应用语言学研究丛书

Phonetic Development of Voiceless Obstruents in Mandarin-Speaking Children

汉语儿童清阻塞音语音发展

马俊周 著

湖南大学出版社

Phonetic Development of Voiceless Obstruents in Mandarin-Speaking Children

汉语儿童清阻塞音语音发展

内 容 简 介

语音发展是一个复杂的过程，受到一系列因素如普遍性、口腔运动肌发育和外部语言环境的影响。本书采用了转写分析法和声学分析法，通过36名3到5岁汉语儿童及12名成人的汉语清阻塞音语音产出实验，研究其习得顺序及语音发展的连续性问题。本书提出了一个组合模型来解释习得顺序，并得出了儿童语音发展是连续的而非离散的这一结论，验证了语音发展的连续性假说。

图书在版编目（CIP）数据

汉语儿童清阻塞音语音发展＝Phonetic Development of Voiceless Obstruents in Mandarin-Speaking Children：英文/马俊周著.—长沙：湖南大学出版社，2021.12
ISBN 978-7-5667-2380-2

Ⅰ.①汉… Ⅱ.①马… Ⅲ.①汉语—发音—现代—英文
Ⅳ.①H116.1

中国版本图书馆CIP数据核字（2021）第249869号

汉语儿童清阻塞音语音发展
HANYU ERTONG QINGZUSEYIN YUYIN FAZHAN

著　　者：马俊周
责任编辑：邓素平
印　　装：长沙云图数字印刷有限公司
开　　本：710 mm×1000 mm　1/16　**印　张**：15.5　**字　数**：274千字
版　　次：2021年12月第1版　**印　次**：2021年12月第1次印刷
书　　号：ISBN 978-7-5667-2380-2
定　　价：58.00元

出 版 人：李文邦
出版发行：湖南大学出版社
社　　址：湖南·长沙·岳麓山　**邮　编**：410082
电　　话：0731-88822559(营销部)，88821691(编辑室)，88821006(出版部)
传　　真：0731-88822264(总编室)
网　　址：http://www.hnupress.com
电子邮箱：820178310@qq.com

CONTENTS

Chapter 1 Introduction

1.1 Research background

Language, as the most powerful tool that human beings have, distinguishes humans from animals. It is generally believed that children can acquire language seemingly effortlessly at a very young age. However, children are not ready to speak to their parents or other caregivers after birth. Instead, each infant has to undergo a long period of struggles before they can successfully communicate with others through speech, which is more often than not neglected by adults. Despite the great importance of speech, the nature of child speech development is poorly understood, and we still do not fully know how human beings develop speech so that they can get involved in the speech community. Fortunately, children provide a chance for us to observe speech in their initial stage (Bühler, 1935). The study of child speech could not only disclose the speech development in children, but also unveil its relationship with languages of the world.

Speech development is a product of a combination of factors such as oromotor and perceptual maturation, and ambient language (de Boysson-Bardies et al., 1989; Ferguson & Farwell, 1975; Ingram, 1988a; Jakobson, 1941/1968; Li, 2012; Locke, 1983; Pye et al., 1987). Abundant research has been carried out to investigate the speech development in children, revolving on the four most controversial issues in the realm of first language acquisition: ①Is language human-specific or not? ②Is child language development a result of nature or nurture, or nature and nurture interact to promote language development? ③Does child speech development follow a universal developmental path or not? ④Are child language developmental stages continuous or discontinuous? This book is to examine whether there exists a universal order of acquiring obstruents cross-linguistically on the one hand or such acquisition order is influenced by other factors on the other hand, and whether child speech development unfolds in a continuous fashion or in a discrete manner (Jakobson, 1941/1968; MacNeilage & Davis, 1990; Nittrouer, 1995). In addition, previous research showed that children did differ not only in the chronology of acquiring sounds but also in the rate of learning to produce speech sounds (Li et al., 2002; Vihman, 1993). However, due to the limitations of experimental methods and different theoretical positions, they arrived at inconsistent conclusions regarding the phonological acquisition order and the issue of continuity or discontinuity of speech development.

