The Segment Status of the Mandarin Glide: A Language Game Experiment

Boer Fu

Massachusetts Institute of Technology

1 Introduction

The acquisition of a phonological grammar requires the slicing of a continuous speech signal into individual consonants and vowels as a first step, yet it is often taken as a given. In recent years, there has been a growing interest in the segmentation question. Downing (2005) investigated the segment status of homorganic NC sequences by inspecting their role in phonological processes. Béland & Kolinsky (2005) conducted perception experiments to see if French affricate segmentation differs between dialects. Gouskova & Stanton (2021) modeled the segmentation learning process computationally. Their learner calculates the co-occurrence probability of two neighboring sounds to determine whether they are a single segment or two. Shaw et al. (2021) approached the segmentation problem from the perspective of articulatory gesture timing.

The segmentation problem represents a wider theoretical question: what do learners do when they are confronted with ambiguous phonological input? Are they consistent in what they learn? In other words, can speakers arrive at a consistent grammar that account for all the data they hear? More importantly, do speakers all arrive at the same grammar? In segmentation terms, the question is whether speakers of one language, being exposed to roughly the same phonological input, arrive at the same segmentation?

I approach the above questions by zooming in on the prenuclear glide in Mandarin Chinese. The segment status of the glide is subject to perennial debate in the literature on Mandarin phonology. Some treat it as an independent segment. Others argue it is the secondary articulation of the onset. Still others claim that it is merely the natural consonant-to-vowel transition. Each camp has developed their own phonological grammar for Mandarin, based on a phoneme inventory assuming their theory of glide segmentation. Each grammar provides a consistent account for various phenomena in Mandarin phonotactics. If the phonological data of Mandarin can lead linguists into different paths, then it is only reasonable to expect Mandarin learners to be confronted with the same ambiguity when they try to make sense of the phonological input. How speakers of Mandarin segment the glide can contribute to our understanding of phonological learning.

In order to find out which glide segmentation theory best match the mental representation of sounds for individual Mandarin speakers, I have designed a codeword language game, which is based on *fanqie* secret languages (see Chao 1931) that split a syllable into onset and rhyme. I have conducted two experiments, one online and one in-person, making use of the codeword language game. In the experiments, Mandarin speakers are invited to disassemble the syllable in an artificial setting. Their treatment of the glide is indicative of whether they segment the glide as part of the onset or part of the rhyme. The results show that Mandarin speakers' segmentation of the prenuclear glide is largely influenced by the place of articulation of the preceding consonant. Speaker variation, both interspeaker and intraspeaker, is observed as well.

The rest of the paper is organized as follows. Section 2 provides a review of the debate on Mandarin glide segmentation. In Section 3, I introduce the methods of the codeword language game. Results from the first online experiment are reported in Section 4, along with a discussion on its limitations. Section 5 includes information on the second in-person experiment, which addresses some of the problems of the online experiment. Discussion on the observed speaker variation is also provided. Section 6 concludes.

* Many thanks to Adam Albright, Edward Flemming, Michael Kenstowicz, and Donca Steriade for discussion and feedback. All remaining mistakes are my own.

2 The Mandarin glide segmentation debate

A Mandarin syllable follows a strict template: CGVX, where G labels the prenuclear glide and X stands for a coda that can either be an offglide or a nasal. Apart from the nucleus vowel V, all three other components, C, G, and X, are optional. The segmentation debate concerns G. The prenuclear glide comes in three forms: the palatal /j/, the bilabial /w/, and the labiopalatal /ɥ/. They are subject to a range of co-occurrence restrictions with the onset. The distribution of the three glides is illustrated in Table 1. The top row lists Mandarin consonants by place of articulation. The leftmost column shows the possible glides that the consonants can combine with, including a glide-less condition. If a CG combination is possible, an example syllable is provided. A shaded cell means the CG combination is unattested.

	Dental Sibilants	Retroflexes	Velars	Labials	Dentoalveolars		Palatals
					[-son]	[+son]	
	ts, tsʰ, s	ʈʂ, ʈʂʰ, ʂ, ɻ	k, kʰ, x	p, pʰ, m, f	t, tʰ	n, l	tɕ, tɕʰ, ɕ
C	[san] *'three'*	[ʈʂʰɑŋ] *'to sing'*	[xej] *'black'*	[fɤŋ] *'wind'*	[tʰow] *'head'*	[la] *'to pull'*	
Cj				[mjɛn] *'flour'*	[tʰjɛn] *'sky'*	[njow] *'ox'*	[ɕja] *'shrimp'*
Cw	[tsʰwo] *'wrong'*	[ɻwan] *'soft'*	[kʰwaj] *'fast'*	[pwo][1] *'wave'*	[twen] *'ton'*	[nwan] *'warm'*	
Cɥ						[lɥe] *'to omit'*	[tɕɥɛn] *'curl'*

Table 1: Distribution of glides

Of special interest to the discussion is the palatal glide. /j/ cannot appear after any of the three onset groups: the dental sibilants /ts, tsʰ, s/, the retroflexes /ʈʂ, ʈʂʰ, ʂ, ɻ/, and the velars /k, kʰ, x/. The palatal glide is contrastive after bilabials /p, pʰ, m/[2] and non-sibilant dentoalveolars /t, tʰ, n, l/, as in (1a & b). After the palatal sibilants /tɕ, tɕʰ, ɕ/, it is obligatory (or a labiopalatal /ɥ/ can take its place), as in (1c & d).

