同济大学 1907-2017
Tongji University

同济博士论丛
TONGJI Dissertation Series

总主编 伍 江 副总主编 雷星晖

夏志华 马秋武 著

汉语对话中韵律趋同的实验研究

Prosodic Entrainment in Mandarin Chinese Conversations: an Experimental Study

同济大学出版社
TONGJI UNIVERSITY PRESS

本研究得到教育部人文社会科学青年基金项目（项目名称：汉语对话中韵律趋同的实验研究 项目批准号：15YJC740105）和江苏师范大学博士学位教师科研支持项目（项目名称：英汉语篇韵律对比研究 项目批准号：15XWR008）的资助，特此感谢。

同济博士论丛
TONGJI Dissertation Series
总主编 伍 江 副总主编 雷星晖

夏志华 马秋武 著

汉语对话中韵律趋同的实验研究

Prosodic Entrainment in Mandarin Chinese Conversations: an Experimentai Study

同济大学出版社
TONGJI UNIVERSITY PRESS

内 容 提 要

本书研究汉语对话中的韵律趋同现象，是会话交流中韵律语用功能实现的研究之一。本书通过实验手段，对自然会话进行分析，研究的主要目的是探索汉语韵律趋同的规律，同时在可行范围内，对英汉韵律趋同规律进行比较，从韵律的语用功能实现上找出两种语言的异同。

图书在版编目(CIP)数据

汉语对话中韵律趋同的实验研究 / 夏志华，马秋武著. —上海：同济大学出版社，2019.10
(同济博士论丛 / 伍江总主编)
ISBN 978-7-5608-8785-2

Ⅰ. ①汉… Ⅱ. ①夏… ②马… Ⅲ. ①汉语—韵律(语言)—研究 Ⅳ. ①H11

中国版本图书馆 CIP 数据核字(2019)第 225442 号

汉语对话中韵律趋同的实验研究

夏志华 马秋武 著
出 品 人 华春荣 责任编辑 熊磊丽
责任校对 徐春莲 封面设计 陈益平

出版发行 同济大学出版社 www.tongjipress.com.cn
(地址：上海市四平路 1239 号 邮编：200092 电话：021-65985622)
经 销 全国各地新华书店
排版制作 南京展望文化发展有限公司
印 刷 浙江广育爱多印务有限公司
开 本 787 mm×1092 mm 1/16
印 张 16
字 数 320 000
版 次 2019 年 10 月第 1 版 2019 年 10 月第 1 次印刷
书 号 ISBN 978-7-5608-8785-2

定 价 74.00 元

前 言

会话趋同是指参与者在对话中进行合作调整，以完成顺畅的交流。它普遍存在于会话交流中，主要表现为句法形式、词汇选择上的相似，发音、语速等韵律上的相仿，肢体表情的接近，等等。本书以其中的韵律趋同为研究重点。

具体而言，韵律趋同指双方在对话中调整其韵律表现，变得相似，以完成顺畅的交流。对话交流中，个体调整韵律表现以适应对方，这种调整过程复杂，涉及认知、心理、语言、社会等因素的综合作用，其中的规律值得探究。

本书研究汉语对话中的韵律趋同现象，是会话交流中韵律语用功能实现的研究之一。本书通过实验手段，对自然会话进行分析，研究的主要目的是探索汉语韵律趋同的规律，同时在可行范围内，对英汉韵律趋同规律进行比较，从韵律的语用功能实现上找出两种语言的异同。据此，本书提出了三个研究问题：韵律在普通话对话中是如何表现趋同的？社会因素如何影响韵律趋同？在跨语言比较中，英汉两种语言的韵律趋同规律有何异同？针对这些问题的分析构成了本书研究的主体：韵律趋同多层次分析、韵律趋同与社会因素关系的分析和韵律趋同跨语言比较分析（详见第 4、5、6 章）。

本书研究有如下两方面意义。

一、本书研究具有一定的理论意义

首先，汉语对话中韵律趋同研究印证和丰富了会话趋同研究的理论框架。现有的韵律趋同研究中已经涉及的语言主要有英语、荷兰语、瑞典语、日语等，极少涉及汉语。本书对于汉语会话中韵律趋同进行了较全面的研究，在印证已有的会话趋同理论的同时，首次提出会话中存在两类趋同，一类是

