Université Sorbonne Nouvelle – CNRS
École Doctorale 622 : Sciences du Langage
UMR7018 / Laboratoire de Phonétique et Phonologie

Thèse en vue de l'obtention du grade de

DOCTEUR EN PHONETIQUE

The apical vowel in Jixi-Hui Chinese: phonology and phonetics

Présentée par

Bowei SHAO

Sous la direction de

Rachid RIDOUANE

Soutenue à Paris le 23 octobre 2020 devant le jury composé de :

La voyelle apicale en Chinois Jixi-Hui : phonologie et phonétique

Les langues chinoises ont un ensemble de segments appelés « voyelles apicales » (舌尖元音 en chinois). Leur nature exacte est à l'origine d'un débat toujours en cours : S'agit-il de consonnes ou de voyelles ? Les « voyelles apicales » ont été analysées dans des études précédentes comme étant de véritables voyelles, des voyelles fricatives, des fricatives syllabiques, ou des approximantes syllabiques. Cette thèse porte sur la voyelle apicale attestée en chinois Jixi-Hui. J'examine ce segment d'un point de vue phonétique et phonologique et montre qu'il est mieux défini comme une consonne fricative voisée (transcrit /z̩/).

Phonologiquement, ce segment est un phonème distinct de /i/. Il est exclusivement attesté en position de noyau de syllabe où il constitue une unité porteuse de ton. Il peut apparaître non seulement après les sibilantes coronales /s ts ts^h/, mais aussi les bilabiales /p p^h/ et les nasales /m n/. Phonétiquement, les caractéristiques acoustiques et articulatoires de ce segment sont examinées. Les résultats montrent que /z̩/ contient dans la majorité des cas un bruit de friction dans sa phase initiale, superposé sur du voisement, avec une structure formantique plus claire apparaissant vers la fin. Les analyses du rapport harmonique/bruit et du taux de passage par zéro confirment cette présence significative du bruit de friction, distinguant clairement ce segment des voyelles. Les généralisations en SS ANOVA des données ultrasoniques montrent que /z̩/ a une forme de langue presque identique à celle de /s/ sur les plans mi-sagittal et coronal, malgré quelques différences spécifiques à chaque locuteur. Cette forme de langue reste constante dans les contextes consonantiques bilabiales et alvéolaires.

La variabilité dans la façon dont /z̩/ est phonétiquement implémentée est considérée comme étant la conséquence de deux contraintes en interaction : une contrainte structurelle liée au statut distinctif de /z̩/ et au rôle qu'il joue dans la structure syllabique, et une contrainte physique liée à l'incompatibilité entre le voisement et le bruit de friction. L'étude souligne également la nécessité de reconnaître les fricatives syllabiques en chinois Jixi-Hui, et probablement aussi dans d'autres langues chinoises.

Mots-clés : Voyelle apicale, Chinois de Jixi, acoustique, ultrason, bruit de friction, fricative syllabique.

The apical vowel in Jixi-Hui Chinese: phonology and phonetics

Chinese languages have a set of segments known as ‘apical vowels’ (舌尖元音 in Chinese). Their exact nature is still the source of an ongoing debate: Are they consonants or vowels? ‘Apical vowels’ have been analysed in previous studies as genuine vowels, fricative vowels, syllabic fricatives, or syllabic approximants. This dissertation is concerned with the apical vowel attested in Jixi-Hui Chinese. I examine this segment from phonetic and phonological perspectives and show that it is best defined as a voiced fricative consonant (transcribed /z̩/).

Phonologically, this segment is a distinct phoneme from /i/. It is exclusively attested in syllable nucleus position where it constitutes a tone-bearing unit. It can appear not only after coronal sibilants /s ts ts^h/, but also bilabials /p p^h/ and nasals /m n/. Phonetically, the acoustic and articulatory characteristics of this segment are examined. The results show that /z̩/ contains in the majority of cases frication noise in its initial phase superposed on voicing, and a clearer formant structure appears towards its end. The harmonic-to-noise ratio and zero-crossing rate analyses confirm this significant presence of noise, clearly distinguishing this segment from vowels. The smoothing-spline ANOVA analyses of ultrasound data show that /z̩/ has a near-identical tongue shape to /s/ on both mid-sagittal and coronal planes despite some speaker-specific differences. This /s/-like tongue shape is constant in bilabial and alveolar consonantal contexts.

The variability in the way /z̩/ is phonetically implemented is argued to be a consequence of two interacting constraints: a structural one related to the distinctive status of /z̩/ and the role it plays within syllable structure, and a physical one related to the incompatibility of voicing and frication. The study further argues for the necessity of recognizing syllabic fricatives in Jixi-Hui Chinese and probably also in other Chinese languages.

Key words: Apical vowel, Jixi-Hui Chinese, acoustics, ultrasound, frication noise, syllabic fricative.

À Florian

Acknowledgement – Remerciements

Je tiens avant tout à remercier mon directeur de thèse, Rachid Ridouane, qui a toujours su trouver la bonne direction au fil des années. Sans son dévouement, sa motivation et son orientation, cette étude n'aurait sans doute jamais abouti.

I am greatly honoured to have Yiya Chen, James Scobbie, Ioana Chitoran, Pierre Hallé and Rudolph Sock on the jury. I am particularly indebted to James Scobbie for his suggestions and criticism of my earlier work, for his examples and demonstrations on ultrasound data processing and for his hospitality when I visited the CASL Research Centre.

Je tiens à exprimer mes sincères remerciements à mes collègues du LPP. Merci en particulier à Cédric Gendrot et Nicolas Audibert pour leurs scripts et conseils. Merci également à Angélique Amelot, Tanja Kocjančič Antolík et Roland Trouville pour leur aide dans l'utilisation du matériel. J'ai beaucoup bénéficié des nombreuses discussions avec Pierre Hallé, Naomi Yamaguchi, Cécile Fougeron, Annie Rialland, Shigeko Shinohara et Didier Demolin. Merci aussi à Jacqueline, Claire, Martine, Léonardo, Jiayin, Shi, Gabriele, Fanny, Jane, Yaru, Daria, Amélia, Alexis, Angélina, Djidji, Emré, Nora, ainsi qu'à tous les phonéticiens, phonologues et ingénieurs du LPP, vous me faites apprendre tous les jours. Un merci tout particulier à Gwénaëlle pour son aide précieuse durant ces années.

I beneficiated a great deal from the discussions with Alan Wrench, on the utilisation of AAA and the ultrasound technique. I would also like to thank Eleanor Lawson, Steve Cowen, Jane Stuart-Smith and Fabienne Westerberg for their technical support and exchanges. Thanks as well to all the members of CASL Research Centre who have made my visit fruitful and enjoyable.

I would like to thank my parents for their support and patience. They speak neither English nor French and have only a vague idea of what I have been doing all these years, but they have chosen to 'let me be'. They are both passionate about local traditions and, of course, the local language. My father helped me to gain access to the facility used for recording sessions. My mother helped in recruiting speakers (herself included), who have been extremely patient and kind. I thank them all sincerely.

Thanks to Danyang, Yunhao, Nana and all of my friends, for their constant support, encouragement and belief.

Ce travail a été financé par un contrat doctoral de l'Université Sorbonne Nouvelle. Il a aussi bénéficié d'une aide du Labex EFL (ANR-10-LABX-0083) et des aides financières du LPP et de l'ED 622.

Merci à Florian, pour avoir toujours eu confiance en moi.

Table of contents

List of figures

List of tables

List of abbreviations

AAA	Articulate Assistant Advanced
HFZ	*Hànyǔ Fāngyīn Zìhuì* 汉语方音字汇
	Phonetic dictionary of Chinese dialects
HMC	Hefei-Mandarin Chinese 合肥话
HNR	Harmonic-to-Noise Ratio
JHC	Jixi-Hui Chinese 绩溪话
QMC	Qinghai-Mandarin Chinese 青海话
SC	Standard Chinese 普通话
SS ANOVA	Smoothing-Spline Analysis of Variance
SWC	Suzhou-Wu Chinese 苏州话
TBU	Tone Bearing Unit
ZCR	Zero-Crossing Rate

Chapter 1
Introduction

> Ask yourself: Why do we talk about “instead of” and “versus” all the time? Why do we partition the world into pairs, contrasting, for example, the genotype and the phenotype, the discrete and the continuous, the individual and the collective, the orderly and the random, the qualitative and the quantitative, the internal and the external, the persistent and the changing, the gradual and the abrupt, the reductionist and the holist—and yes, the certain and the uncertain. The list goes on and on. It is pretty obvious, one intuits, that both represent polarized extremes, and that reality must lie somewhere in between. One might even say that we categorize things and ideas in this polarized fashion in order to be sure that what we are really after will be captured in between. One might even say that all science is about the in-between. One might, indeed. Few have.
>
> J. A. Scott Kelso (2005)

Vowel-consonant dichotomy is a cardinal distinction in the study of speech sounds. Researchers in phonetics and phonology use these notions as if they were obvious and self-evident, despite the fact that different criteria are applied within different approaches. These approaches generally leave a residue of ‘dubious’ sounds (O’connor & Trim, 1953) that raise analytical problems. This dissertation is concerned with one type of these ‘dubious’ sounds: apical vowels.

Chinese languages are known to have ‘apical vowels’, termed 舌尖元音 in Chinese. Their exact nature is still the source of an ongoing debate: Are they consonants or vowels? Their phonetic nature suggests that they are sibilant or fricative consonants, but they phonologically pattern with vowels. This dissertation is about the apical vowel attested in Jixi-Hui Chinese 绩溪话, a language of Hui group 徽语 spoken in southern Anhui 安徽.I examine this segment from phonetic and phonological perspectives and conclude that it is best defined as a voiced fricative consonant (hereafter transcribed /z̩/). Furthermore, I will argue for the necessity of syllabic fricative consonants in Jixi-Hui Chinese (JHC) and probably also in other Chinese languages.

The existing literature on apical vowels is in majority based on Standard Chinese (SC). The terminology proposed for this set of segments was coined based on the description of this language (Karlgren, 1915). Two variants of ‘apical vowels’ are attested in SC ([ɿ] and [ʅ] in non-IPA notations) and occur only following alveolar sibilants. These two sounds are in complementary distribution with [i], and are naturally analysed as allophonic variants of this vowel. Trubetzkoy (1969: 171) described them phonetically as ‘*a type of vowel with a much lesser degree of aperture and with a much more fronted position of articulation than, for example, i, so that a frictionlike noise resembling a humming is audible in its production*.’ This frication noise is generally argued to be a mere consequence of the homorganicity between the apical vowels and the preceding sibilant consonants.

The apical vowel in JHC is a typologically interesting case study, as it displays two structural properties that make it different from the most studied variants in other Chinese languages: It is a separate phoneme contrastive to /i/ and to other vowels and it occurs not only following alveolar sibilants but also after bilabial plosives /p pʰ/, bilabial nasal /m/ and alveolar nasal /n/.

1.1. Aim of the dissertation

This thesis provides phonological and phonetic analyses of the apical vowel in JHC. On the phonological level, a detailed description of this segment is carried out based on common principles of phonological analyses: lexical distribution, phonemic contrast, and function within the syllable. On the phonetic level, two production experiments are conducted in order to examine the (i) acoustic and (ii) articulatory characteristics of this segment.

(i) The acoustic experiment analyses various spectral and temporal properties displayed by apical vowels. The presence or absence of frication noise is analysed qualitatively, its duration is quantified in different contexts, and its formant structure is compared to that of vocalic segments. In a following step, focus is put on the frication noise, as this is the hallmark of apical vowels. Frication noise is examined in different contexts using Harmonic-to-noise ratio (HNR) and Zero-crossing rate (ZCR), and furthermore, the nature of the coarticulatory effects between onset consonants and apical vowels is reported, with a special focus on the labial plosives.

(ii) On the articulatory level, apical vowels are investigated using ultrasound tongue imaging method (Articulate Instruments Ltd., 2012), on both the mid-sagittal plane and the coronal plane. On both planes, the tongue contours are quantitatively analysed using the smoothing-spline analysis of variance (SS ANOVA). The general tongue contours are analysed and different sets of quantitative and qualitative comparisons are conducted in order to determine the articulatory properties of the apical vowel and show how they dynamically vary depending on the nature of the preceding onset consonant.

The presence of apical vowels in Chinese languages has always been related to a historical vowel /i/ in some stage of the evolution (Zhu, 2004; Zhao, 2007). The voiced fricative /z̩/ in JHC could be one step within this evolution, while in other Chinese languages an approximant-like sound may be observed. In this dissertation, I will argue for an extension of the nucleic inventory of JHC from the most vocalic sounds to nasals, approximants and voiced fricatives. All these segments pattern as a class defined using the features [+ continuant, + voiced].

1.2. Structure of the dissertation

Chapter 2 introduces relevant concepts and terminologies as the background for this study. First, I will comment on the vowel-consonant dichotomy to pave the way for discussing the vowel-consonant continuum and the position of the apical vowel within this continuum. Then I will present a general review of the phonetics and phonology of apical vowels in Chinese languages, as well as several 'similar' segments in non-Chinese languages.

Chapter 3 provides a thorough description of the sounds of JHC. Its consonantal, vocalic and tonal inventories are described with a focus on the syllabic consonants. A phonological description of the apical vowel is provided based on common principles of phonological analysis, including distributional, combinatory and functional attributes of the segment.

Following these chapters, two empirical studies will be presented.

Chapter 4 presents the acoustic study. The basic characteristics of the apical vowel will be analysed, including its duration compared to other syllable nuclei and its formant frequencies compared to other vowels. A focus will be placed on the nature of the frication noise and its comparison to sibilant consonants.

Chapter 5 presents the articulatory study. The tongue configurations of JHC /z̩/ will be explored based on ultrasound tongue imaging, on both mid-sagittal and coronal planes. The ultrasound experiment will more specifically seek to determine whether the shape of the tongue is similar to that of the consonant /s/ or the high vowel /i/.

Chapter 6 recapitulates the main findings and contributions of this thesis. I will present an account to explain the variability in the way apical vowels are phonetically implemented, not only in JHC but also in other Chinese languages. Furthermore, I will discuss the concept of the vowel-consonant dichotomy versus a vowel-consonant continuum and the nature of syllable nuclei in JHC. Finally, I will conclude this chapter by pointing out some unsolved questions and proposing possible avenues for future research.

Chapter 2
Background: Apical vowel and the vowel-consonant dichotomy

2.1. Is the vowel–consonant dichotomy universal?

The distinction between ‘vowel’ and ‘consonant’ is a fundamental division in linguistics. As Pike (1971: 66) pointed out, this division is frequently assumed for descriptions of single languages, with no attempt to define it. In this section, the definitions of vowel and consonant are discussed in the light of the mismatch between their phonological function and their phonetic implementation.

There are two basic approaches in defining vowels and consonants, and the distinction between the two classes. One is based on phonological behaviour and function, such as distributional or combinatory facts (Vendryès, 1921; Grammont, 1946; Hjelmslev, as cited in Troubezkoy, 1949: 96–97; O’Connor & Trim, 1953). The other one is based on phonetic observations, describing the acoustic and/or articulatory facts (Passy, 1906; Sapir, 1921; Bloomfield, 1923; Trubetzkoy, 1969; Chao, 1980; Ladefoged & Johnson, 2011). The present discussion is essentially based on these two approaches.

A third approach, which has a functional angle (Nespor, Peña & Mehler, 2003), can also be found in the literature. In this approach, the vowel and consonant are considered different by the division of labour in the lexicon and syntax: the consonant category conveys information about the lexicon while the vowel category conveys information about grammar. This functional point of view, which does not involve an acoustico-articulatory observation, will not be focussed on in the discussion.

2.1.1. Defining vowel and consonant: more than a century of debate

There has been no consensus on the definition of vowel and consonant. The debate on the fuzzy boundary between the two categories seems to be everlasting since 1890s. When this question is addressed, two approaches are generally found in the literature.

2.1.1.1. Defining the vowel–consonant categories based on phonetics

The first approach is phonetically based, motivated by acoustico-articulatory observations. This is the approach adopted by Passy (1906), Sapir (1921), Bloomfield (1923), Trubetzkoy (1969), Chao (1980) and Ladefoged & Johnson (2011). This approach has an empirical, bottom-up view. The categories of vowel and consonant are defined by generalising common acoustico-articulatory specificities of observed speech sounds.

For Passy (1906) the acoustico-articulatory boundary between vowel and consonant is not categorical. His definitions of the two sound-units are based on the opening of the vocal tract and the acoustic character of the sound: a vowel is a musical sound produced with an open mouth, a consonant is a noise sound produced with a closed or nearly closed mouth. Adopting this definition makes it almost impossible to find an exact limit between the two categories of sounds, since the articulatory adjustment could be insensible from a vowel to a consonant as illustrates the pair [u w].

Sapir (1921) categorises vowels as voiced sounds during which the breath is allowed to pass through the mouth without being checked or impeded at any point. But he also notes that the breath may be momentarily checked or allowed to stream through a greatly narrowed passage with resulting air friction (Sapir, 1921: 51). Sapir does not give any specific example of vowels having 'narrowed passage' and 'friction', but his idea is that there is potentially infinite number of vowels. When the sound has an 'oral resonance chamber', then it is a vowel. Following this direction, the approximant sounds, such as [w j], could be considered as vowels with some 'friction'. The consonants are those in which the stream of breath is interfered in some way. The result of the interference is a sharper, more incisive quality. He further divides the consonants into four groups: stops, fricatives, laterals and trills. It is clear that his categorisation is based on acoustico-articulatory characteristics of the speech sounds. It is unclear, however, if Sapir considers the boundary between vowel and consonant to be categorical.

Bloomfield's (1923) phonetic description is based on IPA principles (1912). His categorisation is nearly identical to that of Sapir (1921). The speech sounds are classed into noise-sounds and musical-sounds. The noise-sounds are further divided into stops, trills and spirants. The musical sounds are further divided into nasals, laterals and vowels. He further gives the definition of vowels as '*modifications of the voiced-sound that involve no closure, friction, or contact of the tongue or lips*' (Bloomfield, 1923: 102). This definition suggests that Bloomfield held a categorical view: when there is closure, friction, or contact of the tongue or lips, the speech sound is not a vowel.

For Trubetzkoy (1969), the categorisation of vowel and consonant can only be based on phonetic characteristics. For him, any definition of the two categories based on criteria other than acoustics and articulation is necessarily flawed. For him (1969: 94), '*what characterizes a consonant is the production of an obstruction and the* *overcoming of such an obstruction. A vowel … is characterized by the absence of any obstruction.*' Ladefoged & Johnson (2011) adopt a similar view. For them (2011: 10–19), in order to form consonants, '*the airstream through the vocal tract must be obstructed in some way*', while in order to form vowels, '*the passage of the airstream is relatively unobstructed*'.

The definition of Trubetzkoy (1969) makes the vowel category and the consonant category mutually exclusive and thus dichotomic, since logically one sound cannot be produced with *AND* without an obstruction. Note, however, that Ladefoged & Johnson's (2011) definition is less categorical: the airstream is '*relatively*' unobstructed can be interpreted as if the airstream during the production of a vowel could be obstructed to a certain degree. This definition can be directly linked to Ladefoged & Maddieson's (1996) 'fricative vowel' category (see section 2.2.2.1), since they consider that vowels (namely the apical vowel in Chinese languages) can have frication as a secondary feature. To produce frication, the airstream must be obstructed to a certain degree.

The phonetic approach, based on acoustico-articulatory observations, argues in majority for a non-dichotomic view of the relationship between the vowel category and the consonant category. It is impossible to determine, for example, the exact boundary where the vowel [i] has a narrowed-enough air passage and becomes an approximant consonant [j].

2.1.1.2. Defining the vowel–consonant categories based on phonology

The second approach is phonologically based, motivated by common phonological principles, such as distributional, combinatory and functional attributes of the sounds. This is the approach adopted by Vendryès (1921), Grammont (1946), O'connor & Trim (1953), and Jakobson & Waugh (1979). This approach has a theoretical, top-down point of view. The vowel and consonant categories are defined based on their phonological patterning, and also by the native speakers' intuitions about the phonology of their language.

Vendryès (1921) proposes that vowels and consonants should be divided on the basis of their role within a syllable. But he notes also that one phoneme can play the role of a vowel or of a consonant. If there is a difference in function, there is no, however, difference in nature. He clearly states that, in nature, there is no clear-cut limit between vowel and consonant.

Grammont (1946) adopts a similar view as Vendryès (1921). He points out that there is no 'impassable border' between the vowel category and the consonant category. His observation is that there are languages where some phonemes are sometimes vowels and sometimes consonants depending on the position they occupy, such as when some consonants can form a syllable either alone or in combination with other consonants. Clear enough, the position within a syllable is a fundamental criterion in defining the vowel category and the consonant category.

This categorisation based on syllable structure has been verified in English (O'connor & Trim, 1953). Their distributional study of the English phonemes has shown that there exists a division of function: some phonemes occur more in the nucleus position, and other phonemes occur more in the margins of the syllable (i.e., onset and coda). The former are vowels, and the latter consonants. This division is completely based on the distribution of English phonemes and can be seen as a dichotomic division of function.

Jakobson & Waugh (1979) holds a nuanced view of vowel and consonant division based on their function. They recognise that the vowel-consonant categorisation is the most cardinal and most obvious bifurcation of speech sounds. The main principle of the differentiation between the two categories is related to their function and distribution: the vowel category functions as the syllable nuclei and consonant category as the margins of the syllable. They however also note that there is a necessity to recognise that the liquid sounds are sonorants, which is a 'shifting' class. That is to say, sonorant sounds have both consonantal and vocalic characters. This means that vowel and consonant categories are not mutually exclusive, since sonorants can be considered as both consonants and vowels.

The phonological approach, based on the patterning of speech sounds, may argue for a dichotomic view of the vowel-consonant categorisation. Under this view, the vowel category functions as the nucleus of a syllable and the consonant category as the margin of a syllable. When it comes to ambiguous sounds (e.g., sonorants), the dichotomic categorisation becomes less evident. These three categories (i.e., vowels, consonants, and sonorants) are further discussed in section 2.1.1.3 based on Feature Theory.

2.1.1.3. Defining the vowel–consonant categories in double or multiple terminologies based on both phonetics and phonology

The studies reviewed above, adopting either the phonetic approach or the phonological approach, propose only one set of terminology, namely vowel and consonant. Some studies, while adopting the two approaches, propose multiple terminologies to account for both phonetic and phonological criteria. This is the case in Saussure (1916), Pike (1947), Greenberg (1962) and Chomsky & Halle (1968). Saussure (1916) proposes vowel/consonant for the phonetic description and *sonante*/*consonante* categorisation for the phonological description. Pike (1947) proposes vocoid/nonvocoid for the phonetic description and vowel/consonant for the phonological description. Greenberg (1962) follows Pike's (1947) categories. Chomsky & Halle (1968) propose sonorant, vocalic and consonantal categories, thus revoking the dichotomy between vowel and consonant.

For Saussure (1916), the vowel-consonant dichotomy has been treated separately on the phonetic level and on the phonological level. For him, the nature of a sound and the function of a sound should be treated separately. The nature of a vowel and the nature of a voiced consonant are comparable since they all include voicing in the phonation, and vowels and voiced consonants differ only by aperture. The oral cavity serves as a resonator in the production of a vowel, while in a consonant, the voicing is intercepted in the oral cavity. Following this point of view, there is no identifiable boundary between vowels and voiced consonants. However, their function differs within a syllable: the vowel is the centre of a syllable, thus having the *sonante* function; while a consonant is the sound that surrounds the vowel in the same syllable, thus having the *consonante* function. Under this double terminology (vowel/consonant, *sonante*/*consonante*), the sound [i] is by nature a vowel but has a *sonante* function in *fidèle* [fidɛl] and a *consonante* function in *pied* [pje].

In Saussure's point of view, the phonetic level and the phonological level should be kept distinct. At the phonetic level, there is no acoustico-articulatory boundary between the vowel category and the consonant category. At the phonological level, their functions are categorical. That is to say, one sound of a certain language can have two functions, each function is by itself categorical. The French phoneme /i/ can illustrate this approach. The phoneme /i/ can have a *sonante* function in one syllable and have a *consonante* function in another syllable. It can even have a *sonante* function in one production (e.g., *lier* 'tie' [li.e] (De Carvalho, Brandão, Nguyen & Wauquier, 2010: 159)) and a *consonante* function in another production (e.g., *lier* 'tie' [lje] (id.)). However, the double-functional phoneme /i/ in French does not imply an intermediate function between *sonante* and *consonante*. It only shows that the phoneme /i/ can occupy two different functions according to the phonotactic environment. When the phoneme is phonetically implemented, only one function can be assigned to it.

Pike (1947) has a nearly identical view. He separates the speech sounds into vocoids/nonvocoids based on their acoustico-articulatory aspects, and into vowels/consonants based on their distributional characteristics. His arguments are the same as Saussure (1916), that is some segments, such as /i u/, can have the same structural positions as /a o/ or as /t s/. The segments /i u/ can thus have a vowel function or a consonant function, which implies the separation of the phonetic form of a sound from its phonological function.

Greenberg (1962) also separates the phonetic shape of a sound from its phonological function, following Pike's (1947) double terminologies. He proposes to define the vowel and consonant categories by '*non-phonetic criteria of distribution*'. He defines the vowel category as '*the class of elements (i.e., phonemes) such that at least one member of the class appears in every expression*', and the consonant class is defined as the complement class of that of the vowels. His definitions are somehow 'vague' since he does not specify what an 'expression' is. However, the most interesting point in his proposal is that he further distinguishes vocoids from nonvocoids based on the following groupings: 1-vocoids, 2-frictionless continuants (sonorants), 3-friction continuants, 4-stops. He notes that in one language, the existence of type 2 nucleus (in his terminology, vowel) implies the existence of type 1 nucleus, but not vice versa. The existence of type 3 nucleus implies the existence of type 2 and 1 nucleus, and so forth. This observation clearly states that the nucleus function can be assigned to any type of sound regardless of its phonetic shape.

The categorisation of speech sounds in Chomsky & Halle (1968) is an example of multiple terminologies, with the major class features having phonetic bases. They propose the three non-mutual exclusive categories (see also Jakobson & Waugh (1979)): sonorant, vocalic and consonantal. The sonorant–obstruent distinction is based on whether a 'spontaneous voicing' is possible in the vocal tract. It is thus clear that 'vowels, glides, nasal, consonants, and liquids are sonorants' (Chomsky & Halle, 1968: 302). The sonorant category is also motivated phonologically, since these sounds clearly pattern together as a natural class in English (Chomsky & Halle, 1968: 85, 354–355). The difference between the vocalic category and the consonantal category is based on whether there is a 'radical constriction' in the mid-sagittal region of the vocal tract. It is also, obviously, motivated by phonological patterning, as the two categories undergo respectively different phonological processes.

It is worth noting here that by saying 'vowels are sonorants' and 'vowels and liquids are vocalic sounds', this analysis shows a similar point of view as presented above (Saussure, 1916; Pike, 1947). That is, 'sonorant', 'vocalic' and 'consonantal' are phonological categories, while 'vowel' and 'consonant' are phonetic descriptions: the phonological level and the phonetic level are described by two sets of terminologies.

2.1.1.4. The none-categorical physical world

Categorisation is an elemental epistemological tool. To classify things is to arrange them in groups which are distinct from each other and are separated by clearly determined lines of demarcation (Durkheim & Mauss, 1903). All studies of human related phenomena start with categorisation: life vs. non-life, animal vs. plant, male vs. female, etc. These categories have been considered as dichotomies but are proven to be non-categorical by modern scientific researches.

One of the most fundamental dichotomies is between life and non-life. But the boundary between these two categories seems to be blurred by the discovery of virus. The viruses are considered to be 'at the edge of life' (Rybicki, 1990). Recently, the discovery of virophages (Pearson, 2008), the fact that a specific kind of smaller virus infects and parasitises bigger viruses, suggest that viruses are a kind of life form. If something can be sickened and eventually brought to death, it surely is alive (Pearson, 2008). However, the opposite view has been forcefully propounded as well: viruses cannot be considered alive because of their inability to reproduce without a cellular host (Moreira & López-García , 2009; López-García, 2012; Koonin & Starokadomskyy, 2016). The dichotomic view of life vs. non-life is challenged by (Corona)virus!

In biological taxonomy, the animal vs. plant dichotomy of life forms has been abandoned since early 20th century. Prior to that, concepts of organismal classification had been constrained within narrow boundaries that defined all life as either plant or animal (Scamardella, 1999). It is widely accepted today that all living and extinct organisms belong to two domains (or superkingdoms): Prokaryota and Eucaryota (Woese et al., 1990; Ruggiero et al., 2015a; Ruggiero et al., 2015b). Plants and animals are only two of the five kingdoms under the Eucaryota domain. Parallel to Animalia and Plantae, there are also Fungi, Chromista, Protozoa (Ruggiero et al., 2015b). In this case, the dichotomic view has been replaced by a hierarchical view with multiples classes organised in an arborescent and non-dualistic representation.

The male vs. female dichotomy in human and social sciences has been largely complemented by the gender continuum (Haig, 2004). Gender is the relationship between biological sex and behaviour (Udry, 1994). It is widely accepted today that gender refers to the social, cultural, and psychological traits linked to male and female through particular social contexts, and people of a particular sex category can be assigned with multiple gender traits linked to male and female (Wharton, 2004). In biological sciences, the dichotomy of sex is also questioned. The phenomenon of natural sex reversal and hermaphroditism is reported in vertebrates since early human history (Chan, 1970). It is shown that sex determination in European eel may be metagamic (i.e., non-generic) and that sex inversion may naturally occur when triggered by environmental factors (Wiberg, 1983). It is also proven that in amniote vertebrates, the sex determination is temperature-dependent. For Agamidae lizard, incubation temperature plays a crucial role in the sexual development of the embryos: the fitness of each sex was maximised by the incubation temperature that produces that sex (Warner & Shine, 2008). In flowers, plants and plant populations, the sexuality is clearly a complex phenomenon: some have unisexual flowers, some have bisexual flowers, some have only hermaphrodite flowers, and some have hermaphrodite, pistillate, staminate flowers on the same plant (Dellaporta & Calderon-Urrea, 1993). It is clear that whether in human and social sciences or in biology, the male vs. female dichotomy is not valid. It would seem unnatural that vowel-consonant dichotomy, based on the same complex physical world, could elude the same challenge.

2.2. Apical vowels and similar sounds: a review

2.2.1. Apical vowels: a brief history of names and symbols

The terminology *'apical vowels'* and the non-IPA symbols [ɿ ʅ] used to transcribe them are due to Karlgren (1915: 295). In his pioneer work titled *Études sur la Phonologie Chinoise*, he describes [ɿ] as *voyelle apico-gingivale* and [ʅ] *voyelle apical-alvéolaire*:

> Les voyelles apicales, rares dans les langues européennes, fleurissent en chinois. Nous trouvons d'une part des apico-gingivales, dont l'articulation linguale est produite le plus facilement en élargissant le passage qui se fait entre la langue et les gencives pour prononcer la consonne z, juste assez pour faire disparaître la friction orale. D'autre part, nous avons des apico-alvéolaires dont l'articulation linguale est produite par un élargissement correspondant du *(au)* passage de ʐ. Hautes et pourtant sans élévation prépalatale du dorsum, ces voyelles font, quand elles ne sont pas labialisées, un effet acoustique qui les rapproche considérablement du ы russe. …
>
> [ɿ] Voyelle apico-gingivale, haute, tendue, délabialisée ou à l'ouverture labiale large. … Elle est partout orale, et n'apparaît qu'en syllabe ouverte et après s, z. … L'absence de ɿ dans les autres langues connues rend sa définition très difficile pour des amateurs. Les identifications courantes comme « l'u bref anglais », « l'eu français » etc. sont toutes très incorrectes.
>
> [ʅ] Voyelle apico-alvéolaire, haute, tendue, délabialisée ou à l'ouverture labiale large … Elle est partout orale et n'existe qu'en syllabe ouverte après ʂ, ʐ. … Une voyelle très analogue à ʅ se rencontre dans certains dial(ect) suédois ; on l'appelle « i de Viby ».
>
> (Karlgren, 1915: 295)[1]

---

[1] N.B. In this dissertation, the English translations of all citations in French are mine. [Apical vowels, rare in European languages, bloom in Chinese. On the one hand, we find apico-gingival vowels, whose lingual articulation is most easily produced by widening the passage between the tongue and the gums to pronounce the consonant z , just enough to make the oral frication disappear. On the other hand, we have apico-alveolar ones whose lingual articulation is produced by a corresponding widening of the passage of ʐ. High and yet without any pre-palatal elevation of the dorsum, these vowels make, when not labialised, an acoustic effect which brings them considerably closer to the Russian ы. ¶ [ɿ] Apico-gingival vowel, high, tense, delabialised or with a wide labial opening. ... It is everywhere oral, and appears only in open syllables and after s, z. ... The absence of ɿ in other known languages makes its definition very difficult for enthusiasts. Common identifications such as "short u in English", "eu in French" etc. are all quite incorrect. ¶ [ʅ] Apico-alveolar vowel, high, tense, delabialised or with a wide labial opening ... It is everywhere oral and exists only in open syllable after ʂ, ʐ. ... A vowel very similar to ʅ is found in some Swedish dialects; it is called the "Viby i".]

To the best of my knowledge these cited phonetic descriptions of the apical vowels are the earliest in the literature. Karlgren presents these segments as ‘genuine’ vowels, but his description is more nuanced when studied in detail. He notes that the essential difference between the apical vowels and the corresponding consonants [z ʐ] is the air passage. The air passage is more enlarged in the apical vowels than in [z ʐ], so that the oral frication noise disappears. Following this description, the apical vowels could as well have been analysed as some fricative-like segments, since the articulatory gesture involved is essentially the same as for [z].

For Karlgren (1915: 296), [ɿ] and [z] are two different segments sharing the same place of articulation. He goes on in his description of apical vowels, and states:

> D’autre part, plusieurs savants ont transcrit ɿ par la consonne ‘z’, graphie nullement mauvaise, en réalité. C’est que ɿ, pour la position de la langue, est congénère de z, et que dans tous les dialectes qui possèdent ɿ on pourra trouver bien des individus qui y substituent un z̥[2] (z porteur de la syllabe) : sɿ et sz̥, tsɿ et tsz̥ se permutent individuellement (ts‘z n’existe guère[3]).
>
> (Karlgren, 1915: 296)[4]

Again, the similarity in articulation between the apical vowel [ɿ] and the consonant [z] is clearly stated. It is interesting to see that Karlgren, while he opposes the consonant [z] to the apical vowel [ɿ], indicates that they are exchangeable ‘*permutable*’. It is difficult to know Karlgren’s exact criteria that differentiate the consonant [z] from the apical vowel [ɿ]. But the existence of frication noise is probably an important indicator since he describes the apical vowel [ɿ] as having an enlarged air passage compared to [z], just enough to make the frication noise disappear. Following this understanding, the difference between a [sɿ] syllable and a [sz] syllable would thus be linked to the presence/absence of frication noise on the syllable nucleus: [ɿ] with no frication noise and [z] with frication noise.

In the 1950s, the apical vowel symbols [ɿ ʅ] coined by Karlgren (1915) were introduced into China possibly by German linguists (Pullum & Ladusaw, 1996). Along with the symbols [ɿ ʅ], two other symbols [ʮ ʯ] (Karlgren, 1915: 297) representing the rounded apical vowels were also introduced into China at the same period (Pullum & Ladusaw, 1996). The four symbols [ɿ ʅ ʮ ʯ] appeared in 汉语方言概要 *A compendium of Chinese dialects* (Yuan, 1960: 7). Since then, the terminology

---

2 The consonant z̥ here corresponds to the IPA notation z̩.

3 The affricate consonant ts‘ here corresponds to the IPA notation ts^{h}.

4 [On the other hand, several scholars have transcribed ɿ with the consonant ‘z’, which is in fact not a bad graph. This is because ɿ, for the position of the tongue, is homorganic to z, and in all the dialects that have ɿ one can find many individuals who substitute it with z̥ (z syllabic): sɿ and sz̥, tsɿ and tsz̥ are interchangeable on an individual basis (ts‘z hardly exists).]

‘apical vowels’ and the four symbols have been widely used until today among researchers working on the phonetics and phonology of Chinese languages (to cite a few: Kratochvíl, 1968; C. Cheng, 1973; Howie, 1976; Svantesson, 1984; Zee & Lee, 2007; Shi, Peng & Liu, 2015; Faytak & Lin, 2015; Faytak, 2018).

Most of the studies on apical vowels are based on SC[5], but apical vowels are also attested in other Chinese languages (Wu, 1995; Wang, 2006; Hu, 2007; Hou, 2009), and even in some non-Chinese Sino-Tibetan languages (Baron, 1974; Michaud, 2008; Wang, 2010). The presence of apical vowels in all these languages has always been related to a historical /i/ in some stage of the evolution (Zhu, 2004; Zhao, 2007; Jacques & Michaud, 2011; Gong, 2016). In this review, the status of apical vowels in SC is reported in section 2.2.2.1, before reporting on other Chinese languages in section 2.2.2.3. Similar phenomena observed in non-Chinese languages are reported in section 2.2.3, namely the Swedish Viby-i, the Lendu vowelless syllable, the Mambila fricative vowel and the Ikema fricative vowel.

The name ‘apical vowels’ and the four symbols used to transcribe them are not accepted by all researchers, as various other symbols have been proposed in different descriptions and analyses. The following review uses the original symbols used in each study, as this can reflect the phonetic and phonological nature adopted by each study. A summary of the correspondence between all the names and all the symbols is given in Table 2.1.

This review focuses on the non-rounded apical vowels. Whenever the term ‘apical vowel’ is used, it refers to the non-rounded apical vowels as, for example, the variants of apical vowels in SC. The rounded versions [ʮ ʯ], being much rarer in Chinese languages, are not included in this review.

[5] SC is also called Putong Hua 普通话 ‘common speech’. Other terms for SC are Beijing Mandarin, Standard Mandarin, Mandarin Chinese, or simply Mandarin (Duanmu, 2007: 4). In this study, the term SC is used for simplicity.

Table 2.1. Correspondence between different terminologies used in the description of the apical vowels.

Concerned segment	Terminology	Symbols
The syllable nucleus of [tsz] 字 (Pīnyīn: zì) in SC or the equivalent segment in other Chinese languages.	Apical vowel Apical front vowel Dental apical vowel	ɿ
	Fricative vowel	—
	Syllabic fricative	z̩ z
	Syllabic approximant Syllabic apical post-alveolar approximant Syllabic dental approximant	ɹ̩ ɹ̺
The syllable nucleus of [tʂʐ] 志 (Pīnyīn: zhì) in SC or the equivalent segment in other Chinese languages.	Apical vowel Apical back vowel Retroflex apical vowel	ʅ
	Fricative vowel	—
	Syllabic fricative	ʐ̩ ʐ
	Syllabic approximant Syllabic apico-laminal or laminal denti-alveolar approximant Syllabic retroflex approximant	ɹ̩ ɻ

2.2.2. Different variants of apical vowels in Chinese languages

In this section, the descriptions of apical vowels in Chinese languages are reviewed. The apical vowels in SC, reviewed in section 2.2.2.1, are the most studied variants. Three other variants of apical vowels, different from the SC ones, are selected according to their phonological and phonetic behaviour and reviewed in section 2.2.2.3.

In section 2.2.2.1, the different analyses of the apical vowels in SC are reported in the following order: the vowel analysis, the syllabic fricative analysis and the syllabic approximant analysis. This order is more or less a chronological one, following the advancements in the understanding of these segments. As already stated, the different analyses are presented keeping with the original symbols used. In the cases where the original symbol is not transparent enough, a footnote is given to specify the correspondence.

In section 2.2.2.3, three Chinese languages are presented: the Hefei-Mandarin Chinese (HMC), the Qinghai-Mandarin Chinese (QMC) and the Suzhou-Wu Chinese (SWC). The first two languages (both belong to the Mandarin group as SC) are chosen since their apical vowels are reported to be different from the apical vowels attested in SC. The Suzhou-Wu Chinese is reported to have apical vowels occurring in the same contexts as in SC, but with a different phonetic implementation.

2.2.2.1. Apical vowels in SC: phonology and phonetics

SC has two apical vowels [ɿ] and [ʅ], as the following examples in (1) show:

(1)	丝 'silk' [sɿ˥]		诗 'poetry'	[ʂʅ˥]
	姿 'posture'	[tsɿ˥]	知 'know'	[tʂʅ˥]
	疵 'flaw'	[tsʰɿ˥]	吃 'eat'	[tʂʰʅ˥]
			日 'sun, date'	[zʅ˥]

The two apical vowels are considered to be allophonic variants of /i/: [ɿ] occurs after dental sibilants, and [ʅ] occurs after retroflex sibilants (R. Cheng, 1966; C. Cheng, 1973). This complementary distribution is summarised in Table 2.2. This is the traditional view of the apical vowels in SC (Hartman, 1944; Hockett, 1947; Fu, 1956; R. Cheng, 1966; Kratochvíl, 1968). In this analysis, the apical vowels are considered as vowels. They occur at the nucleus position of the syllable hence function as the tone-bearing unit (TBU) of the syllable (Hartman, 1944).

As shown in Table 2.2, the high vowel /i/ does not occur after dental and retroflex sibilants. It is proposed that there is a co-occurrence restriction (R. Cheng, 1966) between the dental and retroflex sibilants and the high front vowel /i/, the syllable nuclei after the dental and the retroflex sibilants are thus the apical vowels.

Table 2.2. Distribution of apical vowels and [i] in SC.

s ts tsʰ	ʂ tʂ tʂʰ	ɕ tɕ tɕʰ
*	*	ɕi tɕi tɕʰi
sɿ tsɿ tsʰɿ	*	*
*	ʂʅ tʂʅ tʂʰʅ	*

As discussed in the previous section, since the earliest description of the apical vowels (Karlgren, 1915), the homorganicity between the apical vowels and the sibilant consonants has been noticed and has always been at the centre of the discussion. The apical vowel [ɿ] is described as the apical front vowel and [ʅ] as the apical back vowel (Zhou & Wu, 1963), the radiographs of the two segments together with the consonants [s ʂ] are presented in Figure 2.1. It is probably due to the tongue tip position difference that Zhou & Wu (1963) have named the two apical vowels as front and back. It can be observed in their radiographs that the tongue tip in [ɿ] has a denti-alveolar position, while the tongue tip of [ʅ] has an alveolar position.

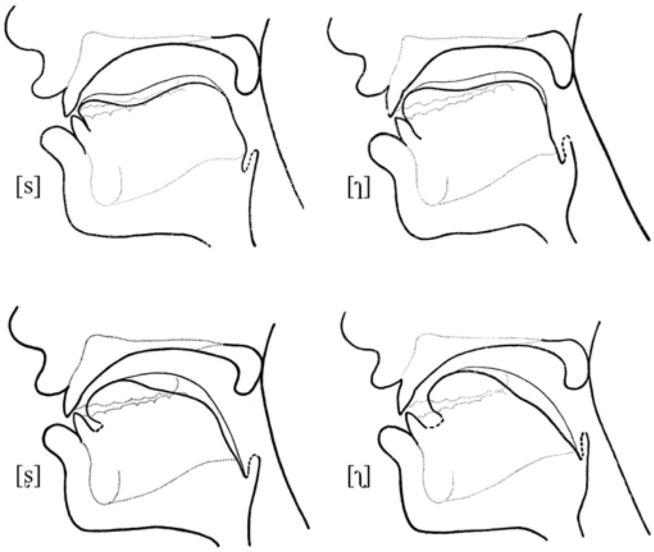


Figure 2.1. X-ray images (Zhou & Wu, 1963: 60, 64, 73, 74) of apical vowels and [s ʂ] consonants in SC. Sibilant consonants [s ʂ] were obtained with [sa ʂa] syllables, apical vowels were obtained with [tsɿ tʂʅ] syllables. The mid-sagittal tongue contours (black lines) are presented with rolled-up tongue sides (grey lines).

Zhou & Wu (1963) also note that the tongue shape of [ɿ] is comparable to the tongue shape of [s], while the tongue shape of [ʅ] can be compared to [ʂ]. They only present the tongue shapes as comparable, but do not give any detailed comments on this comparison. As shown in Figure 2.1, the tongue shape of [ɿ] and the tongue shape of [s] are fairly similar but the tongue shapes of [ʅ] and [ʂ] do show some differences. The tongue apex seems to be more raised in [ʅ] than in [ʂ], and the sublingual cavity is larger in [ʅ] than in [ʂ]. This point may have been the reason why Zhou & Wu (1963) did not claim confidently that the apical vowels were homorganic to the sibilants [s ʂ], but only as 'comparable'.

The homorganicity between the apical vowels and the preceding sibilants is also claimed by C. Cheng (1973). He refers to the apical vowel [ɿ] as 'dental apical vowel' and the apical vowel [ʅ] as 'retroflex apical vowel'. The two apical vowels are variants of the high vowel /i/ and they are clearly described as 'homorganic to the preceding consonant' (C. Cheng, 1973: 13). The dental apical vowel [ɿ] is homorganic to [s] and the retroflex apical vowel [ʅ] is homorganic to [ʂ]. In a similar vein, Ladefoged and Maddieson (1996: 314) note that the apical vowels are '*made with the tongue in essentially the same position as in the corresponding fricatives*' and refer to these segments as 'fricative vowels'.

The other important attribute of apical vowels is that they can contain frication noise (Trubetzkoy, 1969; Ladforged & Maddieson, 1996). Trubetzkoy (1969: 171) describes these segments as having audible frication noise:

> Phonetically, if pronounced clearly, it is a type of vowel with a much lesser degree of aperture and with a much more fronted position of articulation than, for example, i, so that a frictionlike noise resembling a humming is audible in its production.
>
> (Trubetzkoy, 1969: 171)

For Trubetzkoy (1969), as for Karlgren (1915) before him, the two segments are 'a type of vowel', notwithstanding their 'frictionlike noise'. A similar observation can be found in Howie (1976: 6), who also noted that the two apical vowels were produced '*often with continued friction*'.

The descriptions presented above have one major point in common: they all consider the 'apical vowels' to be vowels. But the vowel analysis, although it has a large influence among scholars, is not accepted by all researchers. Chao (1961: 22) argues that the apical vowels are not genuine vowels, given their homorganicity with the preceding sibilants. His observation is particularly interesting since he is among the first researchers to avoid using the term 'apical vowel' and the [ɿ ʅ] symbols:

> The first final, which we represent by the letter **y**[6], is a vocal prolongation of the preceding consonant. It has two qualities. (1) After the dental sibilants: **tz**[7], **ts**[8], **s**, it has a buzzing quality, like a prolonged **z** in buzz. Thus, the syllable **sy** sounds like **s** + vocalized **z**. After the consonant has been pronounced, the vocalic part — the buzzing part — need not, and usually does not, have much frication, but the tip of the tongue remains behind and near the teeth to give the **z**-quality. The lips are open. (2) … After the retroflexes … this final is pronounced as a vocalized **r**.
>
> (Chao, 1961: 22)

In his *A grammar of spoken Chinese* (1968: 24), Chao uses the symbol [z̩] for [ɿ] and the symbol [ɻ̩] for [ʅ], clearly showing that he analyses the apical vowels as syllabic consonants homorganic to the sibilant onsets. But it is interesting to see that he is careful about the presence of frication noise, as he considers the apical vowel [ɿ] as '*need not, and usually does not have much frication*'. That is to say, the apical vowel after dental sibilants can sometimes contain some frication noise, but this frication noise is not systematic. It seems that Chao is faced with three contradictory

---

[6] This notation corresponds to the two apical vowels [ɿ ʅ].
[7] This notation corresponds to the consonant [ts]
[8] This notation corresponds to the consonant [tsʰ]

facts: (i) the apical vowel is homorganic with the preceding sibilant onsets [s ts tsʰ] (ii) it does not always contain the should-be-there frication noise given that it has the same articulatory gesture as the sibilants, and (iii) it still has a 'buzzing quality' and still 'sounds like a prolonged z'. His final choice is to analyse the apical vowels as syllabic consonants [z̩] and [ɹ̩], which reflects correctly the homorganicity between the two segments and the sibilant onsets. But he provides no further discussion on the frication noise.

