prosody, with discussions from both the functional aspect and formal aspect of prosody. Moreover, the chapter reviews the previous relevant studies on Mandarin speech prosody, and presents the objective of the present study by using speech synthesis technology in evaluating the proposal of form-function mapping. Chapter 2 is a review on the important related literature on speech prosody, including the classical studies on intonation of non-tonal languages by the British, the Dutch and the American school. The review aims to provide an overview of background information of prosodic study. Chapter 3 presents the details of the two Mandarin speech corpora employed in the present study: a spontaneous dialogue corpus and a read speech corpus. Discussions of the labeling methodology of the two corpora and the selection criteria of speech data are presented. In Chapter 4, detailed discussions are given on three important functional components of Mandarin prosody: lexical tones, intonation, and accents. The three factors closely interact in contributing to the prosodic form of Mandarin speech, and together play a salient role in the linguistic functional aspect. Chapter 5 discusses the results of speech synthesis derived from the symbolic representation of prosodic form with the INTSINT annotation system. The principle consists in finding a plausible way in which the physical facts of prosodic form can be related to the functional aspect of prosody. The proposal is tested through the implementation of the Prozed tool, with the predicted

symbolic coding of 60 Mandarin utterances synthesized into stimuli, and compared to the original utterances on the melodic features. It is shown that satisfactory synthesized results can be derived: the stimuli can closely follow the surface contour movement of the original ones. Chapter 6 summarizes the mutual interrelations among the three factors in Chinese speech prosody, and concluded that the phonetic makeup of speech could be correlated to the functional components of prosody.

### 1.1.1 Function of prosody

According to Hirst (2005), many of the linguistic functions of prosody are nearly universal, as in all languages prosody is an essential part for the word identity at the lexical level (via tone, accent and quantity); above that, prosody can highlight the key information by marking out certain words from the background for expressing prominence; moreover, prosody can indicate the finality and non-finality of speech unit with boundary tones. The linguistic function of prosody in natural language can be summarized according to its contribution at two distinct levels, as illustrated in Figure 1.1 from Hirst & di Cristo (1998: 4). Such distinction of prosodic function at lexical and non-lexical levels can provide a basic framework for the comparative description of prosodic systems across languages.

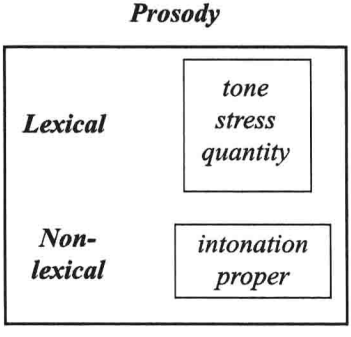


Figure 1.1: Prosodic function at the lexical level and non-lexical level

### *1.1.1.1 Prosodic function at the lexical level*

For the prosodic contribution at the lexical level, there are language-specific ways in the use of prosodic parameters for the distinction of lexical identity. In a tone language, like Mandarin Chinese, the pitch contrast on each syllabic unit contributes to the lexical distinctive function. There are four distinct tones in Mandarin: the first tone (Tone 1), the second tone (Tone 2), the third tone (Tone 3), and the fourth tone (Tone 4). The four lexical tones in Mandarin can be marked in written form with an iconic diacritic above the nucleus of the associated syllable, i.e., zhī, zhí, zhǐ, zhì. Therefore, the identical syllable, when associated with four different tones, leads to four completely different word morphemes, as seen in the following Table 1.1.

| Syllable | zhī | zhí | zhǐ | zhì |
|---|---|---|---|---|
| Tone | Tone 1 | Tone 2 | Tone 3 | Tone 4 |
| Morpheme | n. *"knowledge"*; v. *"to know"* | adj. *"straight"* | n. *"paper"* | adj. *"intelligent"* |

Table 1.1: Identical syllable with four different tones

For the tonal representation, Chao (1930) proposed a numerical notation system of five scaling points, 1 to 5 corresponding respectively to low, half-low, medium, half-high and high position within a speaker's normal pitch range. According to native speaker's perception, the four lexical tones can be represented with a succession of numerals marking respectively the significant points of the pitch contour of each tone. Accordingly, Tone 1 is transcribed as [55], Tone 2 as [35], Tone 3 as [214], and Tone 4 as [51], indicating the high-level pitch form of T1, the high-rising form of T2, the low-falling-rising form of T3, and the high falling form of T4.