Although a large body of literature was devoted to the study of phonemes in general(Hua & Dodd, 2000; Mowrer & Burger, 1991; Pye et al., 1987; Smith, 1973; Templin, 1957; Waterson, 1971), most of them were based on dairy studies using the transcription method, not the acoustic method. The auditory-based transcription method is very likely to obscure the actual trend of speech development and inevitably arrive at conflicting conclusions regarding the order of phonological acquisition. The shortcomings of impressionistic studies have been repeatedly pointed out by many previous studies (Li, 2008; Li & Munson, 2016; Yang, 2014). Firstly, adult transcribers do not agree with each other when transcribing ambiguous speech, especially in the case of unclear speech, and their transcription accuracy is often affected by the phonetic training they received and their language experiences. Secondly, speaker's age and gender, physical and personality characteristics, speech intelligibility and linguistic context will probably affect the transcription accuracy as well (Shriberg & Kent, 2003). Thirdly, the nature of the sound itself will also affect adult's transcription because children could produce fine-grained differences that are quite under adults' categorical perception threshold, known as covert contrast, which refers to the perceptually indistinguishable, but statistically significant differences in the productions of contrastive sounds (Maxwell & Weismer, 1982; Scobbie et al., 2000). Another shortcoming of the transcription method is that it inevitably leads to the conclusion that speech development follows a discrete manner, rather than a continuous fashion because transcription methods depend on adults' perception of target phonemes. However, adults are more likely to perceive sounds categorically. Consequently, it poses greater challenges to perceive speech sounds that are within the categorical boundary than those cross the boundary (Chen et al., 2017; Liberman et al., 1957). As a result, adults are more likely to transcribe those intermediate sounds produced by children based on their subjective judgment, thus obscuring the nature of child speech development. Therefore, an objective method is warranted to complement phonetic transcription to further examine child speech development, rather than solely relying on transcription methods.

Advancement in computer technology has enabled researchers to adopt more efficient ways to find stable and consistent acoustic cues in speech signals (Forrest et al., 1988; Kewley-Port, 1983; Lahiri et al., 1984; Stevens & Blumstein, 1981; Sussman et al., 1991). The continued search for acoustic invariance within the speech signal has been recently aided in part by a relatively new type of analytical tool, the spectral moments analysis. This type of analysis examines the spectral characteristics of discrete time segments of speech signals in terms of multiple statistical moments (i.e., mean, standard deviation, skewness, and kurtosis). By combining these various "statistical snapshots" of perceptually relevant segments of the physical waveform, researchers could identify unique and discrete patterns of acoustic energy within the speech signal and use spectral moments analysis to examine the spectral characteristics of obstruents produced by children and adults (e.g., Forrest et al., 1988; Fox et al., 2001; Jongman et al., 2000; Nissen & Fox, 2005, 2009; Nittrouer, 1995; Tomiak, 1990). In addition, considerable acoustic research provided supporting evidence for the continuous view of speech development, suggesting that children constantly fine-tune their articulatory process and gradually approximate adult-like patterns (Kent, 1976; Li, 2012; Macken & Barton, 1980; Nittrouer, 1995; Yang, 2018; Yu et al., 2015). The acoustic analysis allows for more in-depth investigations of children's acquisition of fine-grained articulatory details that are quite below native adults' perceptual thresholds, and the use of phonetic transcription could provide a frame for the interpretation of the acoustic patterns. This book will adopt both the transcription method and the acoustic method to analyze the speech sounds in order to probe the speech development in children.

1.2 Purpose of the study

This book will document the acoustic developmental patterns of Mandarin voiceless obstruents produced by children aged 3-5 and examine how children's productions differ from or resemble those produced by adults. The purposes of the study are, therefore, threefold: ①It aims to analyze adult's productions to parameterize the acoustic properties of Mandarin voiceless obstruents including stops, fricatives and affricates, respectively, to form the baseline and criterion for children's speech productions. ② It seeks to describe the developmental patterns of Mandarin voiceless obstruents, thus to examine the acquisition order of Mandarin voiceless obstruents based on both transcription and acoustic methods, and to discuss to what extent the factors put forward by previous theoretical frameworks: universal account, oromotor maturation, and language-specific account, could account for the actual acquisition order of Mandarin voiceless obstruents. ③This study will address whether child speech development unfolds in a continuous fashion or in a discrete manner through examining how children's productions differ from or resemble those produced by adults in terms of multiple acoustic parameters.