Similar to Chao's analysis, apical vowels have been analysed as syllabic fricative consonants, and transcribed using [z] and [ʐ] symbols by other authors (Dell, 1994; Wiese, 1997; Yu, 1999; Duanmu,2007). According to Yu (1999) the presence of high-frequency noise in the 4000–7000 Hz region is the definitive indication of the sibilant property of 'apical vowels'. From a phonological perspective, Dell (1994) interprets apical vowels as the voiced prolongation of the preceding syllable onsets, this analysis is fundamentally identical to Chao's (1961, 1968) view. For Wiese (1997) and Duanmu (2007: 44), the homorganic property of apical vowels is a result of feature spreading: the coronal features (including [+fricative]) of the preceding sibilant onset spread into the empty syllable nucleus (see Figure 2.2).

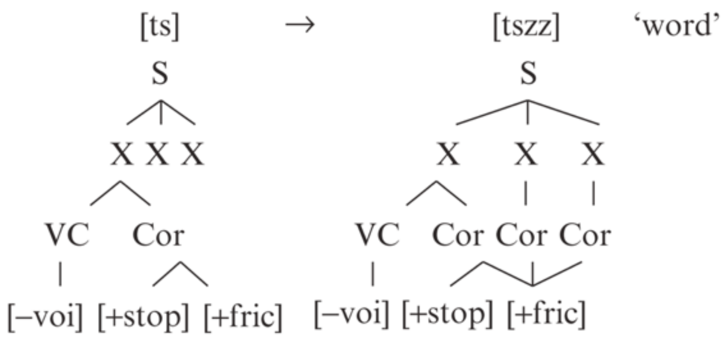


Figure 2.2. Feature geometry representation of the SC syllable [tszz] 'word', reproduced from Duanmu (2007: 44).

Duanmu's analysis can account for the homorganicity between the apical vowels and the preceding sibilant onsets. The onset [ts] occupies a single time unit and the empty nucleus occupies two time units. The onset [ts] has two articulators, Vocal-cords, which dominates the feature [-voice], and Coronal, which dominates [+stop] and [+ fricative]. [ts] must be linked to the onset. The empty slots in the rhyme trigger the spreading of [+fricative], which activates Coronal. The result is [ts] in the onset and [zz] in the rhyme. The voicing of the rhyme [zz] comes from the nucleus position, which requires that any segment occupying this position to be voiced. The same analysis could be applied to all the onsets having articulator Coronal-[+fricative]

(Duanmu, 2007: 44). Following this analysis, the examples in (1) are analysed as in (2).

(2) 丝 'silk' /s/ → [szz˥]

姿 'posture' /ts/ → [tszz˥]

疵 'flaw' /ts^{h}/ → [ts^{h}zz˥]

诗 'poetry' /ʂ/ → [ʂʐʐ˥]

知 'know' /tʂ/ → [tʂʐʐ˥]

吃 'eat' /tʂ^{h}/ → [tʂ^{h}ʐʐ˥]

日 'sun, date' /ʐ/ → [ʐʐʐ˥˩]

It is worth noting that Duanmu (2007) does not consider the apical vowels to be allophones of /i/, since he does not propose any underlying form for the apical vowels. In his analysis, as presented above, the apical vowels are triggered by an empty nucleus slot. Their phonological specifications are the result of feature spreading processes. Furthermore, in his analysis, the vowel /i/ has a different distribution compared to the traditional view presented in Table 2.2. He considers the syllables [ɕi tɕi tɕ^{h}i] to have variants [s^{j}i ts^{j}i ts^{hj}i], and they are all derived from the underlying /si tsi ts^{h}i/ syllables. This means that the vowel [i] can occur underlyingly after dental sibilants /s ts ts^{h}/ and that the palatal consonants [ɕ tɕ tɕ^{h}] are not phonemic.

The analysis of 'apical vowels' as fricatives has been questioned by some recent studies, who argue that these segments are best analysed as approximants (Lee & Zee 2003, Lee-Kim 2014). In their IPA description of SC, Lee and Zee (2003), who use the symbol [ɹ̩] to transcribe the two apical vowels, described them as 'syllabic apical post-alveolar approximant' and 'syllabic apico-laminal or laminal denti-alveolar approximant'. This transcription is adopted by Lin (2007: 72) who uses the symbol [ɻ̩] to transcribe the two apical vowels in SC:

> In terms of articulation, the tongue tip stays in the same location within the oral cavity throughout the whole syllable. …the syllable onset consonants … are voiceless, but the nuclear part of the syllable, i.e., the syllabic consonant, is voiced. To learn how to pronounce these syllabic consonants, you basically prolong the pronunciation of the consonant. … produce a voiceless consonant in the first phase of the syllable (i.e., the syllable onset) and a voiced one in the second phase of the syllable (i.e., the syllabic nucleus). During the syllabic nuclear phase, there can be a lesser degree of constriction; that is, the tongue tip can be moved slightly away from the teeth or the post-alveolar region at the end of the syllable with little friction.

(Lin, 2007: 72)

Lin (2007) considers apical vowels to be the voiced prolongation of the syllabic onsets with possibly a lesser degree of constriction, thus yielding little friction. This description is fundamentally identical to what has been proposed by Chao (1968). Similar to Chao (1961, 1968), Lin (2007) is not sure about the presence of the frication noise and argues that there ‘*can be a lesser degree of constriction*’, which should lead to less frication noise.

In Lee-Kim’s (2014) comprehensive revision of SC apical vowels, she shows that the apical vowels are indeed homorganic to the sibilant onsets with a slightly retracted tongue root for [ɿ] and a slightly lowered tongue body for [ʅ]. But they have no or very little frication noise. She does observe very short carryover frication noise that is attributed to a mere gesture overlap between sibilant onsets and the homorganic apical segments.

The spectrograms of the SC apical vowels in Lee-Kim’s (2014) study are reported in Figure 2.3. She presented three speakers showing three different patterns: RJ does not have any frication noise, WY has very short carryover frication noise, and WH has relatively long frication noise on the apical vowel [ɹ̪̩] after sibilant [s] but no frication noise on the apical vowel [ɻ̩] after sibilant [ʂ]. She notes that with the exception of these three subjects, the apical segments behave similarly to other vowels, presenting a periodic waveform and a nearly clear separation of the frication noise from the following vocalic period.

Based on the observations that (i) the apical vowels are homorganic to the sibilant onsets and that (ii) no systematic frication noise is produced, Lee-Kim (2014) argues that the apical vowels are syllabic dental and retroflex approximants, which she transcribes as [ɹ̪̩] and [ɻ̩], respectively.

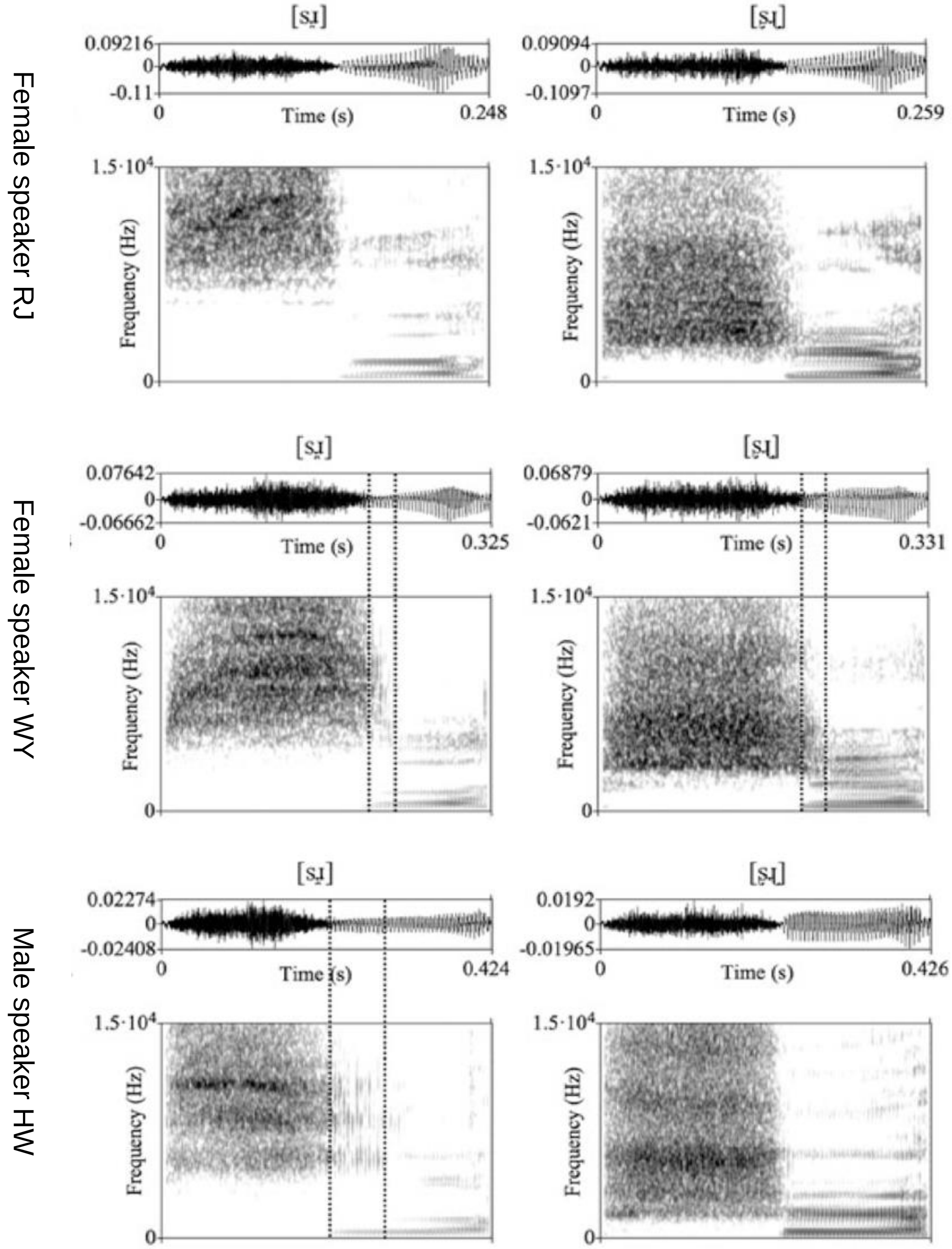


Figure 2.3. Spectrograms of [sɹ̩] and [ʂɻ̩] of two female speakers (RJ and WY) and one male speaker (HW). There is no frication noise in RJ's production, but some frication noise is observed in WY and HW's production. Frication in the vocalic period is represented with dotted vertical lines. This figure is reproduced from Figure 9, 10, 11 of Lee-Kim (2014).

2.2.2.2. Mismatch between phonetics and phonology of apical vowels in SC

The four different accounts on 'apical vowels' (i.e., apical vowel, fricative vowel, syllabic fricative consonant, syllabic approximant consonant) seem to depend on whether phonological or phonetic criteria are applied. In this section, the four accounts will be discussed individually with a focus on the advantages and disadvantages of each account.

The vowel analysis, either ‘apical vowel’ or ‘fricative vowel’, is essentially based on their phonological patterning: (i) they are allophonic to vowel /i/; (ii) they function as syllable nuclei; and (iii) they can be tone-bearing units. This point of view is phonologically convenient since it complies with the habitual syllable structure of SC, that is the nucleus of a syllable should be a vowel[9]. In this view, the acoustic presence of a clear formant structure in apical vowels is considered as a definitive character for vowels, and the raised tongue body is considered to be a vowel gesture (C. Cheng 1973). However, the vowel analysis is phonetically inconvenient since it cannot explain the fact that (i) the apical vowels are homorganic to the sibilant onsets, and (ii) the possible presence of frication noise. These two points argue strongly for a syllabic consonant analysis (whether fricative or approximant), since a vowel, a priori, should not be homorganic to a fricative consonant and should not contain any frication noise.

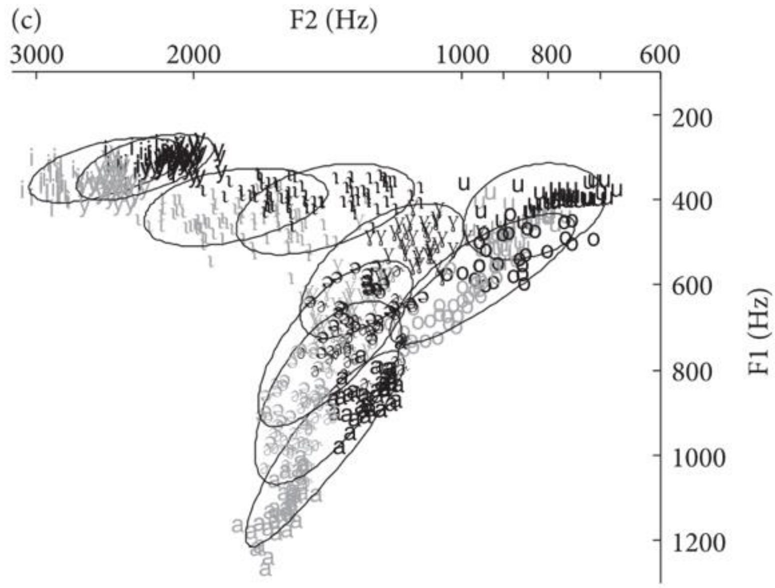


Figure 2.4. Formant (F1-F2) plot of the ten vowels in continuous speech of SC. This figure is reproduced from the Figure 34.2(c) from Shi et al. (2015). Black symbols for male speakers and grey symbols for female speakers.

Furthermore, if the two apical vowels are called ‘apical front vowel’ for [ɿ] and ‘apical back vowel’ for [ʅ] (Zhou & Wu, 1963), their formant values do not match their description. The apical back vowel [ʅ] has a higher F2 than the apical front vowel [ɿ],

---

[9] The nucleus of a syllable in SC is almost always a vowel, only very marginal examples of syllabic consonants such as [m n ŋ] (Duanmu 2007: 34) can be found.

as shown in Figure 2.4. This means that the correlation between the F2 and the anteriority of vowels is not applicable to the SC apical vowels (Lee, 2005) since the F2 values indicate that [ʅ] should be the front one and [ɿ] the back one.

The 'fricative vowel' analysis argued for by Ladefoged and Maddieson (1996: 314) is an attempt to unify the phonological behaviour (i.e., 'vowel') and the phonetic implementation (i.e., 'fricative'). This analysis acknowledges the 'vowel' status of the apical vowels and adds the 'fricative' property as an additional vowel feature. But Ladefoged and Ferrari Disner (2012: 26) stated that a vowel would be '*any sound occurring in the middle of a syllable, provided that it is produced without any kind of obstruction of the outgoing breath*'. Clear enough, the 'fricative vowels', if produced with frication noise, would not comply with this definition. Ladefoged[10] changed his earlier observation and noted '*it is certainly true that the fricative noise carries over into the beginning of the vowel. But the greater part of each of these vowels does not have any fricative turbulence.*' This new observation should rule out his earlier 'fricative vowel' analysis. It is evident that in this analysis, the presence of frication noise is again at the centre of the discussion. If the frication noise exists, then the apical vowels cannot be vowels since vowels should not have any kind of obstruction of the outgoing breath. If the frication noise does not exist, then the category 'fricative vowel' should not exist. Either way, the 'fricative vowel' analysis is not ideal.

The analysis of apical vowels as 'syllabic fricatives' is based mainly on phonetic observations. It can reflect correctly the acoustic presence of frication noise and the homorganicity with the coronal sibilants. But this analysis, it could be argued, is phonologically unnatural, since it assumes that an underlying vowel (i.e., /i/) has fricative consonants as allophonic variants (Wiese, 1997). This point of view is not adopted by Duanmu (2007: 44). In his analysis, the syllabic fricatives are not derived from an underlying /i/ or any other underlying vowel, but rather triggered by an empty nucleus slot. When adopting Duanmu's point of view, the unnaturalness of having allophonic fricative consonants and vowel /i/ no longer exists. Still, the syllabic fricative analysis also assumes that the fricative consonants in [sz̩ ʂʐ̩] occupy the nucleus position of the syllables and function as tone-bearing units, in a striking exception to the behaviour of other obstruents in SC.

The 'syllabic approximant' analysis is also based on phonetic observations. It captures the tongue shape's slight articulatory difference between the sibilant onsets and the 'apical vowel', and thus explains the absence of frication noise in some cases. But such absence of frication noise is not consistent in all studies. Lee-Kim (2014) observes no or little frication noise on apical vowels, while other researchers report important interspeaker variation (Yu, 1999; Faytak & Lin, 2015). When frication noise is observed, it is generally limited to the very beginning of the segment and lasts a very short period of time. This suggests, according to Lee-Kim (2014), that

[10] This citation can be found on the web site of UCLA Phonetics Lab data: http://www.phonetics.ucla.edu/appendix/languages/chinese/chinese.html

this frication noise is an outcome of gestural overlap between sibilants and the following homorganic ‘apical vowels’, so that the frication noise cannot be directly related to a fricative nature. Phonologically, however, the syllabic approximant analysis is not more ‘natural’ than the syllabic fricative analysis if the syllabic approximants are analysed as derived from an underlying /i/.

In a nutshell, the main difficulty encountered in the analysis of ‘apical vowels’ in SC is that their phonetic implementation does not match their phonological behaviour: the former provides evidence for a consonant analysis while the latter provides arguments for a vowel analysis. The same dilemma was faced by researchers working on ‘apical vowels’ in other Chinese languages.

2.2.2.3. Apical vowels in other Chinese languages

In addition to SC, other Chinese languages present numerous examples of apical vowels, especially the non-retroflexed non-rounded [ɿ] variant. It can be termed ‘apical vowel’, ‘apical front vowel’ or ‘dental apical vowel’ in different studies. This is the variant that occurs after sibilants [s ts ts^h] in SC. In this section, only this variant is reviewed, since it is more widespread than the retroflexed variant [ʅ] and the rounded ones. As observed in the description of Chinese languages, when there is only one apical vowel, it is the [ɿ]. The other variants only occur when there is already the [ɿ] in the inventory.

In this section, apical vowels from Hefei-Mandarin Chinese, Qinghai-Mandarin Chinese and Suzhou-Wu Chinese are reviewed. These three Chinese languages are chosen according to the availability of the descriptions and the characteristics of the apical vowel. The two Mandarin Chinese languages are chosen to show that even within the Mandarin group, apical vowels can behave differently.

Traditionally, the Chinese languages are divided into ten groups (Li et al., 1987): Mandarin 官话, Jin 晋语, Wu 吴语, Hui 徽语, Xiang 湘语, Gan 赣语, Min 闽语, Yue 粤语, Ping 平话, and Hakka 客家话; each group can be further divided into subgroups. In order to provide more information on the cited Chinese languages, the group name is given with the name of the city/region where the language is spoken. For example, Hefei 合肥 is the city where the Hefei Chinese is spoken. Hefei Chinese is a Mandarin group Chinese language, so it is referred to as Hefei-Mandarin Chinese.

The tones in Chinese languages can be transcribed using tone letters (Chao, 1930) as shown in examples (3), (4) and (5), or be transcribed using IPA tone symbols, as shown in examples (6) and (7). The original transcriptions are retained in the cited examples.

2.2.2.3.1 Hefei-Mandarin Chinese

Hefei-Mandarin Chinese 合肥话 (HMC) is a Mandarin Chinese language of Jianghuai-Mandarin 江淮官话 subgroup (Li et al., 1987). It has two [ɿ ʅ] (Department of Chinese of Peking University, 1989) or three apical vowels [ɿ ʅ ʮ] (Wu, 1995; Wan, 2014) depending on the descriptions. A complete analysis of the HMC phonemic inventory is not available. Concerning the apical vowels, [ʅ] occurs only after [ʂ tʂ tʂʰ] but the distribution of [ʮ] is not clear (the case of [ɿ] is dealt with below). When the consonants /ɕ tɕ tɕʰ/ are followed by /i/, they are realised [s ts tsʰ] respectively, and the vowel /i/ is realised [ɿ], as shown by the following examples from Wu (1995) in (3). The apical vowel [ɿ] in HMC is thus phonologically analysed as a contextual variant of /i/ (Wu, 1995), similar to the traditional view of apical vowels in SC.

(3)	西 'west'	/ɕi³¹/	→ [sɿ³¹]
	鸡 'chicken'	/tɕi³¹/	→ [tsɿ³¹]
	骑 'ride'	/tɕʰi³¹/	→ [tsʰɿ³¹]

Unlike in SC, however, [ɿ] in HMC is not always preceded by a homorganic onset. It can be preceded by [p pʰ m z] in addition to [s ts tsʰ] consonants (Wu, 1995). The following examples in (4) are also from Wu (1995):

(4)	比 'compare'	[pɿ²⁴]
	皮 'skin'	[pʰɿ⁵⁵]
	米 'rice'	[mɿ²⁴]

At the phonetic level, the acoustic study from Hou (2009) reports that the HMC apical vowel has strong high-frequency frication noise at the 3000 – 5000 Hz region and the frication noise does not continue to the end of the segment. Hou (2009) considers the apical vowel as a fricative vowel 带擦元音 but does not provide further discussion on its frication noise. Wan (2014) has also conducted an experimental study on the HMC vowels. She treats the 'apical vowels' as genuine vowels and also concludes that [ɿ ʅ] have 'strong frication noise', without however providing any further measurements on this property. Kong et al. (2019) follow Ladefoged and Maddieson's (1996) analysis and consider the apical vowels as 'fricative vowels'. Their study shows that strong frication noise can be observed on the HMC apical vowels even when it is preceded by bilabial and nasal consonants, though not on the entire duration of these segments. The frication noise is superposed on voicing and is observable only on the first 15 – 20% of the entire duration. They conclude that the frication noise is a secondary feature of the HMC apical vowel, and argue that this feature may be an important perceptual cue for the HMC apical vowels.

2.2.2.3.2 Qinghai-Mandarin Chinese

Qinghai-Mandarin Chinese (QMC) 青海方言 is a Mandarin dialect of Lanyin-Mandarin 兰银官话 group (Li et al., 1987). It has one apical vowel [ɿ] which can be preceded by [p pʰ m l s ts tsʰ] consonants (Wang, 2006). The phonological analysis of this dialect is not available but Wang describes the phoneme /i/ as having three allophones [i j ɿ]. The two allophones, namely [ɿ j], coexist in regional variants of QMC. The following examples in (5) are from Wang (2006):

(5)	鸡 'chicken'	[tɕj^{44} tɕj^{44}]
	资 'fund'	[tsɿ^{44}]
	一 'one'	[ɿ^{44} j^{44}]
	梯 'ladder'	[tsʰɿ^{44}]
	礼 'politeness'	[lɿ^{53}]
	地 'soil'	[tsɿ^{213}]
	皮 'skin'	[pʰj^{24}]
	七 'seven'	[tɕʰj^{44} tsʰɿ^{44}]
	李 'apricot'	[lɿ^{53}]

Wang (2006) does not provide any argument on why [j] can be the nucleus of a syllable, but it seems that he uses the symbol [j] to represent a [i] vowel with a closer lingual-palatal distance. This means that his [j] symbol can also be considered as a raised [i̝], hence [ɿ i̝] and [i] are the three allophonic variants of the phoneme /i/.

No experimental study on QMC apical vowels is available but as examples in (5) show, similar to HMC, [ɿ] is not systematically homorganic to the preceding onset consonant, as it does not share the same place of articulation with [p pʰ m]. The same remark can be made here: if the apical vowel is analysed as a syllabic consonant, then this syllabic consonant (whether a syllabic [z̩] or a syllabic [ɹ̩]) cannot be analysed as always homorganic to the onset.

Interestingly, the syllabic consonant [v̩] occurs after [p l tʂ] in QMC, as the following examples in (6) from Dede (2006) show. If the apical vowel is analysed as a syllabic consonant [z̩], then it will not be the only syllabic obstruent in QMC, since [v̩] is also a syllabic obstruent.

(6)	路 'road'	[lv̩˧]
	炉 'stove'	[lv̩˦˥]

	猪 ‘pig’	[tʂɿ̩˥]
	厨 ‘kitchen’	[tʂʰɿ̩˧˥]
	布 ‘cloth’	[pɿ̩˧]

2.2.2.3.3 Suzhou-Wu Chinese

An interesting case of apical vowels has been observed in Suzhou-Wu Chinese 苏州话 (SWC), a Wu 吴 group Chinese language (Li et al., 1987). The description of this dialect shows that it has two apical vowels, a rounded [ʮ] and an unrounded [ɿ], all occurring after sibilant onsets [s ts tsʰ z] (Ye, 1993). Interestingly, the apical vowels /ɿ/ and /ʮ/ are independent phonemes which contrast with /i/, as shown by the following minimal triplets in (7) (Ye, 1993):

(7)	四 ‘four’	[sɿ˥˩˧]
	世 ‘world’	[sʮ˥˩˧]
	细 ‘small’	[si˥˩˧]
	紫 ‘purple’	[tsɿ˥˩]
	主 ‘main’	[tsʮ˥˩]
	姊 ‘sister’	[tsi˥˩]
	此 ‘this’	[tsʰɿ˥˩]
	鼠 ‘mouse’	[tsʰʮ˥˩]
	取 ‘take’	[tsʰi˥˩]

The examples in (7) show that even though the apical vowels in SWC occur only after dental sibilants, they are distinct phonemes contrastive to /i/. A consequence of this is that they cannot be analysed as derived from an underlying high vowel /i/.

At the phonetic level, Faytak (2018: 45) notes that they ‘*have an apico-alveolar constriction similar to a* /z/ *and could be transcribed as syllabic rounded and unrounded alveolar fricatives with a loose degree of constriction, i.e., syllabic, lowered* [z], [zʷ]*; both exhibit noticeable strident frication with a* [z]*-like quality*’. Ling’s (2009) acoustic and EMA analysis show that the two vowels display similar F1 and F2 values, display frication noise in the 3000 – 8000 Hz region, and have a similar flat or concave tongue shape and a same apical-alveolar constriction. Faytak (2018: 94) also reports that the apical vowels and the onset consonant [s] in SWC have a similar articulatory gesture. He argues that this articulatory similarity is expected owing to the co-occurrence restrictions that require the apical vowels to occur immediately following an alveolar fricative or affricate.

2.2.3. Similar phenomena observed in non-Chinese languages

Since Karlgren (1915), the apical vowels are compared to similar segments in non-Chinese languages. For example, the apical vowel [ʅ] is said to be similar to the Swedish 'Viby-i' (Karlgren, 1915: 295). Bell (1978) considers the apical vowels in Sino-Tibetan languages as '*coronal, strident, and voice — in other words, z-like sounds*', and he also notes that '*the only non-Sino-Tibetan language with syllabics of this type is Lendu*'. Ladefoged and Maddieson (1996) consider the apical vowels in SC as fricative vowels, and they report having observed similar phenomenon in the north-western part of the Bantu area (as in Mambila). More recently, the Ikema language of the Japonic family is also reported to have a 'fricative vowel' similar to the apical vowel, to the extent that some researchers use the apical vowel symbol [ɿ] to transcribe it (Fujimoto & Shinohara, 2018). In order to provide a comparison to similar phenomena occurring in non-Chinese languages, the four cases said to be similar to the Chinese apical vowels are reviewed here: the Swedish Viby-i, the Lendu vowelless syllable, the Mambila fricative vowel, and the Ikema fricative vowel.

2.2.3.1. Swedish Viby-i

Swedish (Glottocode: swed1254, ISO 639-3: swe) is a North Germanic language. It has 17 vowels /iː ɪ yː ʏ ʉ̟ ɵ eː ɛː ɛ øː œ ɑː ɑ oː ɔ uː ʊ/ (Engstrand, 1990). The 'Viby-i' is an allophone of /iː/ found in many parts of Central Sweden. It is described as having a 'thick, buzzing, damped' quality (Westerberg, 2019). The Viby-i and the normal [iː] of the same phoneme /iː/ are two allophones that cannot coexist within a speech community, and speakers tend to use either Viby-i or standard [iː] categorically. The Viby-i is also subject to many of the same phonological processes as [iː] (e.g., diphthongisation and end-frication). In the case of the end-frication, it often takes a [z] or [ð] rather than a [j] offglide. At the phonotactic level, the Swedish Viby-i is not conditioned by the consonantal environment (Westerberg, 2016).

The Viby-i has been the object of some acoustic studies. It is characterised by a markedly low F2, and a high F1 and F3, in comparison to standard [iː] (Westerberg, 2016, 2019). This formant pattern is also considered as 'centralised' compared to standard Swedish /iː/ (Björsten & Engstrand, 1999). As shows the Viby-i has a clear formant structure from the beginning of the segment until the offglide. After the release of the [p] onset, there is no visible frication noise. Figure 2.6 shows that on a F1/F2 plane, the Viby-i is centralised, compared to [y e] vowels. The F3 of the Viby-i is also lower than the standard [iː].

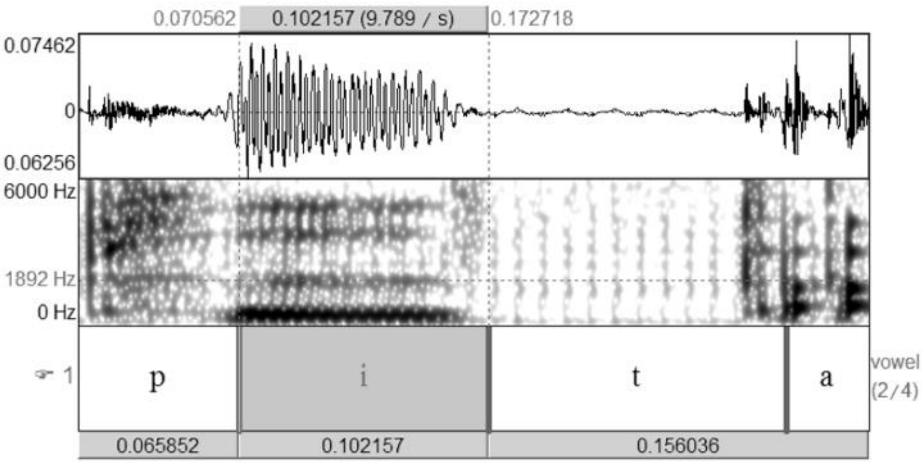


Figure 2.5. Acoustic signal and spectrogram of Viby-i from a young female speaker saying *pita* [pʰɨᶨtɑ̰]. The vowel segment includes the fricated offglide. This figure is reproduced from Figure 2 in Westerberg (2019).

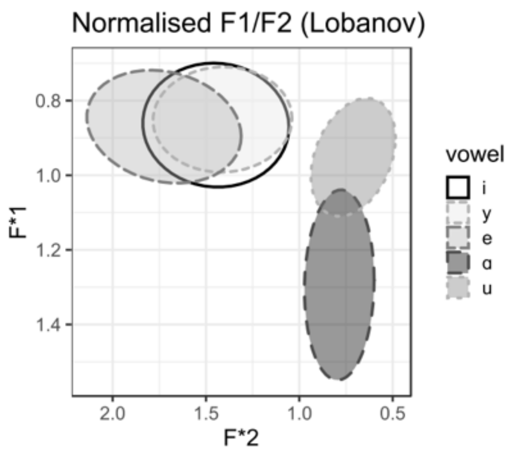


Figure 2.6. Normalised F1/F2 for 34 speakers, the [i] denotes the Viby-i. This figure is reproduced from Figure 3 in Westerberg (2019).

Table 2.3. Mean F1, F2, F3 (Hz) of the Swedish Viby-i from different studies, in comparison with the standard /i/. This table is reproduced from Table 2 in Westerberg (2019) and Table 1 in Björsten & Engstrand (1999).

Viby-i	Females	Males	Males
F1	398	337	350
F2	1946	1709	1590
F3	3209	2741	2860

Standard /i/	Males
F1	291
F2	2107
F3	3135

Articulatorily, the Viby-i is found to be frequently produced with a double-bunched tongue shape as shown in Figure 2.7 (Westerberg, 2016). The tongue dorsum forms the first bunch and the tongue root forms a second bunch. The tongue dorsum has clearly a convex shape, but the highest point is lower than for [e]. Based on this observation, the Viby-i is also described to have a lowered tongue gesture compared to standard [iː], and only 5 out of 18 speakers analysed in Westerberg (2019) have similar tongue heights for /iː/ and /eː/.

In sum, the Viby-i is analysed as a vowel phonologically and phonetically which may be represented with the symbol [ɨ] (Björsten & Engstrand, 1999). It has a centralised formant structure compared to [iː] and is frequently produced with a lowered tongue dorsum compared to other high vowels. A double-bunched tongue shape is found in its articulation. The articulatory characteristics of the Viby-i do not suggest a fricative analysis, but a lowered and backed vowel compared to high vowels. The role of the tongue in producing frication or 'buzzing' is still unclear (Westerberg, 2019).

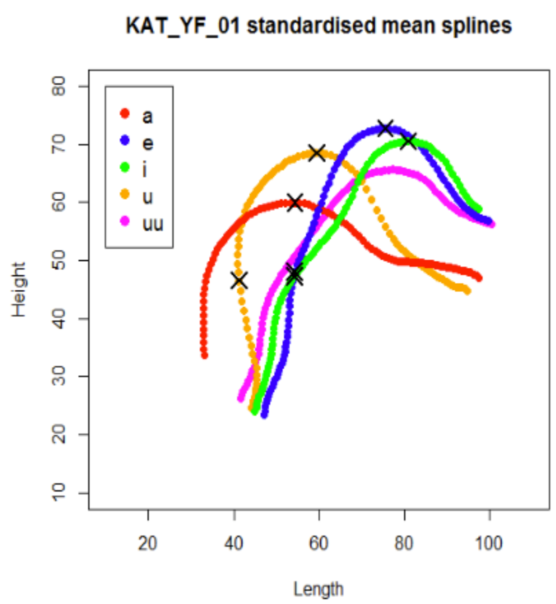


Figure 2.7. Tongue splines showing a double-bunched tongue shape of the Swedish Viby-i. This figure is reproduced from Figure 66 in Westerberg (2016). The [i] denotes the Viby-i in this figure.

The Viby-i seems to be different from the apical vowels in both phonological behaviour and phonetic implementation. Phonetically, the most noticeable difference is that the frication noise is not a typical characteristic of this segment; the ‘damped’ or ‘buzzing’ quality is not systematically discussed in the literature and its tongue contour does not argue for a consonant analysis. Phonologically, the Viby-i is not conditioned by the consonantal environment, and its distribution is not linked to any sibilant sounds.

The resemblance between the Viby-i and the apical vowel [ʅ] is firstly reported by Karlgren (1915: 295). He does not provide any description or argument on this resemblance, only stating that the apical vowel [ʅ] is ‘*très analogue*’ to the Viby-i. His comparison only concerns the apical vowel [ʅ] but not [ɿ]. It is reasonable to assume that apical vowel [ɿ], which is the variant attested in JHC, is not analogous to the Viby-i.

2.2.3.2. Lendu vowelless syllables

Lendu (Glottocode: lend1245, ISO 639-3: led) is a Central Sudanic language within the Nilo-Saharan language family (Kutsch Lojenga, 1989). It has eight underlying vowels, but the exact phonetic characteristics of each sound are not clear. Kutsch Lojenga (1989) presents three [+ATR] vowels /i u ə/ and five [−ATR] vowels /ɪ ɛ U ɔ ɑ/ as underlying vowels for Lendu. The segment in Lendu that resembles to the apical vowels in Chinese languages is the nucleus of the so-called ‘vowelless syllables’. The description of the vowelless syllables by Kutsch Lojenga (1989) can be compared to the descriptions of the SC apical vowels presented earlier in section 2.2.2.1, they are nearly identical from a phonotactic point of view:

> … Lendu (has) on the surface so-called ‘vowelless syllables’. That is, in certain instances the vowel nucleus of the syllable seems to consist of a continuation of the consonantal syllable onset, without displaying a clear vowel quality.
>
> …these vowelless syllables are manifested with the following consonants as onset: s, z, and r, and all complex consonants ending in any one of these three: ts, dz, n(d)z, tr, dr, ndr … pr, kr, kpr, gr, gbr, mbr, ngr, ngbr, ɗr, ɓr and ƭr.
>
> We should note that when the syllable onset s or ts spreads to the nucleus, it acquires voicing in order for the tone to be realised. Phonetically it is as if the s or ts onset is followed by a nucleus consisting of z. The old spelling systems of Lendu … reflect this (z nucleus). … In the following examples the first z or r belongs to the onset, the second one to the nucleus.

kàzz ‘fire’ rr ‘ medicine’
dzz ‘ground’ ndrr ‘goat’
ndzz̀ ‘yesterday’

(Kutsch Lojenga, 1989: 119–120)

As this citation shows, the Lendu vowelless syllables parallel the SC apical vowels in three aspects:

(i) In Lendu and SC, there exist two types of segments. The first one is a [z]-like segment and the second one is a [r]-like segment (or a rhotic-like segment).
(ii) The [z]-like segment occurs after [s ts ts^h] in SC and after [s ts z dz n(d)z] in Lendu. The [r]-like segment occurs after [ʂ tʂ $tʂ^h$] in SC and after [r tr dr ndr pr kr kpr gr gbr mbr ngr ngbr dr βr ɾr] in Lendu. They are arguably homorganic to their onsets.
(iii) They are all obligatorily voiced since they all serve as tone-bearing units.

The phonetic nature of the [z]-like segment[11] in Lendu is studied in Demolin (2002). He reports that there is obvious continuity in the frication from the sibilant onset to the z-like nucleus, as can be seen in Figure 2.8.

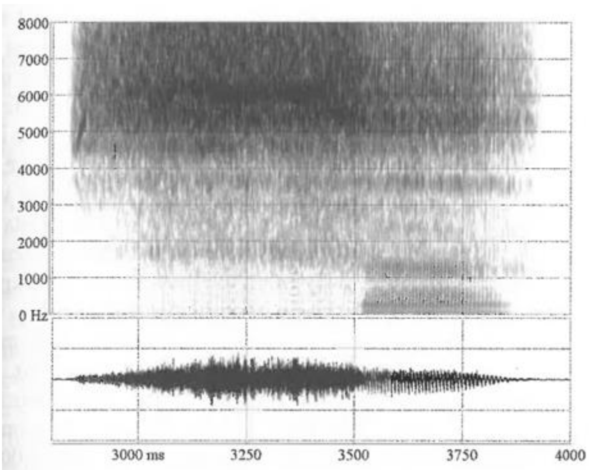


Figure 2.8. Spectrogram and acoustic signal of the Lendu word *sz̀* ‘to prepare beer’, it has an onset [s] and a [z]-like nucleus. This [z]-like nucleus is described as phonologically vowelless. This figure reproduces the Figure 17 in Demolin (2002: 485).

Demolin (2002) reports that in a [z]-like nucleus preceded by an onset [z], the noise is concentrated between 3000 Hz and 5000 Hz. He also notices that when the [z]-like nucleus is preceded by an onset [s], as shown in Figure 2.8, the frication noise

[11] The [r]-like segment was not studied in Demolin (2002).

continues into the syllable nucleus but with less amplitude. The tone is realised during, and phonologically associated with, the voiced nucleus.

Demolin's (2002) aerodynamic data further show that during the word *sɚ̀* in Lendu, the intra-oral pressure behaves in a different way compared to other syllables with normal vowel nuclei. The intra-oral pressure shows that '*after an increase of pressure corresponding to the narrowing of the vocal tract, pressure decreases gradually up to the end of the syllable*' (Demolin, 2002: 487). This pattern is shown in Figure 2.9. Demolin believes that this gradual decrease of intra-oral pressure is correlated to a gradual release of the constriction in the vocal tract, and some constriction is maintained during the vowel. Based on this observation, he further argues that the nucleus in the Lendu word *sɚ̀* has two gestures: an alveolar fricative and a second 'apical dorsal vocoid' gesture.

The case of Lendu vowelless syllables is particularly relevant in the analysis of Chinese apical vowels since they are phonotactically and phonetically similar. The resemblance between the [z]-like nucleus in Lendu and the apical vowel [ɿ] in JHC will be further highlighted in the general discussion (see section 6.3).

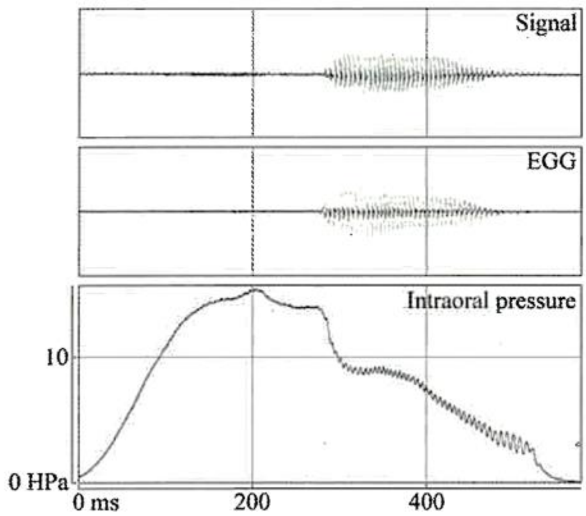


Figure 2.9. Acoustic signal, EGG signal and the intra-oral pressure of the Lendu word *sɚ̀*. This figure is the Figure 21 in Demolin (2002).

2.2.3.3. Mambila fricative vowel

Mambila (Glottocode: mamb1312) is a Bantoid language of the Volta-Congo language family (Connell, 2017). It comprises several dialects or languages straddling the Nigeria-Cameroon border, some of which are mutually intelligible (Connell, 2007). The Len dialect of Mambila, as described in Connell (2007), has nine phonological vowels /i ɨ ɯ u e ɛ o ɔ a/. The fricative vowels in Len are traditionally analysed as the result of a spirantisation process affecting the vowel /ɨ/

when it is preceded by palatal or labiodental consonants. They are noted using digraphs [ʑ͜i] and [v͜ɯ], reflecting their complementary distribution: [ʑ͜i] following labial plosive [b] and postalveolar fricative [ʃ], [v͜ɯ] following consonants [t d k f v] and pre-nasalised consonants [ɱf ɱv].

The fricative vowel [ʑ͜i] can have a clear front vowel quality [i] or a more centralised quality [ɨ] depending on the speaker. The following examples of the syllable [bʑ͜i] are from two different speakers, illustrating the presence of frication noise on this fricative vowel.

The two examples are described as having alveolopalatal frication (Connell, 2007). It is notable that the frication noise of these two fricative vowels patterns differently. In the first case (Figure 2.10 left image), the frication seems to appear right after the labial release and disappears in the middle of the syllable nucleus. The syllable onset being a [b], it is not possible to analyse this frication noise at being part of (or influenced by) the onset. In the second case (Figure 2.10 right image), the frication noise is observable after the labial release at around 5000 Hz and at the end of the syllable nucleus in a range of 5000 – 10000 Hz. It is still not possible to attribute the origin of the frication noise to the syllable onset. Connell (2007) does not provide acoustic signals of the two examples, but he notes that when asked to do several repetitions of the same word, the frication noise varies in intensity and frequency, sometimes it may also disappear entirely and this disappearing of frication noise has no noticeable effect on vowel quality.

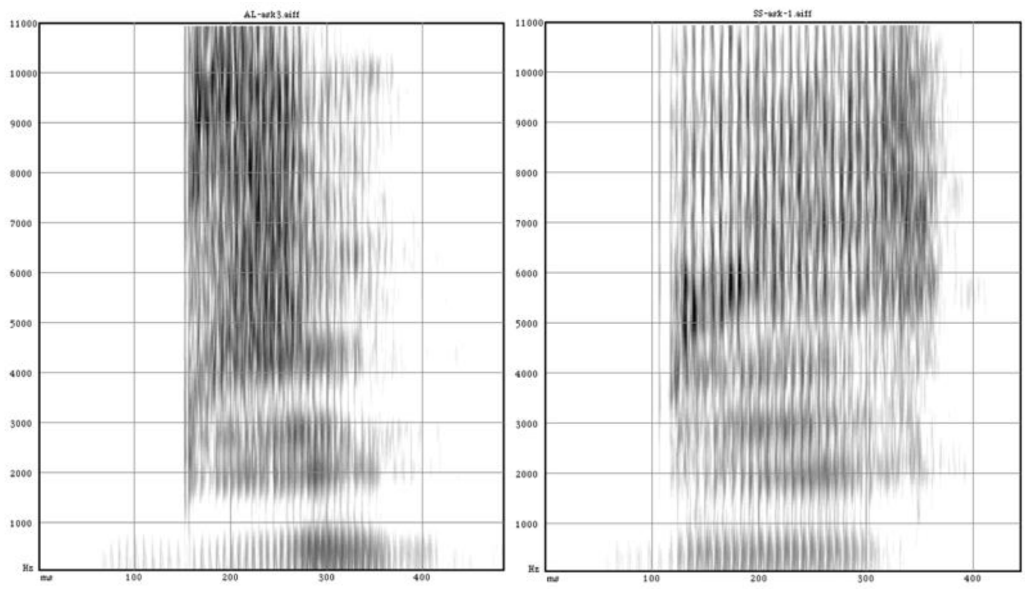


Figure 2.10. Spectrograms of Len (Mambila) fricative vowel with alveolopalatal frication. These two syllables are both [bʑ͜i] 'ask' and are from two different male speakers. This figure is reproduced from the Figure 5a and the Figure 5b in Connell (2007).

As observed in Figure 2.10, the fricative vowel [ʑ͜i] has a F1 at around 500 Hz and a F2 at around 2000 Hz. Considering that the examples are from male speakers, this formant structure is different to what is observed in the apical vowels in Chinese languages. Moreover, the end-frication observed in [ʑ͜i] is not superposed on voicing.

The fricative vowel [v͜ɯ] has a high central unrounded vowel quality, and it is described as having labiodental frication noise (Connell, 2007). An example having [v͜ɯ] as a nucleus is presented in Figure 2.11. As observed in Figure 2.11, the [kv͜ɯ] syllable has a slightly aspirated [k], and the frication noise seems to appear at the end of the nucleus, not at its beginning. Connell (2007) notes that there is no intervening period of frication between the consonant and the vowel. Again, this end-frication, similar to the end-frication observed in [ʑ͜i], cannot be argued to be part of the syllable onset [k] and is not superposed on voicing.

As Connell (2007) summarises, the vowel qualities of the two fricative vowels are not schwa-like, and the end-frication cannot be directly associated to an onset consonant. He further argues that the frication associated with the fricative vowels in Len is best seen as a vowel feature. This analysis is identical to the proposal of Ladefoged and Maddieson (1996: 314) presented above.

The fricative vowels in Len do not resemble the apical vowels in Chinese languages, neither phonologically, nor phonetically. It seems that the 'fricative vowel' proposal of Ladefoged and Maddieson (1996: 314) is more suitable for the fricative vowels of Mambila than for the apical vowels of Chinese languages.

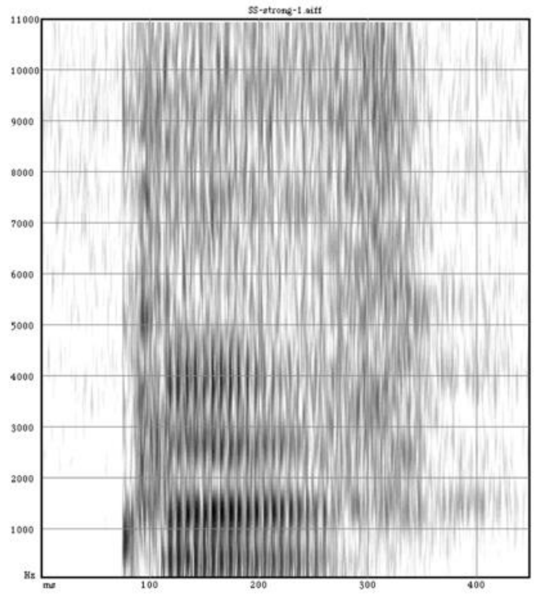


Figure 2.11. A spectrogram of Len (Mambila) fricative vowel with labiodental frication. The syllable [kv͜ɯ] 'strong' is from a male speaker. This figure is reproduced from the Figure 6a in Connell (2007).