绝对趋同，另一类是相对趋同。会话中两类趋同的提出丰富了已有的会话趋同理论，有助于会话趋同过程的描述，同时增进人们对该过程的理解。

其次，韵律趋同规律的跨语言比较具有语言类型学意义。本书比较了汉英两种语言中韵律趋同规律，帮助人们更深入理解不同类型语言中对话双方如何借助韵律手段实现会话趋同。从语言类型上来讲，英语和汉语属于两种不同类型的语言。在不同类型语言中进行韵律语用功能实现方式的比较具有语言类型学上的意义。

二、本研究具有一定的实践意义

首先，以本书结论为基础，可总结出有效的交流策略和方法，其具有实践指导意义。韵律趋同是对话双方在对话中韵律上的相互合作调整。本书关于时长、音高、音强韵律特征在不同层面上表现突出趋同的结论可以用来指导有效交流。

其次，本书的结论期待为语言合成带来些启发。让合成的语言和人类语言一样自然是言语工程研究的一大挑战。本书研究中所发现的对话趋同中韵律的调整是改进语言自然度的一个方面，所得出的研究结论也希望能为自然语言韵律合成研究带来些启发。

本书语料库和主要分析如下：

本书研究基于同济大学游戏语料库。该语料库包含 115 个即兴对话，每个对话平均时长为 6 分钟。这些即兴合作对话由本书研究设计的实验所引发。受试根据要求完成两类实验——图片排序和图片分类。70 组受试者参加实验，每组两人，其中有 14 组根据要求参加了两次实验。

本书选取韵律的 3 个主要方面（时长、音高、音强），共计 7 个声学参数为变量，它们分别是：时长特征（语速），基频特征（基频最低值、基频平均值、基频最高值），音强特征（音强最低值、音强平均值以及音强最高值）。

IPU（停顿间隔单元，本书研究中停顿间隔设定为 80 毫秒）被认定为最小的数据提取和分析单元。通过软件 SPPAS 对语料进行 IPU 的自动识别和标注，然后对被识别的 IPU 边界进行人工核查。在标注好的语料上运行 Praat 脚本，并提取每个 IPU 单元上的 7 个声学参数为本书的分析数据。当韵律趋同分析涉及了更大单元时（即该单元包含多个 IPU），本书计算 IPU

的加权均值(以每个 IPU 的时长为权重),然后针对加权均值进行分析。

第 4 章的主要内容是多层次上韵律趋同的分析。从宏观到微观,本章在会话、话轮和声调单元三个不同层次上进行韵律趋同分析,其中每个层次上的分析主要从接近性、融合性和同步性三个角度展开。基于这些分析结果,本章针对相同角度不同层次上韵律趋同规律进行了从宏观到微观的跨层次比较分析。

通过接近性的比较分析发现,时长和音强特征在会话、话轮和声调单元三个不同层次上都表现出整体的相似性。通过融合性的比较分析发现,在话轮层次上,对话双方表现出显著的融合性,即随着对话的推进,两者的韵律表现越来越接近,但是在会话和声调单元层次上,均未发现类似的规律。通过同步性的比较分析发现,对于音高特征而言,即使整体的接近性和融合性不存在,对话双方在微观层次上也表现出了显著的同步性。

通过声调趋同分析发现,汉语对话中,混合性别对话者在声调产出上存在相对趋同,即混合性别对话者在声调的相对调域上存在趋同。

第 5 章的主要内容是韵律趋同与社会因素的分析,包含两方面:韵律趋同与性别关系的分析以及韵律趋同与角色关系的分析。

韵律趋同与性别关系的分析包括不同性别组合上的接近性分析和趋同程度分析。分析发现,在男女组合的对话中参与趋同的韵律参数最多,而在男男组合的对话中韵律参数最少;相同性别组合(女女组合和男男组合)中时长特征和音强特征上都有趋同表现,在混合性别组合(男女组合)中,包含以上两方面的韵律特征外(时长、音强),音频特征上也有趋同表现;男男组合的趋同度表现出最小的趋势。

韵律趋同与角色关系的分析包括角色影响检测和角色方向检测。在实验设计中,图片排序游戏中对话双方角色有区分(信息给予者和信息接收者),而图片分类游戏中对话双方角色无区分。角色影响检测结果表明角色对于韵律趋同程度有显著影响,即图片排序游戏中的韵律趋同程度显著大于图片分类游戏中的趋同程度。角色方向检测结果表明,对话双方在趋同的方向上的表现是:信息给予者更多地趋同于信息接收者。