1.3 Significance of the study

Firstly, this study systemically examines the acoustic properties of Mandarin voiceless obstruents and bridges the gap of literature on the phonetic development of Mandarin voiceless obstruents produced by children by means of both temporal and spectral parameters, allowing for in-depth examination of the nature of child speech development. Secondly, in addition to the traditional transcription method, the acoustic method is used to analyze the phonetic development of Mandarin voiceless obstruents to solidify the phonological acquisition order and to evaluate the contributions of the factors proposed by related theoretical frameworks. Thirdly, affricates are first included to further our understanding of the phonetic development of Mandarin voiceless obstruents. Fourthly, an integrated model, incorporating various factors with different weightings, is proposed to account for the acquisition order. Fifthly, raw data can be used as normative data for speech pathology in order to assess speech development and rehabilitation for speech-impaired children, though the current study is analytical.

1.4 Research questions and hypotheses

Grounded in the above brief review of child speech development, this book is to address the following three questions:

①What are the acoustic properties of Mandarin voiceless stops, fricatives and affricates?

②What is the respective acquisition order of Mandarin voiceless stops, fricatives and affricates produced by the children aged 3-5 years old based on the transcription and acoustic methods, and how does the observed acquisition order relate to pertinent acquisition theories?

③Whether children's speech develops in a continuous or discrete fashion by examining the acoustic similarities and differences between children's and adults' speech productions?

Given that the first research question is descriptive in nature, hypotheses are formulated only for the second and third research questions based on pertinent acquisition theories regarding the order of phonological acquisition and the issue of continuity, which will be elaborated in Chapter 2.

According to Jakobson's (1941/1968) universal account for the phonological acquisition order, the following three hypotheses are formulated:

H_1: Stops: velar > bilabial > alveolar

H_2: Fricatives: alveolar > labiodental > velar > retroflex > alveopalatal

H_3: Affricates: alveolar > retroflex > alveopalatal

Because oromotor maturation could only apply to lingual obstruents, the following hypotheses are formulated:

H_4: Stops: velar > alveolar

H_5: Fricatives: alveopalatal > alveolar / retroflex

H_6: Affricates: alveopalatal > alveolar / retroflex

According to the language-specific account for the phonological acquisition order，the following three hypotheses are formulated：

H_7：Stops：alveolar ＞ velar ＞ bilabial

H_8：Fricatives：retroflex ＞ alveopalatal ＞ velar ＞ labiodental ＞ alveolar

H_9：Affricates：alveopalatal ＞ retroflex ＞ alveolar

A final hypothesis was formulated according to the continuous theory of speech development.

H_{10}：Child speech development unfolds in a continuous fashion，not in a discrete manner.

Chapter 2 Literature review

This chapter consists of four sections. Section 2.1 describes the articulation and acoustics of obstruents. Section 2.2 introduces the theoretical frameworks related to the phonological acquisition order of speech sounds, and the theories regarding the discontinuity and continuity of speech development. Section 2.3 provides a detailed review of literature regarding the phonetic development of obstruents. In Section 2.4, there is a summary of the literature review.

2.1 Articulation and acoustics of obstruents

Speech sounds produced with a significant degree of frication or transient noise excitation are classified as obstruents. Obstruent speech segments may contain these sound sources in isolation or in combination with voicing. Stops, fricatives, and affricates all fall into this class of speech sounds, and are differentiated in virtue of the place of the vocal tract constriction. In this section, the articulation and acoustics of Mandarin voiceless obstruents will be elaborated. Section 2.1.1 describes the articulatory mechanism of stops, fricatives and affricates, respectively. In Section 2.1.2, speech acoustics of stops, fricatives and affricates are provided on the basis of articulatory mechanism.