2.2.3.4. Ikema fricative vowel

Ikema (Glottocode: ikem1234) is a Ryukyuan language belonging to the Japonic language family. It is spoken on Ikema Island, Sarahama (Irabu Island) and Nishihara (main Miyako Island) in Miyako-jima City of Okinawa Prefecture. Like other Ryukyuan varieties, Ikema is generally not spoken by the younger generations (Hayashi, 2010). Ikema has four main vowels /i u a ɨ/ and two other vowels /e o/ which appear only in interjection of sentence final particles (Hayashi, 2010). The vowel argued to be the 'fricative vowel' is /ɨ/. Ikema is a language with a word-tone system. Lexical tone appears at the right-most position of the lexical word. Two tones (a rasing/high level tone and a falling tone) are observed in the Ikema tonal system (Hayashi, 2010).

The /ɨ/ vowel must be preceded by a consonant, restricted to /s z c f/. It is argued that this vowel /ɨ/ contains frication noise and can be transcribed using the /ɯ/ symbol or even the apical vowel symbol [ʅ] (Fujimoto & Shinohara, 2018). Fujimoto & Shinohara (2018) studied this vowel /ɨ/ with MRI (Magnetic Resonance Imaging) method. Their results show that the articulation of this vowel does not resemble that of a close central vowel. As Figure 2.12 clearly shows, /ɨ/ has a narrowed oral cavity at the alveolopalatal area, with a raised tongue tip for both speakers and a raised tongue dorsum for the first speaker (left image). The F1/F2 pattern shows that this vowel occupies the place of a close central vowel in the acoustic space (F1 around 380 Hz and F2 around 1380 Hz).

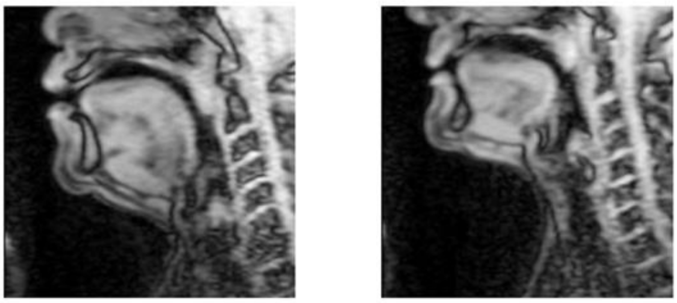


Figure 2.12. MRI images of the central vowel (fricative vowel) of Ikema language from two speakers. This figure is reproduced from the Figure 1 in Fujimoto & Shinohara (2018).

The Ikema vowel /ɨ/ has several similarities with the apical vowels in Chinese languages, these similarities can be found on both phonological and phonetic levels. Firstly, the vowel /ɨ/ appears only in nucleus position and can only be preceded by fricative consonants. This restricted distribution mirrors the SC and SWC apical vowel (see section 2.2.2.1 and section 2.2.2.3.3), which can only be preceded by alveolar fricatives. One small difference though is that the onset fricative consonants in Ikema, namely /s z c f/[12], do not have the same place of articulation. However, recall that the apical vowel in JHC, HMC and QMC (see section 3.5, section 2.2.2.3.1 and section 2.2.2.3.2) can all be preceded by labial consonants (though not /f/). Secondly, Ikema is a tonal language. The vowel /ɨ/ serves as a tone-bearing unit, just like any apical vowel in Chinese languages. Thirdly, the vowel /ɨ/ contains frication noise, and it is also called 'fricative vowel' and transcribed using the apical vowel symbol [ɿ]. Its F1/F2 pattern resembles that of an apical vowel. Its articulation is studied for only two speakers but the results obtained are comparable to the JHC apical vowel. The two speakers in Figure 2.12 present two different gestures: the first speaker (left) has a slightly raised tongue tip and a high raised tongue dorsum, the second speaker (right) has a raised tongue tip but the tongue dorsum is flat. This similarity is also reported in 5.2.2 with comments.

Overall, the Ikema fricative vowel provides a new case of 'fricative vowel' comparable to the apical vowel in Chinese languages. The similarity, which merits further investigation, lies in both the phonological function of the segment and its articulatory configuration.

2.3. Summary of the review

In this review, apical vowels in Chinese languages and similar sounds in non Chinese languages are presented with a focus on their phonological status, and their phonetic manifestation. As sections 2.2.2 and 2.2.3 show, apical vowels can vary both in phonological behaviour and phonetic implementation.

Phonetically, the presence of frication noise on apical vowels is at the centre of the debate since it is one of the main factors that help determine their phonetic nature. The presence of frication noise argues for a fricative/sibilant analysis and the absence of frication noise argues for an approximant/vowel analysis. Depending on the dialect, and sometimes on the speaker, apical vowels may exhibit a more or less important frication noise or even no frication noise at all. Any generalisation on the phonetic nature of apical vowels needs to take this variability into account.

The articulation of the apical vowels is less known to researchers, as experimental studies have only been done on the variants that occur after alveolar sibilants (i.e., SC and SWC). Indeed, there are no articulatory studies conducted on apical vowels occurring after consonants other than alveolar sibilants. It is natural to relate the articulatory similarity to the phonotactic constraint between the apical vowels and the sibilant onsets and argue that the onset and the nucleus (i.e., the apical vowels) are homorganic. But this is simply not applicable to all apical vowels in Chinese languages, since not all apical vowels occur after alveolar sibilants only. The

---

[12] The /c/ consonant has two allophones [ts tɕ] (Hayashi, 2010), there is no description on their distribution. The vowel /ɨ/ seems to occur only after the allophone [ts] (Fujimoto & Shinohara, 2018). It may be the case that the /ɨ/ vowel occurs only after coronal fricatives and affricate [s z ts] and labial fricative [f].

articulatory characteristics of the apical vowels after labial consonants should reveal their true articulatory nature since the labial onset should not have any influence on the tongue gesture of the following nucleus.

Phonologically, apical vowels do not display similar behaviour across Chinese languages. Depending on the dialect, they can have a more or less restricted distribution and they can be phonemic or non-phonemic. It seems that the apical vowels always occur after alveolar sibilants but can also occur additionally after labial consonants and even lateral consonant; the occurrence of the latter implies the occurrence of the former.

Last but not least, the similarity between apical vowels and other segments in non-Chinese languages is not always based on solid grounds. It seems that the apical vowels in Chinese languages resemble only to the vowelless syllables in Lendu and the Ikema fricative vowel as discussed in section 2.2.3. The reported similarity between the apical vowel and the other two segments (i.e., the Swedish Viby-i and the Mambila fricative vowels) is overestimated.

This review raises several questions about the phonetic and phonological nature of the apical vowel in JHC:

- Are apical vowels produced with frication noise? If so, is it systematic or occasional?
- What constriction in the vocal tract is responsible for the generation of the frication noise?
- How are the apical vowels articulated? More precisely, are they articulated with a fricative gesture or a vowel gesture?
- The apical vowels are not always homorganic to their onset consonants since they are not always preceded by alveolar sibilants. How do they behave articulatorily when preceded by non-alveolar onsets?
- It is also clear that the apical vowels do not have the same phonological status in every Chinese language. How do they behave when they are not allophonic to high vowel [i] as in SC?

There is only one apical vowel in JHC, the one that is described as an apical front vowel or dental apical vowel with the symbol [ɿ]. I adopt the /z̩/ symbol to transcribe this segment in JHC, following Dell (1994), Wiese (1997), Yu (1999), Duanmu (2007) and the IPA symbol guidelines (IPA, 2010). This symbol, as I shall show, reflects the phonetic nature and the phonological function of the apical vowel in JHC: a voiced alveolar fricative that occupies the nucleus position of a syllable. When mentioning the other apical vowels in Chinese languages, the [z] notation is adopted for simplicity.

Chapter 3
Jixi-Hui Chinese

In this chapter, a detailed description of the sound inventory of JHC is provided. The chapter is organised as follows: consonants, vowels and tones are described, with a focus on syllabic consonants /m̩ n̩/ [v̩]. The apical vowel /z̩/ and its phonological behaviour are presented at the end of the chapter.

The present description of the sound inventory of the language is mainly based on Zhao's (1989, 2003) work, but I also provide a description of the allophonic variants based on my own observations.

3.1. General introduction

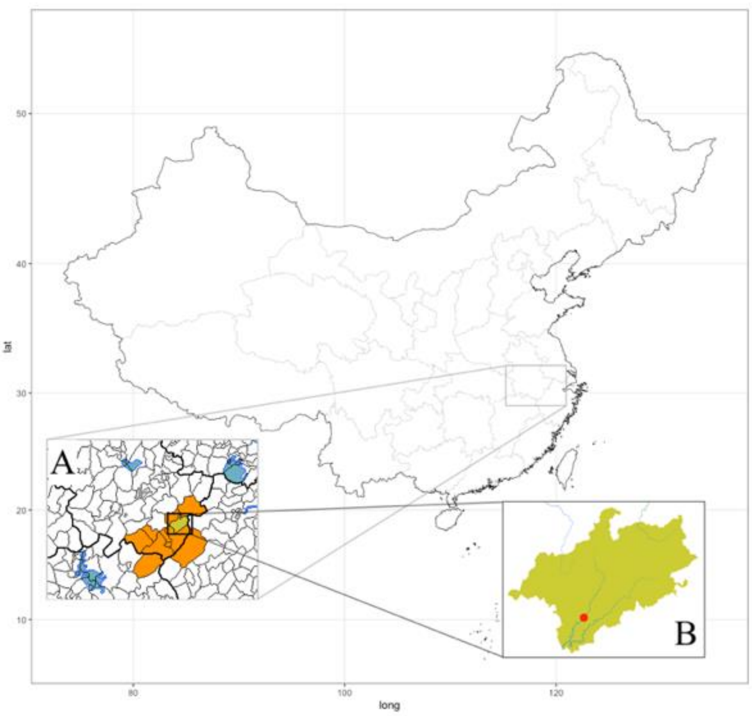


Figure 3.1. Location of Jixi county 绩溪县 in China. The coloured zone in A corresponds approximately to the area of Hui 徽 group Chinese language (Li et al., 1987) , the red dot in B corresponds to the town Huayang 华阳镇, the administration centre of Jixi county.

JHC (Glottocode: jixi1238) is a Hui 徽 group Chinese language (ISO 693-3: czh, Glottocode: huiz1242), spoken in the Jixi county 绩溪县 in Anhui province 安徽省 (Li et al., 1987). It has two major variants, the Lingnan 岭南 variant and the Lingbei 岭北 variant (Zhao, 1989, 2003). The administration centre is located at the town of Huayang 华阳镇 where the Lingnan variant is spoken. The recent descriptions (Hirata, 1998; Zhao, 1989, 2003) and the present study are all based on this variant. The most recent population census reported 160 000 inhabitants in Jixi County (Anhui Bureau of Statistics, 2019). Although JHC is still widely spoken in both professional and familial contexts, the youngest generations predominantly speak SC, which is the language of education. Figure 3.1 shows where the Jixi county can be located on a map of China. The red dot in Figure 3.1-B represents the town Huayang, where all our recording sessions took place.

3.2. Consonants

Table 3.1. Consonant inventory of JHC.

	Bilabial	Labiodental	Alveolar	Palatal	Velar	Glottal
Plosive	p pʰ		t tʰ		k kʰ	ʔ
Affricate			ts tsʰ			
Nasal	m		n		ŋ	
Fricative		f	s		x	
		v	z			
Approximant	w			j ɥ		

The examples shown in Table 3.2 are chosen according to the following criteria: a nucleus /a/ is used whenever possible; the tone /˧˥/ is used whenever possible, otherwise the tone /˧/ is used, or the tone /˨˩˨/ is used if no lexical item is available. Exceptions are [nz̩˧], [ɥã˧˥], [kwa˧˥] and [ɔʔ˥˧]: /n/ occurs as onset only in [nz̩˧] and before nasal vowels; /w/ occurs only in combination with /k kʰ/ consonants; /ɥ/ has a very limited distribution, /ɥɔ ɥã ɥĩ ɥɑʔ ɥeʔ/ are the only syllables in which /ɥ/ occurs as onset; /ʔ/ appears as syllable coda in closed syllables with checked tone and only idiosyncratically as syllable onset (see below). The '-' marks the cases where the corresponding orthographic transcriptions are not available.

Table 3.2. A list of JHC items with orthographic forms and glosses illustrating the consonants.

	Phonetic	Orthographic	Gloss
p	pɑ˦˥	背	‘back’
pʰ	pʰɑ˦˥	配	‘match’
m	mɑ˦˥	妹	‘child’
f	fɑ˧	回	‘return’
v	vɑ˧	-	‘make a fuss’
t	tɑ˦˥	对	‘pair’
tʰ	tʰɑ˦˥	退	‘recede’
ts	tsɑ˦˥	再	‘again’
tsʰ	tsʰɑ˦˥	菜	‘vegetable’
n	nz˧	泥	‘soil’
s	sɑ˦˥	碎	‘broken’
z	z̩˦˥	腻	‘greasy’
w	kwɑ˦˥	-	‘spill’
j	jɑ˨˩˨	也	‘too’
ɥ	ɥã˦˥	润	‘moisten’
k	kɑ˦˥	盖	‘lid’
kʰ	kʰɑ˨˩˨	隑	‘stand’
ŋ	ŋɑ˦˥	爱	‘love’
x	xɑ˧	孩	‘child’
ʔ	ɔʔ˦	阿 (as in 阿姆)	‘paternal grand-mother’

In this section, the allophonic variants of the consonants, their distributions, and their possible position in a syllable are presented. There is no experimental study known to the author on the phonetic nature of the consonants in JHC.

3.2.1. Plosives and affricates

All plosives and affricates occur only as syllable onsets and are paired as aspirated and non-aspirated, except for the glottal plosive /ʔ/. Aspirated /pʰ tʰ kʰ tsʰ/ are strongly aspirated with a much longer VOT compared to their non-aspirated counterparts /p t k ts/. As mentioned above, the glottal stop occurs as syllable coda when the syllable has a checked tone (as in [iɑʔ˦] ‘leaf’). It also occurs idiosyncratically in onsetless syllables, as for example in the form /ʉ˨˩/ ‘speech’ which can be realised as [ʔʉ˨˩].

The bilabial plosives /p pʰ/ can precede the apical vowel /z̩/, as well as oral vowels and nasal vowels /ʉ u ɤ o ɔ a ɑ̃ õ ẽ/ in open syllables (e.g., [pa˦˥] 'back', [pz̩˦˥] 'close'). In closed syllables, they can precede only /ɤ o ɔ/ vowels (e.g., [poʔ˦] 'to shell' [pɤʔ˦] 'no', [pɔʔ˦] 'eight'). They can also be followed by the /j/ glide. Zhao (1989, 2003) presented one syllable [pi˥˧] 'stela' in which /p/ can be followed by a /i/ vowel; this syllable is considered illegal by the speakers recorded in the present study and the word for 'stela' 碑 is pronounced [pa˥˧] instead.

The alveolar plosives /t tʰ/ are produced with the tongue tip against the alveolar ridge; but they can be produced with the tip of the tongue against the upper teeth when preceded by /i a/ vowels. They can occur before the glide /j/ and are palatalised in that context. They can also occur before /ʉ u ɤ ɔ ɑ̃ ẽ õ/ vowels in open syllables and before /ɤ o/ in closed syllables (e.g., [tʉ˨˩˧] 'hide', [toʔ˦] 'dot').

The velar plosives /k kʰ/ can also occur before /ʉ u ɤ o ɔ ɑ̃ ẽ õ/ vowels in open syllables and before /ɤ o ɔ/ in closed syllables (e.g., [kʉ˨˩˧] 'wrap', [koʔ˦] 'angle'); they are palatalised when followed by front vowel /i/. However, they can only precede /w/ glide, and the lip rounding starts at the same time as the velar closure, the entire consonant being rounded.

The alveolar affricates /ts tsʰ/ have the same place of articulation as their plosive counterparts, and they can also be dentalised when preceding /i a/ vowels. They can precede /i ʉ u ɤ o ɔ a z ɑ̃ ẽ õ/ vowels in open syllables and /ɤ o ɔ/ in closed syllables (e.g., [tsʉ˨˩˧] 'left', [tsoʔ˦] 'catch'), but cannot precede any glide. In the context of /y ĩ/ vowels or /j ɥ/ glides, they are realised as alveolo-palatal affricates [tɕ tɕʰ] (e.g., /tsy˨˩˧/ [tɕy˨˩˧] 'mouth'). These allophones are produced with the tongue apex against the alveolar ridge and the anterior of the hard palate.

3.2.2. Fricatives

JHC has four fricatives in four different places of articulation: /f v s x/. The labial-dental fricatives /f v/ occur as onsets preceding oral and nasal vowels /i u ɤ ɔ a ɑ̃ ẽ/ in open syllables and preceding /ɔ/ in closed syllable; but they cannot be followed by any glide (e.g., [fa˥˧] 'dust', [fɔʔ˦] 'law'). [v] occurs also as an allophonic variant of /u/ vowel, as will be discussed in section 3.2.5.

The alveolar fricative /s/ has an allophonic alveolo-palatal fricative [ɕ] variant; they have the same place of articulation as their plosive and affricate counterparts, and display the same distribution as them: the palatal fricative [ɕ] occurs only before vowels /y ĩ/ and /j ɥ/ glides, while the alveolar fricative [s] occurs before /i ʉ u ɤ o ɔ a z ɑ̃ ẽ õ/ (e.g., [si˨˩˧] 'wash', /sy˨˩˧/ [ɕy˨˩˧] 'water'). The alveolar [s] can also be dentalised when preceding /i a/ vowels.

The velar /x/ occurs as an onset preceding oral and nasal vowels /i ʉ ɤ o ɔ a ã õ ẽ/ but cannot be followed by any glide (e.g., [xa˨˩] 'sea'); it is palatalised when preceding the vowel /i/. This velar is highly uvularised when preceding back vowels and could thus arguably have an allophonic [χ] variant in /xo xɤ xɔ xã xõ/ syllables (e.g., /xo˨˩/ [χo˨˩] 'down', /xɤ˨˩/ [χɤ˨˩] 'good').

3.2.3. Nasals and lateral

JHC has three nasal consonants /m n ŋ/; the alveolar nasal /n/ has three allophonic variants [n ɲ l]. The nasal /ŋ/ and the allophonic variants [ɲ l] can only occur as syllable onsets, while /m/ and [n] can occur as syllable onsets or as syllabic consonants. The syllabic /m/ and [n] will be discussed in section 3.2.5.

When it is an onset, the labial nasal /m/ can be followed by the apical vowel /z̩/, the oral and nasal vowels /ʉ ɤ o ɔ a ã õ ẽ/ (e.g., [mʉ˨˩] 'someone', [mz̩˨˩] 'rice'), and it can also be followed by the /j/ glide. Zhao (1989, 2003) presented a [mi˨˩] 美 'beautiful' syllable in which [m] can be followed by /i/. This syllable is considered illegal by the recorded speakers in the present study. They argued that the Chinese character 美 'beautiful' is rather pronounced [mẽ˨˩].

The alveolo-palatal [ɲ] is produced with the tongue apex against the alveolar ridge and the anterior of the palate, akin to [tɕ tɕʰ ɕ] consonants. Chinese dialectologists have been using the non-IPA symbol [ȵ] to transcribe this sound. The different variants of the alveolar nasal /n/ have a complex distribution: the alveolar [n] occurs as onset with the apical vowel /z̩/ and can also function as a syllabic consonant (e.g., [nz̩˨˩] 'in'); the alveolo-palatal [ɲ] occurs as onset only with /y/ and can precede /j ɥ/ glides (e.g., /ny˨˩/ [ɲy˨˩] 'women'); the lateral [l] occurs before /i ʉ u ɤ o ɔ a/ as onset and cannot be followed by any glide (e.g., /ni˨˩/ [li˨˩] 'willow'). However, if the rime is a monophthong nasal vowel, the alveolar [n] and the lateral [l] can be in free variation, thus [lã] and [nã], [lẽ] and [nẽ], [lõ] and [nõ] are equally acceptable.

The velar nasal /ŋ/ occurs as onset with oral and nasal vowels /i ʉ ɤ o ɔ ɑ ã ẽ/, it cannot be followed by any glide (e.g., [ŋi˨˩] 'lotus root', [ŋã˨˩] 'cover'). /ŋ/ is palatalised when followed by /i/ and /ẽ/, produced in this case with the tongue dorsum against the posterior of the hard palate.

3.2.4. Approximants

In recent IPA descriptions of Chinese languages (Chen & Gussenhoven, 2015; Li et al., 2019; Zeng, 2020), the segments [j ɥ w] are denominated by two different terminologies: approximant and glide. Although ‘approximant’ and ‘glide’ are often used interchangeably, a distinction could be made depending on their role within a syllable: ‘glide’ when it is part of the nucleus position and ‘approximant’ when it is at the margins of the syllable. Duanmu (2007) and Li et al. (2019) consider for SC and Tianjin Mandarin, that the [j ɥ w] are part of the syllable onsets. In Shanghai Chinese (Chen & Gussenhoven, 2015), the glides after alveolo-palatals are argued to be mere transitional elements. Traditionally, the JHC [j ɥ w] are analysed as being part of the syllable nucleus and form diphthongs with the following vowel (Zhao, 1989, 2003).

In JHC, the exact phonological and phonetic nature of these sounds will not be dealt with in the present study. The phonotactics are presented without further phonological analysis. The described common practice is followed. The [j ɥ w] segments are presented in syllable initial position and referred to as approximant onsets. They are referred to as glide in post-consonantal, pre-vocalic position. None of them contrast with their high vowel counterparts in both positions.

Amongst [j ɥ w], only [j ɥ] can occur syllable initially (e.g., [ja˨˩˧] ‘too’, [ɥã˨˩˧] ‘ever’). These initial approximant onsets are alveolo-palatal, with lip rounding for [ɥ]. The onset [j] occurs before [ɔ o ʉ e ã ĩ õ] in open syllables and before [a o e] in closed syllables. The onset [ɥ] occurs before [a ɔ ã ĩ] in open syllables and before [a e] in closed syllables.

All three [j ɥ w] occur at post-consonantal, pre-vocalic position (e.g., [pje˨˩˧] ‘watch’, [pjã˨˩˧] ‘bread’, [ɕɥɑ˨˩˧] ‘toss’, [kwɑ˨˩˧] ‘turn’). The glide [w] only occurs after /k kʰ/ consonants (e.g., [kwɑ˦˥] ‘spill’, [kʰwɑ˦˥] ‘fragment’). They can be followed by /i ɔ a ã ẽ/ in open syllables and by /ɔ ɤ/ in closed syllables. The glide [ɥ] only occurs after /ts tsʰ s n/ consonants. These consonants are palatalised and realised as their allophonic variants [tɕ tɕʰ ɕ ɲ] (e.g., /sɥa˨˩˧/ [ɕɥa˨˩˧] ‘toss’). [ɥ] can be followed by /a ɔ ã ĩ/ in open syllables and by /a e/ in closed syllables.

The glide [j] has the largest distribution, it occurs with /p pʰ m t tʰ ts tsʰ s n/ consonants (e.g., [pje˨˩˧] ‘watch’, /sjã˨˩˧/ [ɕjã˨˩˧] ‘awake’). It triggers the palatalisation of /ts tsʰ s n/ which are realised as their allophonic variants [tɕ tɕʰ ɕ ɲ]. When preceded by labials /p pʰ m/ and alveolar plosives /t tʰ/, the glide [j] can be followed by /e ã/ in open syllables and by /e/ in closed syllables. When preceded by [tɕ tɕʰ ɕ ɲ], the glide [j] can be followed by /ɔ o ʉ e ã õ/ in open syllables and by /a o e/ in closed syllables.

More detailed phonological analysis is necessary in order to fully understand the behaviour of the three glides in JHC. Based on phonemic economy, co-occurrence restrictions and rhyming patterns in poetry, a more adequate analysis would probably be to consider the glides as part of the onset. For example, in rhymed folksong N°20 of Luo’s (1936) collection, [mã˥˦], [jã˥˦], [tsʰã˥˧] and [kwã˥˧] are the rhymes of four verses. Even though the [j] in [jã˥] is arguably the onset of the syllable, it is clear that the presence of glide [w] does not change the rhyming pattern. Hence [w] cannot be analysed as part of the rhyme. As far as this example is concerned, it is more convenient to analyse the glide as part of the onset of the syllable and the rhyme is thus /ã/.

Table 3.3. The distribution of the onset consonants and the glides [j w ɥ] in JHC.

	labial		alveolar			velar	occur as onset
	plosives	nasal	plosives	nasal	sibilants		
j	yes	yes	yes	yes	yes	no	yes
ɥ	no	no	no	yes	yes	no	yes
w	no	no	no	no	no	yes	no

The phonological status of the glides is crucial in establishing the syllable structure of JHC. Indeed, a maximal Chinese syllable, regardless of the dialect, is often thought to contain four positions, or CGVX, where C is a consonant, G a glide, V a vowel, and X either a consonant or the second part of a long vowel or diphthong (Duanmu, 2011). In JHC, the glides' phonological status defines the maximal syllable: if the glide is affiliated to the onset, then the maximal syllable is CVʔ, where the C could be a consonant with double articulation (e.g., /mʲ/)[13]; if the glide is affiliated to the rime, then the maximal syllable is CGVʔ. In any case, the onset is optional, the nucleus is obligatory, the coda is optional and can only be a glottal plosive /ʔ/.

3.2.5. Syllabic consonants /n̩ m̩/ and [v̩]

In this section I present the phonological behaviour of /n̩ m̩/ and [v̩] consonants; the apical vowel /z̩/ is presented in section 3.5. The two syllabic nasals /n̩ m̩/ are presented in the first place, before dwelling on the more complicated case of the syllabic fricative [v̩].

Syllabic nasals are rather common in Chinese language. In SC, the syllabic nasals are used in interjections (Duanmu, 2007), but lexical syllabic nasals are attested widely in other Chinese languages (Duanmu, 2011) (see also section 6.4). In JHC, /m̩ n̩/ are lexically developed on several tones. All lexical items with syllabic /m̩ n̩/ according to the JHC dictionary (Zhao, 2003) are presented in Table 3.4. Despite the limited lexicon, they are full lexical items and are extremely recurrent in daily speech.

[13] This analysis follows Duanmu's Consonant-Glide combination (CG combination) proposal (Duanmu, 2007: 25), since he argues that there is only one time-slot in the onset which C and G must share. The other possible analysis would be that CG form a complex onset. In this case the maximal syllable would still be CGVX where C and G are both affiliated to the onset.

Articulatorily, the syllabic nasals /n̩ m̩/ are produced with maintaining the consonantal gesture during the whole duration of the syllable: the tongue does not move in a /n̩/ syllable after the alveolar construction; the lips do not open in a /m̩/ syllable. Since they are pronounced without opening of the articulators, it is hard to claim that they contain a vowel (Duanmu, 2011). Therefore, they are the rhyme of the syllable and the tone-bearing unit.

Table 3.4. A list of JHC items with syllabic /m̩ n̩/, presented with orthographic forms and glosses (Zhao, 2003).

Phonetic	Orthographic	Gloss
m̩˧	无	'no, none'
m̩˨˩˧	母	'female animal'
m̩˨˩˧	姆	'mother'
n̩˧	儿	'son'
n̩˨˩˧	你	'you'
n̩˥˩	二	'two'
n̩˥˩	尔	'this, here'

In addition to the syllabic nasals, the /u/ vowel in JHC has an allophonic [v̩] variant, which occurs exclusively in syllables with a /f v/ onset. Zhao (1989, 2003) noted that when onset /v/ is followed by /u/, the syllable is realised as [v̩]; the same phenomena is observed with a /f/ onset. Phonologically, [v̩] is in complementary distribution with [u] in the syllable nucleus position, and they are both the rhymes of the syllable. In rhymed folksong N°38 of Luo (1936)'s collection, [v̩˥˩] and [ku˥˩] are the rhymes of two verses. This suggests that syllabic [v̩] is the rhyme and the tone-bearing unit of the syllable, just like [u].

Articulatorily[14], in /fu vu/ syllables, when the upper teeth and the lower lip achieve the fricative gesture for onsets, they do not move until the end of the syllable. It also seems that the tongue achieves the /u/ position during the onset and maintains this position throughout the entire duration of the syllable. A consequence of this is that the syllabic [v̩] is highly velarised, and can thus be transcribed more appropriately as [v̩ˠ]. Table 3.5 provides a list of items with [v̩] in nucleus position. As we can see, this syllabic consonant can occur with all tones; but occurs only in open syllables and not in closed syllables (i.e., not on the checked tone).

[14] There is no articulatory study available on the syllabic [v̩] in JHC. To the best of my knowledge there is no articulatory study on any syllabic [v̩] in Chinese languages. The articulatory description given here is based on my impressions as a native speaker.

Table 3.5. A list of JHC items having [v̩] nucleus with orthographic forms and glosses.

Phonetic	Orthographic	Gloss
fv̩˧˨	夫	‘husband’
fv̩˦	浮	‘float’
fv̩˧˩˧	虎	‘tiger’
fv̩˧˥	富	‘rich’
fv̩˧˨	户	‘household’
v̩˧˨	乌	‘black’
v̩˦	无	‘none’
v̩˧˩˧	五	‘five’
v̩˧˥	焐	‘warm up’
v̩˧˨	姆	‘paternal grand-mother’

Syllabic [v̩] is less common in Chinese languages compared to syllabic nasals, however it is attested in different groups of Chinese languages (see section 6.4 for more details). Descriptions show that, not only in JHC but in all consulted descriptions of Chinese languages, the syllabic [v̩] is always allophonic to /u/, it occurs when /u/ is preceded by [f v] (Zhu, 1995: 21; Hirata, 1998: 52, 69, 146). This particular distribution parallels that of the apical vowel in SC. As discussed earlier in section 2.2.2.1, the SC apical vowels occur after dental and retroflex sibilants while [i] occurs elsewhere. There exists a co-occurrence restriction between the vowel [i] and the apical vowels. Based on this phonological patterning, the apical vowels have been argued to be ‘vowels’ since they serve as the nucleus of the syllable and the tone-bearing unit, akin to the vowel [i].

The exact same behaviour can be found for the JHC syllabic [v̩] and [u][15]. The syllabic [v̩] occurs after labio-dental fricatives while [u] occurs elsewhere. The syllabic [v̩] has the same phonological function as [u]: it serves as the nucleus of a syllable and as the tone-bearing unit. Based on the phonological patterning and following the apical vowel example, one could possibly argue that the syllabic [v̩] is a ‘labial vowel’ in complementary distribution with the vowel [u]. To the best of my knowledge, no researcher has ever argued in this direction for the syllabic [v̩] in JHC or in any Chinese language that contains this segment.

On the phonological level, the onset consonant /v/ is phonemic, it occurs as onset and is not conditioned by any specific phonotactic constraint (i.e., it has exactly the same context of occurrence as its non-voiced counterpart /f/). Its phonetic

[15] The phones [v̩ u] with the same distribution as in JHC are found in Shanghai-Wu Chinese 上海话 (Zhu, 1995, 2004), Shexian-Hui Chinese 歙县话, Qimen-Hui Chinese 祁门话, Wuyuan-Hui Chinese 婺源话 (Hirata, 1998), and Meijiang-Hakka Chinese 梅江客家话 (Yuan, 1983).

implementation (the tongue shape during the frication) is not velarised but may be influenced by the following nucleus due to coarticulatory effects. The nucleus [v̩], however, occurs only after /f v/ onsets. It is not a phoneme but an allophone of the phoneme /u/. Its context of occurrence is highly constrained by the onset consonants and its phonetic implementation is systematically velarised.

The syllabic [v̩] may be accounted for by a feature spreading processes. That is, the labial feature of /u/ is delinked and the labial feature of the onset /f/ spreads to /u/. This analysis is presented in Figure 3.2, only relevant nodes and features are shown in the feature geometry (Clements & Hume, 1995; Duanmu, 2007).

This feature spreading analysis follows the basic assumption of the syllable structure of Chinese languages (Duanmu, 2011). In a syllable, the onset occupies one time-slot and the rhyme two time-slots. Since the rhyme consisted the nucleus /u/ only, it occupies the two time-slots of the rhyme. The onset /f/ is linked to the C-place node which dominates the place feature [Lab], which corresponds to the labiodental constriction. The rhyme /u/ has two place features under the V-place node: [Lab] and [Dor], which correspond to the lip rounding and the retracted and raised tongue dorsum. In a /fu/ syllable, the feature [Lab] of the onset /f/ may spread from the C-place node of the onset to the C-place node of the following rhyme /u/, thereby delinking the rhyme's [Lab] under the V-place node due to incompatibility. This results in a syllabic rhyme that shares the [Lab] specification of the onset /f/ but keeps its [Dor] specification. The phonetic output of the rhyme is thus a velarised syllabic fricative [v̩].

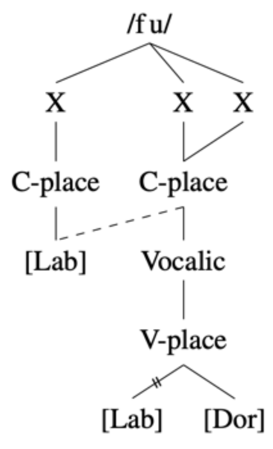


Figure 3.2. Feature spreading in a /fu/ syllable of JHC, only relevant nodes and features are shown. [Lab] stands for [Labial], [Dor] stands for [Dorsal].

3.3. Vowels

3.3.1. Oral vowels

JHC has nine monophthong oral vowels /i y ʉ u e o ɤ ɔ a/, all of which occur in open syllables, and /ɑ ɔ o ɤ/ may also occur in closed syllables.

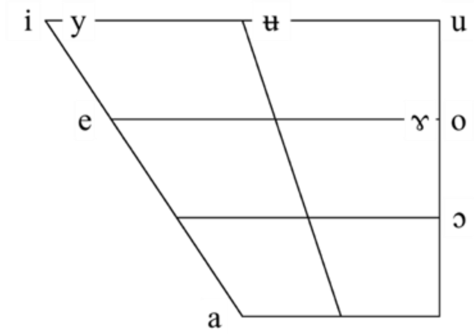


Figure 3.3. Vowel chart of JHC.

The examples below are chosen according to the following criteria: Onset /p/ onset and tone /˨˩˦/ are used whenever possible; onset /f/ or a syllable without onset is used when onset /p/ is not possible; the tone /˥˩/ is used if the tone /˨˩˦/ does not produce the desired lexical item; the /je/ syllable is an exception due to the limited distribution of /e/.

Table 3.6. A list of JHC items with orthographic forms and glosses illustrating the vowels.

	Phonetic	Orthographic	Gloss
i	fi˨˩˦	匪	'bandit'
y	y˨˩˦	雨	'rain'
ʉ	ʉ˨˩˦	-	'loiter'
u	pu˨˩˦	补	'complete'
e	je˨˩˦	有	'have'
o	po˨˩˦	把	'handle'
ɤ	pɤ˨˩˦	饱	'full'
ɔ	pɔ˨˩˦	板	'board'
ɑ	pɑ˥˩	杯	'cup'

The vowel /i/ occurs after /f v t t^h n ts ts^h s k k^h ŋ x/ onsets and /w/ glide (e.g., [fi˨˩˦] 'bandit', [tsi˨˩˦] 'walk', [ŋi˨˩˦] 'lotus root', [kwi˨˩˦] 'ghost'). It occurs only in open syllables. When it is preceded by /n/, the onset consonant is realised as the allophonic variant /l/; so the syllable /ni/ is produced as [li]. The vowel /i/ in JHC is slightly diphthongised, the tongue apex starts at a slightly lower position than for a typical /i/ and raises at the beginning of the vowel to reach the /i/ position, it can thus be transcribed as [ᶦi] or [ᶤi]. This diphthongisation exists on every /i/ vowel but is more noticeable when /i/ is preceded by alveolar stops, affricates and lateral consonants (i.e., /t t^h ts ts^h/ and [l]) (Zhao, 2003). It is worth noting that in Zhao's (2003) description, the vowel /i/ occurs also after /p m/ onsets. He proposed the following entries in the JHC dictionary: 碑 [pi˥˩] 'stele', 美 [mi˨˩˨] 'beautiful'. These two syllables do not exist in the 'city accent' according to consulted JHC speakers. In the acoustic study presented in Chapter 4, the recorded speakers produced the Chinese characters 碑 as [pa˥˩] and 美 as [mẽ˨˩˨] respectively. This observation is also mentioned in section 3.5.2.

/y/ has a limited distribution, it occurs only after /ts ts^h s n/ consonants, and these onsets are realised as alveolo-palatals [tɕ $tɕ^h$ ɕ ɲ] (e.g., /tsy˨˩˨/ [tɕy˨˩˨] 'mouth'). It occurs only in open syllables and can occur in a syllable without onset (e.g., [y˨˩˨] 'rain'). Despite its limited distribution, it is a distinct phoneme. It is contrastive to other vowels in syllables without onsets (e.g., [y˨˩˨] 'rain' vs. [ʉ˨˩˨] 'loiter' vs. [ɔ˨˩˨] 'I' vs. [z̩˨˩˨] 'already').

/ʉ/ is a rounded central high vowel, which occurs after /p p^h m t t^h n ts ts^h s k k^h ŋ x/ onsets and in syllables without onset (e.g., [pʉ˥˩] 'wave', [tsʉ˨˩˨] 'left', [xʉ˨˩˨] 'fire'); it occurs only in open syllables (e.g., [ʉ˨˩˨] 'loiter'). /ʉ/ is transcribed as /ɵ/ in earlier descriptions (Luo, 1936; Zhao, 1989, 2003), but the tongue body position and the F1 value of this vowel all indicate that /ʉ/ is a more appropriate transcription. Its F1/F2 pattern is presented in section 4.2.3.

/u/ is a round back vowel, occurring after /p p^h f v t t^h n ts ts^h s k k^h/ onsets and only in open syllables (e.g., [pu˨˩˨] 'repair', [tsu˨˩˨] 'group', [ku˨˩˨] 'valley'). The syllable /nu/ is realised as [lu]. As discussed in section 3.2.5, when /u/ occurs after /f v/ onsets, it is realised as its allophone, a velarised [v̩].

/o ɤ ɔ/ vowels occur in both open and closed syllables. They all occur after /p p^h m t t^h n ts ts^h s k k^h ŋ x/ onsets in open syllables (e.g., [po˨˩˨] 'handle', [pɤ˨˩˨] 'full', [pɔ˨˩˨] 'board'). Additionally, /ɔ/ occurs after /f v/ in open syllables, while /o/ can be found in onsetless syllables (e.g., [fɔ˨˩˨] 'turn over', [ɔ˨˩˨] 'I, me'). In closed syllables, they all occur after /p p^h m t t^h n ts ts^h s k k^h ŋ x/ onsets and are followed by /ʔ/ (e.g., [poʔ˦] 'to shell' [pɤʔ˦] 'no', [pɔʔ˦] 'eight'). Additionally, in closed syllables, /ɤ ɔ/ occur after /f v/ onsets and /ɔ/ occurs in closed syllables without onset. As already stated, the syllables /no/, /nɤ/ and /nɔ/ are realised as [lo], [lɤ] and [lɔ].

/e/ is a front close-mid vowel, which has a limited distribution. It occurs in both open and closed syllables but only after glides [j ɥ] (e.g., [je˩˧] 'have' vs. [ja˩˧] 'also' vs. [jo˩˧] 'elegant' vs. [jɔ˩˧] 'wild' vs. [jʉ˩˧] 'friend'). In earlier descriptions, /e/ was not analysed as a monophthong vowel. Instead, it was analysed as part of diphthong vowels /ie ye/ (Zhao, 1989, 2003) since the glides [j ɥ] were analysed as vowels. The present analysis takes a different stance and does not treat /ie ye/ as diphthongs. See also section 3.2.4.

The /ɑ/ vowel occurs after /p pʰ m f v t tʰ n ts tsʰ s k kʰ ŋ x j ɥ/ in open syllables and after /j y/ in closed syllables.

3.3.2. Nasal vowels

JHC has four nasal vowels /ɑ̃ õ ẽ ĩ/. All nasal vowels occur only in open syllables; they cannot be followed by the coda /ʔ/.

Table 3.7. Nasal vowels in JHC and examples with orthographic forms and glosses.

	Phonetic	Orthographic	Gloss
ɑ̃	pɑ̃˩˧	本	'root'
õ	põ˩˧	绑	'tie'
ẽ	pẽ˩˧	扁	'flat'
ĩ	tɕĩ˩˧	检	'inspect'

The vowel /ɑ̃/ has a tongue position similar to that of the back vowel /ɑ/, the lip is not rounded. It can occur after /p pʰ m f v t tʰ n ts tsʰ s k kʰ ŋ x/ onsets, and it can also occur after [j w] glides (e.g., [pɑ̃˩˧] 'root', [ŋɑ̃˩˧] 'cover', [jɑ̃˩˧] 'shadow').

The vowel /õ/ has the tongue position and the lip protrusion of /o/. It occurs after /p pʰ m t tʰ n ts tsʰ s k kʰ x/ onsets and in syllables without onset. It can also occur after [j] glide (e.g., [põ˩˧] 'tie', [õ˩˧] 'shake').

The vowel /ẽ/ is more open than the oral vowel /e/, it could be transcribed as [ɛ̃]. It occurs after /p pʰ m f v t tʰ n ts tsʰ s k kʰ ŋ x/ onsets and in syllables without onsets. It can also be found after the /w/ glide (e.g., [pẽ˩˧] 'flat', [ŋẽ˩˧] 'ant', [e˩˧] 'ear').

The vowel /ĩ/ is produced with the tongue position of /i/. It occurs after /ts tsʰ s n/ onsets, which are realised as [tɕ tɕʰ ɕ ɲ] (e.g., /tsĩ˩˧/ [tɕĩ˩˧] 'check'). It occurs also after [ɥ] glide and in syllables without onset. Although /ĩ/ has a limited distribution, it is a distinct phoneme contrastive to other vowels, especially to /ẽ/ (e.g., [ĩ˩˧] 'perform' vs. [ẽ˩˧] 'ear' vs. [õ˩˧] 'swing'). Luo (1936) and Zhao (1989, 2003) state that

JHC has nasal diphthong and triphthong vowels /ẽi iẽi/, but they are not attested in this study. The words proposed to have /ẽi/ rhyme are produced with /ẽ/ rhyme (e.g., [pẽi˨˩˧] is produced as [pẽ˨˩˧]), while the words proposed to have /iẽi/ rhyme are produced with /ĩ/ rhyme (e.g., [iẽi˨˩˧] is produced as [ĩ˨˩˧]).

3.4. Tones in citation form

JHC has six lexical tones, traditionally noted in tone letters (Chao, 1930) as follows: tone 1 as 31, tone 2 as 44, tone 3 as 213, tone 4 as 35, tone 5 as 22, tone 6 as 32 (Zhao, 2003). The previous impressionistic description does correspond to the pitch contour described by Li (2007) who observed speaker specific pitch variation in JHC tones and described the tones as follows: tone 1 as high-falling, tone 2 as high-level, tone 3 as low-dipping, tone 4 as mid-rising, tone 5 as low-falling, tone 6 as mid-falling and short. Li's (2007) description is adopted here.

Table 3.8. Tones in JHC with orthographic forms and glosses.

	Phonetic	Orthographic	Gloss
Tone 1	ŋɔ˧˩	安	'safe'
Tone 2	ŋɔ˦	崖	'cliff'
Tone 3	ŋɔ˨˩˧	矮	'short'
Tone 4	ŋɔ˧˥	晏	'late'
Tone 5	ŋɔ˨	岸	'shore'
Tone 6	ŋɔʔ˧˨	鸭	'duck'

Tone 1 is high-falling; the pitch contour falls from the high range to the low end of the speaker's pitch range. In traditional tone letters, the tone 1 is described to be a mid-falling tone (i.e., 31), but Li (2007) shows that the starting point of tone 1 is higher than the average pitch of tone 2, which is a high-level tone. It is thus necessary to describe tone 1 as high-falling rather than mid-falling.

Tone 2 is high-level. The pitch contour is flat and maintained at a mid-high range. This tone can be produced with breathy voice.

Tone 3 is a low-dipping tone in citation form. The turning point is often marked by glottalisation. The falling and the rising part of tone 3 have a similar pitch range (Li, 2007).

Tone 4 is mid-rising; the pitch contour raises from the middle and ends at a very high range compared to other tones. The end of the tone could be realised with falsetto voice.

Tone 5 is low-falling; it starts at a low range and ends at an even lower range. The end of this tone could be realised with glottalisation.

Tone 6 is the ‘checked tone’ as it ends with a glottal plosive. The glottal plosive /ʔ/, which ends the voicing rapidly, makes the duration of this tone considerably shorter than the other tones. This glottal plosive is the only coda permitted in JHC phonotactics and it can only be preceded by /a e o ɤ ɔ/ nuclei. Zhao (2003) notices that the /ʔ/ sound is more audible when tone 6 is on the first syllable of a dissyllabic word.

3.5. Apical vowel [z] in JHC

JHC has four syllabic consonants: /n̩ m̩ z̩/ and [v̩]. In this section I present the ‘apical vowel’ /z̩/. Compared to other syllabic consonants, the apical vowel /z̩/ is not only a distinct phoneme but is lexically more productive. Syllables containing the apical vowel account for 7.2% of the monosyllabic entries of the JHC dictionary (Zhao, 2003)

3.5.1. Lexical distribution

The apical vowel in JHC can be preceded by /p pʰ m n s ts tsʰ/ onsets. /z̩/ can also stand for a syllable on its own, without a phonological onset. Examples are shown in Table 3.9.

Table 3.9. A list of JHC items having apical vowel /z̩/ as nucleus, with orthographic forms and glosses.

Phonetic	Orthographic	Gloss
pz̩˧˩˨	比	‘compare’
pʰz̩˧˩˨	被	‘quilt’
mz̩˧˩˨	米	‘rice’
nz̩˧˩˨	里	‘in’
tsz̩˧˩˨	紫	‘purple’
tsʰz̩˧˩˨	此	‘here’
sz̩˧˩˨	死	‘die’
z̩˧˩˨	椅	‘chair’

The JHC apical vowel occurs also on different tones (e.g., [pʰz̩˥˩] ‘criticise’, [pʰz̩˦] ‘skin’, [pʰz̩˧˩˨] ‘quilt’, [pʰz̩˦˥] ‘nose’, [pʰz̩˦˨] ‘prepare’). In a /pʰz̩/ syllable for example, the only voiced segment is /z̩/. Therefore, it is natural to assume that /z̩/ is the rhyme of the syllable and the tone-bearing unit.

/z̩/ does not occur in closed syllables. Hence, it cannot be followed by the coda /ʔ/. This limited distribution is however not specific to the apical vowel. Among all possible syllable nuclei in JHC (/i y ʉ u e o ɤ ɔ a ɑ̃ õ ĩ ẽ n̩ m̩/ [v̩]), only /e o ɤ ɔ a/ can be followed by the /ʔ/ coda. Recall that the other syllabic fricative [v̩] presented in section 3.2.5 has the same tonal distribution as the apical vowel /z̩/.

3.5.2. Phonemic contrast

The JHC apical vowel /z̩/ is a distinct phoneme. It is contrastive to vowels in all possible distributions and in particular, it is contrastive to vowel /i/. A few minimal pairs and triplets are given in Table 3.10 and Table 3.11.

Table 3.10. Minimal pairs between /i/ and /z̩/ in JHC, with orthographic forms and glosses.