In most cases, each monosyllabic morpheme of Chinese aligns with one of the above citational tones. There are also morphemes which

carry no tones, or so called 'neutral' tones. According to Chao (1968), such neutral tones occur on two types of syllables. First, there are inherently toneless syllables, which are usually grammatical morphemes, such as particles. Second, there are tonal syllables whose canonical tonal features are intentionally neutralized by speakers, especially, when such lexical morphemes are located on the second syllables in disyllabic words, such as the kinship term *ma1 ma0* (n."mother"), where the second repetitive syllable *ma* (with citational Tone1) is actually neutralized. Besides the above two types of toneless syllables, neutral tone also occurs in certain lexical words, where the tone of the second syllable is neutralized, and such lexical words usually form a minimal pair with the identical phonotactic ones without tonal neutralization, such as *dong1 xi0* (n. "thing") vs. "*dong1 xi1*" (n./adj. "east-west"). These neutral-tone syllables are phonetically analogous to unstressed syllables in English, with features such as vowel centralization, short duration and weak intensity. According to Yip (2002), all the neutral-tone syllables are regarded as default unstressed syllables in contrast with the full-tone syllables in Chinese speech.

In a non-tonal language, such paradigmatic opposition of tonal pitch does not contribute to the lexical identity, for instance in English or Italian. The two languages employ quantity and accent, instead of tone, in making the distinctions of lexical identity. There are also languages in which prosody is not required to perform any function at the lexical level. In Hirst (2004), it was mentioned that modern standard French does not distinguish lexical identity according to either accent, or tone, or quantity; thus, the perceptual prosodic contrast in French fluent speech is not derived from the underlying lexical level, but rather from the surface sentential level. The discussion of prosodic typology at the lexical level can be seen in the study of Hyman (2009), in which it was proposed that word-prosody should be conducted in a property-driven approach and the author argued against the use of the term

"pitch-accent" for a language like Japanese, which has an intermediate property of accentual tone between English (a stress language) and Mandarin (a tone language).

*1.1.1.2 Prosodic function at the non-lexical level*

For the prosodic contribution at the non-lexical level, there exist a large quantity of discussion in the literature on the language intonation systems. Such kind of literature has especially flourished in English language, inspired by the pedagogical purpose of English-as-a-second-language (ESL) acquisition. It can be seen in many earlier works, such as Palmer (1922), Armstrong & Ward (1926), Kingdon (1958), Cystal (1969), Halliday (1967, 1970), O'Connor & Arnold (1973) and etc., which formed the well-developed approach of the British school in the study of intonational functions and forms.

Despite the fact that intonation may present particular forms in different languages, it is generally agreed that the linguistic functional role of intonation in expressing prominence, called the *weighting function* of intonation by Gårding (1989), and the function of marking connective and demarcative features in speech, called *grouping function* of intonation (ibid), are nearly universal in all languages.

In this study on Mandarin prosody, the nature of abstract prosodic properties, both at the functional and formal level will be explored. The discussion on prosodic function is focused on the lexical distinctive function of tones and the weighting and grouping function of intonation in connected speech. The functional analysis of prosody will pave the way for the interpretation and modelling of the acoustic data of prosodic form. The plausibility of the form-function mapping proposal is evaluated through the MOMEL-INTSINT system (Hirst & Espesser 1993, Hirst & di Cristo 1998, Hirst 2005). The aim is to examine the underlying and potential factors, which contribute to the surface phenomena, by means of combining acoustics with interpretation, and

evaluating proposals with the synthesis result.

In the following section, I shall discuss the physical events related to the form of prosody.

### 1.1.2 Form of prosody

The form of prosody mainly concerns the salient sound events perceived by listeners, which are indicated by parameters such as pitch and length, which will be discussed in details in the present section, as the two serve as important correlates of prosody in contributing to the linguistic functions in speech.

#### *1.1.2.1 Pitch*

Pitch is acknowledged as the primary parameter in listeners' perception of speech melodic form, which generally consists of intonation and also lexical tones in certain languages. Previous studies on intonation forms, such as the British School, greatly depend on listeners' impression of the "rising" or "falling" tendency of pitch contour movement especially at the sentence end, which are believed to be associated with the expression of modality, in particular, statement is stylized with a falling pitch contour, while question is expressed with a rising pitch pattern. Pitch, as the primary acoustic parameter in the perception of intonation, was also discussed in Hirst & di Cristo (1998: 4).

In a tone language like Mandarin, speech melody is a combined carrier of pitch information at the lexical level as well as the sentential level. The pitch movement of the component syllables in each utterance contributes to the basic melodic form of a Mandarin utterance, while the intonation pattern at the sentential level interact with the lexical pitch contours, which can be in a way of either modifying (broadening or narrowing) the pitch range of the local lexical contour, or adding intonation boundary tones at the initial and final boundaries