(1) *Palatal glide after different onsets*

a.	la	'wax'		
b.	lja	'two people'	→	*Palatal glide contrastive after non-palatal onsets*
c.	*ɕa	*Unattested*		
d.	ɕja	'shrimp'	→	*Palatal glide obligatory after palatal onsets*

The codependent relation between the palatal onset consonants and the palatal glide has invited much interest in the field of Mandarin phonology. Various analyses have been proposed to account for the palatal onset-glide sequence, with the aim to develop a consistent segmentation for the Mandarin prenuclear glide in general. I discuss 3 main proposals, each representative of a larger body of literature with similar views. Different segmentation hypotheses for the prenuclear glide leads to different transcription in IPA for the same Mandarin syllable. This is demonstrated for three monosyllabic words [lje] 'to hunt', [ɕje] 'shoe', and [ɕɥe] 'snow' in Table 2. They serve as concrete examples for the glide segmentation proposals.

Lin (1989) treated the prenuclear glide as an independent segment that is separate from the onset. If this is the case, the three example words ought to be transcribed as [lje], [ɕje], and [ɕɥe]. Duanmu (2002), on the other hand, argued that the glide is the secondary articulation of the onset, with no segmental representation independent of the onset. The triplets are accordingly written as [lʲe], [ɕʲe], and [ɕᶣe].

In addition, Duanmu analyzed the palatal series /tɕ, tɕʰ, ɕ/ as a result of synchronic palatalization, derived from the dental sibilant series /ts, tsʰ, s/, when they come into contact with a palatal secondary articulation.

[1] [pwo] is often transcribed as [po] in the literature. Here, the syllable is transcribed with a bilabial glide, since there is an audible glide period between the onset and the vowel. It is not equivalent to French *peau* [po] 'skin'.
[2] The labiodental /f/, although usually grouped with the bilabials, never appear before /j/ due to diachronic reasons.

Since there is no morphophonological processes concerning the onset in Mandarin, any claim on its synchronic alternation in different phonological environments is to be treated with caution. Chao (1934) pointed out that the palatals are especially tricky for phonological analysis, since they are in complementary distribution with three groups of consonants, the dental sibilants, the retroflexes, and the velars.

Hypothesis	Independent Segment?	*'to hunt'*	*'shoe'*	*'snow'*	Syllable Segmentation
Independent segment (Lin 1989)	Yes	[lje]	[ɕje]	[ɕɥe]	CGV
Secondary articulation of onset (Duanmu 2002)	No	[lʲe]	[ɕʲe] (/sʲe/)	[ɕᶣe] (/sᶣe/)	C^GV
Natural palatal CV transition (Ladefoged & Maddieson 1996)	No	[lje]	[ɕe]	[ɕe]	CGV/CV
Double representation of glide	Yes	[lʲje]	[ɕʲje]	[ɕᶣɥe]	C^GGV

Table 2: Prenuclear glide segmentation hypotheses

In Ladefoged & Maddieson's (1996) view, the palatal glide after a palatal onset is merely the natural CV transition. They pointed out that the formants of /j/ is no different from what one would expect from the transition between a palatal consonant and a vowel. Therefore, their transcription has [ɕe] for 'shoe', with no symbol for the glide. This analysis might capture the regular palatal glide /j/ with adequacy, but it cannot account for the labiopalatal glide /ɥ/ in 'snow'. With no separate IPA symbol for the glide between the onset and the vowel, 'snow' ought to be written as [ɕe], indistinguishable from [ɕe] 'shoe'. The information on labialization in the syllable for 'snow' is lost. One solution to the missing labialization problem is to ascribe it to the onset, as in [ɕᶣ], without creating a separate labiopalatal segment after the onset. Note that this makes the natural CV transition hypothesis indistinguishable from Duanmu's (2002) secondary articulation account. Therefore, I have decided to exclude the natural CV transition hypothesis from consideration.

Finally, I consider the possibility that the glide is both a secondary articulation of the consonant and an independent articulatory target that exists outside the onset. Under this double representation hypothesis, the three example syllables are transcribed as [lʲje], [ɕʲje], and [ɕᶣɥe] respectively.

The choice between these glide segmentation hypotheses is not only an analytical problem for phonologists, but also a learning problem for children acquiring Mandarin.

As a note of transcription, throughout this paper, when Mandarin examples are cited in general, the glide is always written as a separate symbol, as in "[ɕje]". This is not to say that the independent segment hypothesis is the correct solution. The glide transcription is simply done for ease of presentation.

3 Codeword language game

In order to find out how Mandarin speakers segment the prenuclear glide, I have designed a codeword language game that invites them to disassemble the syllable in an artificial setting. In the game, the speaker is asked to encode a disyllabic Mandarin word by swapping the onsets of the two syllables. For example, the original word (2a) [kʰa fej] 'coffee' is encoded as the nonce word (2b) [fa kʰej]. The rhymes of the two syllables stay put, while the onset [kʰ] and [f] have swapped.

(2) *Codeword language game task*

a.	**kʰ**a **f**ej	'coffee'	→	b.	**f**a **kʰ**ej	*Nonce word*
	Original word		*Swap onsets*		*Codeword*	

In the training phase, the participants are only exposed to glide-less items like [kʰa fej]. No instruction is provided on whether the prenuclear glide counts towards the onset or the rhyme. In the experiment phase, there is a glide in one of the syllables in a disyllabic test item. What the speaker does with the glide can inform us on how they segment it. If the speaker considers the glide to be part of the onset, then they will move it along with the onset. On the other hand, if they see the glide as an independent segment, it will be left behind with the vowel. This is demonstrated for the example test item [ta ljaw] 'star anise' in (3).