Phonetic	Ort.	Gloss		Phonetic	Ort.	Gloss
sz̩˨˩˦	死	'die'	vs.	si˨˩˦	洗	'wash'
tsz̩˨˩˦	紫	'purple'	vs.	tsi˨˩˦	走	'walk'
tsʰz̩˨˩˦	此	'here'	vs.	tsʰi˨˩˦	丑	'ugly'
pz̩˥˩	屄	'female genitals'	vs.	pi˥˩	碑	'stela'
mz̩˨˩˦	米	'rice'	vs.	mi˨˩˦	美	'beautiful'

Table 3.11. Minimal triplets of /z̩ ɑ ʉ/ in JHC with orthographic forms and glosses.

Pho.	Ort.	Gloss		Pho.	Ort.	Gloss		Pho.	Ort.	Gloss
pz̩˦˥	闭	'close'	vs.	pɑ˦˥	背	'back'	vs.	pʉ˦˥	簸	'winnow'
pʰz̩˦˥	屁	'fart'	vs.	pʰɑ˦˥	配	'match'	vs.	pʰʉ˦˥	破	'broken'
mz̩˨˩˦	米	'rice'	vs.	mɑ˨˩˦	每	'every'	vs.	mʉ˨˩˦	某	'someone'
nz̩˧	泥	'dirt'	vs.	nɑ˧	来	'come'	vs.	nʉ˧	罗	'sift'

The pairs /pz̩/ versus /pi/, /mz̩/ versus /mi/ were reported by Zhao (1989, 2003). Whilst [pz̩˥˩] and [mz̩˨˩˦] are accepted without hesitation, the speakers recorded in this study did not pronounce [pi˥˩ mi˨˩˦] but [pɑ˥˩ mẽ˨˩˦], respectively. These syllables are said to not exist in the 'city (i.e., Huayang county) accent'. It seems that the /i/ vowel cannot take labial consonants in the city variant of JHC. But this variation does not affect the fact that the vowel /i/ and the apical vowel /z̩/ contrast after coronal sibilants.

3.5.3. Phonological behaviour

The apical vowel in JHC behaves like any other vowel, it is the nucleus of the syllable thus the rhyme of the syllable. The following passages of rhymed folk songs are retranscribed based on Luo (1936) using symbols proposed in this study. The tones are omitted. The (8)-a folk song has two sections. The paragraph presented here contains the first six verses of the second section. It has three syllables per verse. The syllables with apical vowel nucleus have the status of a stressed syllable, identical to [pʰɔ], [tɕjo] and [ɲjeʔ] syllables which have normal vowel nuclei.

The (8)-b folk song has two sections. The paragraph presented here is the first three verses of the first section. In this example, the rhyme is /z̩/. Luo's (1936) collection of 38 folk songs has no examples of apical vowel nucleus rhyming with any other nucleus.

(8) a. tsʰz̩ tɕjoʔŋi ‘(She is) riding a cow’
pʰz̩ tɕjoʔ tsʰi ‘covered by a piece of silk’
pʰɔʔ tsz̩ ɕɥĩ ‘with a white paper fan’
tɕjo ɲjeʔ tʰi ‘(that) shields the sunlight’

b. tɕjɑ̃ sʉ sz̩ ‘Golden key(s)’
ɲjɑ̃ sʉ sz̩ ‘silver key(s)’
ko xɑ̃ kuɔ ko tsʉ ɕje z̩ ‘marry to an official to be his concubine’

The above examples confirm that apical vowel /z̩/ in JHC is the nucleus of a syllable, and a tone-bearing unit (TBU). As a TBU it undergoes tone sandhi processes, as shown in Table 3.12.

Table 3.12. Examples of tone sandhi on apical vowels in JHC, with underlying forms, surface forms and glosses.

/sz̩˧˩/	‘western’	+	/tsõ˧˩/	‘clothes’	→	[sz̩˧ tsõ˧˩]	西装	‘suit’
/tsʰz̩˦˥/	‘gas’	+	/tɕʰɔ˧˩/	‘vehicle’	→	[tsʰz̩˦ tɕʰɔ˧˩]	汽车	‘car’
/pʰz̩˧/	‘prepare’	+	/tʰɑ˧˩/	‘tyre’	→	[pʰz̩˥˧ tʰɑ˧˩]	备胎	‘spare tyre’

In a nutshell, the phonological behaviour and function of the apical vowel in JHC is equivalent to any other vowel in the vocalic inventory of the language.

Chapter 4
Acoustic study of the apical vowel [z] in JHC

The aim of this chapter is to understand the basic acoustic characteristics of the apical vowel in JHC: its duration compared to other syllable nuclei, its formant frequencies compared to other vowels and to other apical vowel variants, and most importantly the nature of the turbulence noise it displays compared to sibilant consonants.

4.1. Methodology

4.1.1. Speakers

Speakers of JHC were recruited according to strict criteria to limit possible dialectal variations: they must be born and raised in the town of Huayang, with both parents also born in the same town; they must live in the town of Huayang and speak JHC in a daily basis both in their professional and non-professional contexts; their age should be around 50.

Nowadays it is highly challenging to find anyone who lives in a city in mainland China and does not speak (more or less fluently) SC. Since JHC speakers all understand and speak SC, only those who speak JHC in both professional and non-professional contexts were selected to limit any potential influence. Five female speakers (labelled FS1 to FS5) and five male speakers (MS1 to MS5) corresponding to these criteria were chosen to participate in the recording sessions. The mean age of the speakers was 49 (±3.8). None of them reported to have speech-related anomalies. They all know each other and consider themselves native speakers of JHC with no accent.

4.1.2. Acoustic data acquisition and segmentation criteria

Acoustic data were recorded with a hypercardioid headset microphone (AKG C520), an external sound card (Edirol UA25), and Audacity (v 2.1.0) on a portable computer. I had access to the sound attenuated studio of the local television channel for the recording sessions. The speakers were sat in a chair and told to read a word list embodied in a frame sentence at a normal speech rate. The word list for acoustic data acquisition (see A.1) contained monosyllables with [p pʰ m n s ts tsʰ] onsets and [z i u a ʉ] nuclei[16]. With different tones, they form real words, presented to speakers in Chinese characters. The speakers were told to read the words in a frame sentence [ki˧ɕɔ˨˩˦ __ ɕɔ˨˩˦sɔ˧fa˧] 'He/She writes __ three times'. The whole list was repeated five times for each speaker, yielding 2150 target items ([z]: 550, [i]: 400, [u]: 450, [a]: 350, [ʉ]: 400). The recorded data were segmented and annotated using Praat (Boersma & Weenink 2018).

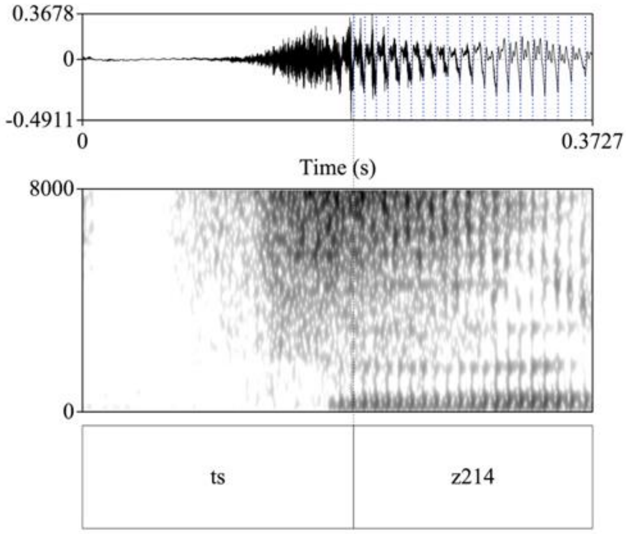


Figure 4.1. Illustration of the segmentation between a sibilant onset and an apical vowel with syllable [tsz˨˩˦] pronounced by FS3. The blue dotted lines represent pulses detected by Praat.

Determining the exact boundary between a sibilant onset and the following apical vowel was not straightforward, as there was no clearly observable frontier in-between. The data were labelled by taking the first pulse detected by Praat as the beginning of the voiced apical vowel. This is illustrated in Figure 4.1. The other segments were annotated following the criteria presented in Table 4.1.

---

[16] In this chapter, I present the phonetic realisations of the concerned segments and syllables, and some of them are not phonemic (e.g., the onsets [n] and [l] are allophones, see section 3.2.3). Hence all transcriptions are enclosed in square brackets, and the syllabic diacritic '̩' for apical vowel [z] is omitted for simplicity

Table 4.1. Segmentation criteria of the target syllables in the acoustic study.

	Start point	End point
[p t]	End of F2 of the preceding vowel on spectrogram	End of burst on acoustic signal and spectrogram
[m n l]	End of F2 of the preceding vowel, starting of nasal zero on spectrogram	End of release on acoustic signal and spectrogram
[ts tsʰ]	End of F2 of the preceding vowel on spectrogram	Onset of periodic voicing
[s]	End of F2 of the preceding vowel on spectrogram, start of non-periodic signal	
[z]	End of the preceding onset consonant or first detected pulse	Offset of periodic voicing
[i ʉ u a]	End of the preceding onset consonant	

4.1.3. Acoustic parameters and statistical processing

The following acoustic parameters were examined:

(i) Presence or absence of frication noise (based on visual examination of acoustic signals and spectrograms)
(ii) Duration of all syllables and syllable nuclei
(iii) The formant frequencies of [a i u ʉ] and [z] (F1, F2 and F3)
(iv) The harmonic-to-noise ratio (HNR) of [i] and [z]
(v) The zero-crossing rate (ZCR) of [a i u ʉ] and [z]
(vi) The centre of gravity of (COG) of [s] and the aspiration phase of [pʰ]

The presence or absence of frication noise was based on visual examination. I categorised the speakers' productions of [z] into three classes: [z] having frication noise on more than half of its duration, [z] having frication noise on less than half of its duration, and [z] having quasi-no visible frication noise. The first two classes are reported in section 4.2.1.1, and section 4.2.1.2 reposts thoses having quasi-no frication noise.

To obtain the formant frequency values, a script trisected all nuclei segments and calculated mean formant values of the middle portion of the target segments. The maximum frequency of formant calculation was set to 5500 Hz for female speakers and 5000 Hz for male speakers. Harmonic-to-noise ratio, zero-crossing rate, and centre of gravity were used to examine the frication noise on the apical vowel [z]. These parameters are widely used when it comes to turbulent sounds (see Fuchs, Toda & Żygis, 2010 for a review).

HNR is defined as the ratio of the harmonic components to the noise. This was obtained for [z] and [i][17] using Praat HNR algorithm (Boersma & Weenink 2018). Basically, the greater the noise component of the signal, the lower the ratio is. This measurement is often used when both frication and voicing need to be considered. HNR was calculated with a 40 ms window that slided every 10 ms on the entire duration of each recording. If t0 was the start point of each recording, then the first window covered the inverval from t0 to t0+40 ms. The second window covered the interval from t0+10 ms to t0+50 ms, and so forth. The mean HNR was obtained for each window. These mean HNR values for each window were then saved as a binary file corresponding to each recording, yielding a set of '.sound' files that contains the changing mean HNR values of each recording (see Figure 4.2). Data points were then obtained on the HNR results at 52 time points for each target segment, including vocalic onsets, offsets and 50 points in-between.

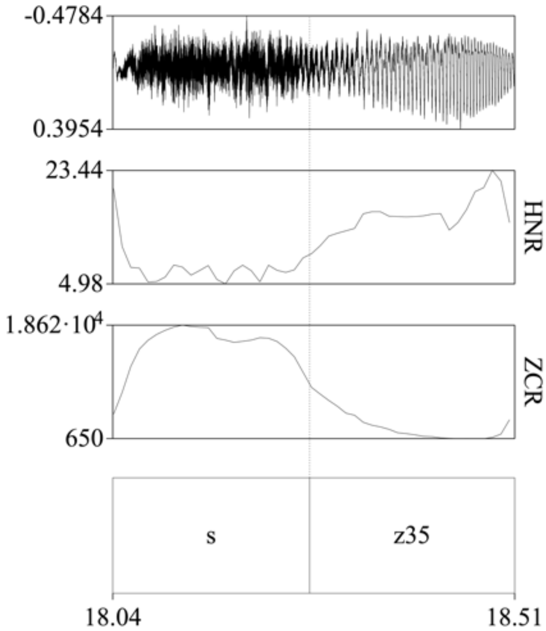


Figure 4.2. An example showing the acoustic signal of the form [sz˥]. as well as the HNR and ZCR values. This item was produced by FS1.

The ZCR, defined as the number of times in a given time-interval the speech signal passed through the value of zero, was measured for all syllable nuclei. Given the strong correlation between ZCR and energy distribution with frequency, high frequencies imply high ZCRs, and low frequencies imply low ZCRs (Rabiner & Schafer, 1978: 128). The signal of a vowel is periodic and the number of zero-crossings in a given time-interval is consequently low. When the signal is aperiodic, as for a fricative, a large quantity of zero-crossings in a given time-interval is

[17] I chose to compare [z] to [i] in order to juxtapose my results with the results obtained by Faytak (2015) on SC.

observed due to the high frequency energy. This measurement is considered to be a simple and reliable measurement of the intensity of frication noise (Shosted, 2006). It has been applied to fricative sounds (Ito & Donaldson, 1971), nasalised fricative sounds (Bombien, 2006), or aspirated vowels (Gordeeva & Scobbie, 2010). In this study ZCR per second was calculated with Praat using the same method as for the HNR (see Figure 4.2). Data points were obtained on the ZCR results at 52 time points, including nuclei onsets, offsets and 50 points in-between. These data points represent the general tendency of ZCR during the target segments.

An interesting point was observed during the segmentation of the acoustic data: The aspiration phase of [pʰ] in [pʰz] displays very different characteristics compared to the aspiration phase of the same segment when preceding other nuclei segments. The aspiration phase of [pʰ] in [pʰz] presents energy concentration at higher frequencies, much similar to a fricative [s]. In order to dig more into this I compared the general energy distribution of the aspiration phase of [pʰ] to frication noise of [s], using the COG measurement. The COG reflects on average how high the frequencies are in a spectrum. This method has been used in studying the SC fricatives and affricates, for example (Svantesson, 1986 among other studies).

Basic descriptive statistics were applied to durational values, formant values and COG values, including ANOVA, TukeyHSD, Student-Newman-Keuls, Welch's *t*-test, etc. For HNR and ZCR parameters, the values obtained, as already stated, were obtained from 52 points on each target segment. For each segment, these 52 points were evenly spread from onset to offset. When presenting the 52 data points on the x-axis and the values corresponding to the data points on the y-axis, these values form a variable during a normalised time range. The y-axis values (i.e., HNR and ZCR) were then smoothed into a moving average with LOESS (locally estimated scatterplot smoothing) method in R with the package Tydiverse (Wickham et al., 2019). It is important to note that the HNR data were particularly difficult to interpret for the JHC apical vowel, since this segment contained abundant frication noise, and HNR values were thus not always correctly detected by Praat. In processing the HNR values, the data were filtered (see section 4.2.4 for more details). The ZCR values were considered to be more reliable and thus not filtered.

4.2. Results

4.2.1. Presence of frication on the apical vowel [z]

4.2.1.1. [z] with abundant frication noise

Abundant frication noise is observed for all speakers, and in all contexts, counting for 87% of the total data. Some illustrative examples are shown in Figure 4.3, with apical vowels displaying frication noise following [p], [m], and [n]. These illustrative examples are chosen on purpose to make clear that turbulence can be observed on [z] segments even following non-sibilants.

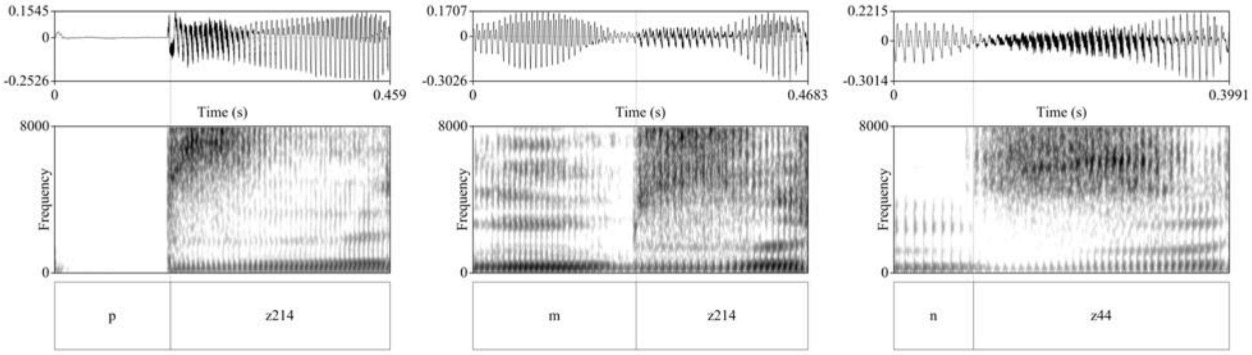


Figure 4.3. Apical vowels containing frication noise in JHC. [pz˨˩˦], [nz˦˦], [mz˨˩˦] syllables come from FS1, MS3 and MS2 respectively.

Frication noise is present at high frequency regions, and is systematically superposed on voicing. This turbulence is not steady. Instead it dynamically evolves during the time course of [z]. Intense frication is observed at the beginning of the segment and at the first half of its duration ([pz˨˩˦] in Figure 4.3), and can sometimes extend to the second half of the apical vowel ([mz˨˩˦] and [nz˦˦]). Frication noise however never extends until the end of the apical vowels. When it diminishes, periodic waveforms become clearer, and a clear formant structure is visible on the spectrograms.

4.2.1.2. [z] with less to non-visible frication noise

13% of [z] productions were produced with much less or quasi-no frication noise. Figure 4.4 presents illustrative examples in which formant structure is clear from the beginning of the apical vowel. These realisations correspond to what has been observed in SC apical vowel [z] (Faytak & Lin, 2015; Lee-Kim, 2014). They show a nearly clear separation between the onset consonants and the following [z], with much less frication noise compared to those presented in Figure 4.3.

These less fricated variants and the fricated ones can co-occur in one speaker's productions. For example, MS2, shown both in Figure 4.2 and Figure 4.3, can produce both variants in the same context. The individual differences concerning the amount of frication noise produced are discussed in more detail in section 4.2.5.3.

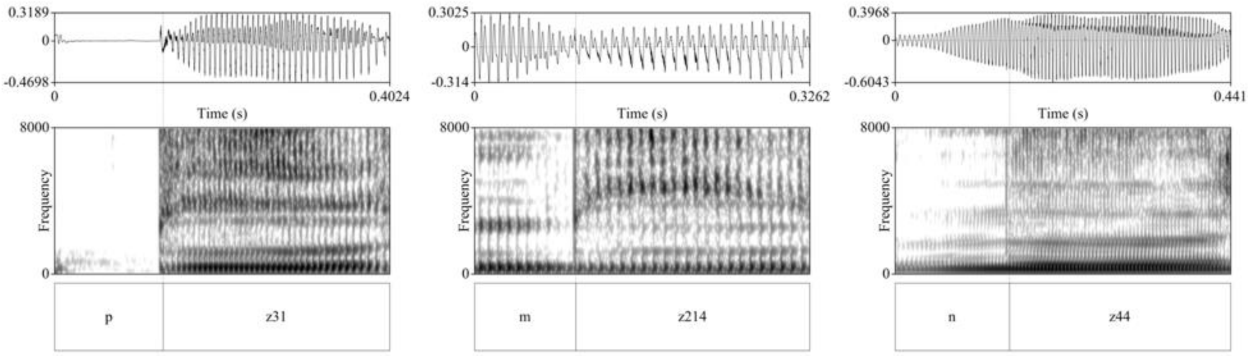


Figure 4.4. Apical vowels containing less to quasi-no frication noise in JHC. [pz˧˩], [mz˨˩˦], [nz˦] come from MS2, FS1 and MS4 respectively.

4.2.2. Duration of syllables and syllable nuclei

In this section, the durations of all recorded syllables and syllable nuclei are reported. I first show whether syllables vary in duration according to tonal contexts, as different tones may imply different durations (Howie, 1976; Ho, 1976), and then show how duration of [z] and other nuclei segments varies depending on the preceding consonant. One of the aims of this section is to justify the time-normalising method used in sections 4.2.4 and 4.2.5.

4.2.2.1. Duration of all syllables in different tonal contexts

The duration of syllables according to the tones they carry is presented in Table 4.2. Results show that the durations of these syllables vary depending on tone: tone 3 (low-dipping) has the longest duration and tone 1 (high-falling) the shortest. Mean values, minimal and maximum values, one-way ANOVA and Student–Newman–Keuls (SNK) post-hoc tests results are reported in Table 4.2.

The mean durations of the 4 tones displayed in Table 4.2 show that the difference between the longest tone (3, low-dipping) and the shortest tone (1, high-falling) is limited to 32 ms. Much important differences occur within one tonal group. The durational differences caused by tone will not be taken into account in the comparison of the general tendency of HNR and ZCR evolution during syllable nuclei.

Table 4.2. Mean durations of all recorded syllables in different tonal contexts, with minimum and maximum durations, one-way ANOVA and SNK post-hoc tests results. SNK result is presented as the hierarchy of mean duration according to tones.

Tone	Mean (ms)	Min and max (ms)	ANOVA and SNK
1 [˥˩]	378	239–595	$F(3, 2036)=16.42$, $p<0.001$ Tone 3 > Tone 2 = Tone 4 > Tone 1
2 [˦]	397	262–635	
3 [˨˩˧]	410	237–791	
4 [˧˥]	396	287–603	

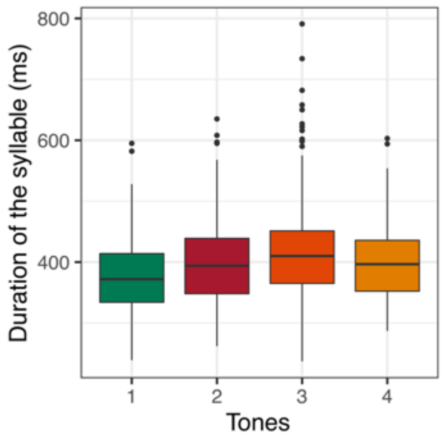


Figure 4.5. Duration of all recorded JHC syllables in different tonal contexts. Numbers on x-axis correspond to tones: tone 1 is the high-falling tone [˥˩], tone 2 is the high-level tone [˦], tone 3 is the low-dipping tone [˨˩˧], tone 4 is the mid-rising tone [˧˥]. The y-axis represents duration in milliseconds.

4.2.2.2. Duration of [i u a ʉ] and [z] in different consonantal contexts

Before presenting the results on duration of syllable nuclei it is important to recall that the onset /n/ has two allophones [l] and [n]. The syllables /na ni nʉ nu/ surface as [la li lʉ lu], and /nz/ is realised [nz]. Durations of the vowels [a i ʉ u] after /n/ are not included in the results.

Table 4.3 presents the duration data for the 5 syllable nuclei [i u ʉ z] in six different consonantal contexts. As can be seen, syllable nuclei in JHC are particularly long, displaying mean durations varying from 176 ms to 252 ms. Setting the apical vowel [z] aside, there are two notable remarks concerning the durations of [a] and [i]: [a] has a significantly longer duration after [m] compared to other consonantal contexts, and [i] has a significantly shorter duration after [tsʰ] compared to [s ts]. I fail to find a plausible explanation to these two differences, as it seems that all nuclei tend to be longer after [m] and shorter after [tsʰ]. This general tendency can be more clearly observed in Figure 4.6.

The most interesting result concerns the apical vowel [z] which displays significantly different durations according to onset consonants: a longer duration after [n m p], and a shorter duration after [pʰ s ts tsʰ]. This consistent pattern, shown in Figure 4.7 and confirmed by the SNK post-hoc test presented in Table 4.3, is puzzling as none of the two groups of segments forms a natural class.

Table 4.3. Mean durations (ms) of [a i ʉ u z] after [m p pʰ s ts tsʰ], with one-way ANOVA and SNK post-hoc tests results. SNK result is presented as the hierarchy of mean duration according to onsets.

	m	p	pʰ	s	ts	tsʰ	ANOVA	SNK
a	252	220	226	224	242	221	$F(5, 266)=3.48$, $p<0.001$	m≥ts=pʰ=s=tsʰ=p
i	–	–	–	227	225	199	$F(2, 190)=5.9$, $p<0.01$	s=ts>tsʰ
ʉ	252	238	236	238	241	–	$F(4, 334)=0.908$, $p=0.5$	m=p=pʰ=s=ts
u	–	233	215	220	226	216	$F(4, 272)=0.928$, $p=0.45$	p=pʰ=s=ts=tsʰ
z	241	234	191	192	180	176	$F(5, 415)=17.73$, $p<0.001$	m=p>s=pʰ=ts=tsʰ

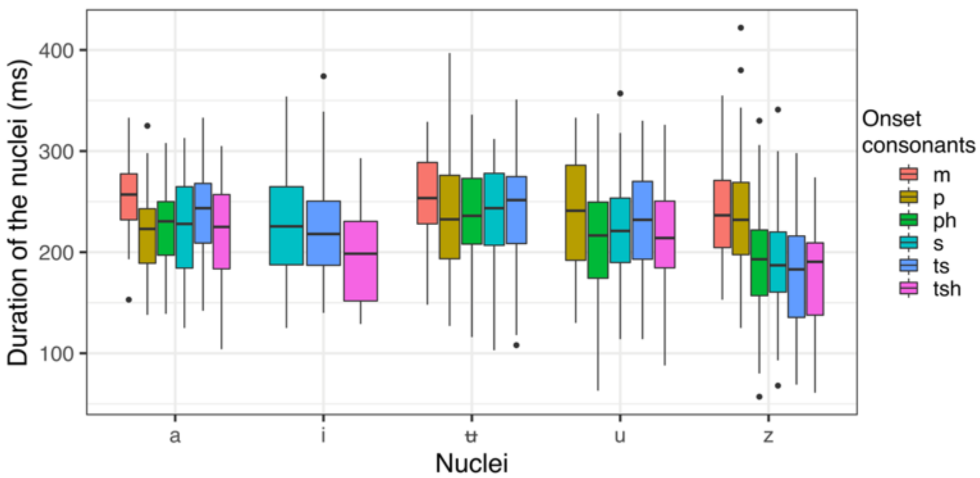


Figure 4.6. Duration of [a i ʉ u z] following [m p pʰ s ts tsʰ]. Data are grouped according to nuclei; in each nuclei group, data are separated by onset consonant types.

The shorter duration of [z] when it is preceded by [s ts tsʰ] is most probably due to gestural overlapping between the sibilant onsets and the apical vowel, making the exact boundary between these segments hard to identify. An example of the transition portion between [s] and [z] in a [sz] syllable is given in Figure 4.8. The acoustic signal and the spectrogram are visibly in a continuity before and after the first detected pulse. There is some visible voicing before the first pulse, but not clear enough to be detected as periodic. In this case, as I have already reported, decision was made to take the first pulse as the indicator of voicing, and thus the start point of the rhyme. Given that [z] is homorganic to the coronal sibilant onsets, it could be the case that part of the frication noise at the offset of the sibilant onsets is actually the onset of the frication noise of the [z] nucleus (suggesting that [z] is probably longer than what the segmentation criteria used may suggest). The sibilant onset [s] demands high intraoral air pressure to generate frication noise; while rhyme [z] demands low intraoral air pressure for voicing to occur. One possibility is that at the offset of [s], the intraoral air pressure is still too high for voicing to be properly generated, making it impossible to detect the pulses during this period.

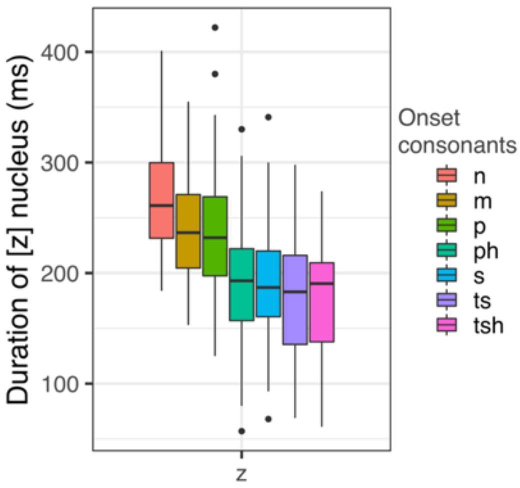


Figure 4.7. Duration of [z] nucleus in all consonantal contexts.

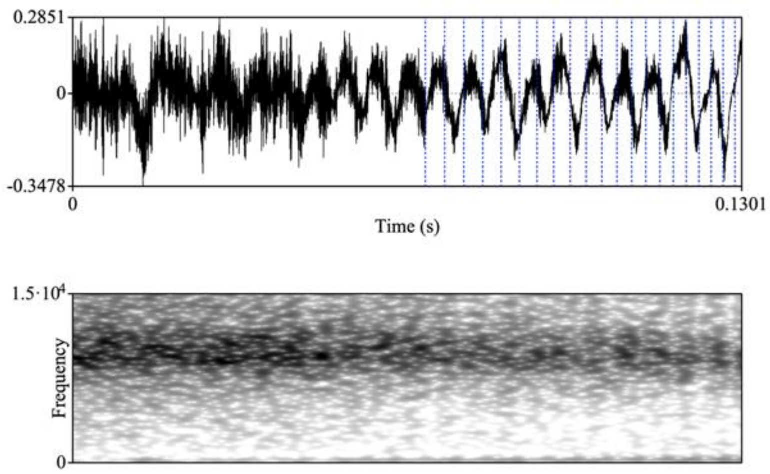


Figure 4.8. Transition portion between [s] and [z] nucleus from FS1. The dotted blue lines represent pulses detected by Praat with the default setting.

The same argument could be used to account for the shorter duration of [z] in [pʰz] compared to [pz mz nz]. The argument here is that the gesture for [z] is already achieved during the release phase of [pʰ], so that the glottal sourced aspiration is masked by the oral sourced frication noise. The result is that the release phase of [pʰ] is more like a [s] sound acoustically. The onset [pʰ] in this particular context could be transcribed as [pˢ]. This aspect will be touched upon in section 4.2.6 and examined in more detail in section 5.2.6 from an articulatory perspective.

As stated above, a time-normalising method is used in sections 4.2.4 and 4.2.5 to account for the variability that characterizes the durations of syllable nuclei. This is achieved by selecting 50 points during each segment, the interval between each two points is determined based on the duration of the segment (interval = the duration of the segment / 51). The end points of the nuclei are easy to define. The start points chosen, though they are not perfectly delimited for [z], do not result in increasing the presence of frication noise on this segment (the 'missing' portion in the segmentation is most probably the most fricative portion of [z]). The 50 points can thus reliably reflect the general pattern of these segments regardless of their absolute duration.

4.2.3. Formant structure of [i u a ʉ] and [z]

The mean formant values are calculated at the middle portion of the apical vowel [z]. This is the portion where the frication noise starts diminishing and a clear formant structure becomes visible. The values obtained are compared to the formant values of the other nuclei segments [i ʉ u a]. The results are also compared to what has been reported for apical vowels in other Chinese languages, as well as to formant values of fricative /z/ in none Chinese languages.

4.2.3.1. Overlap in formant space between [ʉ] and [z]

Figure 4.9 presents the formant values of [i u a ʉ z] for male and female speakers. It shows that the frequencies of [z] overlap those of [ʉ] and [u], especially for male speakers who have comparatively smaller vocalic space. The same observation can be made from Figure 4.10.

Several statistical tests were conducted on the formant values of [z u ʉ] segments. One-way ANOVA tests were conducted on F1, F2, F3 values respectively, then Student-Newman-Keuls and TukeyHSD post-hoc tests are used to bring out the inner pattern of the three segments. The results are reported in Table 4.4. They show that [u] and [z] present significantly similar F1 values for all speakers, and that both [u] and [z] have higher F1 values than [ʉ]. They also confirm that for male speakers, the F2 values of [ʉ] and [z] are not significantly different. The similarity in formant structure between [z] and [ʉ] is interesting in that it may result in perceptual similarity between the two categories. I will come back to this issue in section 6.2.

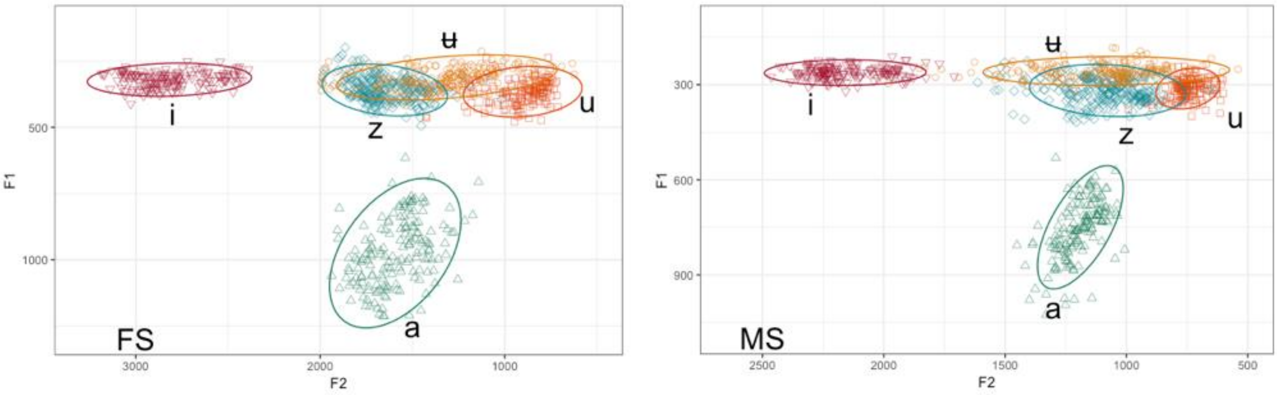


Figure 4.9. Scatter plot of formant values of [i u a ʉ z] of JHC. Data points represent mean values of the central part of each segment, with 95% confidence ellipses. Female speakers are on the left and male speakers on the right.

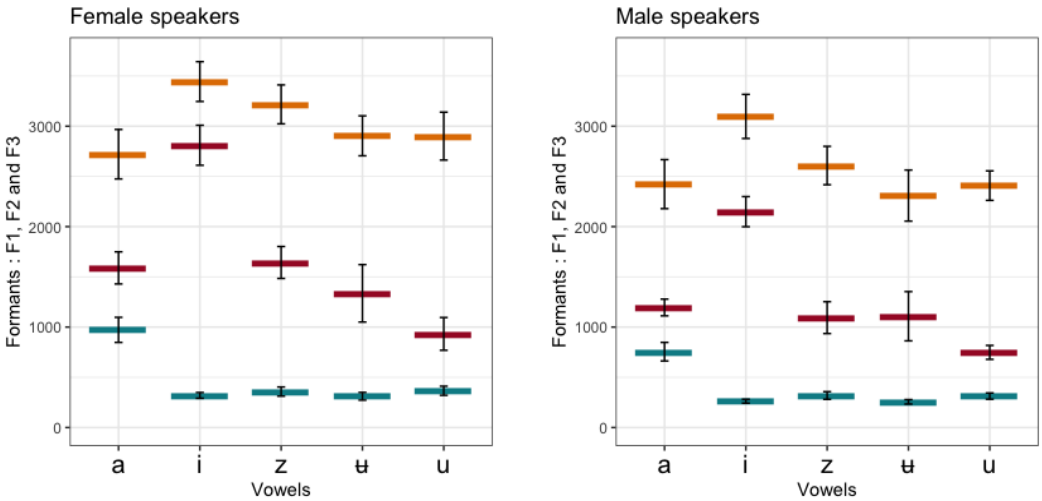


Figure 4.10. Structure of formant values of [i u a ʉ z] of JHC. The thick horizontal bars represent F1, F2, and F3 respectively for each segment; the error bars represent the standard deviation of each formant.

Table 4.4. Results of one-way ANOVA, SNK post-hoc tests and TukeyHSD post-hoc tests conducted on formant values of [z u ʉ] in JHC. FS stands for female speakers and MS for male speakers; SNK post-hoc tests are presented following the hierarchy of the mean formant values (Hz).

		one-way ANOVA	SNK	TukeyHSD Pair	*p*
FS	F1	*F*(2,630)=89.12, *p*<0.001	[u]366 = [z]358 >[ʉ]311	z-ʉ	<0.001
				z-u	=0.14
				u-ʉ	<0.001
	F2	*F*(2,630)=625.2, *p*<0.001	[z]1643 > [ʉ]1336 > [u]932	z-ʉ	<0.001
				z-u	<0.001
				u-ʉ	<0.001
	F3	*F*(2,630)=172, *p*<0.001	[z]3217 > [ʉ]2904 = [u]2901	z-ʉ	<0.001
				z-u	<0.001
				u-ʉ	=0.99
MS	F1	*F*(2,616)=239.4, *p*<0.001	[z]319 = [u]313 > [ʉ]256	z-ʉ	<0.001
				z-u	=0.19
				u-ʉ	<0.001
	F2	*F*(2,616)=244.5, *p*<0.001	[ʉ]1108 = [z]1094 > [u]748	z-ʉ	=0.69
				z-u	<0.001
				u-ʉ	<0.001
	F3	*F*(2,616)=128.7, *p*<0.001	[z]2608 > [u]2408 > [ʉ]2309	z-ʉ	<0.001
				z-u	<0.001
				u-ʉ	<0.001

4.2.3.2. Formant values of JHC [z] compared to other [z] sounds

The formant structure of different variants of apical vowels in Chinese languages has been examined in different acoustic studies (see also 2.2.2). The data of the present study are compared to data obtained from some of these studies, to verify whether apical vowels in these languages share the same formant structure. I also report results concerning the formant frequency of English and Polish /z/.

As shown in Table 4.5, the formant structures of all [z]'s are virtually identical across variants, although slight differences can still be observed. However, these differences do not exceed the acoustic space of the apical vowel shown in Figure 4.9 and Figure 4.10. The formant structure of all the variants can thus still be seen as consistent. There is also a similarity between the apical vowel [z] and the consonant [z] measured in English and Polish. Based on this qualitative similarity – at least as far as formant structure is concerned – it can be argued that there is a phonetic basis for identifying apical vowels in Chinese languages as one segment using the same symbol.

Table 4.5. Mean formant values (Hz) of different variants of apical vowel [z] in Chinese languages, compared to consonant [z] in English-US and Polish. Question marks indicate unknown speaker gender. The values of the present study are in bold. SC stands for Standard Chinese, HMC stands for Hefei-Mandarin Chinese, and SWC stands for Suzhou-Wu Chinese.

		F1	F2	F3	Language	
Apical vowel [z]	?	389	1232	2692	SC	Svantesson, 1984
	Male	320	1380	3140	SC	Howie, 1976
		300	1500	3140		
		340	1320	3280		
		397	1296	2923	SC	Zee & Lee, 2001
		378	1463	–	HMC	Hou, 2007
		344	1179	2741	SWC	Ling, 2009
		416	1190	2932	SC	Lee-Kim, 2014
		353	1558	–	HMC	Wan, 2014
		374	1373	–	SC	Shi et al., 2015
		383	1548	2990	HMC	Kong et al., 2019
		319	1094	2608	JHC	Present study
	Female	376	1680	3501	SC	Zee & Lee, 2001
		403	1723	–	HMC	Hou, 2007
		378	1405	3388	SWC	Ling, 2009
		515	1396	3272	SC	Lee-Kim, 2014
		443	1751	–	HMC	Wan, 2014
		475	1581	–	SC	Shi et al., 2015
		367	1501	2938	HMC	Kong et al., 2019
		357	1643	3217	JHC	Present study
Consonant [z]	?	–	1570	2720	En-US	Jassem, 1965[18]
		–	1770	2870	Polish	

4.2.4. Harmonic-to-noise ratio of [z] and [i]

The results of HNR measurements are presented in Figure 4.11. They show the evolution of the HNR during the duration of [z] and [i], with a clearly lower HNR for [z] reflecting its greater noise component. Specifically, while the HNR of [i] goes up directly after the onset, and maintains a high level throughout the total duration before the falling portion, the HNR of [z] goes up gradually until it reaches its highest level at the second half of its duration and then, after a short plateau, falls rapidly[19].

[18] F1 could not be measured accurately, hence not reported in Jassem's study.

[19] The falling portion of the two segments [i z] is probably the result of coarticulation with the following fricative [ɕ] contained in the frame sentence. The presence of this fricative also explains the rising phase of the ZCR curves as will be shown below.

The mean HNR values for the two segments, also shown in Figure 4.11, confirm this pattern for both males and females, with [i] displaying a higher HNR value than [z]. This is not the case for SC as reported in Faytak (2015), the [i] and [z] in SC seem to have similar HNR values. This difference indicates that JHC [z] is probably more fricated than SC [z].

While the HNR measurement gives expected results, it is not ideal for a thorough investigation of frication noise in JHC apical vowel (this will be done using ZCR measurement below). One reason is that the JHC apical vowel is far too fricated for the HNR values to be perfectly reliable. Since the HNR is based on the periodicity of the signal, and that the beginning phase of the apical vowel is much more fricated than periodic, the periodicity of the signal is not always correctly detected. This results in variable negative HNR values (many extreme values can be smaller than −200 dB). Taking these values into consideration would result in unreadable generalisations. In Figure 4.11, only positive HNR values are taken into account. Although the results still show that the beginning phase of [z] contains more frication than [i], they are generalised without the most fricated tokens of [z] since they would contain the most negative HNR values.

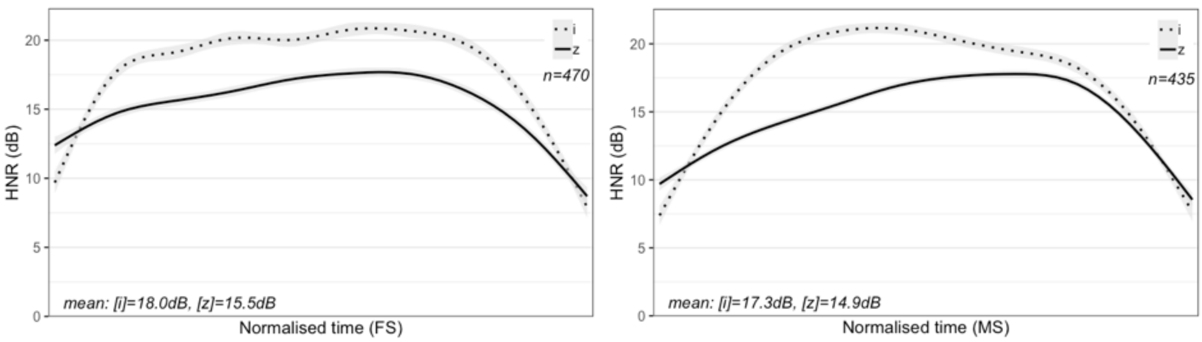


Figure 4.11. HNR of [i z] in JHC. The curves were generated with loess smoothing method, x-axis represents normalised time of the vocalic segments and y-axis represents the HNR values. Female speakers (FS) on the left and male speakers (MS) on the right.

The other reason is that it is difficult to compare all nucleic segments with HNR values. The vowel [u], for example, has few harmonics, which results in low HNR values. For many speakers, [u] is only slightly rounded with a very small labial opening. The articulators open only enough to produce an audible oral release, but the lips do not form a strong lip-rounding, resulting in the oral cavity being considerably reduced compared to a normally rounded [u][20].

[20] My feeling is that this vowel may be undergoing a sound change. The phoneme /u/ has two allophones [u] and [v̩] (see section 3.2.5)

4.2.5. Zero-crossing rate of [i u a ʉ] and [z]

In this section, the ZCR patterns of [z] are presented in detail. I first compare this segment to the other syllable nuclei [i ʉ u a], then I examine how these patterns vary depending on the nature of the preceding consonant and depending on speakers.

ZCR, as already stated, calculates the number of times in a given time-interval the speech signal passes through the value of zero, and can thus directly reflect the presence of frication noise on acoustic signals. Unlike HNR it measures the times of zero-crossings in a given time-interval, which does not involve the detection of voicing or pitch. The ZCR measurement can be used for all syllable nuclei regardless of their phonetic nature (e.g., lack of harmonics in [u] vowels), and it is virtually independent of speaker volume and less speaker-dependent than spectral analyses (Ito & Donaldson, 1971).

It is expected to observe a high ZCR when there is more frication noise (i.e., non-periodic signal) and a low ZCR when there is less frication noise (i.e., periodic signal). [z] is thus expected to have higher ZCR at the beginning of its duration and lower ZCR at the second half of its duration. Same as for the HNR measurement, the ZCR results are presented as a variable in time. This method of presentation gives a direct access to the dynamic evolution of ZCR during the nuclei segments.

4.2.5.1. General pattern of ZCR for all syllable nuclei

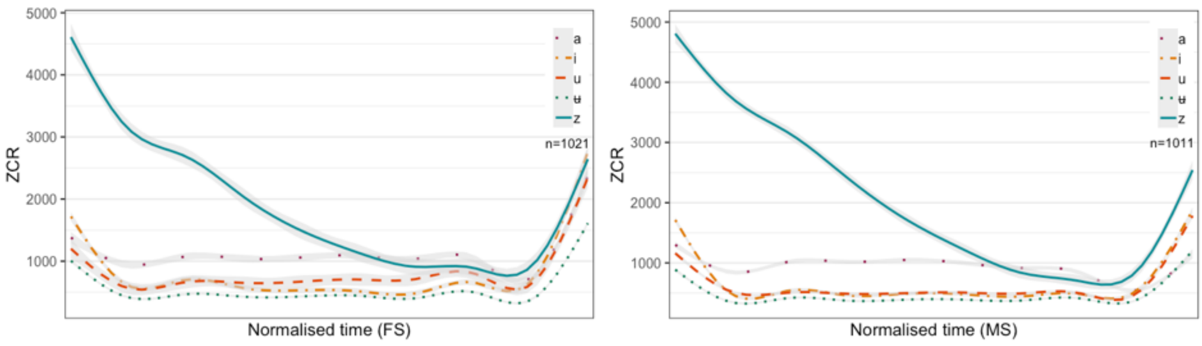


Figure 4.12. ZCR of [a i u ʉ z] in JHC. The curves were generated with loess smoothing method, x-axis represents normalised time of the vocalic segments and y-axis represents the zero-crossing times per second. Female speakers (FS) on the left and male speakers (MS) on the right.

The general pattern of ZCR for all syllable nuclei is reported in Figure 4.12. As this figure clearly shows, [z] behaves in a different way compared to other vowels: it starts with a very high ZCR, corresponding to the frication noise observed at the beginning of the segment. The diminishing ZCR during [z] corresponds to the slow disappearing of this turbulence.

This general pattern can also be observed from the raw data presented in Figure 4.13. The ZCR of the apical vowel [z̩] is presented separately. It can be observed that for [i ʉ u a] [21] vowels, the ZCR is lower and more constant, mostly concentrated below 2000 times per second, while for the apical vowel [z̩] the ZCR is much higher and more variable, varying from 2000 times per second to 10000 times per second for the majority of the tokens.

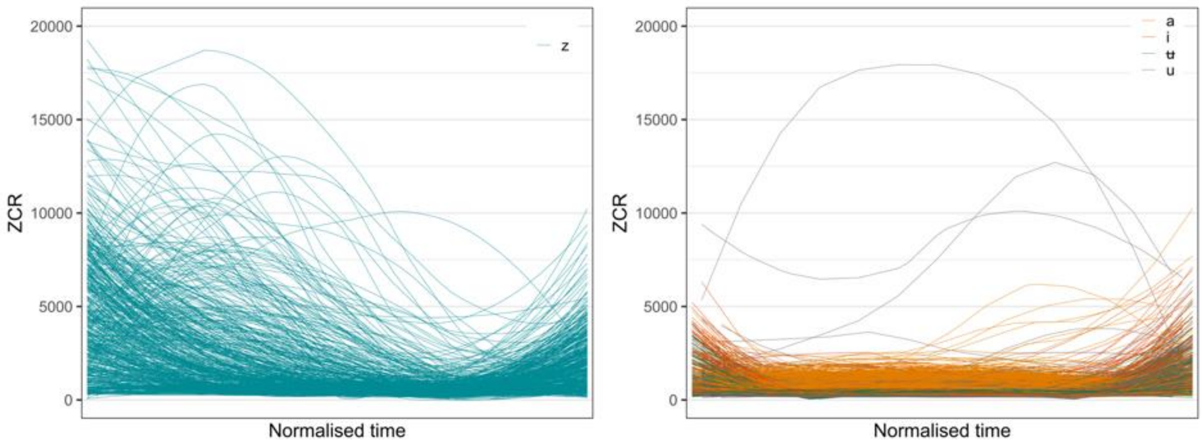


Figure 4.13. ZCR of all tokens of [z̩] and [a i u ʉ] in JHC. The x-axis represents normalised time of the vocalic segments and y-axis represents the zero-crossing times per second. [z̩] tokens on the left and [a i u ʉ] tokens on the right.

