中国社会科学院创新工程学术出版资助项目

中国社会科学院文库 · 文学语言研究系列
The Selected Works of CASS · Literature and Linguistics

宁波话元音的语音学研究

A PHONETIC STUDY OF THE VOWELS IN NINGBO CHINESE

胡方 著

中国社会科学出版社

中国社会科学院创新工程学术出版资助项目

中国社会科学院文库 · 文学语言研究系列
The Selected Works of CASS · Literature and Linguistics

宁波话元音的语音学研究

A PHONETIC STUDY OF THE VOWELS IN NINGBO CHINESE

胡 方 著

中国社会科学出版社

图书在版编目(CIP)数据

宁波话元音的语音学研究/胡方著.—北京：中国社会科学出版社，2014.4

ISBN 978-7-5161-4192-2

Ⅰ.①宁… Ⅱ.①胡… Ⅲ.①吴语—元音—方言研究—宁波市—英文 Ⅳ.①H173

中国版本图书馆 CIP 数据核字(2014)第 078169 号

出 版 人 赵剑英
责任编辑 张 林
特约编辑 贾 媛
责任校对 张冬梅
责任印制 戴 宽

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号（邮编 100720）
网 址 http://www.csspw.cn
中文域名:中国社科网 010-64070619
发 行 部 010-84083685
门 市 部 010-84029450
经 销 新华书店及其他书店

印 刷 三河市君旺印务有限公司
装 订 三河市君旺印务有限公司
版 次 2014 年 4 月第 1 版
印 次 2014 年 4 月第 1 次印刷

开 本 710×1000 1/16
印 张 28.5
插 页 2
字 数 448 千字
定 价 68.00 元

凡购买中国社会科学出版社图书,如有质量问题请与本社联系调换
电话:010-64009791

《中国社会科学院文库》出版说明

《中国社会科学院文库》（全称为《中国社会科学院重点研究课题成果文库》）是中国社会科学院组织出版的系列学术丛书。组织出版《中国社会科学院文库》，是我院进一步加强课题成果管理和学术成果出版的规范化、制度化建设的重要举措。

建院以来，我院广大科研人员坚持以马克思主义为指导，在中国特色社会主义理论和实践的双重探索中做出了重要贡献，在推进马克思主义理论创新、为建设中国特色社会主义提供智力支持和各学科基础建设方面，推出了大量的研究成果，其中每年完成的专著类成果就有三四百种之多。从现在起，我们经过一定的鉴定、结项、评审程序，逐年从中选出一批通过各类别课题研究工作而完成的具有较高学术水平和一定代表性的著作，编入《中国社会科学院文库》集中出版。我们希望这能够从一个侧面展示我院整体科研状况和学术成就，同时为优秀学术成果的面世创造更好的条件。

《中国社会科学院文库》分设马克思主义研究、文学语言研究、历史考古研究、哲学宗教研究、经济研究、法学社会学研究、国际问题研究七个系列，选收范围包括专著、研究报告集、学术资料、古籍整理、译著、工具书等。

中国社会科学院科研局

2006 年 11 月

Abstract

This dissertation investigates the production of the vowels in Ningbo Chinese from the acoustical, articulatory and aerodynamic perspectives. It interprets the linguistic vowel features, such as vowel height, vowel backness, lip rounding, nasality, and duration by examining the acquired acoustic, articulatory, and aerodynamic data. The diphthongs in Ningbo are investigated by examining the diphthong targets and the movement between them in terms of their acoustical and articulatory characteristics. Twenty speakers, ten female and ten male, provided the speech data for acoustical analyses; seven speakers, three female and four male, provided articulatory data; and six speakers, three female and three male, provided aerodynamic data of the nasal vowels and vowel nasalization.

The acoustic results show that vowel height and backness can be best interpreted in a perceptually scaled acoustic F1/F2 vowel plane. Ten normal-length vowels in Ningbo form a pattern of four levels of vowel height and three degrees of vowel backness. Results from the articulatory study show that there is no clear correlation between linguistic vowel height and tongue height based on the individual speaker's data. Nevertheless, a PARAFAC modeling of the lingual articulation successfully decomposes the complicated tongue shapes into two underlying lingual movement mechanisms, namely "retraction and back raising" (retraction of the tongue body and raising of the tongue dorsum towards the velum) and "front raising" (raising of the front part of the tongue towards the hard palate). Using the extracted two tongue movement mechanisms, the linguistic vowel backness and vowel height

can thus be defined in a purely tongue-based and speaker-independent representation of the vowels. Both the acoustic and articulatory data suggest that there is a three-way distinction of lip rounding in Ningbo: spread for [i], horizontal protrusion for [y], and vertical protrusion for [ʏ]. The acoustic data show that the two apical vowels [ɿ ʮ] in Ningbo are in a semi-high central position in the F1/F2 plane, and the articulatory data suggest that the Ningbo apical vowels are best understood as doubly articulated vowels, i.e., being apex and dorsal simultaneously. The acoustic data also show that Ningbo vowels, in particular the high, mid-high, and low vowels, tend to be lowered when acquiring nasality. At the same time, the articulatory data reveal that for most speakers lip protrusion is involved during the production of nasal vowels, possibly enhancing the difference in vowel quality between the nasal vowels and their oral counterparts.