第 6 章的主要内容是汉英韵律趋同跨语言比较分析。基于两种语言中

相对应的研究部分，本章中的跨语言比较主要从宏观（会话层次）、微观（话轮层次）和不同性别组合上展开。

跨语言比较发现，两种语言在韵律趋同方式上表现出显著的相似性。比如，在会话层次，两种语言中韵律参数表现出相似的整体接近性；两种语言中体现话轮层次上接近性的韵律参数相同（它们是语速、音强均值、音强最高值）；在话轮层次上，两种语言表现出相似的融合性和同步性；混合性别组合的对话中，参与趋同的韵律参数最多，男男组合最少。

跨语言比较还发现，两种语言在韵律趋同程度上表现出不同。比如，话轮层次上，两种语言的融合程度不同；话轮和声调单元层次上，两种语言的同步程度不同；对于不同性别搭配而言，两种语言中，三种性别组合的韵律趋同程度不同。

对比还发现，现有的研究中，英语表现绝对趋同，而在本书研究中，汉语表现绝对趋同与相对趋同共存。

两种语言趋同方式的相似证明了，尽管隶属于不同类型的语言，汉语和英语利用相似的方式来实现音长、音高和音强上的趋同。而两种语言韵律趋同程度的差异，应该与两种语言类型上的差异有关。

综上所述，本书研究的主要发现有：

（1）研究结果证实，汉语普通话的对话中，韵律的 3 个主要方面（时长特征、音高特征、音强特征）都有显著的趋同表现。

本书研究选取韵律的 3 个主要方面，共计 7 个声学参数为变量，它们分别是：时长特征（语速），基频特征（基频最低值、基频平均值、基频最高值），音强特征（音强最低值、音强平均值以及音强最高值）。第 4、5、6 章中三项主要分析（韵律趋同多层次分析、韵律趋同与社会因素关系的分析和韵律趋同跨语言比较分析），证实了韵律的 3 个主要方面都有显著的趋同表现。

（2）本书研究表明，汉语韵律在微观层次上的趋同表现比宏观层次上更明显。

首先，本书发现在宏观层次上表现趋同的韵律参数数量多于微观层次上的数量（详见第 4 章）。4 项韵律参数在会话层次上表现趋同，而 5 项在声调单元上体现趋同。6 项韵律参数在话轮层次上表现同步性，而 7 项在声调单

元上表现出同步性。其次，本书研究还发现，对于音高特征而言，尽管在相对宏观层次上，其趋同表现不明显，而在相对微观层次上，它们表现出显著的趋同性。

(3) 韵律参数在不同层次上展现不同的趋同规律。

不论是在相对宏观还是微观层面上，音强特征表现出明显的趋同性；时长特征在相对宏观层次上表现更明显的趋同；音高特征在相对微观层次上表现更明显的趋同。通过第 4 章中的跨层次比较发现，音强特征几乎在分析的所有层次上都表现出显著的趋同性。时长特征在会话层面表现出的趋同性比话轮层面要更明显。音高特征在话轮和声调单元上表现出明显的同步性，而在会话层面上的同步性却不显著。

(4) 本书研究发现，随着对话的推进，话轮层面表现出融合性，即对话双方在韵律参数上的差距随着时间的推移而缩小，但是会话层面和声调单元层面并未发现这样的规律。

(5) 本书研究证明韵律趋同与社会因素具有紧密的联系。

对于性别特征而言，三种性别组合上体现出不同的韵律趋同规律。混合性别组合对话中参与趋同的韵律参数数量最多。从表现出趋同的韵律参数数量和趋同的程度上来讲，都是男性组合趋同最少。对于角色而言，本书研究证实角色影响会话中韵律趋同的程度，而且对话中信息提供者更多地趋向于信息接收者。

(6) 通过跨语言比较发现(详见第 6 章)，英汉两种语言在韵律趋同方式上存在显著共性，在韵律趋同程度上存在显著差别。

两种语言在韵律趋同方式上表现出显著的相似性。比如，在会话层次，两种语言中韵律参数表现出相似的整体接近性；两种语言中体现话轮层次上接近性的韵律参数相同；在话轮层次上，两种语言表现出相似的融合性和同步性；男女组合对话中参与趋同的韵律参数最多，而男男组合最少。

两种语言在韵律趋同程度上表现出不同。比如，话轮层次上，两种语言的融合程度不同；话轮和声调单元层次上，两种语言的同步程度不同；对于不同性别搭配而言，两种语言中，三种性别组合的韵律趋同程度不同。