4.2.5.2. ZCR patterns of [z̩] in different consonantal contexts

In this section, I report the effect of consonantal context on the way frication noise is realised during [z̩]. Specifically I seek to determine whether any difference can be observed depending on whether the preceding consonant is sibilant [s ts tsʰ], labial [p pʰ m] or coronal nasal [n]. The results, presented in Figure 4.14, show that there are observable differences in the way frication is implemented throughout the segment. These differences can be categorised into three patterns depending on the nature of the preceding consonants: [z̩] after [pʰ s ts tsʰ], [z̩] after [m n], and [z̩] after [p].

The most obvious pattern is when [z̩] is preceded by [pʰ s ts tsʰ]. In this context, [z̩] displays much higher ZCR at its beginning. This is interesting since it mirrors the finding reported in section 4.2.2.2, where [z̩] after the same class of consonants [pʰ ts tsʰ s] had shorter duration compared to [z̩] after [m n p]. The same explanation provided to account for the shorter duration of [z̩] also holds here. Specifically, the gesture for [z̩] is already achieved during the release phase of [pʰ],

[21] Three [u] vowels have abnormal ZCR values which may be explained by the devoicing which can affect this high vowel.

so that aspiration is masked by the oral sourced frication noise. This issue will be examined in more detail in section 5.2.6 based on ultrasound data.

The [z] segments after [p] have lower ZCR, but the values are still above 2000 times per second at the beginning of [z]. This shows that even when [z] is completely heterorganic with the onset consonant, it still presents abundant frication noise. The ZCR of [z] segments after [m n] behave in a different way: it starts at a low level and systematically increases to achieve a rather high level before the final falling phase. The lower ZCR at the starting point can be explained by the nasality of the onset consonants. Nasal consonants require an open nasal cavity that prevents high intraoral air pressure. After the release of the nasal consonant, the nasal cavity closes, and intraoral air pressure rises. Similar to the [pʰ] case, the gesture of [z] is most probably achieved anticipatorily during the onset consonants [n m], which means that the fricative tongue shape is already formed when the intraoral air pressure rises after the closure of velopharyngeal port. The fricative tongue shape and the raised intraoral pressure lead to frication noise, shown as the increasing of the ZCR value. This behaviour confirms again that [z] displays frication even when preceded by nasal consonants, and clearly suggests that turbulence noise is inherent to the production of the apical vowel. There is a small difference between [z] after [m] and [z] after [n], with lower frication noise in [mz] than in [nz]. This difference could be attributed to the same alveolar constriction shared by [n z] sounds, but not by [m z]. It could be that the alveolar constriction was achieved later and was less constricted after [m].

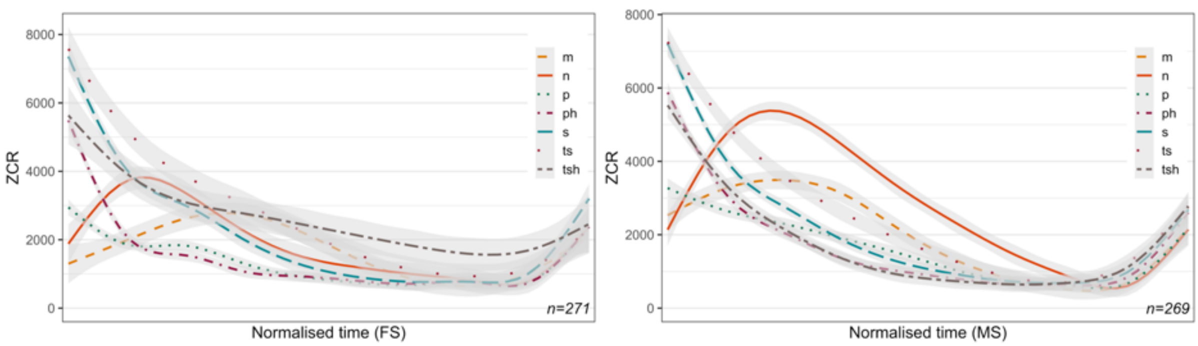


Figure 4.14. ZCR of [z] in different consonantal contexts in JHC. The curves were generated with loess smoothing method, x-axis represents normalised time of the [z] segments and y-axis represents the zero-crossing times per second. Female speakers (FS) on the left and male speakers (MS) on the right.

4.2.5.3. ZCR patterns of [z] for different speakers

Although the presence of frication noise is prevalent during the production of [z] in JHC, speaker-specific differences still exist, with some speakers producing more frication noise than others. The ZCR patterns of [z] for all speakers are presented in Figure 4.15. It is observed that 9 speakers out of 10 have higher than 2000 times per second ZCR at the beginning of [z]. Only FS5 behaves differently, as she has just above 1000 times per second ZCR at the beginning of her productions.

An interesting aspect is seen when FS5 is compared to MS2. These two subjects pattern together and exhibit similar mid-sagittal tongue contours during the production of [z] (as will be shown in section 5.2.2), yet MS2 has much higher ZCR values than FS5. This difference indicates that in addition to tongue configuration, the aerodynamic dimension also plays an important role in the generation of frication noise. For MS2, the intraoral pressure may decrease more slowly than for FS5, which could result in more frication noise for the former and less frication noise for the latter. The importance of aerodynamic adjustments in generating frication noise during the production of [z] is discussed in section 6.3.

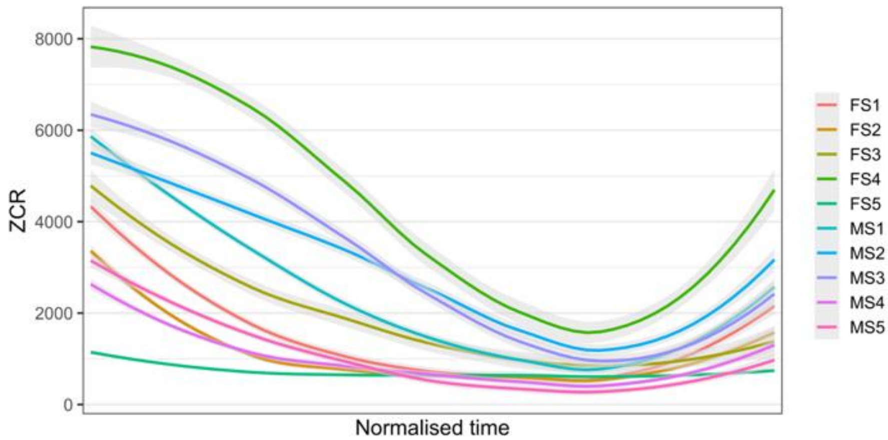


Figure 4.15. Individual ZCR patterns of [z] for the ten speakers. The curves were generated with loess smoothing method, x-axis represents normalised time of the [z] segments and y-axis represents the zero-crossing times per second.

Although FS5 does not have higher than 2000 per second ZCR values, her productions of [z] still exhibit higher ZCR values than her vowels (Figure 4.16). This speaker has only a short period of frication noise at the beginning of the apical vowel, which diminishes rapidly when voicing starts (see Figure 4.17 for an example).

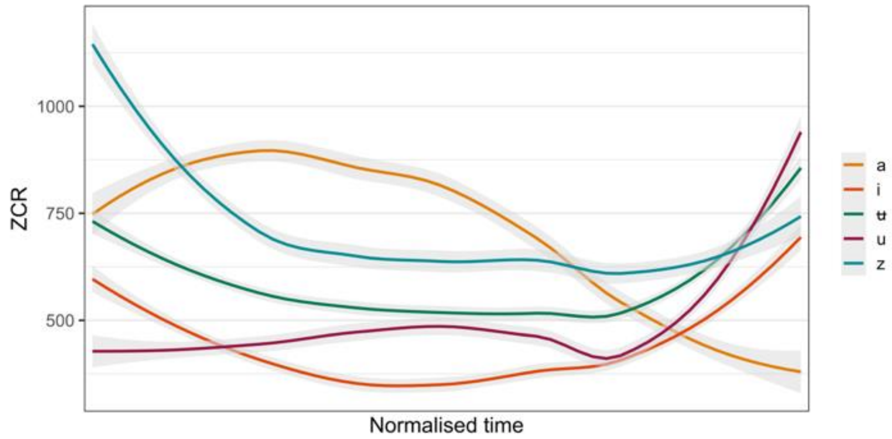


Figure 4.16. ZCR pattern of all nuclei from FS5. The curves were generated with loess smoothing method, x-axis represents normalised time of the [z] segments and y-axis represents the zero-crossing times per second.[22]

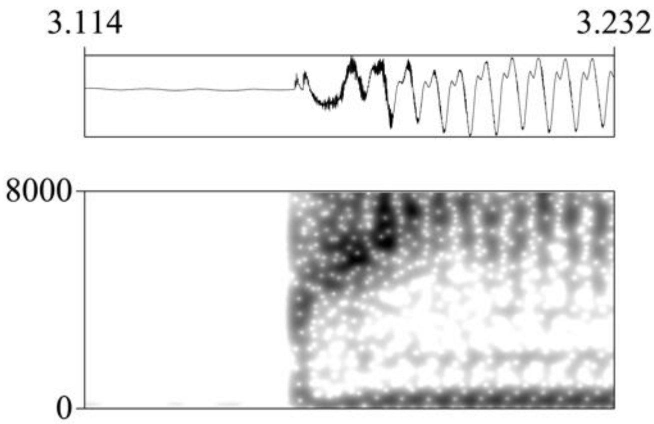


Figure 4.17. Acoustic signal and spectrogram of the form [pz˩˧] produced by FS5.

4.2.6. Centre of gravity: comparing the aspiration in [pʰ] to frication in [s]

As shown in section 4.2.2.2 and section 4.2.5.2 concerning the duration and the COG values of [z], the aspirated labial stop [pʰ] patterns with sibilants [s ts tsʰ], rather than with the other labial consonants. The explanation provided for this 'unnatural' patterning was that the alveolar constriction of [z] is achieved during (or even before) the release phase of [pʰ]. The consequence of this overlap is that the frication noise generated by [z] gesture dominates the glottal sourced aspiration noise of [pʰ], resulting in a phase that has acoustic characteristics resembling those of a [s] consonant (i.e., energy concentration at higher frequencies). Two examples of [pʰz] are given in Figure 4.18 for illustration.

[22] Here the vowel [a] seems to have highe ZCR, but as shown in Figure 4.12, it is the normal value for this vowel.

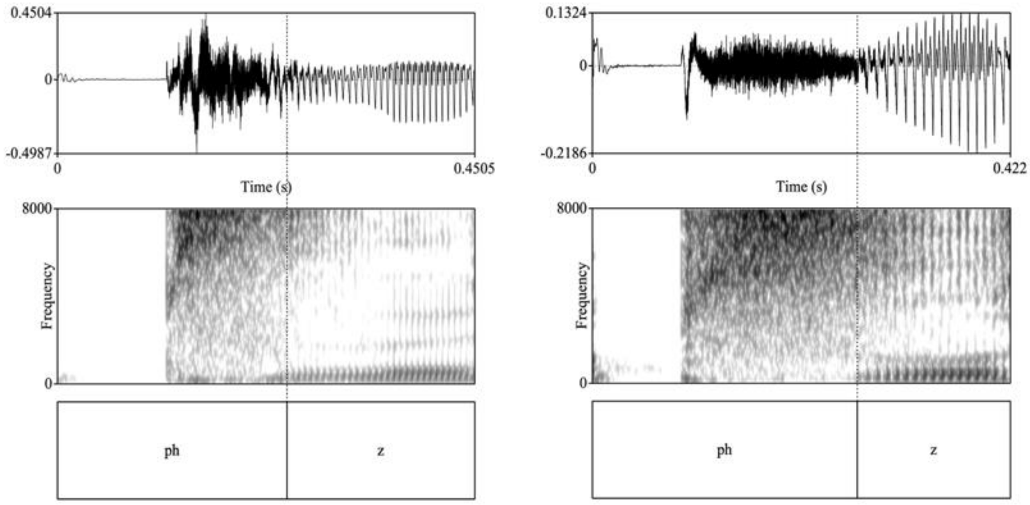


Figure 4.18. Acoustic signals and spectrograms of two realisations of [pʰz] by FS1 (left) and MS2 (right). These two examples contain frication noise only at the beginning of [z].

COG measurement was used to examine the acoustic similarity between the aspiration phase of [pʰ] and the frication noise of [s] in the same vocalic contexts. The results are presented in Figure 4.19. They show that when followed by [ʉ u a], the COG of [s] and the COG of the aspiration phase of [pʰ] are different. However, when followed by [z], the aspiration of [pʰ] displays the same COG as the frication noise of [s]. Welch's *t*-tests conducted on each vocalic context confirm this statistically (see Table 4.6). This result shows that the release of [pʰ] contains frication noise generated during the production of fricative [z].

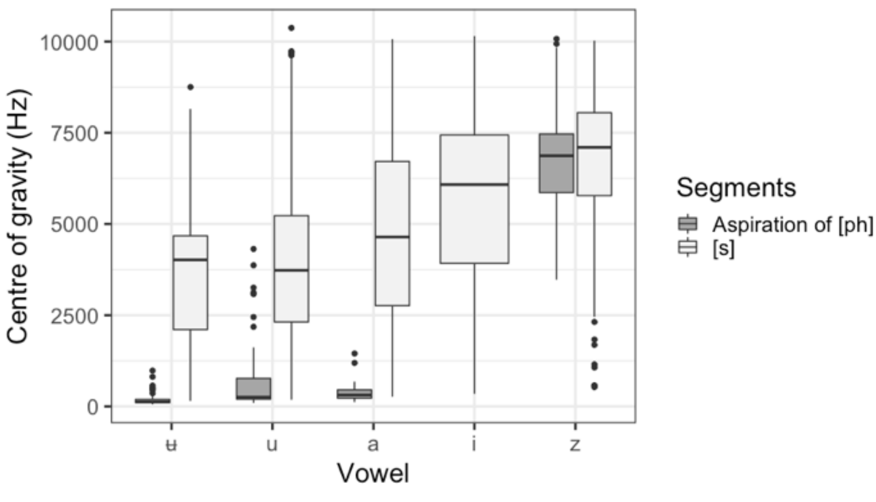


Figure 4.19. COG of the relase phase of [pʰ] and the frication noise of [s] in the contexts of [i ʉ u a z]. Data obtained from all speakers ([i] vowel can only be preceded by [s]).

Table 4.6. Mean COG values (Hz) of [s] and the release phase of [pʰ]. Welch's *t*-test results conducted on the two groups of COG values.

Nuclei vowels	Mean COG (Hz) of the aspiration phase of [pʰ]	Mean COG (Hz) of [s]	Welch's *t*-test
ʉ	176	3670	$t(50.26)=-12.12$, $p<.0001$
u	711	4003	$t(123.09)=-11.84$, $p<.0001$
a	384	4728	$t(50.33)=-11.75$, $p<.0001$
z	6735	6682	$t(164.54)=0.19$, $p=.85$

4.3. Summary of the acoustic study

This chapter presented an acoustic investigation of the JHC apical vowel [z]. The main findings obtained are summarised below:

1) [z] is produced predominantly with frication noise superposed on voicing. The frication noise never continues throughout the entire duration of [z], and stops earlier or later in the second half of the segment. When frication noise diminishes, the formant structure becomes clearer, resulting in an approximant-like configuration.
2) [z] has a shorter duration after [pʰ s ts tsʰ] compared to [m p n]. This difference was explained by the anticipatory realisation of the nucleus [z] coupled with the segmentation criteria used.
3) [z] has similar F1-F2 structure to [ʉ], resulting in important overlap in the vocalic space.
4) [z] displays lower HNR values compared to [i], confirming that the former has more frication noise.
5) The higher ZCR for [z] compared to vowels is additional evidence for the abundant frication noise displayed by this segment especially at its beginning. The presence of this noise is observed after labial stops and nasals, although coronal sibilants induce more frication noise.
6) COG shows that the release phase of [pʰ] in [pʰz] has the acoustic characteristics of a fricative [s], suggesting that the alveolar constriction of [z] is achieved during the release phase of the onset.

Taken together these results show that [z] in JHC has acoustic characteristics of a fricative consonant. The abundant frication noise even when it is preceded by labial stops and nasals is a strong argument in support of this analysis. Given that the fricative gesture of [z] could be achieved anticipatorily during these onset consonants, this gesture can only be considered as inherent to [z] since labial stops and nasals do not objectively involve a fricative gesture. In the next chapter, this aspect and others are explored based on ultrasound imaging.

Chapter 5 Articulatory study of the apical vowel [z] in JHC using ultrasound tongue imaging

The objective of the articulatory study is to examine the tongue configuration of JHC [z][23]. Based on ultrasound tongue imaging, on both mid-sagittal and coronal planes, it specifically seeks to determine whether the shape of the tongue is similar to or different from that of the consonant [s] or vowel [i], and to examine how it is affected by the nature of the preceding onset consonants.

5.1. Methodology

5.1.1. Speakers and recording materials

The ultrasound data were acquired one year after the acoustic data, in the same sound-attenuated room. Seven of the ten speakers recorded for the acoustic experiment participated in this data acquisition (FS1, FS3, FS5 and MS1, MS2, MS3, MS5). The word list used was a shorter version of the word list used in the acoustic study. The list was shortened in order to reduce the time of recording, which took approximately 50-60 minutes for each speaker. As for the acoustic study, this list contained monosyllabic words with [p pʰ m n ts tsʰ s] onsets and [i ʉ u a z] nuclei. The word list, written in Chinese characters, was printed and presented to speakers on a A4 format paper (See A.1). The seven subjects read the words in a frame sentence [ki˧ɕɔ˩˧ _ ɕɔ˩˧sɔ˦fa˦] 'He/She writes _ three times', repeated three times for mid-sagittal recordings and three times for coronal recordings (1218 total target syllables, [z]: 462, [i]: 168, [u]: 168, [a]: 126, [ʉ]: 294).

Note that each recording was saved separately. That is, each carrier sentence (with the target syllable) was produced by the speaker, then saved. The speaker was then

---

[23] As in the previous chapter, I will be presenting the variable phonetic realisations of the concerned segments and syllables, and some of them are not phonemic (e.g., the onsets [n] and [l] are allophones, see section 3.2.3). All transcriptions are thus enclosed in square brackets, and the syllabic diacritic '̩' for apical vowel is omitted for simplicity.

asked to produce the next carrier phrase (with the next target syllable, not a repetition of the same target syllable) and so on. Each recording was preceded by the syllables [kaka] in order to provide extra information on the synchronisation of the ultrasound recording and the acoustic signal, based on the acoustic release of the velar [k].

5.1.2. Ultrasound data acquisition and segmentation criteria

All ultrasound data were recorded with the Ultrasound Stabilisation Headset (Articulate Instruments Ltd., 2008) and the Articulate Assistant Advanced software (AAA, V217.03) (Articulate Instruments Ltd., 2012). The probe used was a microconvex portable ultrasound probe, with a diameter of 40 mm. Before the recording began, speakers were familiarised with the ultrasound probe and headset, as well as with the word list. They also had time to visualise the tongue movements through AAA.

The speakers were sat in a comfortable chair. The headset was then adjusted to the morphological specificities of each speaker, in a way that it would not move during the recordings, while making it as much confortable as possible. The mid-sagittal ultrasound data were recorded first and then coronal ultrasound data, with a small pause between the two sessions. During the pause the headset was removed from the head of the speaker. None of the speakers reported having experienced pain or discomfort.

The corresponding acoustic data were recorded by using the same equipment described in the methodology section of the acoustic study (see section 4.1.2), except that the microphone was attached on the ultrasound headset. Although the acoustic signals obtained were not perfectly suitable for fine acoustic analyses, they were sufficient for segmentation use. The synchronisation of the ultrasound images and the acoustic signals were done automatically using AAA software, and checked manually by observing the form [kaka] which preceded each recording.

A small but important detail in using the headset was that it became impossible for some speakers to wear glasses once it was put on and well adjusted. To avoid experiencing difficulty in reading a screen from a relatively long distance, the option used was to present the word list printed on a paper using large font size.

The mid-sagittal ultrasound recording procedure was rather straightforward. The headset kept the ultrasound probe in the mid-sagittal plane and the probe was pointed to the anterior of the tongue in order to have a better image of the tongue tip. The coronal recording was obtained by turning the probe in a 90° angle and pointing the probe to the anterior part of the tongue. The probe was adjusted in a way that the medial grooving of [s] consonant was easily observable as was shown in Stone (1992). It is important to note however that the direction of the ultrasound probe was not controlled as in Stone et al. (1988, 1992)[24], because a perpendicular base could not be provided with the headset on. There was no object of reference that allowed adjusting the angle of the probe to a fixed degree.

The ultrasound probe used in the data acquisition had a field of view of 92°. The depth was adjusted to have a maximum view of the tongue, causing different frame rates for female and male speakers (82.1 frames per second for female speakers and 81.4 frames per second for male speakers). Ultrasound data were segmented in AAA manually using the corresponding audio, and traced manually with the help of the built-in tracing algorithm. The tracing criteria are presented in detail in sections 5.1.3 and 5.1.4.

The segmentation criteria were the same as those presented in the acoustic study. In the acoustic study, the onset of the apical vowel was set at the first pulse detected by Praat (see section 4.1.2). However, AAA does not provide a pulse detection function, so the starting point of the apical vowel was decided by visual observation of the acoustic signals and the spectrograms. The onset of [z] was thus defined as the beginning of voicing of this segment. This decision will not have much effect on the results presented since the first pulse detected by Praat corresponds to the starting of voicing. The smoothing-spline ANOVA analysis conducted were done based on the ‘nearest midpoint image’ of each segment (see section 5.1.6).

5.1.3. Mid-sagittal tongue contours tracing criteria and data-points extraction

The tracing of the mid-sagittal tongue contour was done on each target syllable with the assistance of the built-in AAA semi-automatic tracking algorithm (Articulate Instruments Ltd., 2012). An advantage of using the semi-automatic tracking is that it reduces the effects of human factor. The tracing procedure started with defining a space in which the tongue moved during the articulation of the target syllable. This space was limited by ‘roof’ and ‘min-tongue’ as shown in Figure 5.1. In the present study, the ‘roof’ limit was not defined as the palate but drawn manually by observing the tongue movement during each target syllable. In the example presented in Figure 5.1, for instance, the ‘roof’ line was drawn manually by including all the ranges over which the tongue surface may be found during the production of the item [pz]. This procedure was necessary for the automatic spline fitting to work. By doing so, AAA could isolate non ultrasound graphics such as menus, scales and axes from the search space (Articulate Instruments Ltd., 2012). The grey ‘min-tongue’ line was defined similarly so that the tongue surface never went under it. The tongue surface could thus only be found between these two lines during the target syllable. The

[24] Stone et al. (1988, 1992) had an ultrasound probe holder which had a perpendicular base. This holder is an inversed L-shaped metal shelf fixed on the floor with the ultrasound probe attached on the top. They were able to adjust the angle of the coronal plane buy adjusting the direction of the probe with goniometers according to the perpendicular mental shelf.

defined ‘roof’ and ‘min-tongue’ was saved as a template so that AAA could use it to track all segments having the same label. For each segment studied, a template was created following the same procedure. That is, for each speaker, five templates were created, one for each syllable nucleus [i ʉ u a z].

AAA offers a semi-automatic function by tracking the tongue surface inside the predefined limits in a template and within the given acoustic segmentation. All ultrasound images within a segmentation were tracked in a batched process. The result was corrected by hand when needed, with the assistance of AAA’s ‘snap to fit’ function (see Articulate Instruments Ltd., (2012) for more details on this function).

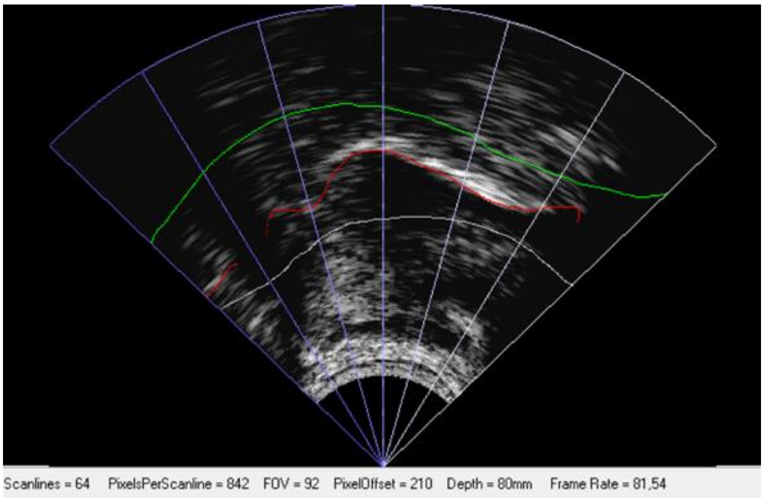


Figure 5.1. One mid-sagittal ultrasound frame of [z] in [pz˧˩] from FS3, illustrating the tracing template (Tongue tip on the right and tongue root on the left.). The green line represents ‘roof’, the grey line represents ‘min-tongue’, and the red line represents the tongue shape traced by AAA without any manual correction.

The semi-automatic tracking result may contain errors. Two such errors are observed in Figure 5.1. The first one relates to the tongue tip. As can be seen from the mid-sagittal ultrasound frame, the tongue tip is not quite visible because of the mandible shadow, so the tracing algorithm provides a pseudo-tongue tip contour. Although this pseudo-tongue tip is probably not far from reality (considering a tongue’s nature shape and position), it has been considered an error and eliminated from the analyses (see Figure 5.2). The other error concerns the tongue root surface, which is not traced correctly because of the hyoid shadow. This part of the tracing was also considered as an error and eliminated in the analysis (see Figure 5.2). As a consequence, only the tracings between the mandible shadow and the hyoid shadow were considered reliable and used for the smoothing-spline ANOVA analysis.

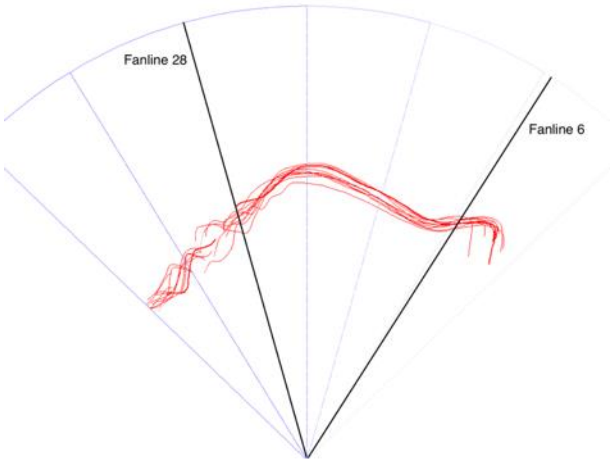


Figure 5.2. Mid-sagittal ultrasound tongue traces as shown in the Spline Workspace of AAA software (tongue tip on the right and tongue root on the left). These traces correspond to the nearest midpoint images from 11 instances of [z], extracted from FS3's productions of [pz˥˩ pz˨˩˧ pʰz˦ pʰz˨˩˧ mz˨˩˧ nz˦ nz˨˩˧ tsz˨˩˧ tsʰz˨˩˧ sz˨˩˧ sz˧˥]. The fanline 6 and the fanline 28 delimit the range of the data points considered as reliable and analysed in this sutudy.

The traces in this figure represent the nearest midpoint images of 11 realisations of [z] by FS3. As these traces clearly show, the semi-automatic traced tongue shapes are highly consistent across tokens. But the tongue tip and the tongue root contours are visibly not consistent. The Spline Workspace shown in Figure 5.2 provides the function to identify each fanline of the ultrasound image. The tongue shapes between fanline 6 and fanline 28 (approximately between the mandible shadow and the hyoid shadow) were considered to be correctly traced across tokens for the [z] segments produced by FS3. The x/y coordinates of the range defined by these two fanlines were used in the smoothing-spline ANOVA (SS ANOVA) analysis presented in section 5.1.6. Since the tongue tip is left out in the SS ANOVA generalisation, I refer to the front part of the tongue as 'tongue front' in the SS ANOVA figures. In order to compare the tongue shapes of [z] to the tongue shapes of other segments, the fanlines limiting the reliable range of [z] were used to determine the reliable range of all segments for one speaker[25]. This procedure was repeated for each speaker. The data points were then extracted for the nearest midpoint image and analysed using SS ANOVA method.

---

[25] Understandably, the two delimiting fanlines were different for each speaker. But for each speaker, the apical vowel [z] was always used as a reference to determine the range within which the tongue shapes were considered as reliable.

5.1.4. Coronal tongue contours tracing criteria and data-points extraction

The tracing of the coronal tongue contours follows the same procedure presented in section 5.1.3[26]. One example of 'roof' and 'min-tongue' is given in Figure 5.3. As the figure shows, the medial part of the tongue is reliably tracked, but not the left and the right edges. These edges were eliminated from the SS ANOVA analysis.

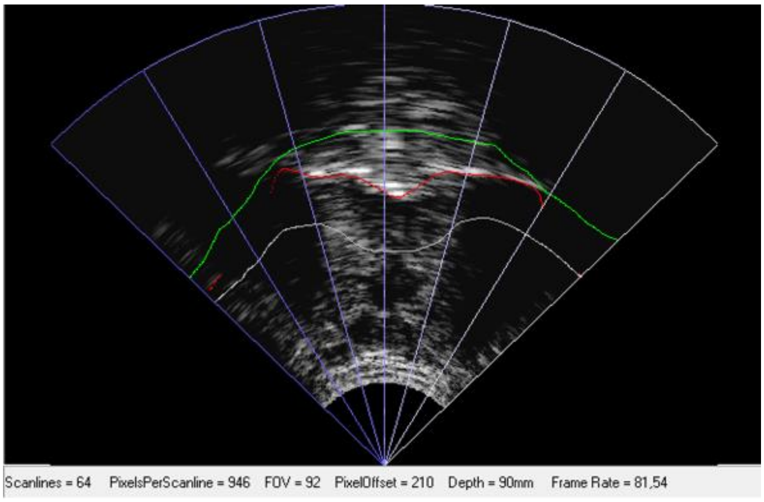


Figure 5.3. One coronal ultrasound frame of [z] in [pzʅ] from FS3, illustrating the tracing template. The green line represents 'roof', the grey line represents 'min-tongue', and the red line represents the tongue shape traced by AAA without any manual correction.

Unlike in the mid-sagittal plane, there was no mandible shadow and hyoid shadow in the coronal plane, so there was no anatomical reference to define a reliable range of the semi-automatic tracking results. The range between fanline 14 and fanline 27 was considered reliable based on the consistency of AAA's tracking results (see Figure 5.4). The visual examination method was conducted on each studied segment and the reliable range was extracted for SS ANOVA calculation.

Another factor that was taken into consideration in defining the reliable range in the coronal plane was the crossing points of the SS ANOVA splines. As shown in Figure 5.5, the splines cross on the left side and the right side (see the crossing between the contours of [z] and [s] and the contour of [ʉ]). The range within these sides was considered reliable. The coronal ultrasound results presented in the following sections focus only on the range within these sides.

---

[26] The built-in tracking algorithm of AAA is designed to track mid-sagittal tongue contours. In this study, it is also used to track coronal tongue contours.

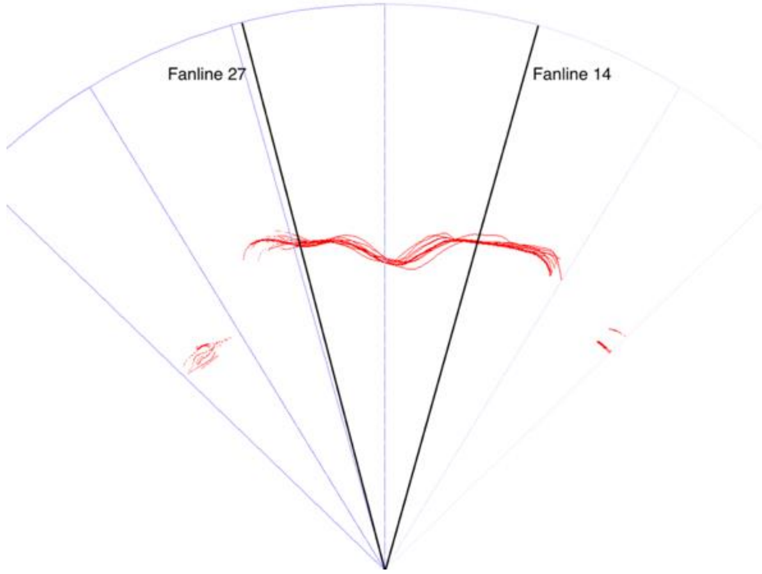


Figure 5.4. Coronal ultrasound tongue traces of the nearest midpoint images of 11 realisations of [z] in [pz˥˨ pz˧˩˧ pʰz˦ pʰz˧˩˧ mz˧˩˧ nz˦ nz˧˩˧ tsz˧˩˧ tsʰz˧˩˧ sz˧˩˧ sz˧˥] by FS3, shown in the Spline Workspace of AAA. The fanline 14 and the fanline 27 are drawn to show the range of the data points considered as reliable.

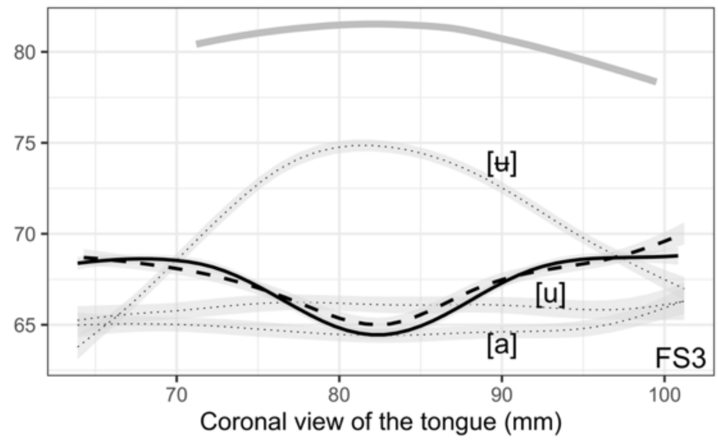


Figure 5.5. Coronal SS ANOVA splines from FS3. Data points are extracted in x/y coordinates from nearest midpoint images of each segment. Dotted grey lines represent [i u a ʉ], dashed black line represents [s] and solid black line [z]. The thick grey line represents the palatal traces and the grey areas represent 95% Bayesian confidence intervals. The vowel [i] is not traced for FS3.

5.1.5. Palate tracing and correction

The palatal contours were obtained by six independent water swallow tasks for each speaker. The results were corrected and ameliorated by using the 'Cave Vocal Tract' function of AAA (version 218.03). This function gives the trajectory of the tongue in the vocal tract and can superpose all trajectories to form an 'articulatory space' for the tongue. An example is given in Figure 5.6.

The black area in Figure 5.6 corresponds to the superposition of all tongue surfaces from three recordings (each recording is a frame sentence with a target syllable). The three embedded target syllables are [pz tsi pʉ]. Recall that the frame sentence contains [k ɕ s i ɔ a] and the target words contain [ts z i ʉ] which induce lingual movement. The tongue surface thus moves in a range consisting of these sounds. Among these sounds, the tongue touched the palate for [k ts], and approached the palate for [ɕ s z i ʉ]. As a consequence, the upper edge of the caved vocal tract is composed of [k ʉ ɕ ts] segments' upper edges ([k] at velar, [ʉ ɕ] at palate, [ts] at alveolar). This upper edge can also be visualised in the Spline Workspace, as shown in Figure 5.7. With the help of this function, the hard palate can be better defined.

Note however that the water-swallow tasks did not always provide easily traceable palatal contours. The palatal contours obtained were corrected using the 'Cave Vocal Tract' function. As presented above, the palatal traces were set to be just higher than the upper edge of the caved vocal tract, as shown in the above figure. The palatal traces were then used to indicate an approximate position of the palate in SS ANOVA results.

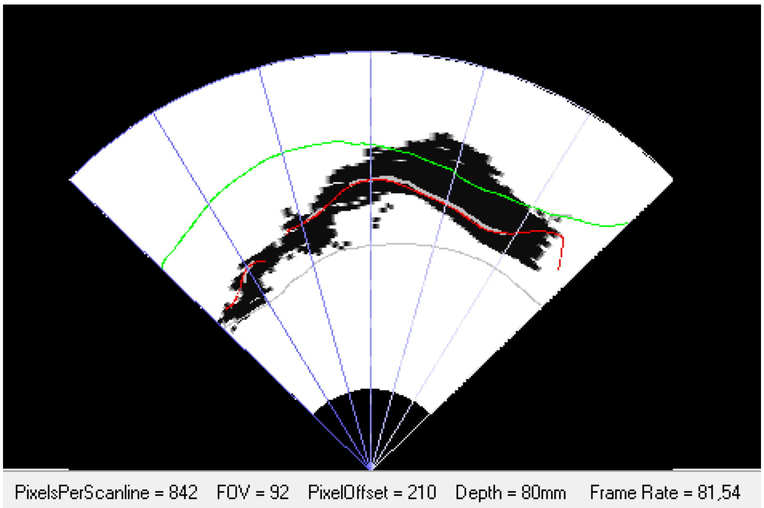


Figure 5.6. Result of the 'Cave Vocal Tract' function in AAA (Articulate Instruments Ltd., 2012: version 218.03) (tongue tip on the right and tongue root on the left). The black area represents the superposition of all tongue surfaces from all ultrasound images of three recordings with [pz tsi pʉ], frame sentence included (Data from FS3). The green line represents 'roof' and grey line represents 'min-tongue'. The red line represents the tongue shape traced by AAA for [z].

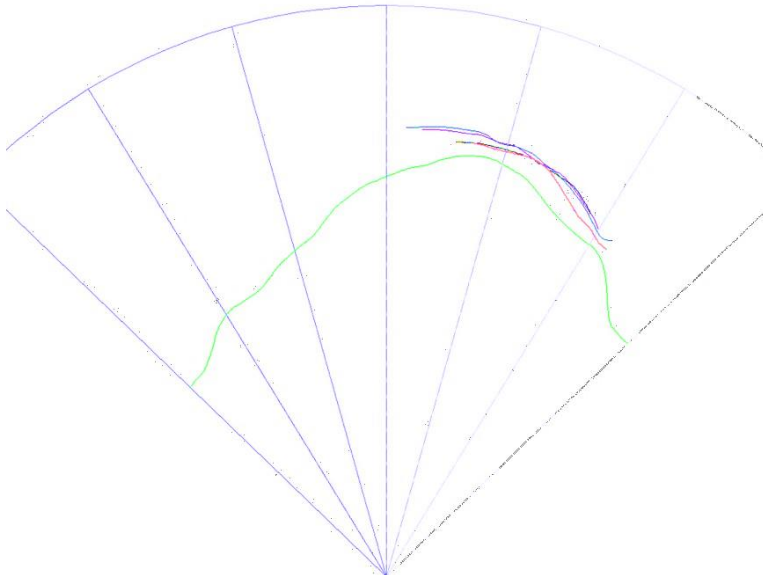


Figure 5.7. The upper edge (green line) obtained in Figure 5.6, presented with hand-traced palatal traces from water-swallow tasks.

5.1.6. Smoothing-Spline ANOVA analysis

As stated above, the nearest midpoint image of each studied segment was extracted in x/y coordinates. The data points in the reliable range were analysed using smoothing-spline ANOVA (SS ANOVA) method. The statistical tool used to calculate the SS ANOVA was RStudio and the Package gss (Gu, 2014). As presented in section 5.1.3 and section 5.1.4, data points from hyoid shadow to mandible shadow in mid-sagittal plane and data within the junction of the traces in coronal plane were taken into consideration by the SS ANOVA estimation.

The SS ANOVA analysis used in the present study follows Davidson's (2006) proposition. She presented the SS ANOVA method as '*a technique for determining whether or not there are significant differences between the smoothing splines that are the best fits for two data sets being compared*'.

The advantage of SS ANOVA is that it can show if two splines are significantly different and where the significant differences lie. If the interaction term of the SS ANOVA model is statistically significant, then the groups have different shapes. Since the interaction may be significant even if only a small section of the curves is different (e.g., the tongue root is the same, but the tip of one group is raised), Bayesian confidence intervals are used to determine which sections of the curves are statistically different (Davidson, 2006). In the present study, the Bayesian confidence intervals are reported for [s]/[z] comparisons and for [z] in labial/coronal context comparisons.

The recordings from MS1 were excluded from analysis since the ultrasound images from this speaker did not provide sufficient information. The tracing was nearly impossible, and the data were considered as not reliable. Similarly, in the coronal ultrasound data from FS3, the [i] vowel was excluded. In her case, when the tongue was raised for [i], there was no sufficient information to trace it reliably. All the remaining data were included in the corresponding SS ANOVA estimations.

5.1.7. Correspondence between mid-sagittal and coronal tongue contours

One difficulty encountered in coronal ultrasound data acquisition was the lack of anatomical references to precisely locate the coronal plane. The method adopted in the recording sessions was to orient the ultrasound probe to the anterior of the tongue and adjust it in a way that the medial grooving of [s] consonant was easily observable as shown by Stone (1992). The recordings proved satisfactory when the mid-sagittal plane and the coronal plane were compared.

By comparing the SS ANOVA splines' relative positioning on mid-sagittal plane and coronal plane, it was possible to determine an approximate range on the mid-sagittal plane in which the coronal plane was taken. This is illustrated in Figure 5.8. The SS ANOVA splines of the coronal ultrasound are on the right side, and the green range in the centre of the coronal view is considered to correspond to where the mid-sagittal ultrasound is taken.

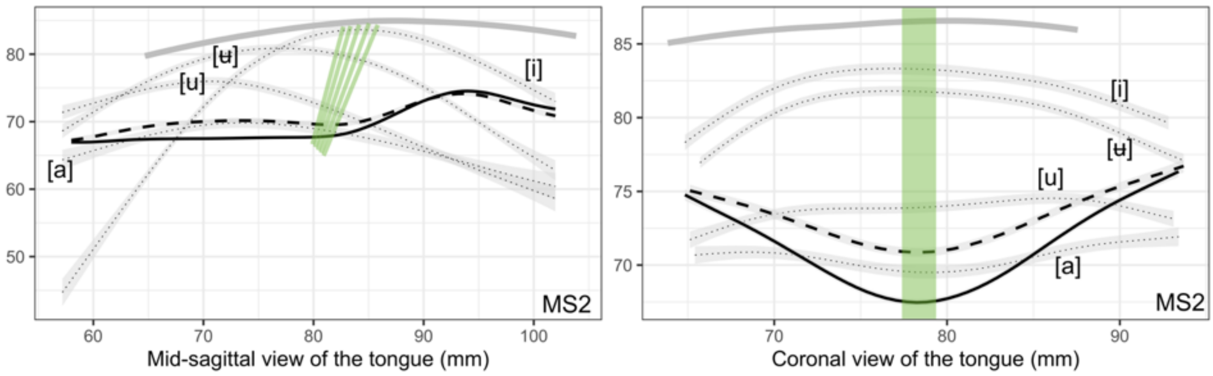


Figure 5.8. SS ANOVA splines showing the approximate correspondence between the mid-sagittal and the coronal plane. The green areas on the mid-sagittal plane correspond approximately to where the coronal plane was obtained, while the green areas on the coronal plane correspond approximately to where the mid-sagittal plane was obtained. SS ANOVA splines of mid-sagittal and coronal tongue contours were extracted in x/y values (mm) at the nearest midpoint image of each segment (the tongue front is on the right and tongue root on the left). Dotted grey lines represent [i u a ʉ], dashed black lines represent [s], and solid black lines [z]. The thick grey lines represent the palatal traces, and the grey areas represent 95% Bayesian confidence intervals. Data are from MS2.

By comparing the relative positioning of the splines in this range, it can be observed that from highest to lowest, the splines are ranked as [i ʉ u s a z]. Taking this ranking into consideration, a possible range is marked in green on the mid-sagittal view in Figure 5.8. Within this range the same ranking as in the coronal plane (i.e., from highest to lowest [i ʉ u s a z]) is also observed. Outside this range, the relative positioning of the splines no longer follows this ranking.

The defined range, determined by visual examination, shows that the coronal ultrasound plane was taken from the anterior part of the tongue dorsum (and not from the tongue tip). The data are still highly informative, as they enabled the comparison between segments on the coronal plane, and between different contexts in one speaker's productions. This comparison can reveal the articulatory gestures of the apical vowel [z] since the medial-grooving shape was observed only in consonant [s] and apical vowel [z], but not in vowels.

5.2. Results

The results of the articulatory study are presented in this section. Qualitative observations of [z] using raw ultrasound images are presented in the first place, before reporting on the general patterns observed based on SS ANOVA analyses. Following these, results from a set of comparisons are presented. The tongue shape of [z] is compared to the tongue shapes of [s i u ʉ] in order to reveal the similarities and/or differences between [z] and [s] on the one hand, and between [z] and the high vowels on the other. The absence of any consistent effect of context on the tongue configuration of [z] is shown by comparing labial and coronal contexts. The section ends with the results on the dynamic tongue shape evolution in [pz pʰz mz] syllables.

5.2.1. The articulation of [z] on mid-sagittal and coronal planes: qualitative observation

A close visual examination of the mid-sagittal tongue contours for [z] reveals two shape-based patterns: pattern 1 (FS1, FS3, MS3, MS5) and pattern 2 (FS5 and MS2). A raised tongue dorsum characterises pattern 1, while pattern 2 shows a flat tongue dorsum and a raised tongue tip. Figure 5.9 illustrates these two patterns and shows how the shape of the tongue changes over time during the articulation of [z] for the six speakers.

It is clear that for all speakers, the tongue does not move significantly during the articulation of [z]. From onset to ¾ point, the tongue shape of [z] remains constant. The image corresponding to the offset of [z] shows the effect of coarticulation between this segment and the following consonant [ɕ] contained in the frame sentence.