(3) *Glide segmentation determines codeword response* (e.g. [ta ljaw] *'star anise'*)

a. [**t**a **l**jaw] → d. [**l**a **t**jaw]
Independent segment: CGVX *GV response*

b. [**t**a **l**^{j}aw] → e. [**l**^{j}a **t**aw]
Secondary articulation: C^{G}VX *CG response*

c. [**t**a **l**^{j}jaw] → f. [**l**^{j}a **t**jaw]
Double representation: C^{G}GVX *GG response*

If the Mandarin speaker assumes the segmentation in (3a) [**t**a **l**jaw], where [j] is an independent segment, then the onset swapping task does not affect the position of the glide. It is left behind with the diphthong [aw]. The speaker will arrive at the codeword response of (3d) [**l**a **t**jaw]. This is labeled the "GV response", since the GV sequence in the original word remains the same in the codeword. However, if the speaker segments the glide as the secondary articulation of the onset, as in (3b) [**t**a **l**^{j}aw], the glide will be moved alongside the onset as a single unit CG. The response will be (3e) [**l**^{j}a **t**aw], labeled as the "CG response" to signify the affinity between C and G in the movement. A third possibility is to treat the glide as both part of the onset and an independent segment of its own. This is the double representation segmentation of (3c) [**t**a **l**^{j}jaw], which results in a codeword response with two glides, as in (3f) [**l**^{j}a **t**jaw]. The label for this type of response is the "GG response", indicating the double copies of G.

Both the GV response and the GG response indicate that the prenuclear glide is an independent segment, or an articulatory target for the speaker. The CG response, on the other hand, shows that the glide is an unalienable component of the onset, and thus not an independent segment.

4 Experiment I: online

Experiment I is conducted during the lockdown period, and therefore assumes an online format. It is run on the PennController IBEX Farm (Zehr & Schwarz 2018).

4.1 *Materials* The disyllabic word list used for the codeword language game includes 48 items containing a palatal glide /j/ in one of its syllables, and another 48 filler items that have no glide. The 48 glide items are divided in half along 3 parameters: onset place (non-palatal vs. palatal), glide position (1st syllable vs. 2nd syllable), and vowel variation (post-glide vowel displays variation depending on the presence or absence of glide vs. no variation). There are also 10 glide-less training items.

The 96-item word list is split into 2 groups. Every participant is randomly assigned to group A or group B. In each group, there are 24 /j/ items and 24 filler items. Within the 24 /j/ items, there are 3 items in each of the 8 categories determined by the 3 binary parameters. The 10 training items are the same for both groups.

The stimuli are presented in written Simplified Chinese on a computer screen. The codeword responses are collected in written form as well. Since only attested syllables in the Mandarin lexicon can be typed as Chinese characters, test items are selected on the basis that their corresponding codeword options can all be written with Chinese characters.

All non-palatal onset items in the experiment have 3 available codeword responses, like [ta ljaw] 'star anise' in Table 3. For palatal onset items, however, only 2 responses are available, namely CG and GG. As shown for the test item [ta tɕ^{h}jaw] 'big bridge' in Table 3, the GV response [tɕ^{h}a tjaw] is unavailable, since the initial syllable *[tɕ^{h}a] is not writable in Chinese. Nevertheless, palatal-initial test items are still interpretable, since either a GV or a GG response qualifies as evidence for independent glide segmentation.

	Non-palatal onset item			*Palatal onset item*		
Test item	[ta ljaw]	大料	'star anise'	[ta tɕ^{h}jaw]	大桥	'big bridge'
GV response	[la tjaw]	辣掉	*available*	[tɕ^{h}a tjaw]	? 掉	*unavailable*
CG response	[l^{j}a taw]	俩到	*available*	[tɕ^{hj}a taw]	恰到	*available*
GG response	[l^{j}a tjaw]	俩掉	*available*	[tɕ^{hj}a tjaw]	恰掉	*available*

Table 3: Availability of codeword responses

An easy solution to the unavailability of Chinese characters problem is to elicit participant response in alphabetical writing. For speakers from China, this is *pinyin*, the romanization script for Mandarin. But since the prenuclear glide is spelled as high vowels in pinyin, there is a concern that participants might resort to simple letter manipulation, without consulting their own phonological grammar. It is for this reason that Chinese characters are chosen over pinyin for text input in the experiment.

4.2 *Participants* 10 native speakers of Mandarin participated in the online experiment. They are all born in China. Their age ranges from 21 to 58. 7 participants identified as male, and 3 as female. Consent was given electronically, prior to the beginning of the experiment. There was no monetary compensation given for participation in the experiment.

4.3 *Procedure* At the beginning of the language game experiment, the participants are told that they are invited to learn a new method of encoding secret messages in Mandarin Chinese. They are given the explicit instruction, written in Simplified Chinese, to swap the “consonants” or “initials” of a disyllabic syllable. “Initial”, or “声母”, is the term used in conventional Chinese grammar for onsets. Two examples are displayed on the screen (see (4)), written in both Simplified Chinese and pinyin. The example in (4b), [xɑŋ2 ʈʂow1] ‘Hangzhou’ → [ʈʂɑŋ3 xow4], is provided so that the participants know that the tones do not have to remain the same in the codeword. Such leeway is put in place, so that the participants have less trouble looking for Chinese characters to represent the syllables in their codeword response. They are informed that the codeword can either be a real word or not. No instruction regarding the prenuclear glide is provided.