Results of formant frequency analysis of the diphthongs in Ningbo show that in general, both the falling and rising diphthongs begin in an onset frequency area close to their target vowels, whereas only the normal-length rising diphthongs end in the frequency region close to the offset targets, while the falling diphthongs and short rising diphthongs do not reach the offset targets. The articulatory data are generally consistent with the acoustic data. With respect to the dynamic aspects of diphthongs, the acoustic data show that the Ningbo diphthongs can be characterized by the formant (F2) rate of change if only the falling diphthongs are concerned. But the F2 rate of change fails to distinguish the rising diphthongs from the falling diphthongs. Meanwhile, the articulatory data of tongue kinematics show that the average velocity and peak velocity can be used to characterize Ningbo diphthongs, given that the falling and rising diphthongs are considered separately. The time to peak velocity serves as a better criterion in distinguishing the falling diphthongs from the rising diphthongs. In short, peak velocities usually occur late in the falling diphthongs, but early in the rising diphthongs. This is because the

peak velocity of the lingual articulator occurs during the gestural change, i.e. during the spectral transition from an acoustical point of view, from diphthong onset to offset. The falling diphthongs in Ningbo have steady states on both the onset and offset elements, while the rising diphthongs only have steady states on the offset elements.

Both the acoustic and articulatory data suggest that the main difference between a short and normal-length vowel or diphthong in Ningbo is in duration, which differs from the tense/lax distinction in English, where lax (short) vowels are mainly correlated with lower tongue positions vis-à-vis their tense (long) counterparts.

The aerodynamic data reveal that nasality can be successfully quantified using the acquired flow data. Results show that in Ningbo the nasal vowel [ɔ̃] has higher nasality than the other nasal vowel [ã]. Regarding the three nasalized vowels, the high vowel [ĩ] has the highest nasality, followed by [õ] and [ə̃] in descending order.

In this study, in addition to searching for the physical and/or physiological correlates for vowel features, attention was also paid to examining articulatory mechanisms and articulatory-to-acoustic relations in vowel production. The articulatory data show that jaw movements are usually coordinated with lingual gestures during the vowel and diphthong production. However, different speakers may employ different articulatory strategies to achieve similar acoustic goals. The data also suggest that the three Ningbo point vowels exhibit a quantal nature during production. It should be noted that the relationship between speech articulation and speech acoustics has been shown to be complex by the comparisons of the articulatory and acoustic data of the diphthong targets in Ningbo. Further studies are necessary in order to have a better understanding of the nature of vowel production and the relationship between articulation and acoustics.

Table of Contents

Chapter One: Introduction

1.1 General theoretical background of vowel production

Since its initiation by Bell (1867) during the mid nineteenth century, it has been the tradition to describe vowels based on certain basic articulatory features: high-low, front-back, and rounded-unrounded. This descriptive model prescribes for each vowel a unique tongue position in terms of the height and backness of the tongue arch in the vocal tract. This model is widely accepted in the field and constitutes an essential base both for the IPA vowel chart and for the distinctive features in phonological theory (Chomsky & Halle, 1968). In the British phonetic tradition, the vowel height and backness are explicitly described in terms of the highest point of the tongue (Jones, 1909). For instance, according to Jones (1956), Cardinal Vowels 1, 2, 3, and 4 ([i e ɛ a]) are "vowels of the 'front' series" and Cardinal vowels 5, 6, 7, and 8 ([ɑ ɔ o u]) are "vowels of the 'back' series". In purely articulatory terms, he defines Cardinal vowel 1 ([i]) as "the sound in which the raising of the tongue is as far forward as possible and as high as possible consistently with its being a vowel" and Cardinal Vowel 5 ([ɑ]) as "a sound in which the back of the tongue is lowered as far as possible and retracted as far as possible consistently with the sound being a vowel" (p. 31). From Cardinal vowel 1 to 2 and to 3 and to 4, "the tongue is lowered through approximately equal intervals"; similarly, from Cardinal Vowel 5 to 6 and to 7 and to 8, "the tongue is raised through approximately equal though smaller intervals" (p. 32).