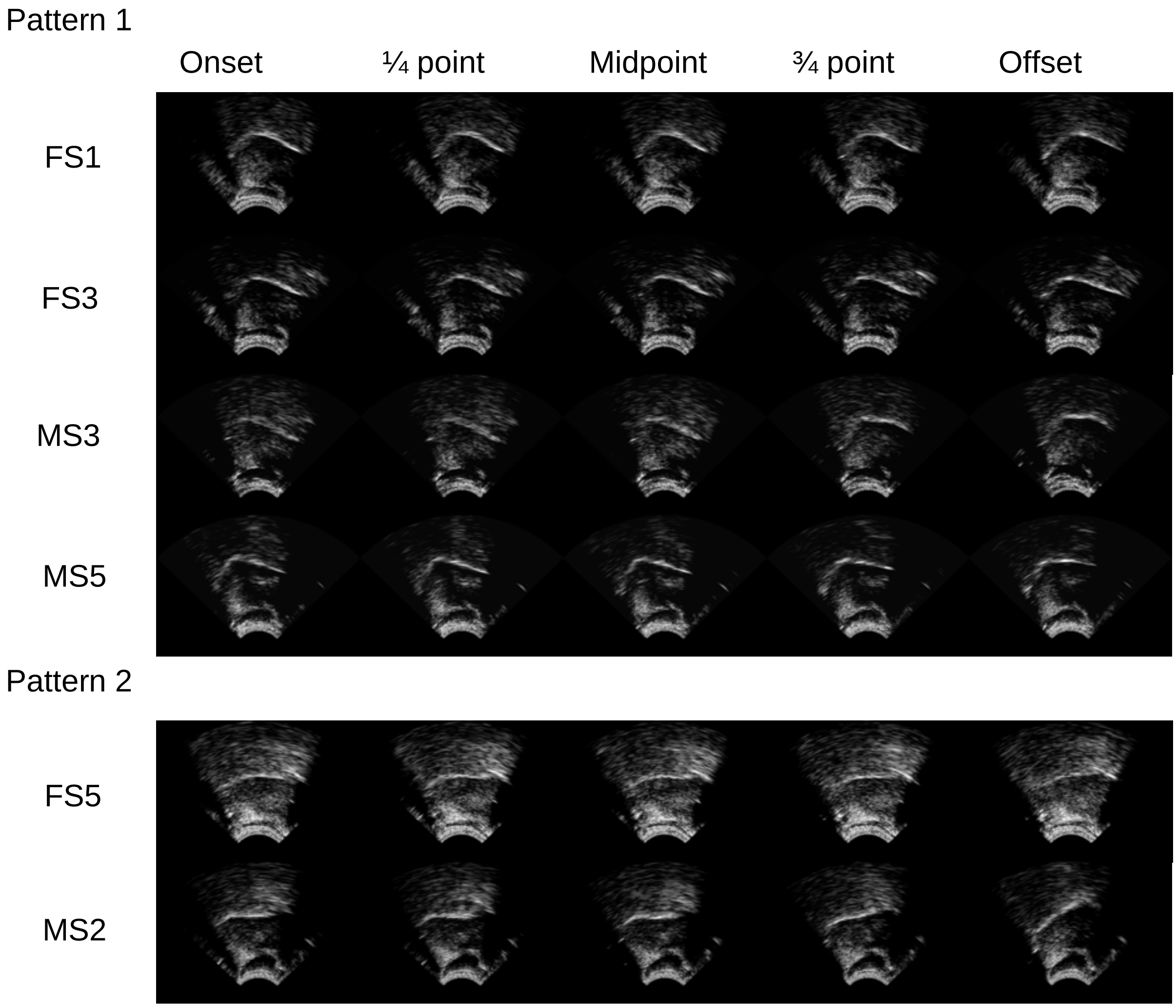


Figure 5.9. Raw mid-sagittal ultrasound images of [z] in [pz] (tongue tip on the left and tongue dorsum on the right). The six speakers are presented in rows. Columns present images obtained at different points during the production of [z]. The onsets, ¼ points, midpoints, ¾ points and offsets are defined according to the acoustic signals and spectrograms corresponding to the ultrasound recordings.

For comparison purposes, the mid-sagittal tongue shape of the SC apical vowel is shown in Figure 5.10. This tongue shape is virtually identical to pattern 1 observed here, suggesting that the two variants can display the same tongue configuration (at least for pattern 1, it is unclear if SC apical vowel has pattern 2 gesture). It should be noted that the mid-sagittal image of the apical vowel in SC was extracted from a [sz] syllable, while the JHC apical vowel images of Figure 5.9 were obtained from [pz] syllables. The two variants thus exhibit the same shape of the tongue although they were produced in two different consonantal contexts. This observation can be related to what was argued for when the formant values of apical vowels were compared in Chinese languages (see section 4.2.3.2).

It is worth mentioning that the Ikema fricative vowel /ɨ/ also has two different articulatory patterns that mirror the JHC pattern 1 and pattern 2. As shown in Figure 2.12 (see section 2.2.3.4), the left image corresponds to pattern 1 and the right image corresponds to pattern 2. It is not clear however whether the Ikema /ɨ/ has the same tongue shape as [s], as Fujimoto & Shinohara (2018) did not provide data on this. It would not be surprising that the vowel /ɨ/ displays the same tongue configuration as for the alveolar sibilant, in roughly the same way as the JHC apical vowel.

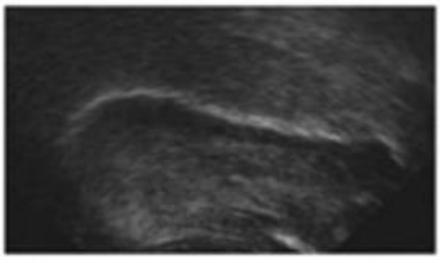


Figure 5.10. Raw mid-sagittal ultrasound image of the apical vowel in SC (tongue tip on the left and tongue dorsum on the right). This figure shows the tongue shape of [z] in [sz], obtained at the acoustic onset of the nucleus [z]. This figure is adapted from the Figure 5 of Lee-Kim (2014).

The ultrasound images of [z] in the coronal plane, shown in Figure 5.11, also show that the tongue shape of the apical vowel is constant during its articulation. Except for FS1, a medial-grooved tongue shape is observed for all speakers: the lateral parts of the tongue are raised, and the medial part of the tongue is grooved. A consequence of this is that the medial part of the tongue is much lower than the lateral parts. This medial-grooved tongue shape closely resembles that observed for the fricative [s], and indicates the presence of a narrowed air channel for both segments. This point will be developed in sections 5.2.3 and 5.2.5.

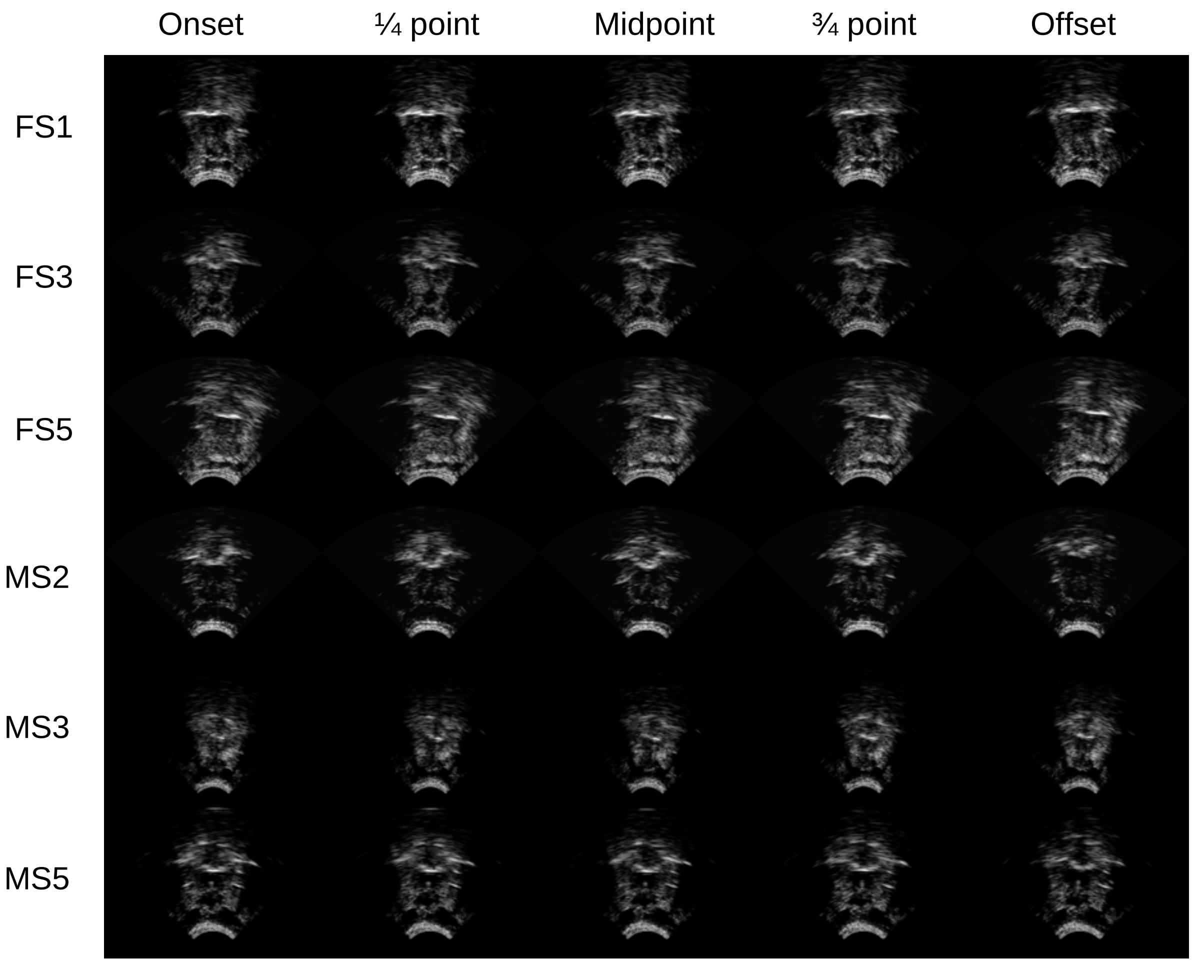


Figure 5.11. Raw coronal ultrasound images of [z] in [pz]. The six speakers are presented in rows. Columns present images obtained at different points during the production of [z]. The onsets, ¼ points, midpoints, ¾ points and offsets are defined according to the acoustic signals and spectrograms corresponding to the ultrasound recordings.

5.2.2. SS ANOVA splines of [i u ʉ] and [z]

5.2.2.1. On mid-sagittal plane

The general mid-sagittal shapes of [z] and the high vowels [i u ʉ], generalised into smoothing-splines using SS ANOVA method, are shown in Figure 5.12 for the six subjects.

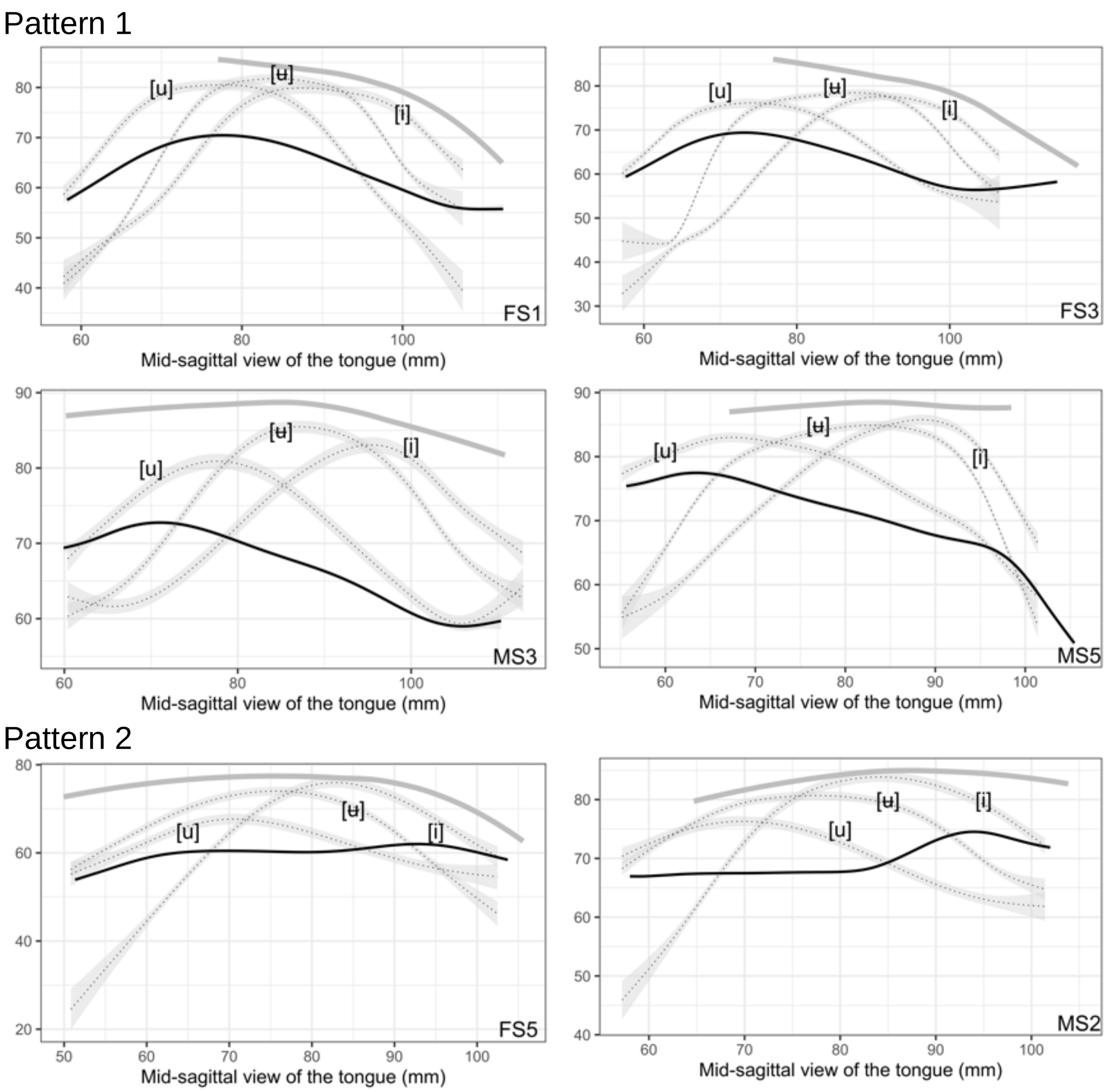


Figure 5.12. SS ANOVA splines of mid-sagittal ultrasound tongue contours, extracted in x/y values (mm) at the nearest midpoint image of each segment (tongue front on the right and tongue root on the left). Dotted grey lines represent [i u ʉ] and solid black lines represent [z]. The thick grey lines represent the palatal traces and the grey areas on the splines represent 95% Bayesian confidence intervals.

The differences in terms of articulatory gestures between the two patterns reported above are easily observable here. Interestingly for both patterns the tongue shape of [z] is very much different from that of high vowels. This difference is more specifically observed at the tongue dorsum, the basic gesture involved in the vowel making process. For high vowels, an arch is observed in the tongue dorsum mass. The position of this arching is what mainly alters the vocal tract and how it resonates, yielding different timbres for [i], [u] and [ʉ]. The front of tongue displays variable configurations depending on speakers. Recall that the tip of the tongue, generally considered neutral for vowels and active for consonants, is not perfectly rendered in our ultrasound data because of the mandible shadow.

5.2.2.2. On coronal plane

The difference in tongue configuration between the apical vowel and the high vowels is even more evident on the coronal plane as shown in Figure 5.13. Unlike high vowels, the apical vowel (except for FS1) displays a medial-grooved tongue shape. This specific tongue shape, which indicates the presence of a narrowed air channel, is directly related to a fricative gesture. As for FS1, there is indeed no medial-grooved tongue shape for [z], but interestingly, as section 5.2.3 will show, she also has the same no medial-grooved tongue shape for [s].

Figure 5.13 shows that [u] may display a slightly medial-grooved tongue shape. This medial grooving, already reported by Stone et al. (1988) and Stone & Lele (1992), is only observed for MS2 and MS5, and the depth of the grooving is not comparable to [s] and [z] segments. Moreover, the laterals of the tongue in [u] are not raised. This grooving shape, with non-raised laterals, reflects the natural grooving of the tongue rather than an articulated tongue gesture.

It is important to draw attention to the interesting difference between [i ʉ] and [u] and this tells us about the coronal ultrasound view recorded in this study. These vowels differ in the anterior [i ʉ] posterior [u] dimension. All speakers have high convex tongue shape for [i ʉ], while [u] is much lower. These shapes correspond to the anterior view of the tongue reported in Stone et al. (1988), and thus prove that the coronal ultrasound view is obtained at the anterior part of the tongue, not the dorsum or the posterior of the tongue. If the coronal view of the tongue were obtained at a more posterior part of the tongue, the tongue contour would be higher for back vowel [u]. This point is also mentioned in section 5.1.7.

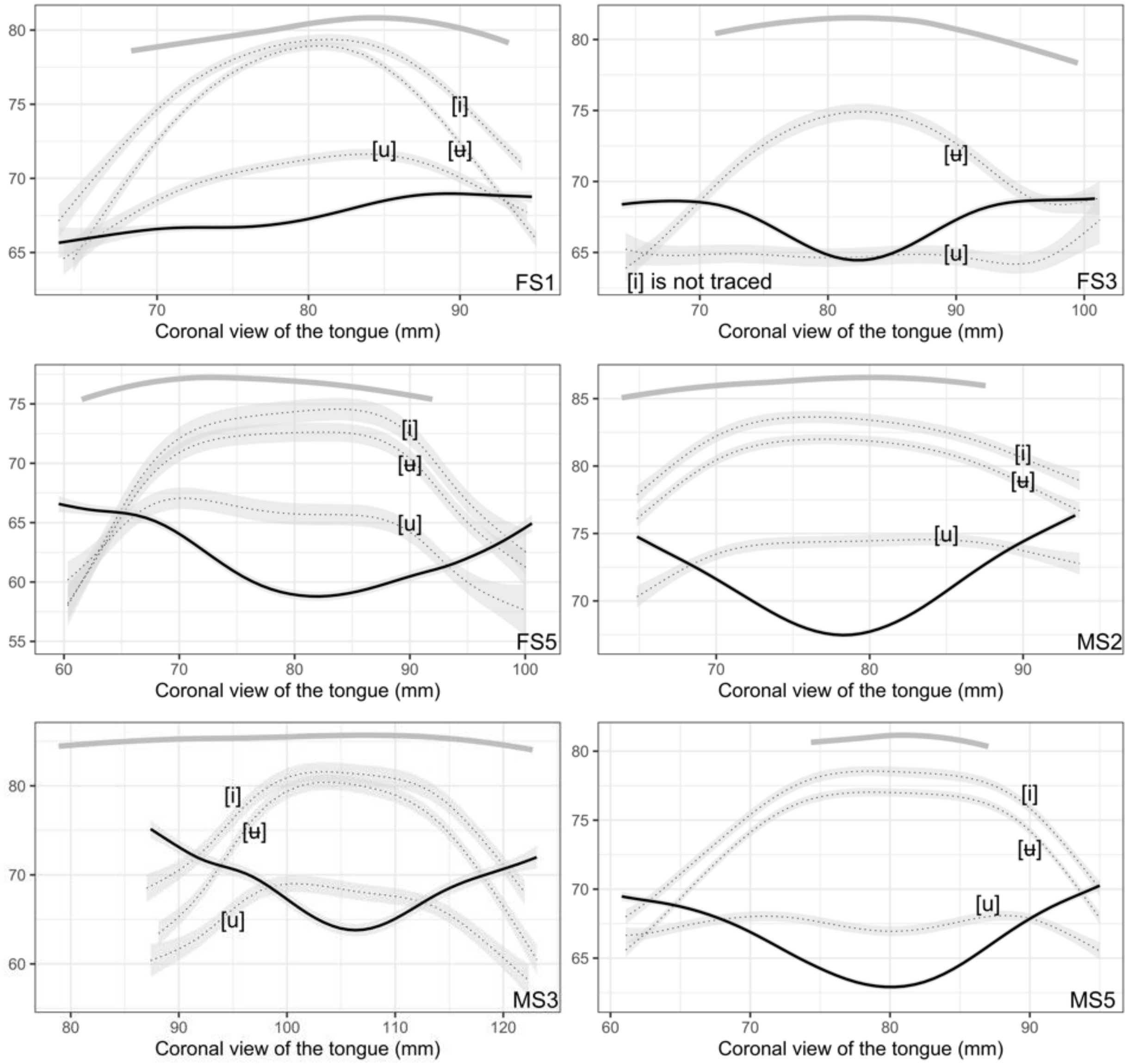


Figure 5.13. SS ANOVA splines of coronal ultrasound tongue contours extracted in x/y values (mm) at the nearest midpoint image of each segment. Dotted grey lines represent [i u ʉ] and solid black lines represent [z]. The thick grey lines represent the palatal traces and the grey areas on the splines represent 95% Bayesian confidence intervals. [i] is not traced for FS3.

5.2.3. Comparing [z] to [s]

This section examines the potential similarities and differences between the tongue configurations of fricative [s] and apical [z]. I examine these on mid-sagittal plane and then on coronal plane.

5.2.3.1. On mid-sagittal plane

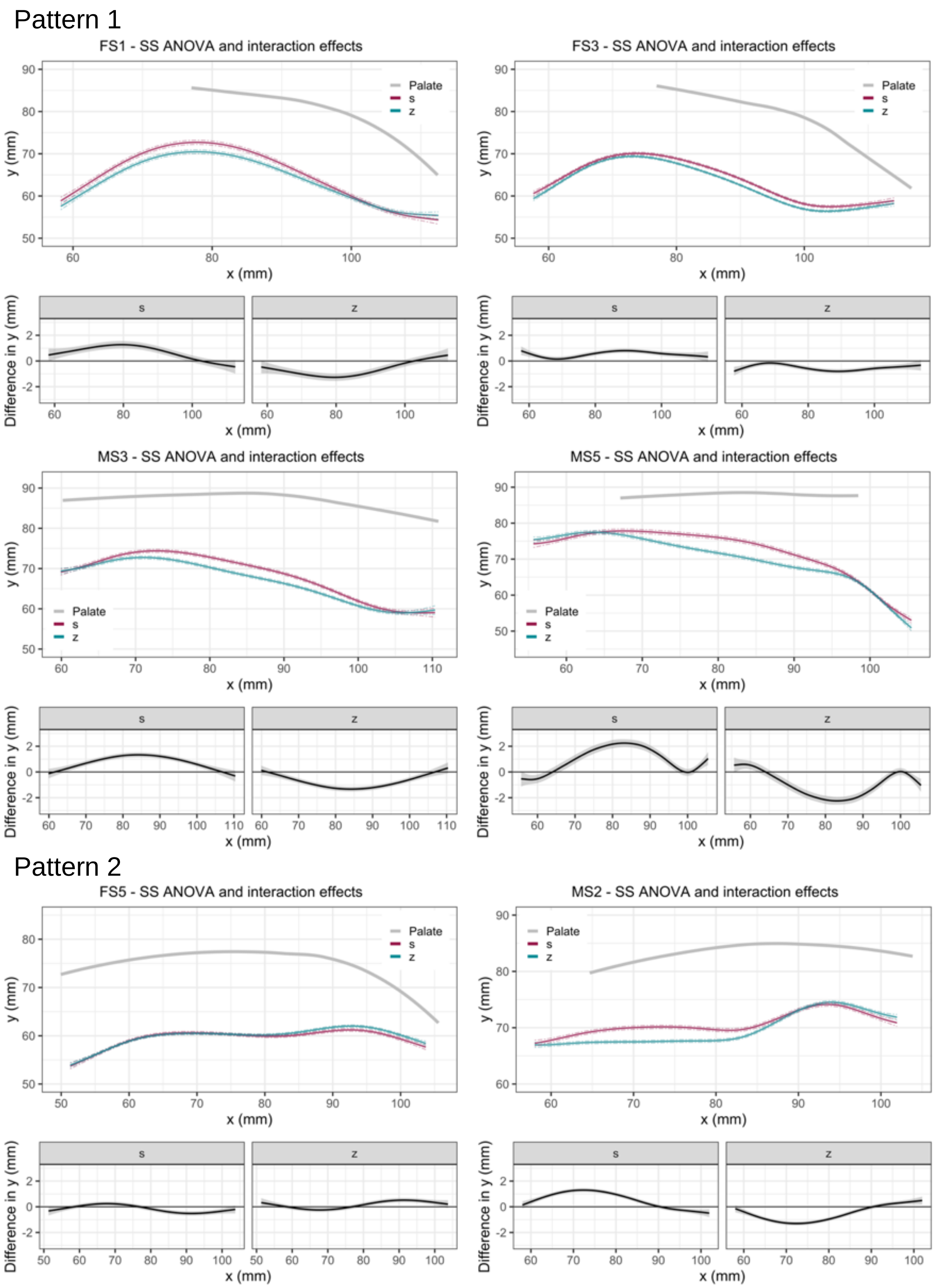


Figure 5.14. SS ANOVA splines for all occurrences of [s] and [z] in our data, and for each speaker (tongue front on the right and tongue root left). The figure also shows the interaction effects, calculated as differences in millimeters between [z] and [s] on y-axis.

The tongue contours of [z] and [s] show a striking similarity. This is true for both pattern 1 and pattern 2. For each speaker, when [z] has a raised tongue dorsum, [s] also has a raised tongue dorsum; and when [z] has a flat tongue dorsum, [s] also has a flat tongue dorsum. The similar lingual configuration of [z] and [s] can’t be attributed to the homorganicity with the preceding onsets since [z] displays the same tongue shape in the context of labials as well.

One noticeable difference between [s] and [z] is observed for all speakers except for FS3: the tongue dorsum in [z] is lower than in [s]. The maximum observable difference varies from less than 1mm (FS3) to nearly 2.5 mm (MS5). As mentioned in our acoustic study, most of the apical vowels [z] have frication noise on the first half of their duration, and the frication noise diminishes while the formant structure becomes clearer. The lowered tongue dorsum in [z] observed here may be responsible for this diminishing of the frication noise: As the tongue dorsum becomes lower, the narrowed air channel is widened, and frication noise diminishes.

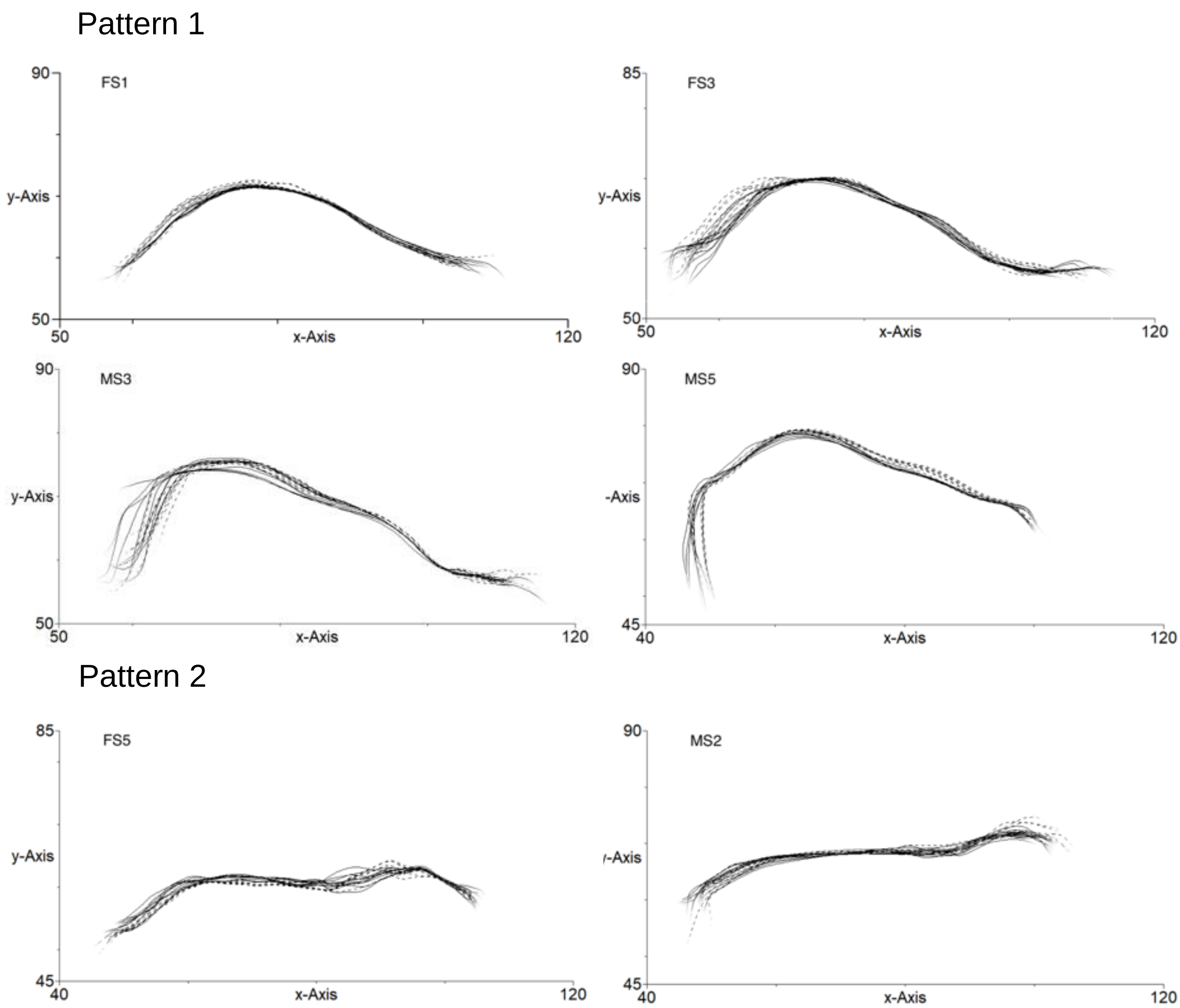


Figure 5.15. Raw mid-sagittal ultrasound tongue contour traces of the syllable [sz˧˥] (tongue front on the right and tongue root left). The figures contain traces from the nearest midpoint image of [s] to the nearest midpoint image of [z]. Dashed lines represent the traces of [s] and solid lines represent the traces of [z].

In order to further illustrate the similarity in tongue configuration between apical vowel [z] and the sibilant [s], raw ultrasound tongue contour traces of the syllable [sz] are shown in Figure 5.15. Albeit some slight differences and minor speaker-specific differences, a virtually identical tongue configuration is maintained across the entire syllable [sz] on mid-sagittal plane. The two segments [s] and [z] clearly share the same tongue shape and the same place of articulation.

5.2.3.2. On coronal plane

The SS ANOVA splines [z] reported in are generated from all occurrences of [s] and [z] in the data. Each speaker has one smoothing-spline figure containing [s z] splines and two figures showing the interaction effects. The interaction effect is reported as differences in millimetres on y-axis.

Except for FS1, a medial-grooved tongue shape for [z] is observed for all speakers: the laterals of the tongue are raised, and the medial part of the tongue is grooved, thus the medial part being much lower than the laterals. This medial-grooved tongue shape indicates the presence of a narrowed air channel for apical [z]. Considering the splines of [z] are from [pz pʰz mz nz tsz tsʰz sz] syllables, the medial-grooved tongue shape observed here cannot be attributed to the preceding consonant. Again, given that labial consonants [p pʰ m] and coronal nasal [n] do not imply medial-grooving of the tongue, the shape exhibited by [z] must be considered as inherent to this segment.

The tongue shape for [z] is strikingly similar to that of [s], although with varying degrees. There could be virtually no difference at all between the two segments as for FS3. There could also be a less than 1 mm difference as for FS1 and FS5. For FS5, the difference is larger at the laterals of the tongue, but the medial part of the tongue is identical. Recall from section 5.1.4 that the laterals were less reliable in the tracing procedure, it is thus difficult to determine whether the laterals in the splines of FS5 reflect the real tongue shapes. A larger difference is observed for MS2 and MS5. For these two speakers, the medial-grooving is deeper in [z] than in [s], with a difference varying from 1 mm to 2 mm. This medial-groove deepening could be related to the tongue dorsum lowering exhibited by these two speakers as shown in Figure 5.14. It indicates the enlargement of the narrowed air channel and could be related to the disappearing of the frication noise reported in the acoustic study. The speaker MS3 behaves in a unique way. The medial part of his tongue is deeper in [z] than in [s], akin to MS2 and MS5, but the laterals of the tongues show a reverse pattern. As already stated, it is difficult to know whether the laterals in this case are reliable.

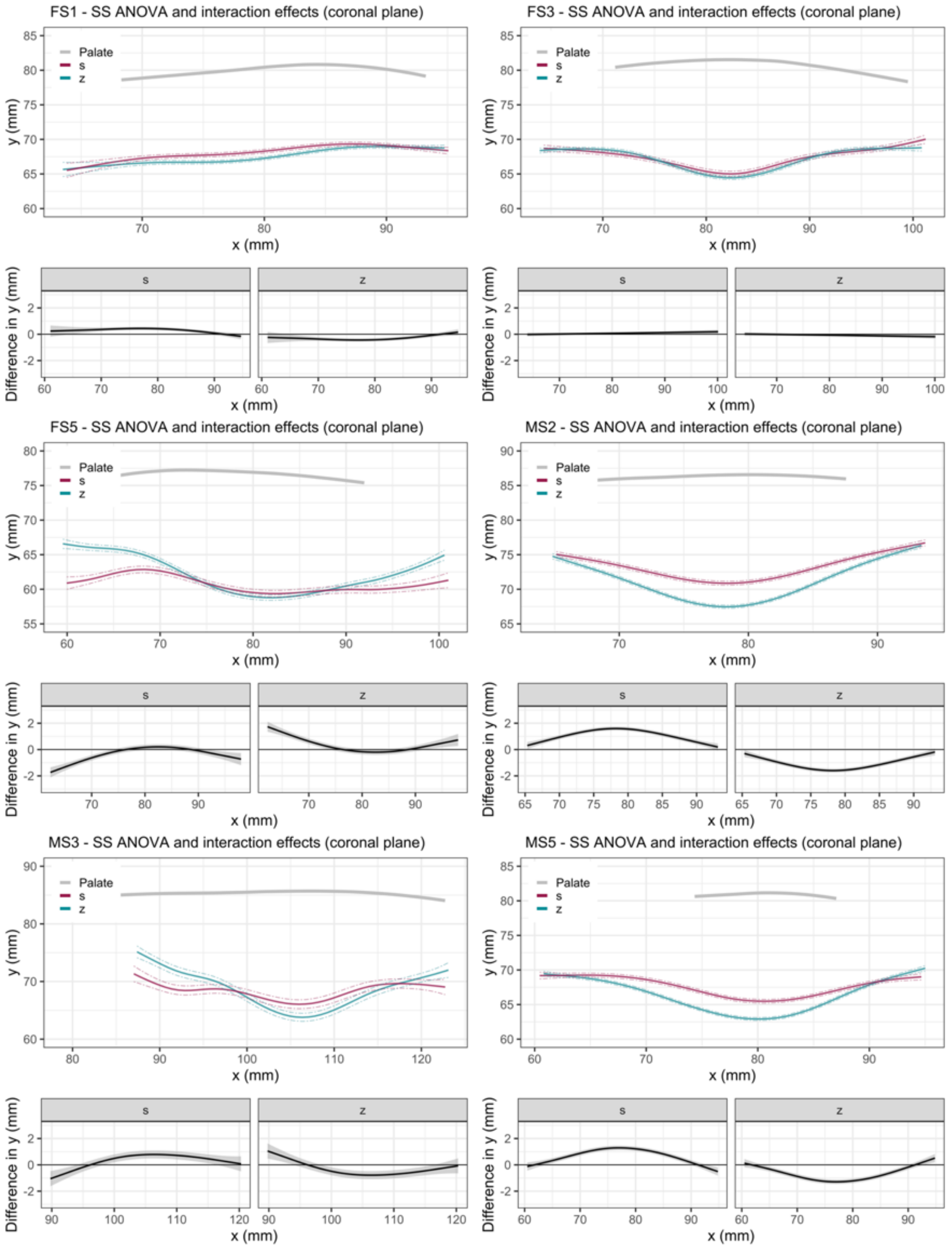


Figure 5.16. Coronal plane SS ANOVA splines with 95% Bayesian confidence intervals for all occurrences of [s] and [z] in our data, and for each speaker. Splines are presented with interaction effects with Bayesian confidence intervals.

In order to further illustrate the similarity in tongue configuration between [z] and [s], Figure 5.17 shows raw ultrasound tongue contour traces of the syllable [sz]. Again, from onset [s] to nucleus [z], the medial-grooved tongue shape does not change significantly (except for FS1 who has a different articulatory strategy). For MS2, this medial-grooving is deeper in [z] than in [s], and for MS3, it is wider in [z] than in [s]. For other speakers, the medial-grooving does not show observable changes of tongue shape. Clear enough [s] and [z] share the same gesture and display the same narrowed air channel, characteristic of fricative consonants.

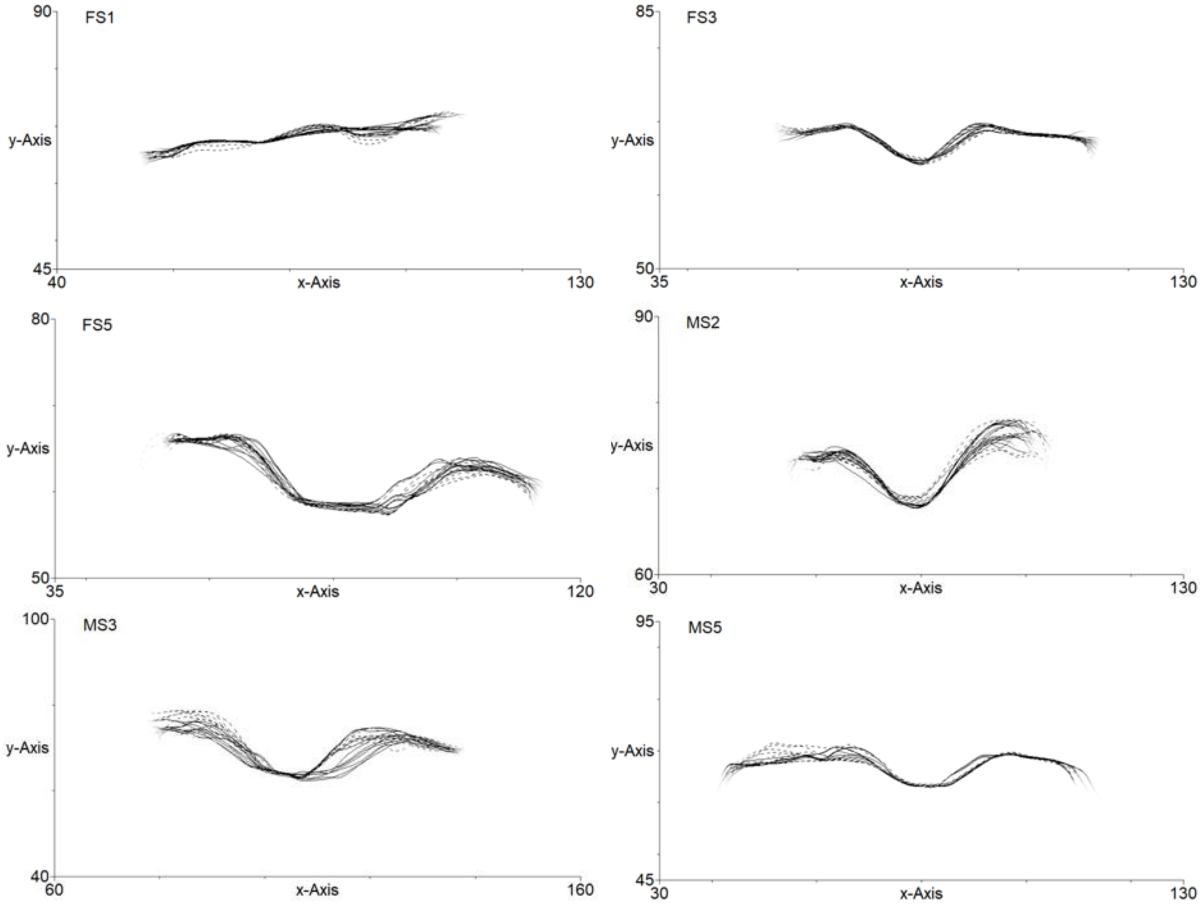


Figure 5.17. Raw coronal ultrasound tongue contour traces of the syllable [sz˥]. The figures contain traces from the nearest midpoint image of [s] to the nearest midpoint image of [z]. Dashed lines represent the traces of [s] and solid lines represent the traces of [z].

5.2.4. Articulation of [si sʉ su] compared to [sz]

The articulation of [z] is shown above to be virtually identical to that of the onset consonant [s]. This gesture for [z] is clearly different from vowel gestures. I further dwell on these differences by comparing the articulation of [si sʉ su] to that of [sz] on both mid-sagittal and coronal planes.

5.2.4.1. On mid-sagittal plane

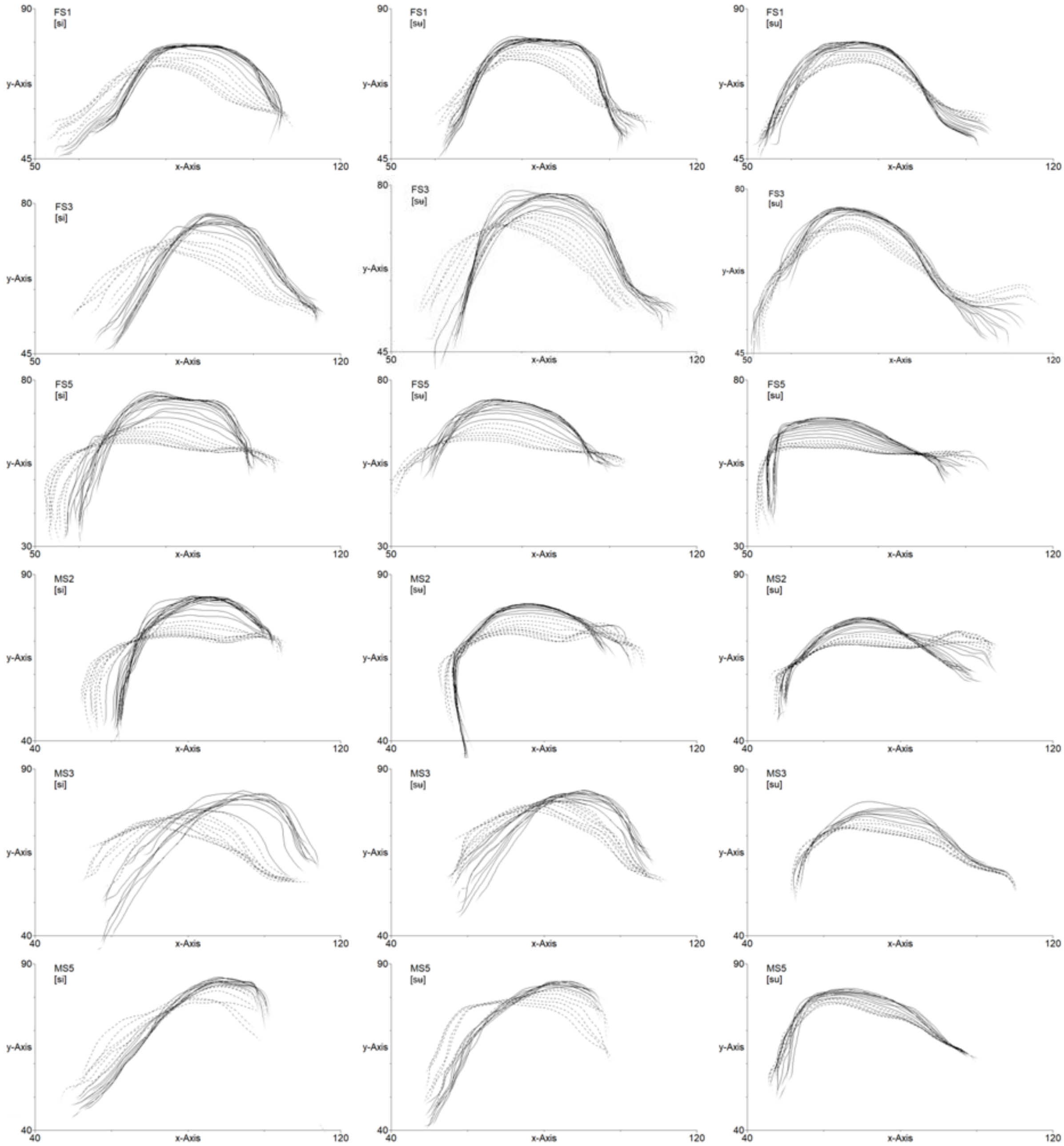


Figure 5.18. Raw mid-sagittal ultrasound tongue contour traces of [si sʉ su] (tongue front on the right and tongue root left). The images contain traces from the midpoint image of [s] to the midpoint image of nuclei [i ʉ u]. Dashed lines represent the traces of [s] and solid lines represent the traces of nuclei.

As shown in Figure 5.18, from onset [s] to nuclei [i ʉ u], the tongue forms high convex shapes which correspond to the three high vowels. The dorsum of the tongue fronts and rises for [i], rises for [ʉ] and rises and retracts for [u]. It is clear that these tongue movements are very different from what was observed in [sz] sequence presented in Figure 5.15, where there was virtually no observable movement from onset [s] to [z]. The absence of theses gestural changes in a [sz] syllable indicates that [z] does not have a high vowel gesture (despite the similarity in formant structure reported in section 4.2.3).

5.2.4.2. On coronal plane

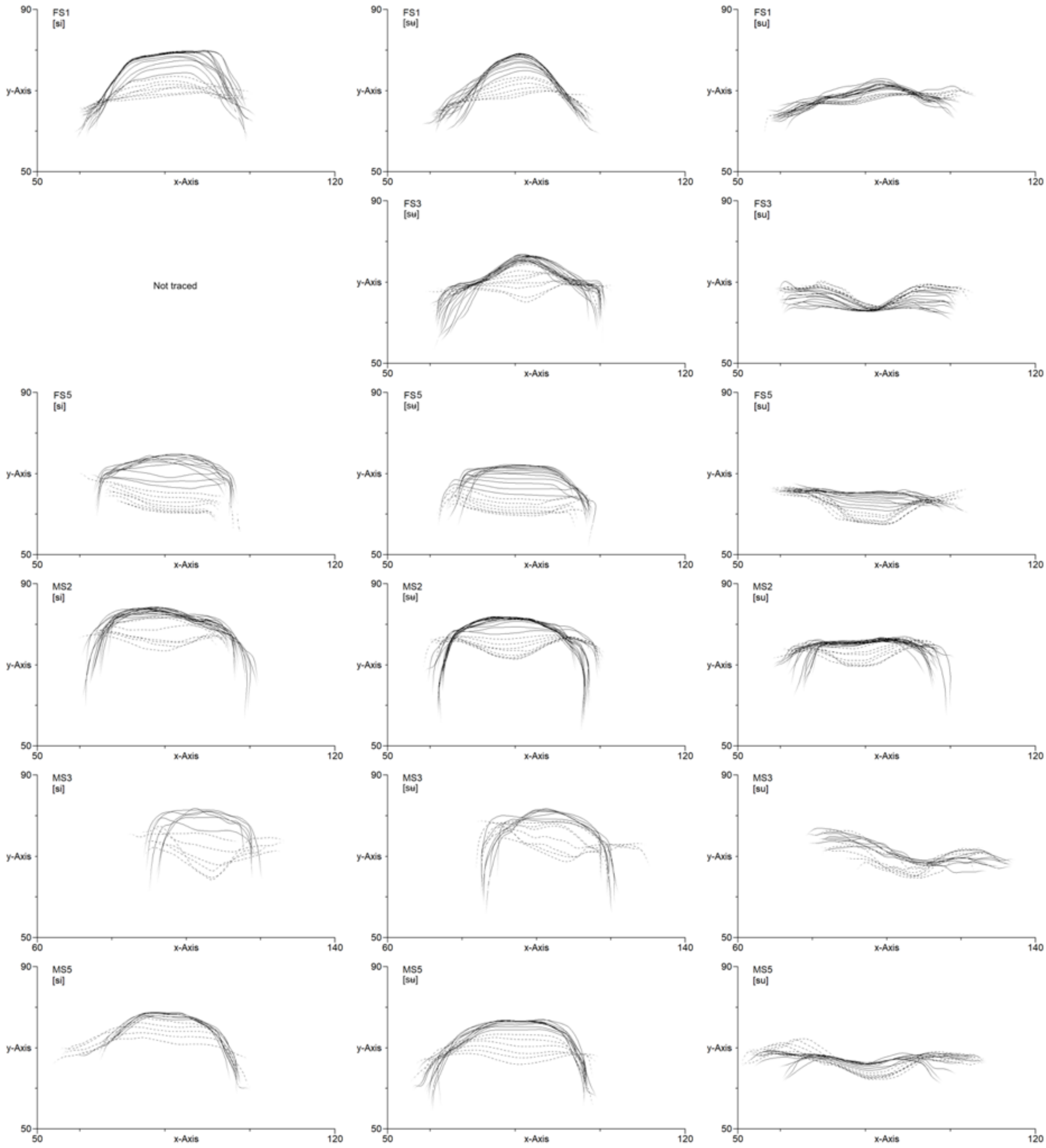


Figure 5.19. Raw coronal ultrasound tongue contour traces of [si sʉ su]. The images contain traces from the midpoint image of [s] to the midpoint image of nuclei [i ʉ u]. Dashed lines represent the traces of [s] and solid lines represent the traces of [i ʉ u].