(4) *Codeword language game instruction examples*

a.	xaj3 nan2	‘Hainan’	→	naj3 xan2	*Tones remain the same*
b.	xɑŋ2 ʈʂow1	‘Hangzhou’	→	ʈʂɑŋ3 xow4	*Tones have changed*

After the instruction phase, the participants are given the chance to try their hand on encoding a few words. Each trial in the training phase is presented as a forced-choice task. For every disyllabic training item to be encoded, two choices are displayed on the screen. One is the onset-swapping codeword. The other is the result of swapping entire syllables. Sometimes, the two choices both correspond to real words. At other times, only one of them is a real word. This is to show participants that they ought to ignore whether the codeword is a real word or not in their own encoding. The order in which these two choices appear is randomized. The participant makes their selection by clicking on the button with the codeword choice. This leads them to a feedback page, which tells them what the correct codeword is.

The training phase is followed by the experiment phase, where the 48 experiment items (24 glide items, 24 filler items) appear in random order, broken into 2 sessions. There is a break between the two sessions, in which the participants are invited to read about a Chinese language fun fact. In each trial, the test item appears on the screen in written form. Participants are asked to type in their codeword. The task is self-paced.

At the end of the experiment, there is an exit survey, in which information on age, gender, and place of birth are collected. Participants are also asked to rate the difficulty of the experiment and optionally comment on which words they find the most difficult to encode. They are also asked if they wrote down any notes in pinyin to help them complete the encoding task. This is done to check if the explicit use of pinyin has an effect on their response.

4.4 *Results* The participants’ responses in Simplified Chinese text were converted into pinyin using the *Talking Chinese to Pinyin/Zhuyin Converter* (Purple Culture 2021), which is then translated into IPA. After uninterpretable responses (syllable swap, wrong onset, wrong rhyme, deletion of glide, etc.) were discarded, the interpretable responses were labeled as GV, CG, and GG accordingly. A generalized linear model (GLM) with a binomial distribution is fitted to the data, using RStudio (R Core Team 2022). Since both GV and GG responses pointed to an independent segment hypothesis, they are both coded as 1, while CG responses are coded as 0. Predictors include onset place, vowel variation, and glide position. Significance level is taken at $p < 0.05$. Only onset place passes the significance threshold ($p = 0.0003$). Neither vowel variation ($p = 0.73$) nor glide position ($p = 0.45$) is a significant predictor for participant behavior.

Figure 1 shows the participants’ written responses ordered by type of onset in two bar plots. For non-palatal onset items (left bars), GV responses (black) are the overwhelming favorite among Mandarin speakers. There are a few GG (gray) and CG (white) responses as well. For palatal onset items (right bars), where the GV response is no longer available, speakers are evenly split between the GG and the CG response.

Recall that both GV and GG responses indicate that the speaker has parsed the glide as an independent segment. A CG response, on the other hand, show that the speaker treats the glide as part of the onset. It can be observed that the participants are significantly more likely to opt for a GV or GG response for non-palatal onset items, compared to palatal onset items. The palatal onset items lead to significantly more CG responses. These results show that the palatal glide /j/ is significantly more likely to be treated as an independent segment by Mandarin speakers when it follows a non-palatal onset, compared to a palatal onset.

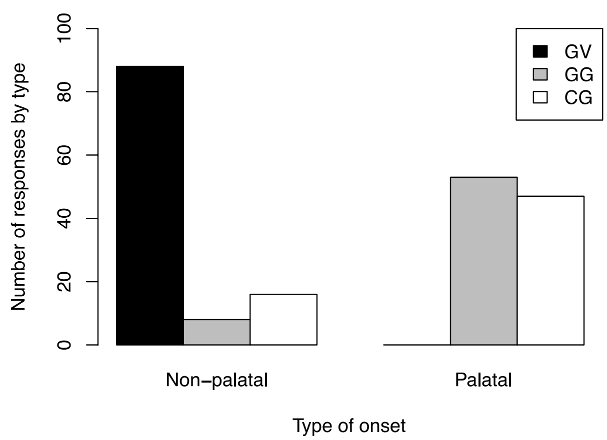

Figure 1: *Participants written codeword responses by type, listed side by side for non-palatal onset items (left) and palatal onset items (right).*

4.5 *Problems* The form in which codeword responses are collected in the online experiment creates problems. Firstly, written Chinese allows no room for any unattested syllables. When a speaker wants to produce a GV response for a palatal onset item like [ta tɕʰjaw] ‘big bridge’, they have no way of representing the unattested syllable *[tɕʰa], even if they believe it is the correct way of encoding the word. Another problem is that the process of typing Chinese characters, for most speakers, involves the medium of pinyin. Participants need to think about pinyin before they can key in their codeword in written Chinese.

To address these problems, I have conducted a second, in-person, edition of the experiment. In experiment II, the stimuli are presented in audio form and spoken responses are collected, as opposed to written ones. This way, speakers get the chance to produce unattested syllables. The interfering effect of pinyin is reduced as well.

5 Experiment II: in-person

The second experiment largely follows the methods of the first. It is conducted in a soundproof recording booth at the MIT Phonetics Lab.

5.1 *Materials* The word list for the second experiment consists of 64 glide test items and 36 glide-less filler items. There are another 20 glide-less items used for the training phase. The 64 glide test items are made up of 24 /j/ items, 24 /w/ items, and 16 /ɥ/ items. This paper, like the poster presentation at AMP, focuses on the /j/ items. These 24 items are divided evenly using 3 parameters, namely onset place, glide position, and vowel variation. Onset place (non-palatal vs. palatal) and glide position (1st syllable vs. 2nd syllable) each has two levels, following the online experiment. Vowel variation, on the other hand, is controlled slightly differently in the in-person experiment. In experiment I, vowel variation is only controlled for the syllable containing the glide. The vowel in the other syllable, which has no glide, is not controlled. This is schematized in Table 4, where a vowel that displays variation is in **boldface** and underlined. In level 1, the post-glide vowel in CG**V**X shows variation, but the vowel in the glide-less syllable can either vary (C**V**X) or not (CVX). In experiment II, however, the vowels in both syllables are controlled for. There are 3 levels. In level 1, only the post-glide vowel varies, whereas the vowel in the other syllable shows no variation. Level 2 is the mirror image of level 1, where the vowel in the glide-less syllable varies, but not the vowel in the glide syllable. Level 3 has no vowel variation in either syllable.