(7) 本书首次提出会话中存在两类趋同，一类是绝对趋同，另一类是相

对趋同。现有的研究中，英语表现绝对趋同，而在本书研究中，汉语表现绝对趋同与相对趋同共存。如汉语中，相对宏观单元上的韵律特征体现绝对趋同，混合性别对话者声调相对调域内体现相对趋同。

（8）本书在汉语会话韵律趋同研究中使用的方法上具有参考价值。

本书中有参考价值的研究方法主要有两类：声调单元上相对趋同的分析方法和会话与话轮层次上绝对趋同的研究方法。

韵律趋同现象复杂，一项研究并不能涵盖其所有内容，本书研究在如下几方面存在，它们可成为后续的研究点。

第一，对于韵律趋同过程的研究需要进一步展开。由于韵律趋同性质难以捉摸，随着对话的进行，韵律趋同的过程在不断变化，是这研究的难点。在本书研究中，除了在话轮层次上，随着对话的推进对话双方在一些韵律参数表现上越来越接近以外，在会话层次和声调单元层次上均未发现类似规律。因此，后续需要进一步展开对于韵律趋同过程的研究。

第二，韵律趋同研究需囊括更多的社会因素。对于韵律趋同与社会因素的关系，本书研究着重于性别和角色这两个因素。事实上，与韵律趋同有关的社会因素众多，包括身份、年龄、性别、角色、区域、种族、环境，等等，这些都值得研究。研究难点在于如何在实验中凸显目标因素，同时控制其他因素。

第三，音高与韵律趋同的关系值得进一步研究。本书研究以二字组一声组合的声调单元为研究对象，从接近性、融合性和同步性三方面考察了音高和韵律趋同之间的关系，其他类型声调组合的韵律趋同研究需进一步展开。同时，对于汉语声调与语调的关系问题，现有研究多数是概念性描述，汉语本体研究还未给出明确清晰的答案，因此需要进一步展开相关研究，探索音高与韵律趋同之间的关系。

目 录

Chapter 1

Introduction

In conversation, the major goal of both interlocutors is to achieve mutual intelligibility. To reach this goal, both parties coordinate their communication ways. Therefore, conversations are considered as joint activities in which two interlocutors share or synchronize their mental states and performances. The joint nature of language processing in communication requires the interpersonal coordination in minds and actions (Brennan *et al*. 2010). This coordination is called entrainment.

Speech communication involves not only an exchange of propositional contents but also an interchange of emotions, attitudes, intentions, etc. of the speakers. Prosody is an effective tool to realize such pragmatic functions in interaction.

This book focuses on the prosodic entrainment and aims to explore how the prosody works in entrainment in Mandarin Chinese conversations, and to find out how prosodic entrainment in Chinese are different or similar to that in English.

1.1 Key Terms

1.1.1 Entrainment

Entrainment in speech means that speakers adapt their communicative behavior to their conversational partners. Social interaction involves participants' mutual coordination or adaptation. Many studies has focused on or touched

this, and different terms are used for this phenomenon, such as entrainment (Brennan 1996; Cummins 2009; Van Der Wege 2009; Lee *et al*. 2010; Levitan & Hirschberg 2011; Levitan *et al*. 2012), alignment (Pickering & Garrod 2006), accommodation (Beňuš *et al*. 2011), convergence (Giles *et al*. 1991; Pardo 2006), synchrony (Edlund *et al*. 2009), mimicry (Couper-Kuhlen 1996; Pentland 2008) and chameleon effect (Chartrand & Bargh 1999). Other terms are also made but in less frequent use (De Looze *et al*. 2014). Child-directed speech or motherese (Fernald *et al*. 1989) is used to describe speakers' accommodation when having conversations with children; foreign talk or foreigness (Ferguson 1975; Zuengler 1991; Smith 2007) is used when talking with non-native speakers; Lombard effect (Summers *et al*. 1988; Zeine & Brandt 1988) is used on the occasion of accommodating to a noisy environment.

There is much evidence that entrainment is critical to humans' assessment of dialogue success, overall quality, and their evaluation of conversational partners. Goleman (2006) points out that humans' ability to synchronize their communicative behavior with that of their conversational partners is critical to successful communication.

Entrainment in speech happens at non-linguistic levels including gestures, facial expressions, body movements, etc. Chartrand & Bargh (1999) find in their studies of mannerism and facial expression entrainment that subjects display strong unintentional entrainment, and that greater entrainment leads subjects to report they liked the confederate more and that the overall interaction was progressing more smoothly. Some other studies also find out that the similar facial expressions and semblable body movements are used in conversations (Condon & Sander 1974; Meltzoff & Moor 1977; Maurer & Tindall 1983; Bavelas *et al*. 1986; Bernieri & Rosenthal 1991; Bernieri *et al*. 1994; Richardson *et al*. 2007; Shockley *et al*. 2007; Shockley *et al*. 2009; Hess & Blairy 2011).