Similar to what was reported in the previous section, the gestural changes from the onset consonant [s] to the nuclei [i ʉ u] are clearly observable in Figure 5.19.

For [si sʉ] syllables, the anterior of the tongue rises from [s] to [i ʉ] to form a high convex shape. For [su], the anterior of the tongue does not rise, but the medial-grooving disappears, which corresponds to the disappearing of the narrowed air channel for [s]. Compared to Figure 5.17, where the shape of the tongue remains virtually the same across the syllable [sz], the tongue shapes in [si sʉ su] syllables clearly belong to two different gestures: a fricative gesture for [s] and a vowel gesture for [i ʉ u]. These different configurations provide additional evidence that [z] is produced with different tongue gestures compared to genuine vowels.

5.2.5. Comparing [z] after labial and coronal onsets

Previous sections have shown that the tongue does not move significantly from onset [s] to nucleus [z] in a [sz] syllable. To determine whether this tongue configuration is inherent to [z] or whether it is a consequence of coarticulation with preceding sibilant [s], I compared the smoothing-splines of [z] after labial onsets to the smoothing-splines of [z] after coronal onsets (the interaction effects were also calculated). Results, presented in Figure 5.20 for the mid-sagittal plane, show that the tongue contours of [z] preceded by coronal consonants [s ts tsʰ n] and by labial consonants [p pʰ m] are virtually identical. The [s]-like tongue shape of [z] is thus observed when [z] is preceded by labials as well, suggesting that this tongue configuration is inherent to apical [z] in JHC.

The same comparisons were conducted on the coronal plane. The results are shown in Figure 5.21. Here too, the medial-grooved tongue shape of [z] is consistent regardless of the nature of the preceding consonant. The fact that [z] maintains the same tongue configuration whether preceded by sibilants or not mirrors the results obtained in the acoustic study (see section 4.2), with [z] displaying frication noise after both labial and coronal consonants. Taken together these results clearly show that, independent of the context, [z] systematically displays acoustic and articulatory characteristics of a fricative consonant.

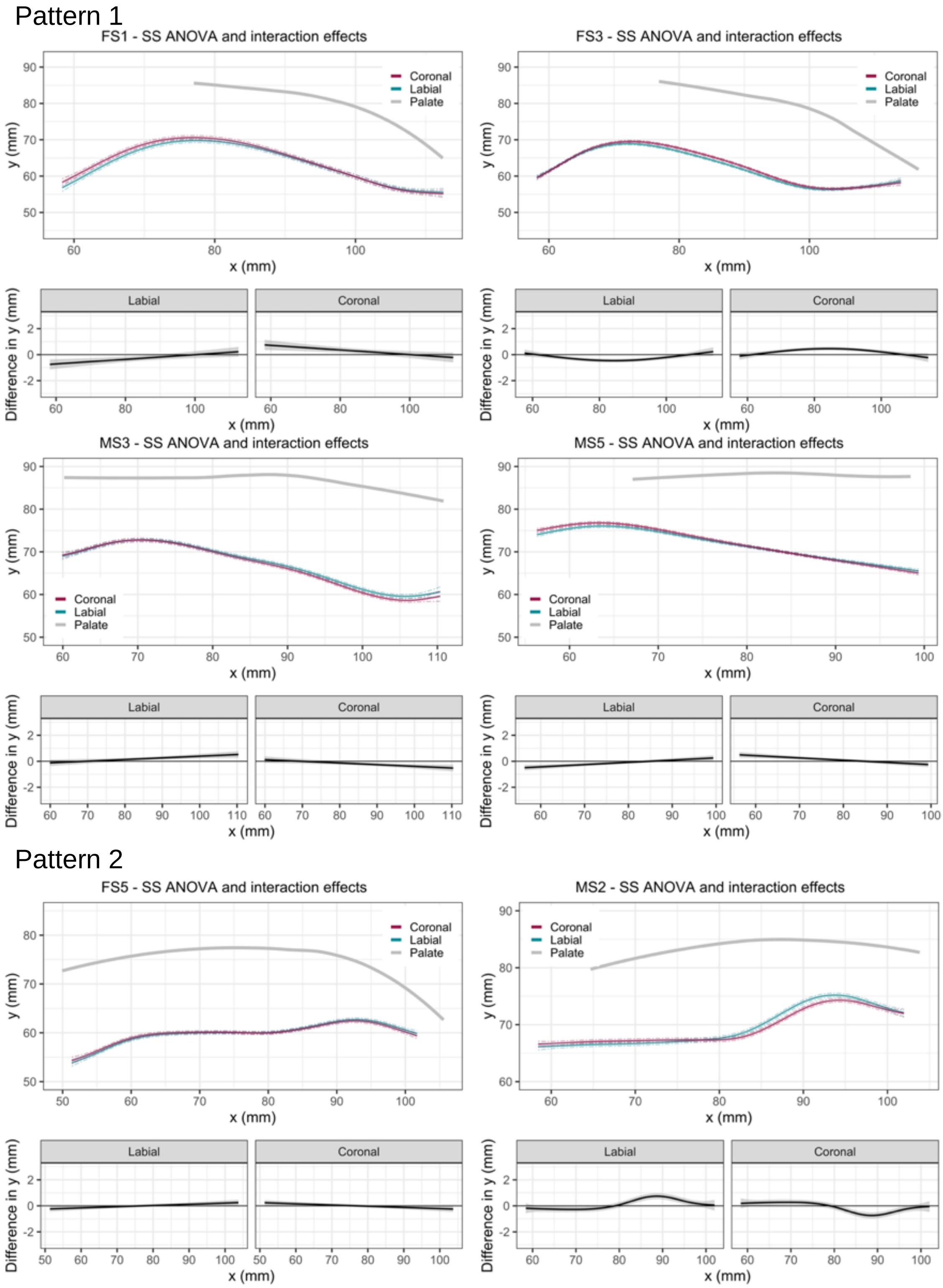


Figure 5.20. SS ANOVA splines with 95% Bayesian confidence intervals for [z] after labial onsets (blue) and after coronal onsets (red) (tongue front on the right and tongue root left). Splines are presented with interaction effects with Bayesian confidence intervals.

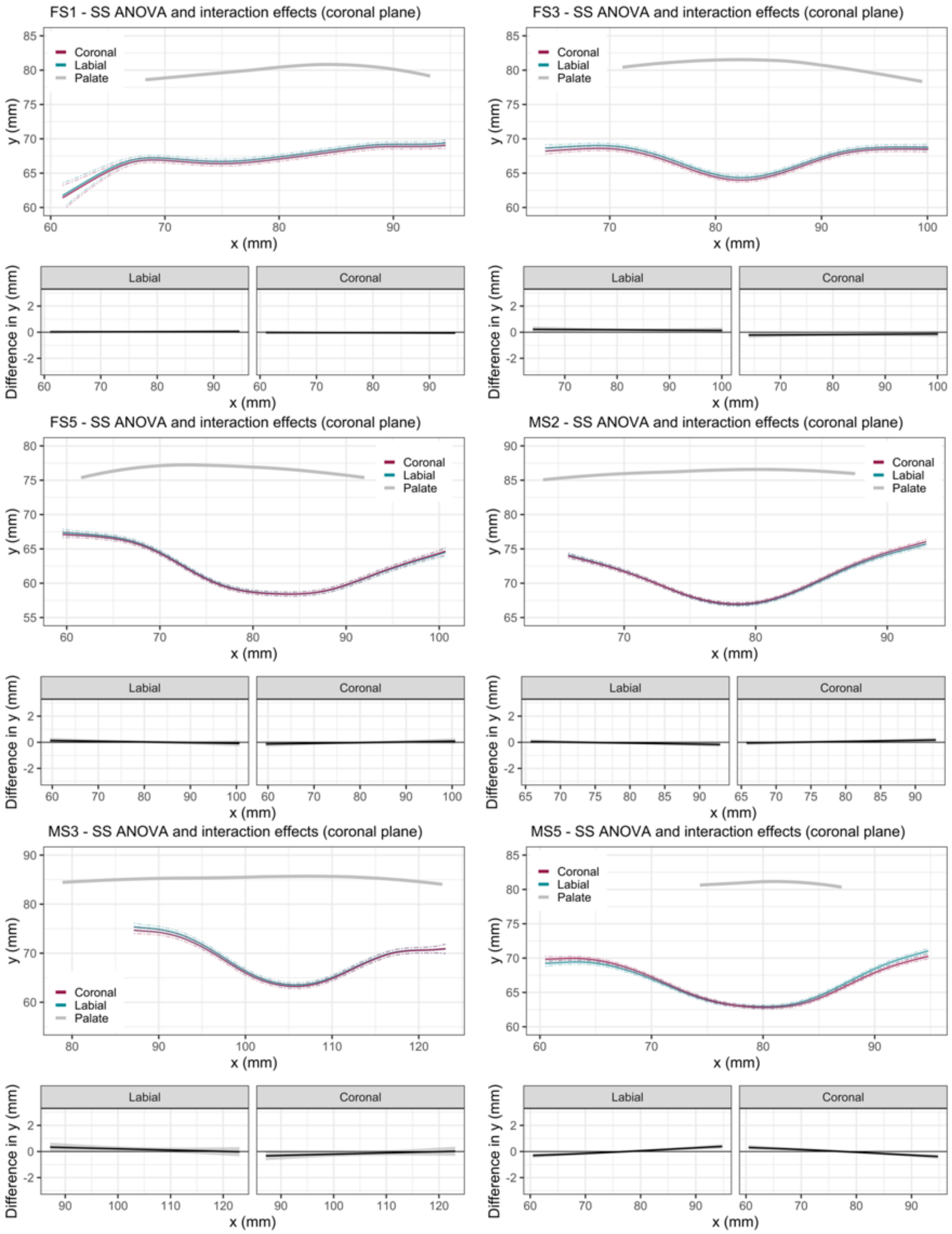


Figure 5.21. Coronal plane SS ANOVA splines with 95% Bayesian confidence intervals for [z] after labial onsets (blue) and after coronal onsets (red). Splines are presented with interaction effects with Bayesian confidence intervals.

5.2.6. Dynamic tongue shape evolution in [pz p^hz mz] syllables

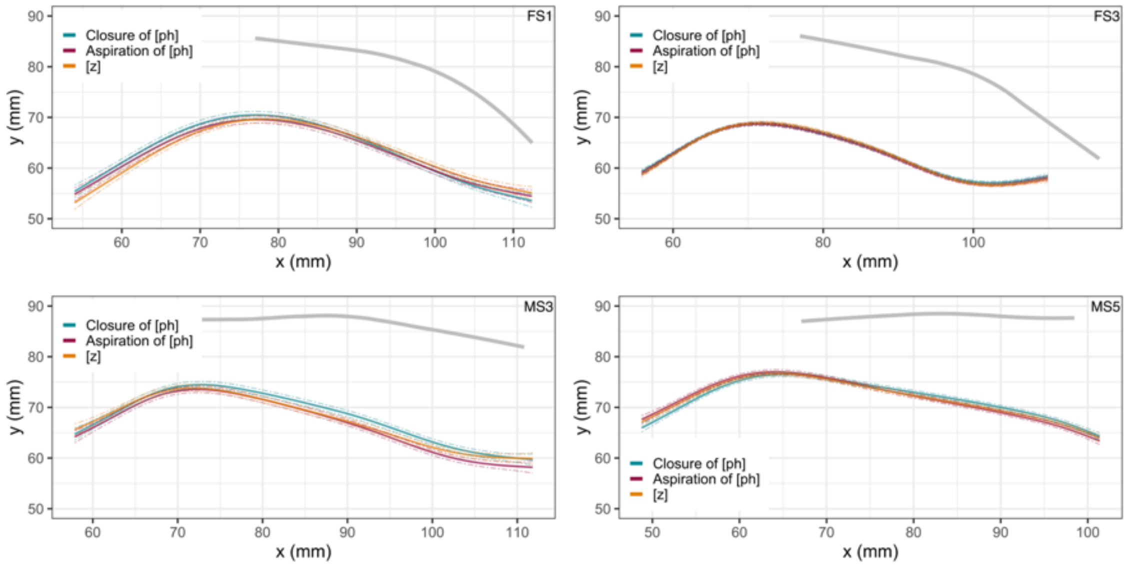
I have shown in the acoustic study (see section 4.2.6) that [p^h] patterns with sibilant consonants [s ts ts^h] rather than with the labial consonant [p]. This unusual patterning was explained by arguing that the gesture for [z] was already achieved during the release phase of [p^h]. In order to assess this explanation, the dynamic tongue shape evolution in [pz p^hz mz] syllables was examined. Figure 5.22 presents the results obtained for the form [p^hz]. This figure shows the tongue shapes from three different points: the nearest midpoint image of the closure phase of [p^h], the nearest midpoint image of the aspiration phase of [p^h] and the nearest midpoint image of nucleus [z]. Clearly, the gesture for [z] is achieved as early as during the closure phase of [p^h]. The tongue keeps this configuration during the entire [p^hz] syllable. This is shown by the fact that all confidence intervals overlap for all speakers (except for MS3 where the tongue dorsum seems to be higher during the closure of [p^h]).

Similar to [p^h], the fricative gesture of [z] is achieved during the onset consonant [p m] in [pz mz] syllables (See A.3 for SS ANOVA figures). When the apical vowel [z] is preceded by voiceless labial stops, the fricative consonant target is achieved earlier than the starting point of the voicing.

The coronal tongue shape evolution in the syllable [p^hz] was also examined. The nearest midpoint image of the closure phase of [p^h], the aspiration phase of [p^h] and the nucleus [z], generated in smoothing-splines, are shown in Figure 5.23. Except for FS1, the medial-grooving typical of [z] is achieved during the closure phase of the labial stop. This medial-grooved tongue shape is maintained during the release phase of [p^h] and continues into the nucleus [z]. The same pattern is also observed in [pz mz] (The SS ANOVA splines of these two syllables are reported in A.3).

The results reported in this section provide an explanation to why the release phase of [p^h] has the same acoustic characteristics as [s], as reported in the acoustic study (see section 4.2.6). The articulatory scenario is that the tongue reaches its target as early as the closure of [p^h] is formed. Then the tongue maintains this alveolar fricative gesture until [z]. The consequence of this is that during the release phase of [p^h], the air flux passes through the alveolar constriction, yielding frication noise that displays the same characteristics as fricative [s] (in terms of centre of gravity for instance).

Pattern 1



Pattern 2

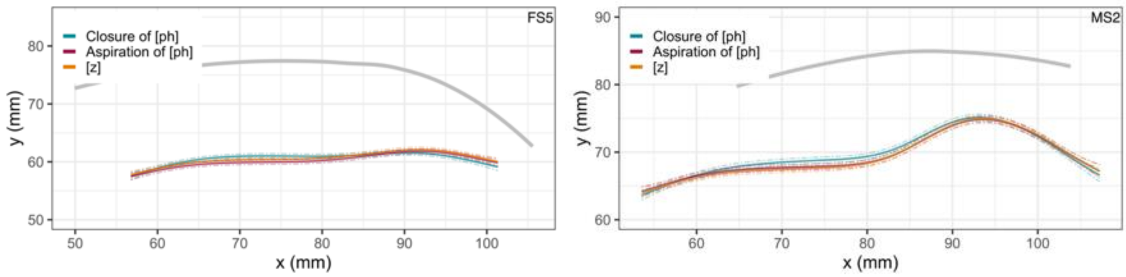


Figure 5.22. SS ANOVA splines of mid-sagittal ultrasound tongue contours, extracted in x/y values (mm) (tongue front on the right and tongue root on the left). Blue lines represent the nearest midpoint images of the closure of [pʰ], red lines represent the nearest midpoint images of the release phase of [pʰ], and yellow lines represent the nearest midpoint images of [z]. The splines are presented with 95% Bayesian confidence intervals. The thick grey lines represent the palatal traces.

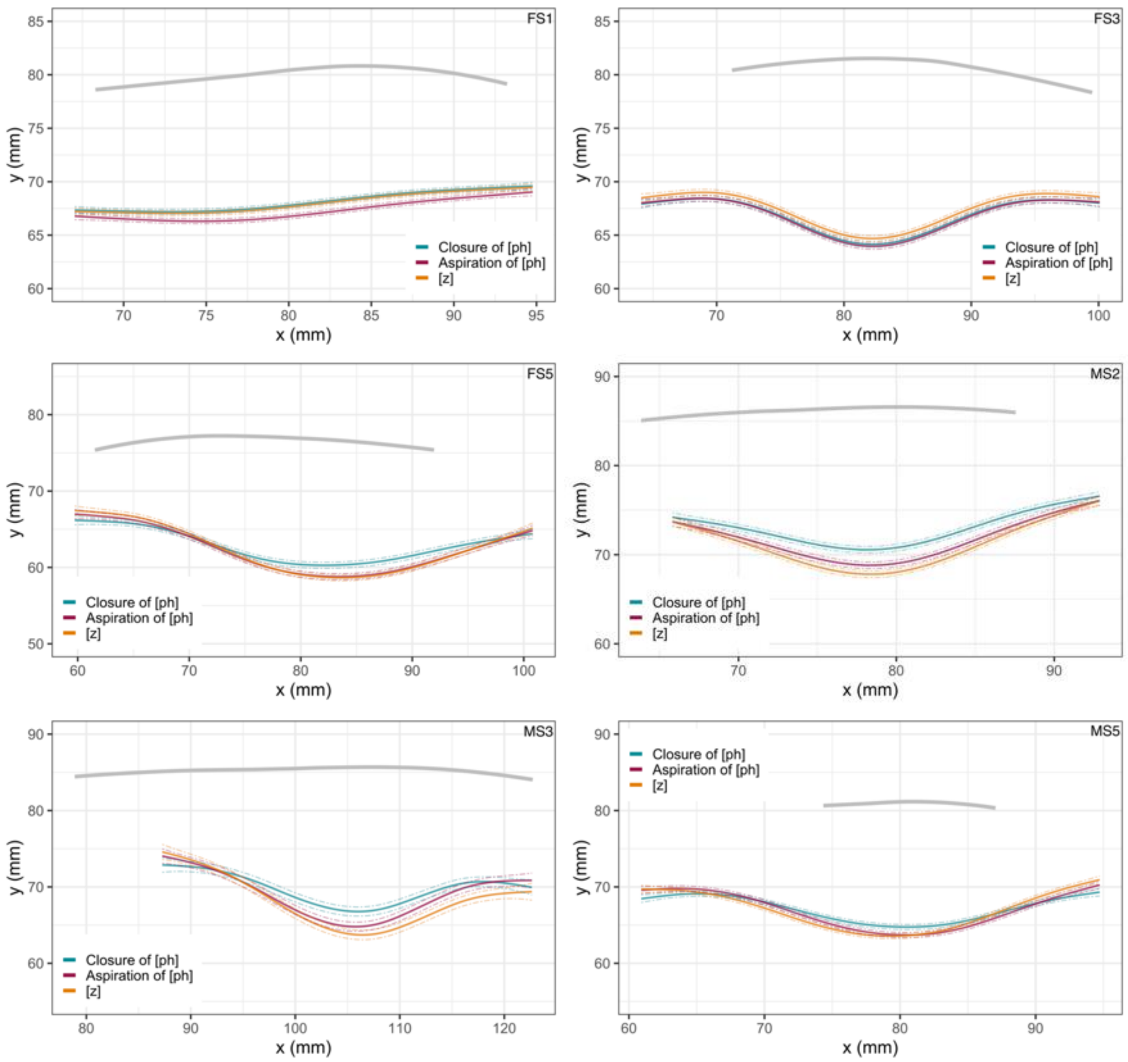


Figure 5.23. SS ANOVA splines of coronal ultrasound tongue contours extracted in x/y values (mm). Blue lines represent the nearest midpoint images of the closure of [pʰ], red lines represent the nearest midpoint images of the release phase of [pʰ], and yellow lines represent the nearest midpoint images of nucleus [z]. The splines are presented with 95% Bayesian confidence intervals. The thick grey lines represent the palatal traces.

5.3. Summary of the articulatory study

The results of this study presented evidence that the apical vowel [z] displays articulatory characteristics of a fricative consonant. The ultrasound data show that the tongue shape of [z] is very much identical to that of an alveolar sibilant [s]. This similarity was observed on both mid-sagittal and coronal planes. The medial-grooved tongue shape of [z] is particularly important, as it is a fundamental indicator of a narrowed air channel typical of a fricative gesture. The fricative-like tongue shape of [z] is observed regardless of the nature of the onset consonants (be them labials or sibilants). This is another important finding since it clearly shows that the tongue configuration of [z] is not a mere consequence of gestural overlap between homorganic sibilants and apical vowels. While the tongue configuration of [z] is similar to that of [s], it is very different from the configuration of high vowels. This difference is observed on the mid-sagittal plane, most notably regarding the arching of the dorsum typical of vowel articulation. It is even more evident on the coronal plane, with [z] displaying a medial-grooved configuration typical of fricatives. The fricative gesture of [z] is inherent to this segment as data from the labial contexts showed. When preceded by labial onsets, the gesture of [z] is achieved anticipatorily during the labial closure and maintains this configuration during the nucleus [z]. A consequence of this anticipatory gesture is that the release phase of [p^h] is realised with similar acoustic characteristics as the frication of [s].

It is important to note however that the tongue configuration of [z] is not constant and evolves during its time course. This dynamic evolution is observed on both the mid-sagittal plane and coronal plane. On the mid-sagittal plane, the tongue dorsum gets lower in [z] than in [s] for almost all speakers. And on the coronal plane, a deeper medial-grooving for [z] as compared to [s] is observed for some speakers. The medial-groove deepening and the tongue dorsum lowering are most probably related, and indicate the enlargement of the narrowed air channel that could be responsible for the disappearing of the frication noise reported in the acoustic study. A consequence of this is that [z] may display a hybrid configuration, being more fricative at the onset and more approximant towards the offset. This hybrid configuration will be discussed in section 6.1.

Chapter 6
Discussion and conclusion

This thesis presented an investigation into the nature of apical vowels in JHC. It contributes to the growing body of literature on this specific set of segments, a residue of 'dubious' sounds not clearly vowels but not totally consonants. The apical vowel in JHC proved to be a typologically interesting case study, as it displays two structural properties that make it different from the most studied variants in other Chinese languages: It is a separate phoneme and it occurs following a wider range of consonants, including non-sibilants.

The results of the investigations carried show that JHC apical vowel is best defined as a voiced alveolar fricative [z]. At the acoustic level, it exhibits abundant frication noise for the overwhelming majority of speakers recorded. This frication is systematically superposed on voicing. The presence of frication is independent of the nature of the preceding consonant, as it occurs when preceded by sibilants as well as by labials or nasals. The ultrasound data show that the tongue configuration of [z] resembles that of an alveolar sibilant [s] on both mid-sagittal and coronal planes. On the mid-sagittal plane, when speakers adopt different strategies for [s], they also adopt the same strategies for [z]; so that the two segments always display similar tongue configurations. Similarly, on the coronal plane, [z] displays a medial-grooved tongue shape, similar to the fricative [s]. This fricative-like tongue shape is also observed regardless of the nature of the onset consonants.

While having acoustic and articulatory characteristics of a fricative consonant, apical vowel [z] in JHC almost never displays frication throughout its entire duration. Indeed, frication tends to gradually disappear during the time course of the apical vowel. And similarly, the tongue dorsum gets lower and a deeper medial-grooving is observed compared to [s]. This dynamic evolution of [z] during its time course makes it often realised as a hybrid segment with the first part fricative-like and the second part more approximant-like. I argue that this hybrid configuration is a consequence of two interacting constraints: a structural one related to the distinctive status of [z] and the role it plays within syllable structure, and a physical one related to the incompatibility of voicing and frication.

6.1. The phonetic implementation of [z]: the trade-off between frication noise and voicing

The phonetic implementation of JHC apical vowel requires a compromise between voicing and frication. This is due to the incompatibility of two aerodynamic requirements: a rise in the intraoral pressure for frication and a certain transglottal pressure difference to maintain voicing. As Ohala (1983: 201) put it:

> For the sake of continued voicing, the oral pressure should be low, but for the sake of frication, the oral pressure should be high, that is, the difference between oral pressure and atmospheric pressure should be high enough to cause high air velocity through the consonantal constriction. Meeting both of these requirements simultaneously may be difficult. To the extent that the segment retains voicing it may be less of a fricative, and if it is a good fricative it runs the risk of being devoiced.

Considering Ohala's aerodynamic voicing constraint, JHC apical vowel, and voiced fricatives in general, should imply a dichotomy between a voiceless fricative and a voiced approximant. In many languages, the incompatibility between frication and voicing is achieved at the expense of voicing. This is the case in Tashlhiyt Berber where geminate fricatives tend to devoice (Ridouane, 2007), or in English (Smith, 1997) and Hungarian (Bárkányi & Kiss, 2010). In French variation of /ʁ/ leads to a continuum between unvoiced fricative and voiced approximant (Gendrot et al., 2015). In JHC the way the incompatibility is resolved varies depending on the articulatory strategies adopted by each speaker (see the next section for a discussion of these strategies). Interestingly the compromise is never achieved at the expense of voicing. Why should this be? The reason is that the JHC apical vowel /z̩/ is a tone-bearing unit. As such it has to be voiced throughout to carry the tone. The JHC apical vowel then, being obligatorily voiced throughout its duration, can hardly display strong frication noise throughout. This is all the more so that apical vowels, similar to other syllable nuclei, are particularly long (with a mean duration varying from 176 ms to 252 ms).

Why, a reasonable person asks, is frication maintained? Why not produce a voiced approximant? A plausible answer, grounded on both structural and acoustico-auditory properties, is that the JHC apical vowel /z̩/ needs to differentiate itself from the close central vowel /ʉ/. Since the close central vowel /ʉ/ has a formant structure which is similar to the apical vowel, frication noise, at least at the beginning of the apical vowel, serves an important perceptual cue to recover the sound. Analysing the JHC apical vowel as a voiced fricative makes it possible to explain the lost frication observed during the last half (the aerodynamic voicing constraint discussed above). Conversely, defining this apical vowel as an approximant makes it problematic to explain why frication is added including in the context of labial and nasal consonants.

I propose, tentatively, to generalise this analysis into other Chinese languages and show that the variability observed is not random but systematic by seeking the reasons for which some variants prefer a fricative-like type while others prefer an approximant-like type.

6.2. Accounting for the variability of [z] in Chinese languages

Four different analyses could be found in the literature concerning the nature of apical vowels in Chinese languages: apical vowel, fricative vowel, fricative consonant and approximant consonant. Section 2.2.2 presented a review of these variable analyses based on data from SC, HMC, QMC and SWC.

While the vowel analysis has not been supported by recent studies (Zee & Lee, 2001; Lee-Kim, 2014), the variability reported in SC still needs to be explained. The phonetic implementation of SC apical vowels could be either approximant-like (Zee & Lee, 2001; Lee-Kim, 2014) or fricative-like (Chao, 1968; Yu, 1999; Faytak, 2015; Faytak & Lin, 2015). While the approximant analysis for SC apical vowels has some consensus, the fricative-like implementation is not unreported (Chao, 1968; Yu, 1999; Faytak, 2015; Faytak & Lin, 2015). Given that there was no observable frication in SC apical vowels analysed by Zee & Lee (2001) and Lee-Kim (2014), the approximant analysis was adopted in these studies. There are two advantages for this according to Lee-Kim (2014: 279): approximants do not interfere with the articulation of the preceding sibilants, and additional cues to the sibilant places in the vocalic period further enlarge the perceptual distance between the three SC sibilants [sɹ̩ ʂɻ̩ ɕi]. This analysis, however, can't explain the fricative-like implementation reported by other studies (Chao, 1968; Yu, 1999; Faytak, 2015; Faytak & Lin, 2015). And it can't be generalised into other variants of Chinese languages where apical vowels are attested also after non-sibilants.

The apical vowel in HMC displays similar characteristics as [z] in JHC: (i) It occurs after labial consonants [p p^h m] and coronal sibilants [s ts ts^h], (ii) It displays frication noise regardless of the nature of onset consonants, (iii) This frication is rarely observed throughout the segment, and (iv) voicing is systematically maintained throughout. The frication noise is however only observed on the first 15–20% of the duration of the apical vowel. At the phonological level, HMC apical vowel is defined as a fricative vowel and analysed as a contextual variant of [i]. Frication is analysed by Kong et al. (2019) as a secondary enhancing feature. In other words, frication is not considered as a defining attribute of [z], as our analysis argues, but as a feature associated with vowels. It is not clear however which acoustic attribute of vowels is enhanced by adding frication, and why this frication is added following non-sibilants (see section 6.3). An enhancing feature generally targets the inherent acoustic parameter defining a speech sound and adds a gesture that increases its perceptual salience (e.g., Stevens et al., 1986; Stevens & Keyser, 1989; Diehl, 1991; Clements & Ridouane, 2006a). This is, for example, the case of the lip-rounding which can be added to the English palato-alveolar sibilant [ʃ] to accent the spectral prominence in the region of F3 (see Keyser & Stevens (2006) for many other examples).

The variability displayed by all apical vowels can be accounted for on the same perceptual grounds as for JHC variant: maximise the perceptual distance between [z] and other syllabic nuclei (namely close central vowels). Given that SC has no close central vowel in its phonemic inventory, the apical vowel does not need to differentiate itself from a segment that could have a similar formant structure. Thus, the frication noise is not as much needed to recover the apical vowel. As a result, frication noise could be 'sacrificed' to satisfy the aerodynamic voicing constraint (Ohala, 1983). Unlike in SC, apical vowels in HMC and SWC display characteristics of a voiced fricative consonant, with frication noise displayed in the majority of the data reported (Hou, 2009; Ling, 2009; Wan, 2014; Kong et al., 2019). As for JHC, the phonemic inventories of these two variants also contain a close central vowel or a central-back vowel: [ɘ̝] in SWC (Ling, 2009), and [ɯ] in HMC (Hou, 2009; Wan, 2014). Table 6.1 shows that these vowels and [z] have overlapping formant frequencies. This qualitative similarity can lead to perceptual confusion. Frication needs thus to be maintained to keep the perceptual distance between these syllabic nuclei.

Table 6.1. Comparison of formant values (Hz) between the apical vowels and the close central or close back vowels in SWC (Ling, 2009) and HMC (Wan, 2014).

			F1	s.d.	F2	s.d.	F3	s.d.
SWC	male	z	343.9	60.4	1179.3	89.1	2741.0	294.4
		ɘ̝	365.0	51.9	1094.2	91.7	2478.9	222.0
	female	z	377.7	50.0	1404.7	130.3	3387.6	254.0
		ɘ̝	428.4	83.5	1196.5	103.7	3103.7	208.0
HMC	male	z	352.5	40.3	1557.8	112.9	—	—
		ɯ	334.6	25.8	1213.9	140.3	—	—
	female	z	442.9	58.1	1696.4	145.7	—	—
		ɯ	412.1	71.5	1324.4	250.9	—	—

In sum, the variable realisations of apical [z] in Chinese languages can be accounted for on both structural and physical constraints. The compromise between frication noise and voicing is constrained by the vocalic inventory of the language and by the role [z] plays within this inventory. This trade-off between frication noise and the voicing can be achieved in two possible ways: an aerodynamic adjustment and/or a gestural adjustment.

6.3. Apical vowels: Gestural and aerodynamic adjustments

The present analysis argues that the mid-sagittal tongue shapes observed with ultrasound data in JHC and SC (Faytak & Lin, 2015) are not vowel gestures. Two mid-sagittal tongue shapes are reported in the articulation of the apical vowel in JHC. Pattern 1 has a raised tongue tip and a more raised tongue dorsum, thus the tongue dorsum is higher than the tongue tip. Pattern 2 has a raised tongue tip but the tongue dorsum is flat, so that the tongue dorsum is not higher than the tongue tip. Pattern 1 is observed in the articulation of SC apical vowel but not pattern 2 (Lee-Kim, 2014; Faytak & Lin, 2015). The fact of not observing pattern 2 tongue shape in SC could be purely accidental. Lee-Kim (2014) only recorded one speaker and Faytak & Lin (2015) recorded five speakers. It is thus possible that SC speakers also have pattern 2 tongue shape for apical vowel [z̩].

The raised tongue body observed in the mid-sagittal tongue contour of SC [z̩] could be related to the articulation of coronal consonants. As Stevens et al. (1986: 436) has shown '*a fronted tongue-body presumably provides a favourable posture from which the apico-alveolar constriction can be achieved*'. Figure 6.1 shows that the mid-sagittal tongue contour in [t] could correspond to the pattern 1 tongue shape in the present study of JHC and other studies on SC apical vowels (Lee-Kim, 2014; Faytak & Lin, 2015), and the tongue shape of [t̪] could correspond to the flat tongue dorsum and raised tongue tip (i.e., pattern 2) in the present study. The coronal consonants [t ts ts^h] in SC (Lee & Zee, 2003) and in JHC can be alveolar or denti-alveolar, with no phonemic distinction between the two. Considering that the only difference between [t] and [t̪] is the tongue shape, not the place of constriction, it could be argued that the different tongue shapes are two articulatory strategies used for the same alveolar consonants. This could be taken as additional evidence that the two patterns observed with ultrasound data for the JHC apical vowel are closely related to a coronal consonant articulation, and not to a vowel articulation.

Figure 6.1 Schematised versions of mid-sagittal sections showing positions of the tongue blade for two coronal consonants, [t] is [+anterior] and [ṭ] is [-anterior]. These images are reproduced from Figures 20.4 in Stevens et al. (1986: 433).

The coronal tongue shapes of [z] show that the trade-off between frication noise and voicing can be achieved by minor gestural adjustments. This minor gestural adjustment, shown in Figure 5.16 and Figure 5.17, corresponds to a deeper medial-grooving compared to the onset consonant [s]. This deeper medial-grooving is achieved gradually. That is to say, in a [sz] syllable, the coronal tongue shape is medial-grooved starting from the onset [s] and gradually deepens until reaching the deepest point during the second half of [z]. The deepened medial-grooving in [z] could be linked to a wider, deeper air channel. This expanded air channel could be responsible for a decrease in intra-oral pressure and thus a weaker frication noise (or in some cases, the absence of frication noise) compared to the consonant [s]. This air channel deepening was also observed by Faytak & Lin (2015) in SC apical vowel, and they argued in a similar vein that the cavity expansion is responsible for the absence of frication noise in some SC apical vowel tokens.

The absence of gestural adjustments in the productions of several speakers suggests that other non-gestural factors may be involved in the trade-off between frication and voicing. Aerodynamic adjustments may also be involved, akin to what has been reported for Lendu vowelless syllable. As shown in section 2.2.3.2, the Lendu vowelless syllable is highly similar to the SC apical vowels. The aerodynamic data (Demolin, 2002) on the Lendu [z] nucleus give an interesting pattern in the airflow adjustment (see Figure 2.9). The intra-oral pressure of the [sz] syllable in Lendu has two phases. The first phase is a rapid increase of pressure corresponding to the formation of the narrowed air channel of [s]. The second phase is the gradual decrease of the intra-oral pressure during the syllable nucleus [z]. The apical vowel [z] in JHC could arguably display the same aerodynamic pattern as in Lendu. In JHC or any other mentioned Chinese language, the syllable [sz] has exactly the same structure as the Lendu word *sz̀*: the onset is the coronal sibilant [s] and the nucleus is a [z]. The beginning of the apical vowel contains usually much frication, and the second half is usually less fricated. This observation matches the aerodynamic pattern described above: The frication noise observed at the beginning of [z] is a consequence of the high intra-oral pressure, and as this pressure diminishes gradually, frication noise diminishes gradually as a consequence. In other words, the intra-oral pressure during the second half of [z] could not be high enough to generate and maintain frication noise.

This is not the only aerodynamic adjustment which could lead to a decrease in the amount of frication noise during the time course of [z]. Other possible aerodynamic adjustments include the pharyngeal expansion which decreases the intra-oral pressure by increasing the cross-sectional area of the vocal tract. Similarly, the leakage through velopharyngeal port could also lower the intra-oral pressure.

6.4. Apical vowel: finding its place in the vowel-consonant continuum

The three discussed categories in section 2.1.1.4 (i.e. life vs. non-life, animal vs. plant, male vs. female) were all considered as dichotomic prior to the modern scientific research. But this dichotomic relationship is challenged, given the recent scientific advancement. The vowel-consonant dichotomy, based on the same complex physical world, could not elude the same challenge. And it has been challenged since at least Rousselot (1897):

> Le tort que l'on a, c'est de chercher une différence caractéristique entre les deux portions d'une série *naturelle* dont les extrêmes seuls sont nettement séparés. Il serait très facile de définir la voyelle et la consonne si le type de l'une était simplement l'*a*, et celui de l'autre, le *b* ; mais la distinction, qui nous apparaît très nette aux deux bouts de la série, tend à s'effacer dans la région moyenne, par exemple entre *i* et *j*.
>
> (Rousselot, 1897: 634)[27]

The vowel-consonant dichotomy versus the vowel-consonant continuum meets the widely accepted view of the difference between phonetics and phonology. That is, phonetics is inherently gradient and continuous, while cognitive organisations (phonology) are categorical (Chomsky & Halle, 1968; Clements, 1992). Although the categorical phonology and gradient phonetics is another imperfect dichotomy (Chitoran & Cohn, 2009), it is certain that phonetics, the study of the physical properties of speech sound, must be inherently gradient and continuous.

[27] [The mistake we have is to look for a characteristic difference between the two portions of a natural series whose extremes alone are clearly separated. It would be very easy to define vowel and consonant if the type of one was simply the *a*, and that of the other, the *b*; but the distinction, which appears very clear to us at both ends of the series, tends to fade in the middle region, for example between *i* and *j*.]

Seemingly, there exists no clear-cut boundary between the vowel category and the consonant category, similar to any categorisation based on complex physical phenomena. Between the 'most vocalic sounds', such as [a e], to the 'most consonantal sounds', such as [p k], there is a 'grey area' where sonorants [l r], approximants [j w] and even voiced fricatives [z ʐ] can be found. The speech sounds, when considered as physical entities, form a continuum on both acoustic and articulatory plane, with the 'most vocalic sounds' on one side and the 'most consonantal sounds' on the other side. The middle ground is occupied by approximants, sonorants, and voiced fricatives.

The multiple terminologies used to define speech sounds show a crucial point in the categorisation of vowels and consonants. The phonetic shape of a sound and its phonological function could be different and should be studied with different criteria. A phonetic vowel may have the phonological function of a consonant and a phonetic consonant may have the phonological function of a vowel. The semi-vowel/semi-consonant/glide is in the former case, and the apical vowel in Chinese languages is in the latter case. Following this point of view, the phonetic categories and the phonological categories can have different specificities: the vowel-consonant ensemble in phonetics may form a continuum, while in phonology, the functions of speech sounds may be categorical.

The analysis advocated in this study, one may object, implies that a fricative consonant can be a syllable nucleus in JHC. This is unusual in Chinese languages, or even unattested (as in SC for example where only vowels can be syllable nuclei). The situation in JHC is different. In addition to /z/ three other consonants can be syllable nuclei and thus tone-bearing units: /m n v̩/. Although as shown in section 3.2.5 these consonantal TBUs are not equally productive in the lexis, it is still a fact that they function as syllable nuclei. This analysis thus implies that the segments in JHC that can serve as TBUs are not only vowels, but also nasal consonants and voiced fricatives. One exception seems to be the velar nasal /ŋ/, as there is no lexical item with this velar nasal functioning as a syllable nucleus in JHC. Note, however, that if interjections are taken into account all nasals can be syllabic consonants including velar nasals: for example, [hŋ˦˨] 'showing contempt', [ŋ˧] 'yes', [ŋ˨˦] 'showing doubt'.

JHC syllable nuclei share common features: they are all voiced, and they can all be maintained for a certain duration. These segments can thus be considered a natural class that can be defined using the features [+continuant, +voice], at the exclusion of all other JHC segments. This specification is straightforward, and it is valid for all consulted Chinese languages. As Table 6.2 shows, there are seven possible syllabic consonants in Chinese languages: [z ʐ v m n ŋ l], and they can all be considered as [+continuant, +voice].

Table 6.2. Syllabic consonants in Chinese languages (interjections are included when the description is available). The abbreviation HFZ stands for *Hànyǔ Fāngyīn Zìhuì* "*Phonetic dictionary of Chinese dialects*" (Department of Chinese of Pekin University, 1989). The symbols [z ʐ zʷ] correspond to the traditional apical vowel symbols [ɿ ʅ ʮ], respectively.

Chinese languages	Syllabic consonants		References
	Lexical	Interjection	
JHC	z v m n	ŋ	Zhao (2003), Current study
Shexian-Hui	z v m n	–	Hirata (1998)
Tunxi-Hui	z v m n	–	Hirata (1998)
Xiuning-Hui	z m n	–	Hirata (1998)
Yixian-Hui	z zʷ m n	–	Hirata (1998)
Qimen-Hui	z m n ŋ	–	Hirata (1998)
Wuyuan-Hui	z v m n	–	Hirata (1998)
Standard	z ʐ	m n ŋ	Duanmu (2007)
Hefei-Mandarin	z ʐ m	–	Wu (1995)
Wuhan-Mandarin	z m n	–	HFZ
Tianjin-Mandarin	z ʐ	–	Li *et al.*, (2019)
Chengdu-Mandarin	z ʋ	–	HFZ
Yangzhou-Wu	z m	–	HFZ
Suzhou-Wu	z zʷ l m n ŋ	–	Yuan (1983), HFZ
Wenzhou-Wu	z m ŋ	–	HFZ
Shanghai-Wu	z v m ŋ	–	Zhu (1995: 21, 2004), Chen & Gussenhoven (2015)
Xiangxiang-Xiang	z m n ŋ	–	Zeng (2019)
Changsha-Xiang	z m n	–	Yuan (1983), HFZ
Shuangfeng-Xiang	z ʐ m n	–	HFZ
Nanchang-Gan	z m n ŋ	–	HFZ
Meijiang-Hakka	z v (ʋ) m ŋ	–	Yuan (1983), HFZ, Lee & Zee (2009)
Guangzhou-Yue	m ŋ	–	Yuan (1983), HFZ
Xiamen-Min	m ŋ	–	Yuan (1983), HFZ
Chaozhou-Min	z m ŋ	–	HFZ
Fuzhou-Min	m n ŋ	–	Yuan (1983), HFZ

It is interesting to see that the syllabic consonant [v] is sometimes described as a syllabic approximant [ʋ]. This hesitating notation between [v] and [ʋ] is identical to what was observed in the literature concerning the apical vowel [z]. The reported trade-off between frication noise and voicing in the apical vowel could also be applied for the [v ʋ] case, where [v] would correspond to the fricative-like phonetic implementation and [ʋ] would correspond to the approximant-like implementation.

The existence of syllabic [v] in other Chinese languages, notably in the Hui group Chinese languages, shows that syllabic fricatives, while unusual, are not limited to JHC.

The existence of syllabic fricatives is not typologically uncommon. The vowel-consonant continuum shown in Figure 6.2 illustrates how languages may differ depending on their nuclei inventory. The rectangular represents the speech sounds in given languages, the gradient colour represents the continuum from the most vocalic sounds to the most consonantal sounds, passing by the middle ground with sonorants [l r] and [n m] among others. Each language contains a certain amount of sounds in its phonological system that can serve as the syllable nucleus, represented by the ellipses.

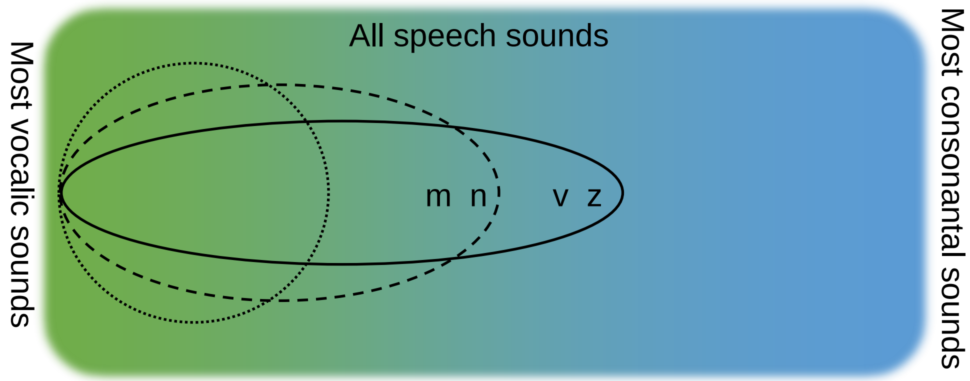


Figure 6.2. Schematic representation of the vowel-consonant continuum. The rectangular represents all possible speech sounds, the gradient colour represents the continuum from the most vocalic sounds to the most consonantal sounds. The dotted ellipse represents the nuclei of French, the dashed ellipse represents English, and the solid ellipse represents JHC. The position of the phonetic symbols is only informative.

Needless to precise that different languages have different nucleic inventory. In French for example only vowels can serve as syllabic nuclei, while other languages allow sonorants in addition to vowels (e.g., /l/ in English, [bɒ.tl] 'bottle'; /r/ in Czech, [br.no] 'Brno'; /n/ in German, [ha.bn] 'to have'). The difference between the French ellipse and the English ellipse does not necessarily mean that the English nucleic inventory is larger than the French one; it only implies that English allows less vocalic sounds in the nucleus position. Following the same logic, the JHC ellipse in Figure 6.2 goes even further into the middle ground, where voiced fricatives such as [z v] are attested. Adopting this analysis, the JHC nucleic inventory is nothing but one step further (than English, for example) into the more consonantal sounds. The case of Berber language (Dell & Elmedlaoui, 1988; Ridouane, 2008) is yet another example where a nucleic inventory even extends into the most consonantal sounds.

One feature shared by all these languages is that all the ellipses necessarily include the most vocalic sounds of their phonemic inventory. As shown by Greenberg (1962), the existence of friction continuant nuclei implies the existence of frictionless continuant nuclei, and the existence of frictionless continuant nuclei implies the existence of vowel nuclei (in his words, vocoids). This is the case for, a priori, all languages. A nucleic inventory of a given language can thus never be formed with only the most vocalic sounds and the most consonantal sounds, but without the middle grounds sounds. JHC makes no exception to this typological generalisation.

6.5. Final remarks and future directions

Future studies will have to complement the present study in order to evaluate the proposed analyses and to increase our understanding of the nature of apical vowels in JHC in particular and in Chinese languages in general. One clear avenue for future studies concerns the aerodynamic characteristics of these segments. To date, there is no study on Chinese apical vowels based on aerodynamic measurements. I have assumed in this study that JHC apical vowel may have the same aerodynamic behaviour as the Lendu vowelless syllable, but this lacks data to back it up.

Similarly, frication noise of the apical vowel is analysed as an important perceptual cue in JHC. This analysis also needs to be confirmed with perception tests. Related to this is the assumption that frication noise may serve for some Chinese languages as an enhancing feature. A possible scenario could be that this frication was first introduced as an enhancing attribute to maximise the contrast between sibilants. Depending on the phonemic inventory this attribute took on a distinctive function (namely in Chinese languages where /z/ occurs not only following sibilants). This is clearly speculative although cases where enhancement gestures become primary (when the defining gestures are weakened or lost) have already been reported in literature. Clements & Ridouane (2006b) for example, based on the interpretation of Mpiranya (1997), argue that the distinction between upper and lower vowels in many Bantu languages has disappeared and assibilation took the distinctive role. Assibilation, which was an enhancing gesture in the historical development of these vowels, preserves the distinction between words with earlier upper high vowels and those with earlier lower high vowels.