	Vowel variation	Syllable with glide	Syllable with no glide	
Experiment I	Level 1	CG**V**X	C**V**X or CVX	*Not controlled*
	Level 2	CGVX	C**V**X or CVX	
Experiment II	Level 1	CG**V**X	CVX	*Controlled*
	Level 2	CGVX	C**V**X	
	Level 3	CGVX	CVX	

Table 4: Vowel variation control across two experiments

Audio stimuli are created for the 100-item experiment word list, the 20-item training word list, as well as 20 codeword responses to the training items. A female native speaker of Mandarin from Shanghai is recorded producing the disyllabic words in isolation. The stimuli speaker is not a linguist and has no knowledge of the purpose of the experiment. The recording session took place in a soundproof booth at the MIT Phonetics Lab.

5.2 *Participants* 42 speakers of Mandarin participated in the in-person experiment. They include 33 native speakers, 8 heritage speakers, and 1 speaker who self-identifies as somewhere in between a native and a heritage speaker. 28 speakers were born in China, 5 in Taiwan, 1 in Singapore, and 8 in the US. The participants signed consent forms prior to the experiment and were compensated for their time.

Out of the 42 participants, the data produced by 33 speakers are analyzed. Reasons for discarding speaker data include incomplete experiment session, misunderstanding of the language game task, high percentage of uninterpretable responses, accidental noise during the experiment session, and technical problems with the recording file.

5.3 *Procedure* In experiment II, participants are not explicitly instructed on how disyllabic words are encoded in the language game. There is no mention of "consonants" or "initials". Instead, they are asked to figure out the encoding method by listening to pairs of original words and codewords on their own. They are also told to pay special attention to the sound of the words. The instructions on a computer screen are written in English, in order to accommodate heritage Mandarin speakers with varying degree of familiarity with written Chinese. The native Mandarin speaker participants also had no problem with reading English.

There are 2 demonstration phases and 2 training phases. In the first demonstration phase, there are 5 pairs of original words and their codewords. For each example pair, participants can listen to them by clicking on the two buttons that appear on the screen. No written form of the stimuli is displayed, either in Chinese character or pinyin. The demonstration phase is self-paced. Participants can listen to the example multiple times before moving on to the next page. There is no glide in any of the demonstration stimuli. The codewords have identical tonal patterns to the original words.

Then, participants are invited to try encoding 5 glide-less words on their own in the first training phase. For each word, the speaker first hears the original word to be encoded. They are asked to repeat the original word verbally first, before saying out loud what they think the codeword is. Afterwards, they can move on to the next page to check if their codeword matches the correct codeword, which can be played by clicking a button. No written form of the training stimuli or correct codeword is provided. The process is repeated in another demonstration phase and training phase, with new glide-less items. This is to make sure that participants who have difficulty figuring out the encoding method get another chance.

Prior to the experiment phase, the participants are told that for some of the items, there is no correct answer, and that they should say what comes to their mind first. During each experiment trial, the participant hears the test stimulus, repeats it out loud, before producing their codeword response. No feedback is given. The entire process is recorded in the soundproof booth. There are 100 trials, split into 4 sessions of 25. The order of test items is not entirely randomized. Measures are put in place to make sure the first 9 items are relatively easy to encode. The first 3 items are all glide-less items like the ones speakers have heard in the training phase. The next 6 items contain are /j/ items and /w/ items, with no palatal onsets. The order within each of these two sets is randomized. The bulk of the experiment trials, with 91 test items, are completely randomized in order. The experiment phase is self-paced. No note-taking (on paper or electronic device) is allowed. During the breaks between the sessions, fun facts about the Chinese language (unrelated to segmentation) are displayed on the screen during the sessions.

At the end of the experiment, participants are asked to fill out an exit survey, which collects information on the participant's place of birth, age, and other Chinese dialect or language they speak. Heritage speaker participants are also asked questions on their level of familiarity with pinyin or *bopomofo* (used in Taiwan).

5.4 *Results* The participant codeword responses are transcribed in IPA by the author, who is also a native speaker of Mandarin. Sometimes, multiple utterances are produced for the same test item, in which case the final response is taken. Uninterpretable responses are discarded from analysis, with the same criteria as the first experiment. If a participant produces fewer than 80 interpretable responses, their data are discarded. The remaining responses are labeled as GV, CG, and GG accordingly.

The statistical analysis for the spoken responses in experiment II is identical to the one used for written responses in experiment I. A generalized linear model with binominal distribution is used. GV and GG responses are coded as 1, and CG as 0. Predictors include onset place, vowel variation, and glide position. Significance level is taken at $p < 0.05$. Onset place ($p = 1.9e^{-9}$) remains a significant predictor for speaker segmentation. In addition, both levels of vowel variation reach significance (CG**V**X CVX vowel variation: $p = 0.02$; CGVX C**V**X vowel variation: $p = 0.006$). Glide position is not a predictor for speaker segmentation ($p = 0.21$).