Entrainment in speech happens at linguistic levels including phonetic elements, prosody, lexicon, syntax, etc. Pickering & Garrod (2004) claims that the automatic alignment at levels of linguistic representations is important

both for production and perception in dialogues, and facilitated interaction. The adaptation of the interlocutors is found in lexicon (Brennan 1996; Pickering & Gorrod 2004; Nenkova *et al*. 2008; Branigan *et al*. 2011; Iwata & Watanabe 2013) and at the grammatical and syntactic structure of the speaking (Cleland & Pickering 2003; Haywood *et al*. 2005; Pickering & Ferreira 2008; Branigan *et al*. 2010). Reitter & Moor (2007) also find that degree of entrainment in lexical and syntactic repetitions occurring in the first five minutes of a dialogue significantly predicts task success in studies of the HCRC Map Task Corpus. The accommodation of the interlocutors is also found in pronunciation (Giles *et al*. 1991; Pardo 2006; Delvaux & Soquet 2007; Aubanel & Nguyen 2010; Bailly & Lelong 2010; Babel & Bulatov 2011).

Interlocutors in conversation also make prosodic adaptation (Natale 1975; Gregory & Hoyt 1982; Gregory *et al*. 1993; Stanford & Webster 1996; Gregory & Dagan 1997; Edlund *et al*. 2009; Levitan & Hirschberg 2011; De Looze & Rauzy 2011; De Looze *et al*. 2011; Levitan *et al*. 2012; Levitan 2013; De Looze *et al*. 2014). The present work focuses on prosodic entrainment in Mandarin conversations. The main reason for doing research on prosodic entrainment is the crucial roles prosody plays in speech communication.

1.1.2 Speech Prosody

Prosody plays crucial roles in interaction. Two definitions of speech prosody are relevant to its functions in interaction. One comes from Crystal (1969), and the other Firth (1957).

According to Crystal (1969), prosody is defined as "sets of mutually defining phonological features which have an essentially variable relationship to the words selected, as opposed to those features which have a direct and identifying relationship to such words". The prosodic phenomena are classified in two layers (Crystal 1969): primary (linguistic), and secondary (paralinguistic). The former includes pitch direction, pitch range, pause, loudness, tempo, rhythm. The latter includes voice qualifiers (for instance whisper, breathiness, creaky, etc.) and voice qualifications (Such as laughter, giggle, sobbing, or

crying).

Most of the West European research tradition follows Crystal's definition. However, because this two-layer division in the classifications of prosodic phenomena is considered to be ambitious, some researchers have doubts on the validity of this definition. For example, one of the doubts comes from the problem of classifying pause between linguistic and paralinguistic phenomena.

Different from Crystal's linguistic and paralinguistic definition, some researchers (Kelly & Local 1989; Couper-Kuhlen 2000) accept a more extensive definition of prosody from Firth. Firth (1957) assumes that prosody is as all types of syntagmatic relationships between syllables, which is not determined by the structure of words and utterances, and that prosody includes syllable structures, stress or accentuation, tone, quality and quantity, and also phenomena like glottalization, aspiration, nasalization, whisper, etc. About Firth's definition, Selting (2010) supplies deeper illustration that all supra-segmental phenomena, which are produced by the interplay of pitch, loudness, duration, and voice quality can be understood as prosodic, and once they are used, they are independent of the language's segmental structure. By Firth's definition, prosody functions as signals for communication. This broader definition is useful for the research on prosody in interaction, in which increased interest has been put.

Why are the functions of prosody in interaction emphasized?

The main reason is that at the importance of prosody in interaction has been recognized and emphasized. A brief review of the research on prosody helps to illustrate how the emphasis is put on prosody in interaction.

In the early research, the perspective on prosody is structural not interactional. In the early part of last century, in the structural tradition, the research on prosody primarily was in structural terms. An analogy was made between prosodic phenomena and meaning-distinctive units (Such as phonemes) and meaning-bearing units (Such as morphemes) (Couper-Kuhlen 2009). Speech prosody have been considered as a part of language competence, analyzed in minimal pairs, and treated by phoneme-or morpheme-like elements with distinctive functions. Pike (1945) assumes that pitch levels in American