JHC has only one apical vowel /z̩/, while SC has two apical vowels [z ʐ]. The variability observed in SC [z] could be explained by the trade-off and the gestural-aerodynamic adjustments discussed above. Compared to [ɹ̩] or [ɹ̩̈] analysis, the /z̩/ analysis is more convenient since it does not introduce a unique sound which is presumably unattested in any other language (Lee-Kim, 2014). A comparison between apical vowels in JHC and in SC (ideally in other Chinese languages as well), using the same experimental setup, is bound to provide some important insight that may allow for better accounting for the variability displayed by these segments.

An interesting aspect concerning SC, as reported by Lee-Kim (2014: 277), is that the retroflex apical [ɻ̩] is an independent phoneme which can occur both in the onset and the nucleus positions. Nothing is known about potential phonetic differences between these two. Following the trade-off analysis presented in section 6.1, the phonetic difference, if any, could be explained depending on its position within the word. A fricative-like implementation may be realised when it functions as the onset of a syllable, while an approximant-like implementation may be realised when it functions as the nucleus of a syllable.

Describing Chinese languages using the IPA symbols [z ʐ] is more informative since it can be directly related to their phonetic implementation, explaining the possible presence of frication noise and the [s ʂ] like tongue shapes. Karlgren (1915) and Yuan (1983: 7) present four different apical vowels in Chinese languages [ɿ ʅ ʮ ʯ]. Following the syllabic fricative analysis advocated here, these symbols may be replaced by [z ʐ zʷ ʐʷ]. There is no available data on the labialised apical vowels, and they are much rarer compared to the non-labialised versions. Yuan (1983: 34, 35, 63) only presented very limited examples of the rounded apical vowels in Huaxian-Mandarin[28], Xianyang-Mandarin[29], Yingcheng-Mandarin[30], Macheng-Mandarin[31] and Suzhou-Wu Chinese. In these examples, the rounded apical vowels only occur after homorganic coronal sibilants [s z ts tsʰ ʂ tʂ tʂʰ] and can also occur in syllables without onset. How do the rounded apical vowels behave in a phonological system and what are their phonetic characteristics? How does the labiality interact with the frication noise (if any)? Current and future work should clearly continue to answer these questions (and possibly others not raised here, such as why apical vowels are concentrated in Sino-Tibetan).

[28] The city of Huaxian 华县 is situated in Shaanxi Province 陕西省.
[29] The city of Xianyang 咸阳 is situated in Shaanxi Province 陕西省.
[30] The city of Yingcheng 应城 is situated in Hubei Province 湖北省.
[31] The city of Macheng 麻城 is situated in Hubei Province 湖北省.

References

Anhui Statistical Bureau. 2019. *Anhui Statistical Yearbook – 2019*. Beijing: China Statistics Press.

Articulate Instruments Ltd. 2008. *Ultrasound Stabilisation Headset User's Manual: Revision 1.5*. Edinburgh, UK: Articulate Instruments Ltd.

Articulate Instruments Ltd. 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Bárkányi, Zsuzsanna, and Zoltán Kiss. 2010. A phonetic approach to the phonology of v: A case study from Hungarian and Slovak. In Fuchs, Susanne, Martine Toda & Marzena Żygis (eds.), *Turbulent sounds: An interdisciplinary guide*, vol. 21, 103–142. Berlin & New York: Walter de Gruyter.

Baron, Stephen. 1974. On the tip of many tongues: Apical vowels across Sino-Tibetan. *7th International Conference on Sino-Tibetan Language and Linguistic Studies.* Atlanta, US.

Bell, Alan. 1978. Syllabic consonants. In Joseph Greenberg (eds.), *Universals of human language*, vol. 2, 153–201.

Björsten, Sven & Olle Engstrand. 1999. Swedish 'damped' /i/ and /y/: experimental and typological observations. *14th International Congress of Phonetic Sciences (ICPhS XIV)*, San Francisco, USA.

Bloomfield, Leonard. 1923. *Language*. London: Ruskin House.

Boersma, Paul & David Weenink. 2018. PRAAT: Doing phonetics by computer. Version 6.0.43, http://www.praat.org (accessed November 2018).

Bombien, Lasse. 2006. Voicing alterations in Icelandic sonorants: a photoglottographic and acoustic analysis. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 37, 63–82.

Chan, S. T. H. 1970. Natural sex reversal in vertebrates. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 259(828), 59–71.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le Maître phonétique* 45, 24–27.

Chao, Yuen-Ren. 1961. *Mandarin primer: An intensive course in spoken Chinese*. Oxford: Harvard University Press.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California

Chao, Yuen-Ren. 1980. *Yǔyán Wèntí* [On Languages], 2nd edn. Beijing: Shangwu Yinshuguan.

Chen, Yiya & Carlos Gussenhoven. 2015. Shanghai Chinese. *Journal of the International Phonetic Association*, 45(3), 321–337.

Cheng, Chinchuan. 1973. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, Robert L. 1966. Mandarin phonological structure. *Journal of Linguistics* 2(2), 135–262.

Chitoran, Ioana & Abigail Cohn. 2009. Complexity in phonetics and phonology: gradience, categoriality, and naturalness. *Approaches to phonological complexity* 16, 21–46.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York, Evanston, and London: Harper & Row, Publishers.

Clements, Geoges Nick, & Rachid Ridouane. 2006a. Quantal phonetics and distinctive features. *ITRW on Experimental Linguistics*, Athens, Greece, 17–24.

Clements, Geoges Nick, & Rachid Ridouane. 2006b. Distinctive feature enhancement: a review. *ITRW on Experimental Linguistics*, Athens, Greece, 97–100.

Clements, George Nick. 1992. Phonological primes: Features or gestures?. *Phonetica* 49(3-4). 181–193.

Clements, George Nick & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (eds.), *The handbook of phonological theory*, 245–317. Oxford: Blackwell.

Connell, Bruce. 2007. Mambila fricative vowels and Bantu spirantisation. *Africana Linguistica* 13(1), 7–31.

Connell, Bruce. 2017. Language history and language contact in the Mambiloid region of Cameroon. In Klaus Beyer and Raija Kramer (eds.), *Language Change under Multilingual Conditions: Case studies from Africa*, 111–129. Cologne: Rüdiger Köppe.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1), 407–415.

De Carvalho, Joaquim Brandão, Noël Nguyen & Sophie Wauquier. 2010. *Comprendre la phonologie*. Paris: Presses Universitaires de France.

Dede, Keith. 2006. Standard Chinese and the Xining dialect: The rise of an interdialectal standard. *Journal of Asian Pacific Communication* 16(2), 319–334.

Dell, François & Mohamed Elmedlaoui. 1988. Syllabic consonants in Berber: Some new evidence. *Journal of African Languages and Linguistics* 10(1), 1–17.

Dell, François. 1994. Consonnes à prolongement syllabique en Chine. *Cahiers de linguistique – Asie orientale* 23(1), 87–94.

Dellaporta, Stephen L. & Alejandro Calderon-Urrea. 1993. Sex determination in flowering plants. *The Plant Cell* 5(10), 1241–1251.

Demolin, Didier. 2008. The search for primitives in phonology and the explanation of sound patterns: The contribution of fieldwork studies. *Laboratory Phonology 7* 4(1), 455–513.

Department of Chinese of Peking University. 1989. *Hànyǔ Fāngyīn Zìhuì* [Phonetic Dictionary of Chinese Dialects], 2nd edn. Beijing: Wenzi Gaige Chubanshe.

Diehl, Randy L. 1991. The role of phonetics within the study of language. *Phonetica* 48(2-4), 120–134.

Duanmu, San. 2007. *The phonology of standard Chinese*, 2nd edn. New York: Oxford University Press.

Duanmu, San. 2011. Chinese syllable structure. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. 1, 2754–2777.

Durkheim, Émile & Marcel Mauss. 1903. De quelques formes primitives de classification : Contribution à l'étude des représentations collectives*. Année Sociologique 1901–1902*. 1–72.

Engstrand, Olle. 1990. Swedish. *Journal of the International Phonetic Association* 20(1), 42–44.

Faytak, Matthew & Susan S. Lin. 2015. Articulatory variability and fricative noise in apical vowels. *18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, UK.

Faytak, Matthew. 2015. Temporal organization of frication in fricativized vowels. *Journal of the Acoustical Society of America* 137(4), 2380.

Faytak, Matthew. 2018*. Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese*, PhD dissertation. University of California, Berkeley.

Fu, Maoji. 1956. Běijīng huà de yīnwèi hé pīnyīn zìmǔ [Phonemes and Pinyin symbols in the Beijing speech]. *Zhongguo yuwen* 5, 3–12.

Fuchs, Susanne, Martine Toda & Marzena Żygis. 2010. *Turbulent sounds: An interdisciplinary guide.* Vol. 21. Berlin & New York: Walter de Gruyter.

Fujimoto, Masako & Shigeko Shinohara. 2018. Articulation of 'central' vowel of the Ikema dialect of Miyako Ryukyuan. *32nd General Meeting of the PSJ*, Okinawa, Japan.

Gendrot, Cédric., Barbara Kühnert, & Didier Demolin. 2015. Aerodynamic, articulatory and acoustic realization of French /ʁ/. *18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, UK.

Gong, Xun. 2016. A phonological history of Amdo Tibetan rhymes. *Bulletin of the School of Oriental and African Studies* 79, 347–374.

Gordeeva, Olga & James Scobbie. 2010. Preaspiration as a correlate of word-final voice in Scottish English fricatives. In Fuchs, Susanne, Martine Toda & Marzena Żygis (eds.), *Turbulent sounds: An interdisciplinary guide*, vol. 21, 167–207. Berlin & New York: Walter de Gruyter.

Grammont, Maurice. 1946. *Traité de phonétique*. Paris: Librairie Delagrave.

Greenberg, Joseph H. 1962. Is the Vowel–Consonant Dichotomy Universal? *Word* 18(1-3), 73–81.

Gu, Chong. 2014. Smoothing Spline ANOVA Models: R Package gss. *Journal of Statistical Software* 58(5), 1–25.

Haig, David. 2004. The inexorable rise of gender and the decline of sex: Social change in academic titles, 1945–2001. *Archives of sexual behavior* 33(2), 87–96.

Hartman, Lawton M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20(1), 28–42.

Hayashi, Yuka. 2010. Ikema (Miyako Ryukyuan). In Thomas Pellard & Michinori Shimoji (eds.), *An introduction to Ryukyuan languages*, 167–188. Tōkyō: Research Institute for Languages, Cultures of Asia & Africa.

Hirata, Shoji. 1998. *Huīzhōu fāngyán yánjiū* [Study on Huizhou dialects]. Tokyo: Kohbun Press.

Ho, Aichen. T. 1976. The acoustic variation of Mandarin tones. *Phonetica* 33(5), 353–367.

Hockett, Charles F. 1947. Peiping phonology. *Journal of the American Oriental Society* 67(4), 253–267.

Hou, Chao. 2009. Héféi fāngyán dàicāyuányīn de shíyàn yánjiū [An experimental study on [ʅ] in Hefei Dialect]*. Journal of School of Chinese Language and Culture (Nanjing Normal University)* 1, 172–177.

Howie, John M. 1976. *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.

Hu, Fang. 2007. Lùn Níngbō fāngyán hé Sūzhōu fāngyán qiángāo yuányīn de qūbié tèzhēng — Jiāntán gāoyuányīn jìxù gāohuà xiànxiàng [On the distinctive features of the high front vowels in Ningbo and Suzhou Wu dialects with reference to sound changes of high vowels]. *Zhongguo Yuwen* 5, 455–465.

IPA. 1912. The Principles of the International Phonetic Association. *The Maître Phonétique supplement* sept.–oct. 1–40.

IPA. 2010. The Principles of the International Phonetic Association (1949). *Journal of the International Phonetic Association* 40(3), 299–358.

Ito, Mabo R & Robert W. Donaldson. 1971. Zero-crossing measurements for analysis and recognition of speech sounds. *IEEE Transactions on Audio and Electroacoustics* 19(3), 235–242.

Jacques, Guillaume & Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4), 468–498.

Jakobson, Roman & Linda R. Waugh. 1989. *The sound shape of language*. Berlin & New York: Mouton de gruyter.

Jassem, Wiktor. 1965. The formants of fricative consonants. *Language and Speech* 8(1), 1–16.

Karlgren, Bernard. 1915. *Études sur la phonologie chinoise*. Uppsala: K. W. Appelberg.

Kelso, J. A. Scott. 2005. 7 The Complementary Nature of Coordination Dynamics: Toward a Science of the In-Between. In Reuben R. McDaniel Jr. & Dean J. Driebe (eds.), *Uncertainty and Surprise in Complex Systems,* 77–85. Berlin & Heidelberg: Springer.

Keyser, Samuel Jay & Kenneth N. Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82(1), 33–63.

Kong, Huifang, Shengyi Wu & Mingxing Li. 2019. Héféi Huà Shéjiān Yuányīn de Mócā Xìngzhi jí Gǎnzhī Zēngqíang Lǐlùn Jiědú [The Frication Property of Apcial Vowels in Hefei Chinese and Its Perceptual Enhancement Account]. *Yuyan Yanjiu* 39(1), 23–33.

Koonin, Eugene V. & Petro Starokadomskyy. 2016. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences* 59, 125–134.

Kratochvíl, Paul. 1968. The Chinese Language Today: features of an emerging standard. London: Hutchinson & Co LTD.

Kutsch Lojenga, Constance. 1989. The secret behind vowelless syllables in Lendu. *Journal of African Languages and Linguistics* 11(2), 115–126.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford & Malden, MA: Blackwell.

Ladefoged, Peter & Keith Johnson. 2011. *A course in phonetics*, 6th edn. Boston, MA: Wadsworth.

Ladefoged, Peter, and Sandra Ferrari Disner. 2012. *Vowels and consonants,* 3rd edn. Malden, MA: Blackwell Publishing Ltd.

Lee-Kim, Sang-Im. 2014. Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3), 261–282.

Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1), 109–112.

Lee, Wai-Sum. 2005. The articulatory and acoustical characteristics of the “apical vowels” in Beijing Mandarin. *The Journal of the Acoustical Society of America* 118(3), 2027.

Li, Huayi. 2007. Jìxī Huáyáng Huà de Shēngdiào Géjú [The tonal pattern of Huayang Dialect in Jixi county]. *Gansu Gaoshi Xuebao* 12(6), 1–6.

Li, Qian, Yiya Chen & Ziyu Xiong. 2019. Tianjin Mandarin. *Journal of the International Phonetic Association* 49(1), 109–128.

Li, Rong, Zhenghui Xiong & Zhenxing Zhang. 1987. *The Language Atlas of China*. Hong Kong: Longman.

Lin, Yen-Hwei. 2007. *The Sounds of Chinese with Audio CD*. Vol. 1. New York, NY: Cambridge University Press.

Ling, Feng. 2009. *A phonetic study of the vowel system in Suzhou Chinese*. PhD dissertation. City University of Hong Kong, Hongkong.

López-García, Purificatión. 2012. The place of viruses in biology in light of the metabolism-versus-replication-first debate. *History and philosophy of the life sciences* 34(3*)*, 391–406.

Luo, Changpei. 1936. Jìxī fāngyīn Shùlüè [Brief introduction to Jixi dialect]. Reprint, In Gao Gengsheng & Sun Hongkai (eds.), *Luó Chángpéi Wénjí* vol 9 [The collected linguistic works of Luo Changpei vol 9], 175–204. Jinan: Shandong jiaoyu chubanshe, 2008.

Michaud, Alexis. 2008. Phonemic and tonal analysis of Yongning Na. *Cahiers de linguistique – Asie Orientale* 37(2), 159–196.

Moreira, David & López-García Purificación. (2009). Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* 7(4), 306–311.

Mpiranya, Fidèle. 1997. Spirantisation et fusion vocalique en bantu: essai d’interprétation fonctionnelle. *Linguistique Africaine 18*, 51–77.

Nespor, Marina, Marcela Peña & Jacques Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio* 2(2), 203–230.

O’Connor, J. Desmond & John L. M. Trim. 1953. Vowel, consonant, and syllable—a phonological definition. *Word* 9(2), 103–122.

Ohala, John J. 1983. The Origin of Sound Patterns in Vocal Tract Constraints. In MacNeilage P.F. (eds), *The Production of Speech,* 189–216. New York, NY: Springer.

Passy, Paul. 1906. *Petite phonétique comparée*. Leipsic and Berlin: B.G. Teubner, Libraire-Éditeur.

Pearson, Helen. 2008. 'Virophage' suggests viruses are alive. *Nature* 454(7205), 677.

Pike, Kenneth L. 1971. *Phonetics: A Critical Analysis of Phonetic Theory and a Technic for the Practical Description of Sounds.* Ann Arbor, MI: The University of Michigan Press.

Pike, Kenneth L. 1947. *Phonemics. A Technique for Reducing Languages to Writing.* Ann Arbor, MI: The University of Michigan Press.

Pullum, Geoffrey K. & William A. Ladusaw. 1996. *Phonetic symbol guide (Second Edition)*. Chicago, IL: The University of Chicago Press.

Rabiner, Lawrence R. & Ronald W. Schafer. 1978. *Digital Processing of Speech Signals.* Englewood Cliffs, NJ: Prentice-Hall.

Ridouane, Rachid. 2007. Gemination in Tashlhiyt Berber: an acoustic and articulatory study. *Journal of the International Phonetic Association* 37(2), 119–142.

Ridouane, Rachid. 2008. Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber. *Phonology* 25(2), 321–359.

Rousselot, Jean-Pierre. 1897. *Principes de Phonétique Expérimentale 1*. Paris: Welter.

Ruggiero, Michael A., Dennis P. Gordon, Thomas M. Orrell, Nicolas Bailly, Thierry Bourgoin, Richard C. Brusca, Thomas Cavalier-Smith, Michael D. Guiry & Paul M. Kirk. 2015a. A higher level classification of all living organisms. *PloS one* 10(4), 1–60.

Ruggiero, Michael A., Dennis P. Gordon, Thomas M. Orrell, Nicolas Bailly, Thierry Bourgoin, Richard C. Brusca, Thomas Cavalier-Smith, Michael D. Guiry, and Paul M. Kirk. 2015b. Correction: a higher level classification of all living organisms. *Plos one* 10(6), 1–54.

Rybicki, Edward. 1990. The classification of organisms at the edge of life or problems with virus systematics. *South African Journal of Science* 86(4), 182.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. Courier Corporation, 2004.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Lausanne & Paris: Payot.

Scamardella, Joseph M. 1999. Not plants or animals: a brief history of the origin of Kingdoms Protozoa, Protista and Protoctista. *International Microbiology* 2(4), 207–216.

Shi, Feng, Gang Peng & Yi Liu. 2015. Vowel distribution in isolated and continuous speech: the case of Cantonese and Mandarin. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 459–473. New York, NY: Oxford University Press.

Shosted, Ryan K. 2006. *The aeroacoustics of nasalized fricatives*, PhD dissertation. University of California, Berkeley.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25(4), 471–500.

Stevens, Kenneth N. & Samuel Jay Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65(1), 81–106.

Stevens, Kenneth N., Samuel Jay Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.

Stone, Maureen & Subhash Lele. 1992. Representing the tongue surface with curve fits. *Second International Conference on Spoken Language Processing*. Banff, Alberta, Canada.

Stone, Maureen, Thomas H. Shawker, Thomas L. Talbot & Alan H. Rich. 1988. Cross-sectional tongue shape during the production of vowels. *The Journal of the Acoustical Society of America* 83(4), 1586–1596.

Stone, Maureen. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s],[ʃ], and [l]. *Journal of phonetics* 20, 253–270.

Svantesson, Jan-Olof. 1984. Vowels and diphthongs in Standard Chinese*. Working Papers of Lund University, Department of Linguistics* 27, 209–235.

Svantesson, Jan-Olof. 1986. Acoustic Analysis of Chinese Fricatives and Affricates. *Journal of Chinese Linguistics* 14(1), 53–70.

Troubetzkoy, Nicolas. S. 1949. *Principes de Phonologie*. Paris: Librairie C. Klincksieck.

Trubetzkoy, Nicolas. S. 1969. *Principles of phonology*. Berkley and Los Angeles, CA: University of California Press.

Udry, J. Richard. 1994. The nature of gender. *Demography* 31(4), 561–573.

Vendryès, Joseph. 1921. *Le langage: introduction linguistique à l'histoire*. Paris: La Renaissance du livre.

Wan, Liyang. 2014. *Xīnpài Hēféi Huà Yuányīn Xìtǒng de Shíyàn Yánjiū* [A acoustic study of the vowel system in New Hefei dialect]. Master thesis. Nanjing Normal University, Nanjing.

Wang, Shuangcheng. 2006. Qīnghǎi fāngyán yuányīn [i] de shéjiānhuà yīnbiàn [Apicalization sound change of the Qinghai dialects' [i] vowel]. *Zhongguo Yuwen* 4, 359–363.

Wang, Shuangcheng. 2010. Ānduō zàngyǔ [ɨ] de shéjiānhuà jí qí lèixíngxué yìyì [Sound Change of [ɨ] in Amdo Tibetan and its Typological Meaning]. *Studies in Language and Linguistics* 30(2), 122–127.

Warner, D. A., and Richard Shine. 2008. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* 451(7178), 566–568.

Westerberg, Fabienne. 2016. An auditory, acoustic, articulatory and sociophonetic study of Swedish Viby-i. Master thesis. University of Glasgow, Glasgow.

Westerberg, Fabienne. 2019. Swedish "Viby-i": Acoustics, Articulation, And Variation. *19th International Congress of Phonetic Sciences (ICPhS XVIIII)*, Melbourne, Australia.

Wharton, Amy S. 2004. The sociology of gender: An introduction to theory and research. Oxford, UK: Wiley-Blackwell.

Wiberg, U. H. 1983. Sex determination in the European eel (Anguilla anguilla, L.). *Cytogenetic and Genome Research* 36(4), 589–598.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43), 1686.

Wiese, Richard. 1997. Underspecification and the description of Chinese vowels. *Studies in Chinese phonology*, 219–249.

Woese, Carl R., Otto Kandler & Mark L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* 87(12), 4576–4579.

Wu, Wei. 1995. Héféi huà "-i"、"-y" yīnjié shēngyùnmǔ qiánhuà tàntǎo [Exploration of the fronting of "-i" and "-y" rhymes in the Hefei dialect]. *Yuyan yanjiu* 3, 58–60.

Ye, Xiangling. 1993. *Sūzhōu fāngyán cídiǎn* [Dictionary of Suzhou Dialect]. Nanjing: Jiangsu jiaoyu chubanshe.

Yu, Alan. 1999. Aerodynamic constraints on sound change: The case of syllabic sibilants. *Journal of the Acoustical Society of America* 105(2), 1096–1097.

Yuan, Jiahua. 1960. *Hànyǔ Fāngyán Gàiyào* [A compendium of Chinese dialects]. Beijing: Wenzi Gaige Chubanshe.

Yuan, Jiahua. 1983. *Hànyǔ Fāngyán Gàiyào* [A compendium of Chinese dialects], 2nd edn. Beijing: Wenzi Gaige Chubanshe.

Zee, Eric & Wai-Sum Lee. 2001. An acoustical analysis of the vowels in Beijing Mandarin. *7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, 643–646.

Zee, Eric & Wai-Sum Lee. 2007. Vowel typology in Chinese. *16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany, 1429–1432.

Zeng, Ting. 2020. The Xiangxiang dialect of Chinese. *Journal of the International Phonetic Association* 50(2), 258–281.

Zhao, Rixin. 1989. Ānhuī Jìxī fāngyán yīnxì tèdiǎn [Phonological characters of Jixi dialect from Anhui Province]. *Fangyan* 2, 125–130.

Zhao, Rixin. 2003. *Jìxī fāngyán cídiǎn* [Dictionary of Jixi dialect]. Nanjing: Jiangsu jiaoyu chubanshe.

Zhao, Rixin. 2007. Hànyǔ fāngyán zhōng de [i] > [ɿ] [Sound change of [i] > [ɿ] in Chinese dialects]. *Zhongguo yuwen* 1, 46–54.

Zhou, Dianfu & Zongji Wu. 1963. *Pǔtōnghuà fāyīn túpǔ* [Articulatory diagrams of Standard Chinese]. Beijing: Shangwu yinshuguan.

Zhu, Xiaonong. 1995. *Shanghai tonetics*. PhD dissertation. Australian National University, Canberra.

Zhu, Xiaonong. 2004. Hànyǔ yuányīn de gāodǐngchūwèi [Sound changes of high vowels in Chinese dialects]. *Zhongguo yuwen* 5, 440–451.

Annexes

A.1 Word list used in acoustic and ultrasound data collection.

	Phonetic	Ortho.	Gloss	Acous.	Ultras.	Note
1	pa˧˩	杯	‘cup’	✓		
2	pʰa˥	赔	‘compensate’	✓		
3	ma˨˩˨	每	‘every’	✓		
4	la˥	来	‘come’	✓		
5	tsa˨˩˨	宰	‘slaughter’	✓	✓	
6	tsʰa˨˩˨	踩	‘step on’	✓	✓	
7	sa˦˥	碎	‘smash into pieces’	✓	✓	
8	pi˧˩	碑	‘stela’	✓		produced as [pa˧˩]
9	mi˨˩˨	美	‘beautiful’	✓		produced as [mɛ̃˨˩˨]
10	li˥	留	‘stay’	✓		
11	li˨˩˨	柳	‘willow’	✓		
12	tsi˨˩˨	走	‘walk’	✓	✓	
13	tsʰi˨˩˨	丑	‘ugly’	✓	✓	
14	si˨˩˨	洗	‘wash’	✓	✓	
15	si˦˥	瘦	‘thin’	✓	✓	
16	pu˨˩˨	补	‘repair’	✓		
17	pʰu˥	葡	‘grape’	✓		
18	pʰu˨˩˨	普	‘general’	✓		
19	lu˥	炉	‘stove’	✓		
20	lu˨˩˨	鲁	‘a family name’	✓		
21	tsu˨˩˨	组	‘group’	✓	✓	
22	tsʰu˨˩˨	础	‘foundation’	✓	✓	
23	su˨˩˨	竖	‘vertical, erect’	✓	✓	
24	su˦˥	漱	‘rinse’	✓	✓	
25	pz˧˩	屄	‘female genital’	✓	✓	
26	pz˨˩˨	比	‘compare’	✓	✓	
27	pʰz˥	皮	‘skin’	✓	✓	
28	pʰz˨˩˨	被	‘quilt’	✓	✓	
29	mz˨˩˨	米	‘rice’	✓	✓	
30	nz˥	泥	‘mud’	✓	✓	

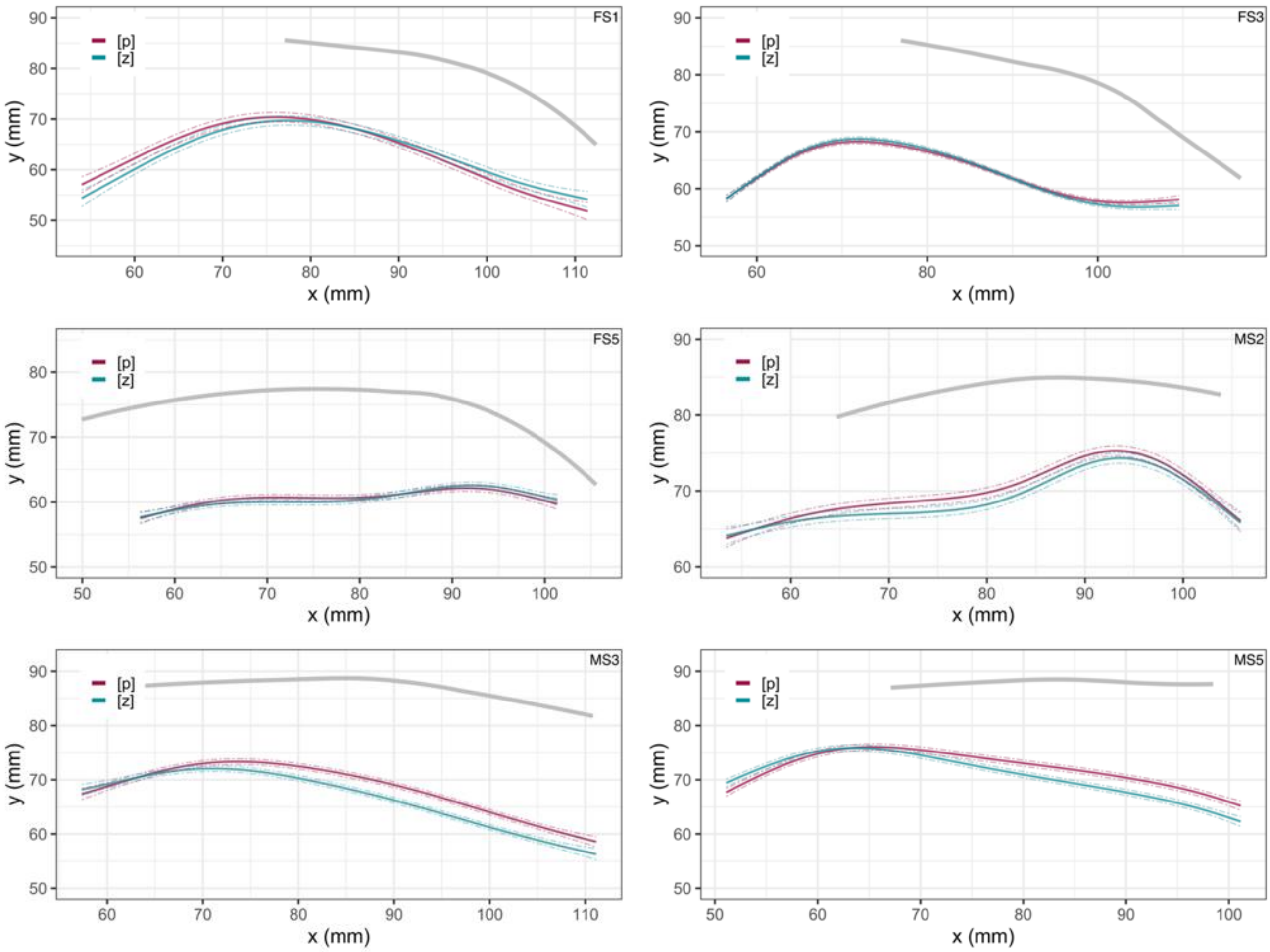
31	nz˨˩˦	里	‘in’	√	√
32	tsz˨˩˦	紫	‘purple’	√	√
33	ts^{h}z˨˩˦	弟	‘younger brother’	√	√
34	sz˨˩˦	死	‘die’	√	√
35	sz˧˥	四	‘four’	√	√
36	pʉ˥˩	波	‘wave’	√	√
37	p^{h}ʉ˥	婆	‘old woman’	√	√
38	p^{h}ʉ˨˩˦	叵	‘impossibly’	√	√
39	mʉ˨˩˦	某	‘someone, something’	√	√
40	lʉ˥	揉	‘knead’	√	
41	tsʉ˨˩˦	左	‘left’	√	√
42	tsʉ˧˥	做	‘make’	√	√
43	sʉ˨˩˦	锁	‘lock’	√	√

A.2 Information on speakers: Recorded speakers' gender, birthyear, highest education and profession at the moment of the recording of acoustic data. The ultrasound data were recorded with FS1, FS3, FS5, MS1, MS2, MS3, MS5 one year after the acoustic recording.

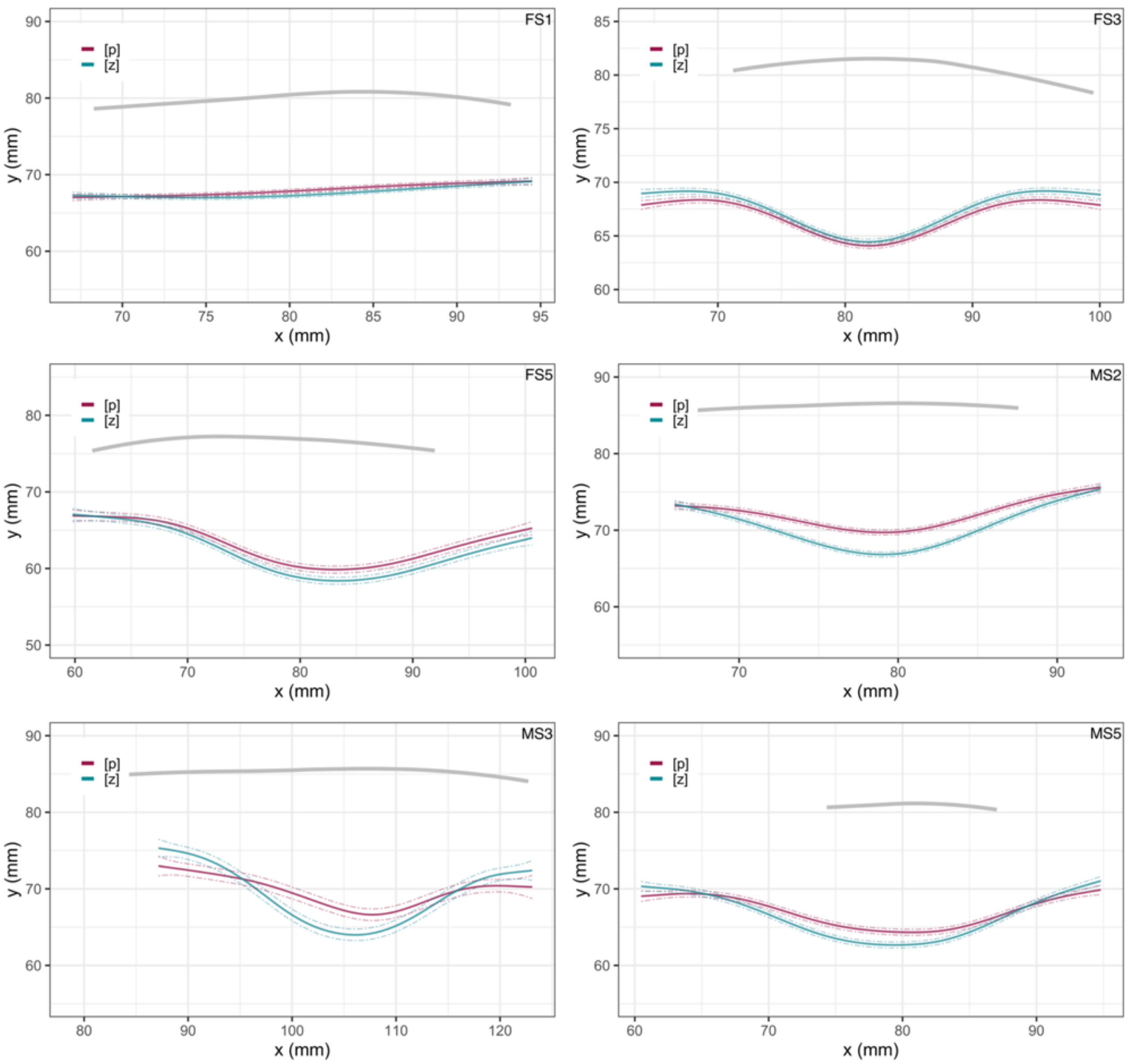
	Gender	Birthyear	Education	Profession
FS1	F	1971	High school	Employee
FS2	F	1964	High school	Housewife
FS3	F	1964	High school	Civil-service employee
FS4	F	1971	High school	Employee
FS5	F	1964	High school	Civil-service employee
MS1	M	1969	College	Medical technician
MS2	M	1971	High school	Entrepreneur
MS3	M	1971	College	Doctor
MS4	M	1970	High school	Employee
MS5	M	1975	High school	Entrepreneur

A.3 Additional figures of SS ANOVA splines of [pz] and [mz] syllables. The data are extracted in x/y values (mm) at the nearest midpoint image of each segment (the tongue front is on the right and tongue root on the left on the mid-sagittal plane). Red lines represent the nearest midpoint images of the closure of [p m] and blue lines represent the nearest centre of [z]. The splines are presented with 95% Bayesian confidence intervals. The thick grey lines represent the palatal traces. Figures (a) and (b) shows [pz] in mid-sagittal and coronal planes, and Figures (c) and (d) show [mz] in mid-sagittal and coronal planes.

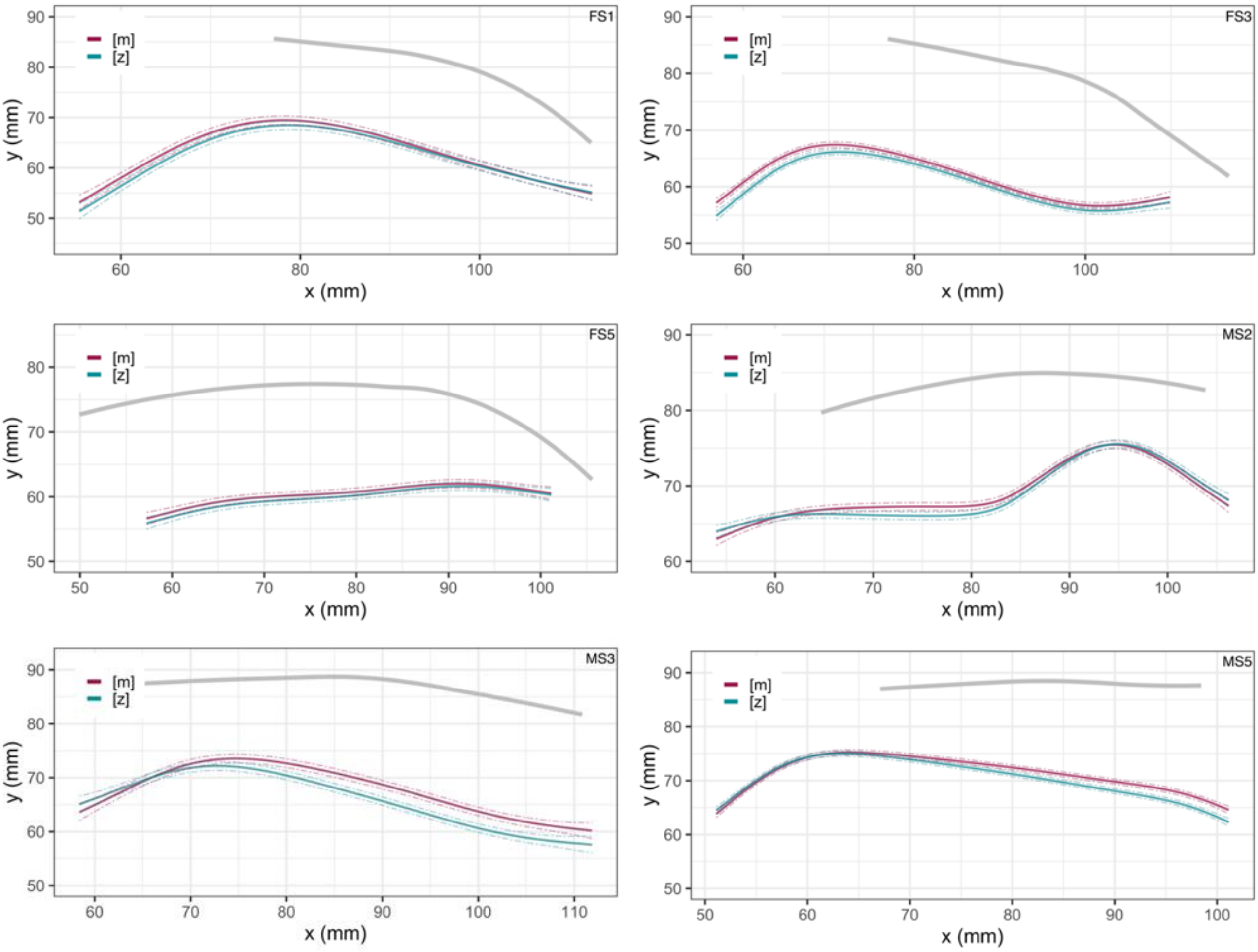
(a)



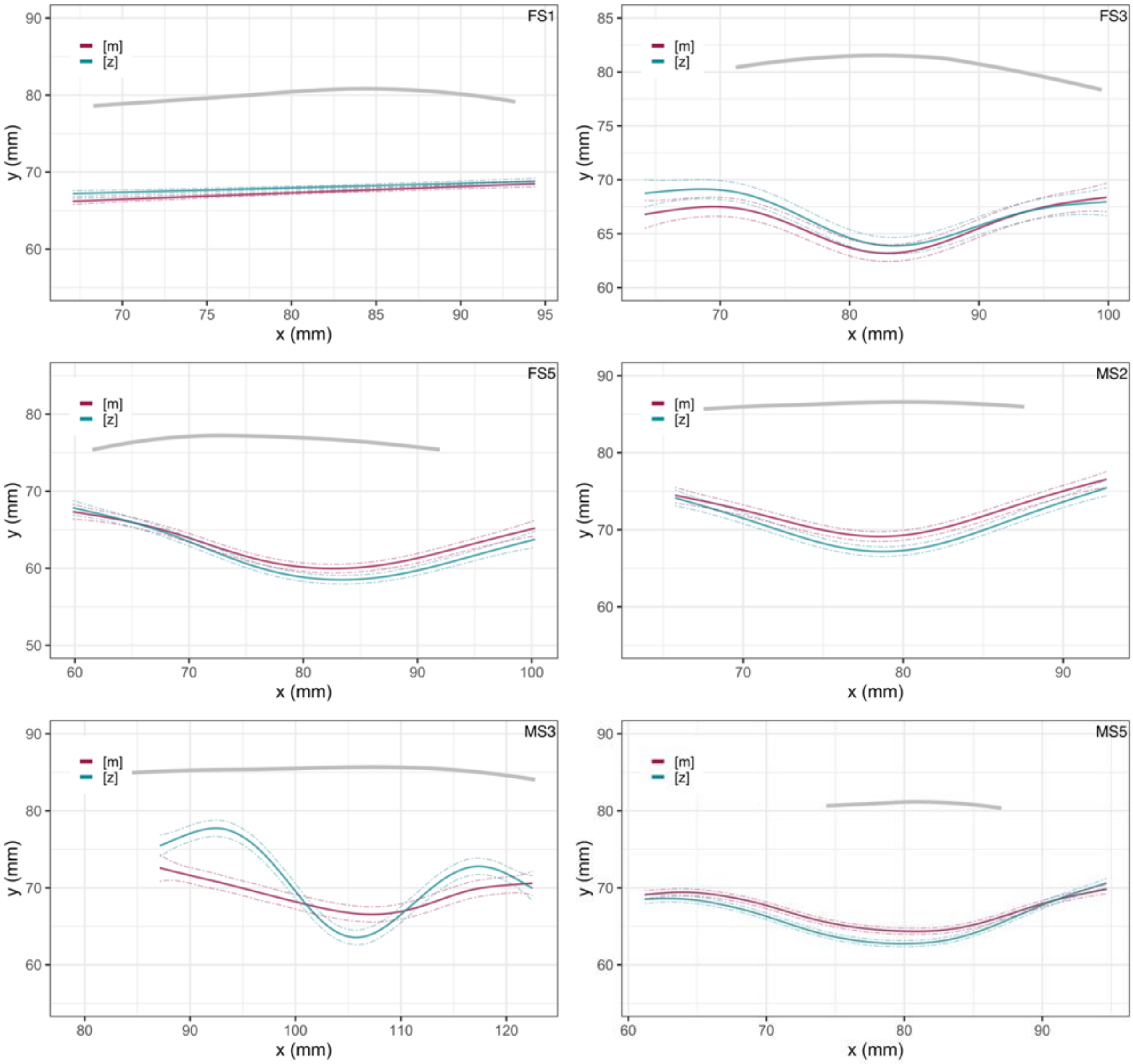
(b)



(c)



(d)



La voyelle apicale en Chinois Jixi-Hui : phonologie et phonétique

Les langues chinoises ont un ensemble de segments appelés « voyelles apicales » (舌尖元音 en chinois). Leur nature exacte est à l'origine d'un débat toujours en cours : S'agit-il de consonnes ou de voyelles ? Les « voyelles apicales » ont été analysées dans des études précédentes comme étant de véritables voyelles, des voyelles fricatives, des fricatives syllabiques, ou des approximantes syllabiques. Cette thèse porte sur la voyelle apicale attestée en chinois Jixi-Hui. J'examine ce segment d'un point de vue phonétique et phonologique et montre qu'il est mieux défini comme une consonne fricative voisée (transcrit /z̩/).

Phonologiquement, ce segment est un phonème distinct de /i/. Il est exclusivement attesté en position de noyau de syllabe où il constitue une unité porteuse de ton. Il peut apparaître non seulement après les sibilantes coronales /s ts ts^h/, mais aussi les bilabiales /p p^h/ et les nasales /m n/. Phonétiquement, les caractéristiques acoustiques et articulatoires de ce segment sont examinées. Les résultats montrent que /z̩/ contient dans la majorité des cas un bruit de friction dans sa phase initiale, superposé sur du voisement, avec une structure formantique plus claire apparaissant vers la fin. Les analyses du rapport harmonique/bruit et du taux de passage par zéro confirment cette présence significative du bruit de friction, distinguant clairement ce segment des voyelles. Les généralisations en SS ANOVA des données ultrasoniques montrent que /z̩/ a une forme de langue presque identique à celle de /s/ sur les plans mi-sagittal et coronal, malgré quelques différences spécifiques à chaque locuteur. Cette forme de langue reste constante dans les contextes consonantiques bilabiales et alvéolaires.

La variabilité dans la façon dont /z̩/ est phonétiquement implémentée est considérée comme étant la conséquence de deux contraintes en interaction : une contrainte structurelle liée au statut distinctif de /z̩/ et au rôle qu'il joue dans la structure syllabique, et une contrainte physique liée à l'incompatibilité entre le voisement et le bruit de friction. L'étude souligne également la nécessité de reconnaître les fricatives syllabiques en chinois Jixi-Hui, et probablement aussi dans d'autres langues chinoises.

Mots-clés : Voyelle apicale, Chinois de Jixi, acoustique, ultrason, bruit de friction, fricative syllabique.

The apical vowel in Jixi-Hui Chinese: phonology and phonetics

Chinese languages have a set of segments known as 'apical vowels' (舌尖元音 in Chinese). Their exact nature is still the source of an ongoing debate: Are they consonants or vowels? 'Apical vowels' have been analysed in previous studies as genuine vowels, fricative vowels, syllabic fricatives, or syllabic approximants. This dissertation is concerned with the apical vowel attested in Jixi-Hui Chinese. I examine this segment from phonetic and phonological perspectives and show that it is best defined as a voiced fricative consonant (transcribed /z̩/).

Phonologically, this segment is a distinct phoneme from /i/. It is exclusively attested in syllable nucleus position where it constitutes a tone-bearing unit. It can appear not only after coronal sibilants /s ts tsh/, but also bilabials /p ph/ and nasals /m n/. Phonetically, the acoustic and articulatory characteristics of this segment are examined. The results show that /z̩/ contains in the majority of cases frication noise in its initial phase superposed on voicing, and a clearer formant structure appears towards its end. The harmonic-to-noise ratio and zero-crossing rate analyses confirm this significant presence of noise, clearly distinguishing this segment from vowels. The smoothing-spline ANOVA analyses of ultrasound data show that /z̩/ has a near-identical tongue shape to /s/ on both mid-sagittal and coronal planes despite some speaker-specific differences. This /s/-like tongue shape is constant in bilabial and alveolar consonantal contexts.

The variability in the way /z̩/ is phonetically implemented is argued to be a consequence of two interacting constraints: a structural one related to the distinctive status of /z̩/ and the role it plays within syllable structure, and a physical one related to the incompatibility of voicing and frication. The study further argues for the necessity of recognizing syllabic fricatives in Jixi-Hui Chinese and probably also in other Chinese languages.

Key words: Apical vowel, Jixi-Hui Chinese, acoustics, ultrasound, frication noise, syllabic fricative.

École Doctorale 622 – Sciences du langage
4, rue des irlandais 75005 PARIS