However, this classical model of vowel description has never been validated by any empirical articulatory data. In fact, it has been contradicted by a large number of the lingual articulation data based on various kinds of techniques in the literature. For example, as early as in 1910, Meyer found that [ɪ] has an unexpected lower tongue position than [e] based on his plastopalatogram data. More convincing is a massive X-ray study from Russel (1928) which conclusively discredited the traditional height/backness articulatory description of vowels. Based on a large number of X-ray data of American English speakers, he found that the tongue position was often not as predicted according to Bell's model. For example, [ɪ] might be lower than [e]; [u] was possibly a front vowel; the articulation position for [ɑ], [ɔ] and [ɒ] was rather in the pharynx cavity.

Russell's work did not attract much attention from his contemporary phoneticians. Several decades later, the problem was highlighted by Ladefoged and his colleagues (Ladefoged, 1967, 1971, 1975, 1976; Ladefoged, De Clerk, Lindau & Papçun, 1972). Like Russell, Ladefoged pointed out that the traditional vowel descriptive model actually described the auditory impressions of vowels but translated them into physiological terms, because the traditional vowel articulation dimensions like high-low and front-back did not correspond to the physiological reality; rather they were correlated with vowels' acoustic features. For instance, Ladefoged et al. (1972) found that the tongue contour is almost identical for the vowels [ɪ e ɛ] in some American English speakers. Furthermore, Ladefoged (1975, 1976) pointed out that the highest points of the tongue in a set of Cardinal Vowels produced by Jones "are not at all similar to the descriptions of that have been given of these vowels". And "there is not even a good separation between front vowels and back vowels"; "and the relative heights are wildly wrong" (1976: 10-12). Similar contradictions between the traditional vowel descriptive model and actual vowel articulation were also found in the cineradiographic data of Perkell (1969) (see Lieberman, 1976). The psycho-acoustical experiment in Ladefoged (1967) showed that trained phoneticians could make fairly accurate judgments of the vowel height and backness even when working from the recorded data, which led him to believe that vowel height and backness has nothing to do with what has done with the tongue; rather, it is just because these trained phoneticians "could label the formant structure of the vowels they heard and imitated" (Fromkin, 1985: 5). As illustrated in Ladefoged (1975, 1976), the traditional Cardinal Vowel chart correlates well with an auditorily scaled acoustic F1/F2 (or F1/F2-F1) vowel chart.

Although to some phoneticians the criticism of the traditional vowel descriptive system was an exaggeration (e.g., Catford, 1981; Fischer-Jørgensen, 1985), the acoustical or acoustical-auditory proposal for vowel description was widely accepted by phoneticians and there was a broad consensus that traditional vowel parameters such as height and backness should be better understood as acoustically or perceptually defined than as articulatorily defined (e.g., Jakobson et al., 1952;

Ladefoged et al. 1972; Lieberman, 1976; Lindau, 1978[1]; Nearey, 1978; Wood, 1982). The apparent inconsistency found in the vowel articulation led researchers to reconsider the articulatory dimensions used in vowel description. As Wood (1979: 25) stated, the confusing pictures obtained from X-ray studies of vowel articulation were due to "the wrong articulatory variables (height and fronting) rather than to articulatory irregularity". In the modeling study of vowel production, Stevens & House (1955) rejected height and backness and replaced them with a new parameterization, namely the degree of vocal tract constriction and the location of constriction relative to the glottis. These dimensions are claimed to be more directly related to vowel acoustics, because the constriction location determines the size of the front and back cavity in the vocal tract and the degree of constriction affects the coupling of the two cavities.

The proposed articulatory dimensions have become the basis of the mainstream in the field of speech production studies. The proposal completely changes the understanding of vowel articulation. While in the general studies of speech production, the length of the vocal tract was treated as a continuum of constriction locations (Stevens & House, 1955; Fant, 1960; Lindblom & Sundberg, 1971), effort has been made to find the explicit constriction locations in vowel production. In the quantal theory of speech production, Stevens (1972, 1989) identified three constriction regions where vowel spectra are relatively insensitive to moderate variations of constriction location: at the palate for [i], at the velum for [u], and in the pharynx for [ɑ]. Based on both the empirical radiographic analysis from many languages and the theoretical three-parameter nomograms by Stevens and House (1955), Wood (1979) found there are four constriction locations for vowels: along the hard palate for [i-ɛ]-like and [y-ø]-like vowels, along the soft palate for [u-ʊ-ɨ]-like vowels, in the upper pharynx for [o-ɔ]-like and [ɤ]-like vowels, and in the lower pharynx for [ɑ-a-æ]-like vowels (see also Gunnilstam, 1974, for a purely theoretical inference of constriction locations from Fant (1960)'s three-parameter nomograms). He stated that these four narrowly