2.1.1 Articulation of obstruents

1)Stops

Stop consonants are characterized by an explosion or burst of air from the vocal tract. This burst is initiated when an active articulator such as the tongue or lip creates a complete constriction within the vocal tract, thereby obstructing the outward flow of air through the oral cavity. The production of stops is so complex that about 70 muscles are involved, and this process consists of three phases: closing, holding and releasing. First, a complete constriction is formed at a specific point in the oral cavity, and the velopharyngeal port prevents air from escaping from the nasal cavity, then the oral closure is held to accumulate the intraoral pressure and speakers simultaneously adjust the status of glottis, and vocal fold to form the voicing feature for voiceless and voiced stops. A burst release occurs when the intraoral pressure reaches a certain degree. Similar to the noise source found in fricatives, the stop burst is also composed of turbulent noise, except that it is transient in nature. Most stop bursts are between 10 and 50 ms in duration (Fry, 1979), whereas the continuous noise source associated with fricative sounds is commonly over 150 ms in length (Jongman et al., 2000). The complete articulation of stops requires delicate coordination of oral and laryngeal articulators. The temporal coordination of the burst release, the expiration of airflow and vocal fold vibration characterizes the voicing and aspiration of stops in a specific language. When the articulators are moving to approximate the vocal tract configuration of the following voiced sound, the stop burst is usually followed by a transitional period.

Mandarin stop system is composed of six categories of stops across three places of articulation, bilabial, alveolar and velar. There is an aspirated and an unaspirated stop in each place. Specifically, they are /p^h, p, t^h, t, k^h, k/. The production of bilabial stops requires full closure of two lips and holding for a certain time until the oral pressure reaches a certain amount and then a burst occurs. If the burst is followed by a long flow of air, an aspirated bilabial stop is produced. If not, an unaspirated bilabial is produced. In order to produce an alveolar stop, the speaker has to raise the tongue blade to the alveolar area to form a complete constriction, and then a burst release occurs after the oral pressure accumulates to a certain amount. Consequently, a voiceless aspirated or unaspirated stop is produced depending whether there is a strong aspiration. Finally, velar stops are produced by raising tongue dorsum against the velar area to form a complete closure, then a burst release occurs, thus a velar stop is produced.

2) Fricatives

The production of fricatives requires more fine-grained control of the tongue relative to oral stops and nasals, over which children have no stable control, and they need to practice their tongues in order to correctly produce the frication (Kent, 1992). Fricatives are characterized with turbulence produced in the airflow passing through the oral cavity or pharynx. In the production of turbulence, there is a close approximation of two articulators, which forms a narrow constriction at a certain point in the vocal tract. As a result, the airstream is partially obstructed and turbulent airflow is produced as a steady airflow passes through the narrow constriction. The resultant fluctuation in airflow is the source for frication, which can be generated at an obstacle or a wall (Shadle, 1990). For example, the frication of /s/, /ʃ/ and /ʂ/, is generated at the rigid body (the upper and lower teeth) as an obstacle, which is perpendicular to the airflow, while a wall source appears when the frication is produced mainly along the rigid part parallel to the airflow, such as /ɕ/, /x/ and /ɤ/. For labiodentals such as /f/ and /v/, the turbulence is generated at the constriction itself, without an obstacle or a wall. Mandarin has five voiceless fricatives in terms of place of articulation: labiodental /f/, alveolar /s/, retroflex /ʂ/, alveopalatal /ɕ/, and velar /x/. Labiodental /f/ is produced by forming a narrow constriction between the lower lip and upper front teeth. The production of /s/ involves the approximation of tongue tip or blade and the alveolar ridge. Therefore, /s/ is an apical or laminal dental alveolar or alveolar fricative. In the production of /ʂ/, a narrow constriction point is made between the tongue tip and the post alveolar area. Retroflex /ʂ/ is not described in terms of place of articulation due mainly to the fact that there are many variations documented in world's languages. Some are produced with a raising tongue tip toward the palatal area, while others are produced with a flat tongue shape (Ladefoged & Maddieson, 1996). Mandarin retroflex /ʂ/ has a much flatter tongue shape compared with retroflex sounds in other languages. Mandarin alveopalatal /ɕ/ is produced by raising the tongue blade toward the alveolar ridge with the bunching of the tongue body. The constriction point is formed at a place between that of /s/ and /ʂ/, forming a long flat channel with the palate. The fricative /x/ is produced furthest back in the oral cavity in Mandarin by raising the back of the tongue towards the soft palate.