In Figure 2, the participants' spoken response types are illustrated in two bar plots, ordered by type of onset. Recall that a key purpose of the in-person experiment is to give participants the opportunity to produce unattested syllables that cannot be written. In Figure 1 for the online experiment, there is no written GV response recorded for the palatal onset items. In contrast, there are many instances of GV responses for palatal onset items in Figure 2. There are 45 palatal GV response tokens in total, in which 12 tokens include an unattested syllable like *[tɕʰa]. Another 33 GV response tokens involve a change of place of articulation for the onset, as in [ta **tɕʰ**jaw] → [**tʂʰ**a tjaw].

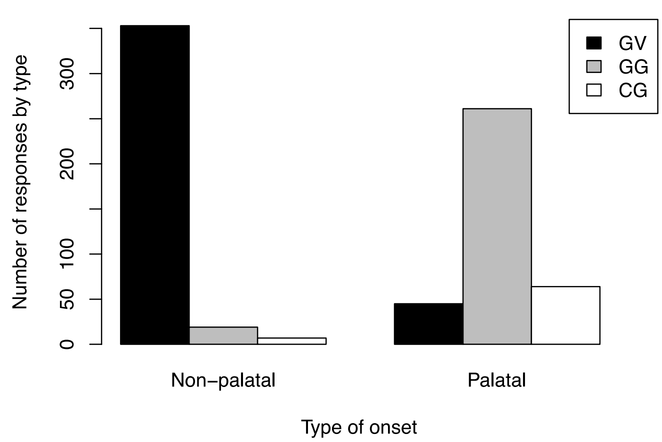

Figure 2: *Participants spoken codeword responses by type, listed side by side for non-palatal onset items (left) and palatal onset items (right).*

For non-palatal onset items, the GV response is most favored by the participants, replicating the results of experiment I. For palatal onset items, on the other hand, the GG response is preferred over the other two options. This is a departure from experiment I, where GG and CG responses are almost equal in number.

5.5 *Discussion* The main finding of experiment I is that non-palatal onsets and palatal onsets trigger different tendencies for glide segmentation by Mandarin speakers. This is replicated in experiment II. Mandarin speakers are significantly more likely to treat the palatal glide /j/ as an independent segment after non-palatal onsets, compared to palatal ones.

There are two points of differences between the results of the two experiments. Firstly, vowel variation is not a predictor for speaker response type in experiment I, but it reaches significance in experiment II. Secondly, for palatal onset items, participants in the online experiment do not have a clear preference between the GG and CG responses, whereas participants in the in-person experiment favor the GG response by quite a margin. It is possible that in the second experiment, the audio stimuli presentation and spoken response collection method have led to a stronger vowel faithfulness effect, which can account for the emergence of vowel variation as a significant predictor and the higher frequency of GG responses.

5.5.1 *Vowel faithfulness effect* Some vowels in Mandarin displays variation depending on whether it is preceded by a glide or not. A notable example is the low vowel raising rule. The low vowel surfaces as [a] before a dentoalveolar nasal coda in a CVN syllable. When it is preceded by a palatal glide in a CGVN syllable, it is raised to a mid [ɛ]. Some examples are shown in (6).

(6) *Low vowel raising*

a.	pan	'half'	c.	pjɛn	'change'
b.	tab	'egg'	d.	tjɛn	'electricity'

The sensitivity of the low vowel to its left-hand side glide neighbor means that in the language game experiment, a decision to remove the glide or not has consequences on the vowel quality. Specifically, it is the CG response that is impacted. If the original word contains a [jɛn] sequence like (7) [tjɛn paw] 'telegraph', its CG response (7e) [***p**ɛn **tʲ**aw] removes the palatal glide trigger for vowel raising. This creates a marked syllable *[pɛn]. Note that neither the GV response (7d) [**p**jɛn **t**aw] nor the GG response (7f) [**p**jɛn **tʲ**aw] has such a problem. To resolve the marked syllable *[pɛn], a speaker might choose to repair it, either by changing its vowel quality, as in (8b) [pan], thus violating IDENT-V, or inserting a palatal glide to satisfy both markedness and faithfulness constraints, as in (8c) [pjɛn], which is equivalent to a GG response.

(7) *CG response removes glide trigger for vowel raising* (e.g. [tjɛn paw] *'telegraph'*)

a.	[**t**jɛn **p**aw] *Independent segment: CGVX*	→	d.	[**p**jɛn **t**aw] *GV response*
b.	[**tʲ**ɛn **p**aw] *Secondary articulation: C^GVX*	→	e.	[***p**ɛ̱n **tʲ**aw] *CG response creates marked syllable*
c.	[**tʲ**ɛn **p**aw] *Double representation: C^GGVX*	→	f.	[**p**jɛn **tʲ**aw] *GG response*

(8) *Repairs to marked syllable* *[pɛn]

a.	[*pɛn tʲaw]	*Marked syllable*	*= Faithful CG response*
b.	[pan tʲaw]	*Not marked, but violates IDENT-V*	*= Unfaithful CG response*
c.	[pjɛn tʲaw]	*Not marked, does not violate IDENT-V*	*= GG response*

The Optimality Theory (OT) tableau in Table 5 illustrates the marked syllable problem and its repairs, by evaluating the codeword candidates for the test item 'telegraph'. Since the marked syllable *[pɛn] is only found in the CG response for the test item 'telegraph', it is only a problem for Mandarin speakers who segment the palatal glide as a secondary articulation. It is for this reason that the tableau assumes the secondary articulation segmentation.