[1] Lindau (1978) was somewhat different in that, based on some new cineradiographic recordings, she claimed that vowel height and backness also correlate with the articulatory data. So, from her point of view, the vowel features of High and Back can be accounted for articulatorily and acoustically and auditorily. However, she admitted that other features such as Peripheral should be best described in the acoustic domain.

constricted locations are those places where F1 and F2 are least sensitive to variations of constriction location and thus it supports Stevens's proposal of the quantal nature of vowels. The new constriction location/degree model provides new dimensions for investigating speech production and sheds light on the uninvestigated aspects of tongue mechanism and even the underlying speech motor control system. In the past several decades, development of new techniques such as x-ray microbeam and electromagnetic articulograph makes it possible to collect large-scale articulatory data, based on which theoretical aticulatory models can be tested by the empirical data. Pilot studies of Perkell & Nelson (1982, 1985) and Perkell & Cohen (1989), using articulatory data obtained from x-ray microbeam, tested the articulatory stability of the production of three quantal vowels [i a u]. Their results basically supported Stevens's quantal theory. More recently, following Perkell and his coworkers' pilot studies, Beckman et al. (1994), using a more quantifiable method, revealed that many vowels, not just the point vowels [i u ɑ], "have target articulatory constrictions and are more or less quantal, in the sense that constriction degree seems to be controlled more precisely than constriction location" (pp. 489). The palatal vowel [i] has the smallest variability, the velar vowel [u] shows the next smallest variability, and the other vowels [ɑ æ ɚ] are less quantal.

1.2 Purpose and the scope of this study

Developments in the field of speech production have shed new light on the understanding of the underlying mechanisms of vowel production and provided us with new insights for examining and reinterpreting the linguistic vowel features. However, what should be pointed out is that previous researches on vowel production are mostly theoretically orientated. Little effort has been made to explore vowel production thoroughly in a specific language. The articulatory mechanisms for vowel distinctions are far from clear. For instance, it still remains unclear how the primary vowel features such as vowel height and backness are correlated with vowel articulation and can thus be interpreted from a linguistic point of view. Although they have been well interpreted in the acoustical or psycho-acoustical domain, vowel height and backness need an explicit explanation in the articulatory domain as well, i.e., through what kinds of articulatory mechanisms are these acoustical distinctions achieved? Thus, a close examination of vowel production in a specific language is of great importance. On the one hand, it gives a complete picture of vowel production in that particular language, making it possible to interpret the linguistic vowel features in both acoustical and physiological domains, and on the other hand, the empirical production data can be used to test the validity of the current theories of vowel production.

The present study is an acoustic and articulatory investigation of vowel production in Ningbo Chinese. The purpose is twofold. On the one First, the study is empirical, and aims to ascertain how the vowel system of a language can be described in the acoustic or articulatory domain or both. Second, it explores the relationship between the acoustic and articulatory aspects of vowel production based on the observed empirical dnta. As mentioned earlier, previous studies have shown that phonological vowel description such as height and backness is based on acoustical or psycho-acoustical characteristics, though it remains unclear how acoustic goals are achieved with articulatory gestures. As only a few vowels (in most cases, only the three point vowels) of a few languages (mainly English or other a few European languages) were examined in past studies, this study is intended to contribute to a more comprehensive knowledge of the articulatory mechanisms and their acoustical consequences during vowel production. The Ningbo Chinese has several advantages for investigating vowel production. First, it has a rich inventory of monophthongs and diphthongs. Second, most of the vowels in Ningbo can constitute meaningful monosyllabic words by themselves and thus are free from consonantal influences.

In addition to the questions concerning vowel features of height and backness, the scope of the present study covers the following issues:

(1) *Vowel distinction in lip gestures.* In most languages of the world, the lip gesture of vowels is either predictable (front vowels are usually unrounded and back vowels are usually rounded) or contrasts in roundedness. It is reported that there is a three-way distinction of lip rounding for the high vowels in Swedish (see 2.3.1.4 for discussion in length and Ladefoged & Maddieson, 1990, for a review). Similarly, the data suggest that the high front vowels in Ningbo contrast in three ways of lip rounding. But unlike the Swedish case, which distinguishes lip protrusion from lip compression (also known as 'inrounded'), Ningbo seems to distinguish in horizontal vs. vertical lip protrusion.