3) Affricates

Affricates are produced with a transient and frication noise. They are produced by stopping the airflow somewhere in the vocal tract, and then releasing it relatively slowly. Mandarin affricates begin with a complete obstruction of the oral cavity near the alveolar ridge. However, the articulators do not widely separate in order to form a burst of air release, which is different from the formation of a burst of stops. Instead, they slightly separate from each other. When the air passes through the narrow opening between the articulators, the frication noise occurs. Therefore, an affricate is considered as a sequence of a stop followed by a fricative. The same acoustic parameters are used to study both fricatives and affricates. Mandarin has six affricates, differing in manner and place of articulation. They are alveolar /ts^h, ts/, post-alveolar/ retroflex /$tʂ^h$, tʂ/ and alveopalatal /$tɕ^h$, tɕ/ (Lin & Wang, 1992).

2.1.2 Acoustics of obstruents

1) Temporal measures

Duration is an important parameter to distinguish voiced from voiceless consonants, aspirated from unaspirated, and to differentiate sibilant from nonsibilant fricatives (Baum & Blumstein, 1987; Behrens & Blumstein, 1988; Crystal & House, 1988; Jongman et al., 2000).

With respect to stops, this temporal cue is employed to cue voicing and aspiration in many languages. Voice onset time (hereinafter: VOT), defined as the time interval between the burst release and the beginning of the following glottal pulse, is regarded as a primary acoustic cue to voicing and aspiration, and is used to distinguish stop consonants in various languages (Cho & Ladefoged, 1999; Lisker & Abramson, 1964). The vibration of vocal fold can occur before, simultaneously with, or after the burst release, by which stops are categorized into three types: pre-voiced (negative values), short-lag (0-25 ms), and long-lag (60-100 ms) stops (Lisker & Abramson, 1964).

According to Lisker and Abramson's(1964) classification, Mandarin has a two-way contrast, short-lag and long-lag stops across three places of articulation: labial, alveolar and velar. Mandarin stops are phonemically and phonetically produced as voiceless aspirated and voiceless unaspirated stops, with no cognate voiced category. Aspiration is the only acoustic parameter used to differentiate the two categories and distinguish meanings. Mandarin stop system is composed of six categories of /p^h, t^h, k^h, p, t, k/ (Ogasawara, 2011; Rochet & Fei, 1991). However, this classification is too general to capture the subtle differences between Mandarin and other languages such as English. Cho and Ladefoged (1999) further classifies voiceless aspirated and unaspirated stops into four categories based on mean VOT values, unaspirated (around 30 ms for velar stops), slightly aspirated (around 50 ms), aspirated (around 90 ms) and highly aspirated (over 90 ms). Mandarin /p^h, t^h, k^h/ belong to highly aspirated category, and Mandarin /p, t, k/ fall into the short-lag category (Chao & Chen, 2008). VOT also varies as a function of place of articulation. The further posterior the constriction is, the longer the VOT will be. Therefore, velar stops have longer VOT values than bilabial and alveolar stops (Cho & Ladefoged, 1999; Morris et al., 2008). However, the difference in VOT values of bilabial and alveolar stops is not consistent. Some studies showed that the VOT of alveolar stops was much higher than that of bilabial stops, whereas others revealed that there was no significant difference in VOTs between them. Perceptual studies of VOT as a cue to place of articulation also arrived at inconsistent conclusions. For example, Lisker (1978), Massaro and Oden (1980) found VOTs as reliable cues to place of articulation, but Tsui and Ciocca (2000) suggested that normal listeners primarily relied on burst release and formant transition information rather than VOT to discriminate Cantonese stops in terms of place of articulation.