Test item: [tʲɛn paw] *'telegraph'*			IDENT-C	*Cɛn	IDENT-V	DEP
a.	CG (faithful)	[pɛn tʲaw]		*!		
b.	CG (unfaithful)	[pan tʲaw]			*!	
☞c.	GG	[pjɛn tʲaw]				*
d.	GV	[pjɛn taw]	*!			

Table 5: *OT tableau for test item* [tʲɛn paw] *'telegraph'* (*secondary articulation segmentation*)

As shown in Table 5, the effect of prioritized vowel faithfulness predicts that if the post-glide vowel in the test item displays variation, the GG response is preferred over the CG response. On the other hand, if the vowel shows no variation, more CG responses are expected.

It is likely that in experiment II, vowel faithfulness effect is stronger than it is in experiment I due to the method of stimuli presentation and response collection. In experiment II, participants listen to audio test items, repeat them, and produce codeword responses by saying it out loud. More attention is given to vowel quality. In experiment I, on the other hand, participants do not hear the test items, and type in their responses with the mediation of a pinyin input system, which does not distinguish between [a] and [ɛ]. For instance, [pan] is written as *ban* in pinyin, while [pjɛn] is written as *bian*. The phonemic writing system might have made it easier for participants to ignore changes in vowel quality, leading to weaker vowel faithfulness effect.

To find out if vowel faithfulness effect plays a key role in experiment II, I have grouped the test items into ones that risk violating IDENT-V by the post-glide vowel and those that do not. The responses are further labeled as being "faithful" to the original vowel or serving as a "repair". The results are shown in stacked bar charts in Figure 3, with non-palatal onset items on the left, and palatal onset items on the right.

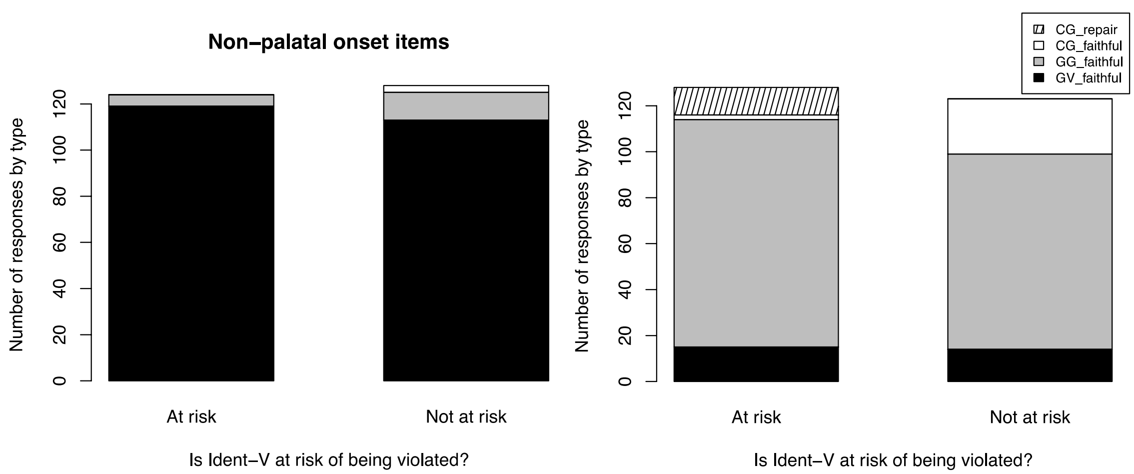

Figure 3: *Participants spoken codeword response types by whether vowel faithfulness is at risk, listed side by side for non-palatal onset items (left) and palatal onset items (right).*

For non-palatal onset items, CG responses are completely absent in test items that risk violating IDENT-V (1st bar from left). For palatal onset items, there are some tokens of CG response in test items that risk violating vowel faithfulness (3rd bar). However, the number of CG responses for this group of items are smaller compared to those test items that do not risk violating vowel faithfulness (4th bar). Note that the number of GV responses are roughly the same between the two types of palatal onset test items, which means that the decrease in CG response tokens in test items that risk violating IDENT-V is roughly converted to an increase in the instances of GG response. This is predicted by a vowel faithfulness effect.

Interestingly, the CG responses participants produce for palatal onset items that risk vowel faithfulness is composed mainly of CG repair responses (e.g. *[pɛn] → [pan]) (3rd bar). They violate IDENT-V. CG responses that maintain vowel faithfulness is represented by a tiny sliver in the graph. Within the speakers who produce CG responses, they seem more concerned with markedness than vowel faithfulness. No matter the relative ranking between markedness constraints and vowel faithfulness, it is clear that post-glide vowel variation presents a dilemma to Mandarin speakers in the experiment, because it potentially generates ill-formed syllables like *[pɛn].

Vowel faithfulness effect seems to play a role in predicting more GG responses than CG responses in experiment II, but it plays only a partial role. Vowel faithfulness effect does not explain why there is such a large increase in GG responses compared to experiment I. Nor does it account for the number of IDENT-V-violating CG repair responses that speakers produce. Further investigation is required.

5.5.2 *Speaker variation* The above discussion deals with the overall performance of the Mandarin speakers in experiment II. In this section, speaker-by-speaker results are inspected. Each speaker is given a score on a scale from 0 to 1 that represents how likely they treat the palatal glide as an independent segment. The independent segment likelihood score is calculated by taking the sum of GV and GG response tokens per speaker, and dividing it by the total number of responses uttered by the speaker. A score of 1 means they treat glides as an independent segment in all test items, using GV or GG responses throughout. A score of 0 suggests that they never treat the glide as an independent segment, always providing CG responses. If the speaker scores 0.8, it indicates that for 80% of the test items, they treat the glide as an independent segment. Each speaker has 3 scores, including a score for all items, one for non-palatal onset items, and another for palatal onset items. The scores of 33 speakers are plotted in Figure 4.

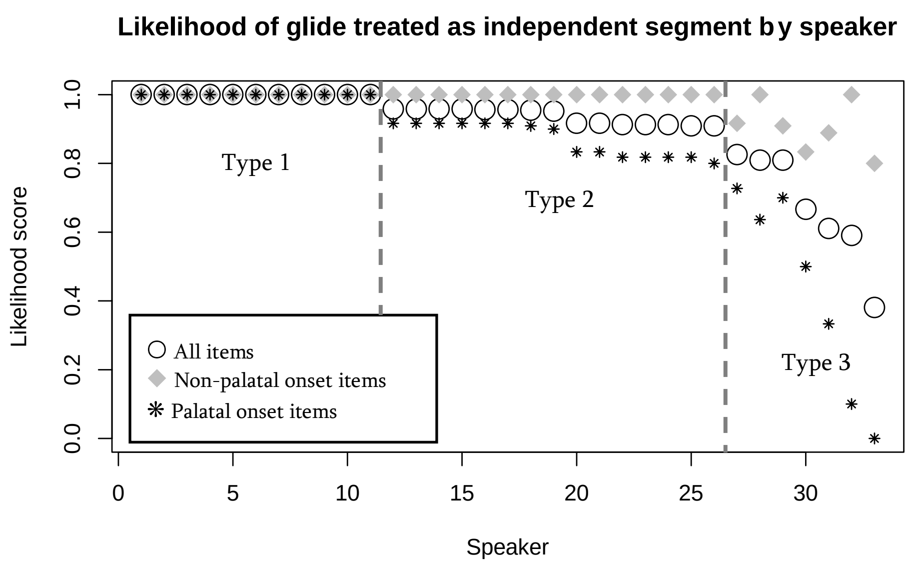

Figure 4: *Scatterplot of likelihood score of independent glide segmentation by speaker, for all items (white circle), for non-palatal onset items (gray diamond), and for palatal onset items (asterisk).*

As shown in Figure 4, there are 3 distinct types of speakers. Type I speakers (11 speakers) have a likelihood score of 1 consistently for all items, which means they treat the palatal glide as an independent segment across the board. Type II speakers (15 speakers) have a score of 1 for all non-palatal onset items, but their likelihood score is in the 0.8-0.9 range for palatal onset items. After non-palatal onset items, /j/ is always an independent segment to these speakers, whereas after palatal onset items, it is mostly, but not always, taken to be an independent segment. Both Type I and Type II speakers are relatively consistent in their glide segmentation. On the other hand, Type III speakers (7 speakers) exhibit much more intraspeaker variation. Not only do they have a relatively low likelihood score for palatal onset items, but they also score less than 1 for the non-palatal onset items (except for 2 speakers, who indeed score 1), which set them apart from the other two types of speakers. Nevertheless, every Type III speaker has a higher likelihood score for non-palatal onset items than they have for palatal onset items. They are more likely to treat the palatal glide as an independent segment when it is after a non-palatal onset than after a palatal onset. The segmentation behavior of Type III speakers is in line with the main finding across the two experiments, namely that onset place is the best predictor of glide segmentation. Note that the onset place effect we observe in the congregate speaker data is the result of the responses made by Type II and Type III speakers. Type I speakers, with their consistency across the board, do not contribute to the observed glide segmentation variation between onset places.

6 Conclusion

The debate on the segment status of the Mandarin prenuclear glide reflects the ambiguity in the phonological input in the language. If phonologists can propose various segmentation grammars to account for the Mandarin syllable inventory, then Mandarin speakers are faced with just as many choices when they are acquiring the language. The codeword language game experiments offer us a chance to see which segmentation grammar corresponds to the one that Mandarin speakers have learned, and whether the speakers converge on the same grammar.

Results from the online and in-person experiments indicate that there is much variation between Mandarin speakers in how they segment the palatal glide. Nevertheless, 3 types of speakers emerge. Type I and Type II speakers have acquired a glide segmentation grammar with relative consistency. They account for 26 out of the 33 speakers in the data. Type III speakers, on the other hand, display a degree of intraspeaker variation. Despite the interspeaker and intraspeaker variation, an onset place effect is detected. A non-palatal onset consonant makes it more likely for Mandarin speakers to segment the palatal glide as an independent segment, whereas a palatal onset makes it less likely. There is not a single speaker who participated in the experiment who shows a segmentation grammar that contradicts this rule.

The next step in this project is to find out how Mandarin speakers have arrived at their various segmentation grammars. Computational algorithms for segmentation learning will be compared, in order to test which model can best account for the widespread speaker variation observed in the codeword language game experiments.

References

Béland, Renée, and Régine Kolinsky (2005). One sound heard as two: The perception of affricates in Quebec French by Belgian French speakers. *Journal of Multilingual Communication Disorders* **3**.110–17.

Chao, Yuen-Ren (1931). 反切語八種 [Eight varieties of secret languages using Fan-ch'ieh]. *Bulletin of the Institute of History and Philology, Academia Sinica* **2:3**. 312-54.

Chao, Yuen-Ren (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* **4:4**. 363-97. [Reprinted in *Readings in Linguistics*, ed. by Martin Joos, 38-54. Washington, D. C.: American Council of Learned Societies, 1957.]

Downing, Laura (2005). On the ambiguous segmental status of nasals in homorganic NC sequences. In Marc van Oostendorp & Jeroen van de Weijer (eds.) *The internal organization of phonological segments*. Berlin: Mouton de Gruyter. 183–216.

Duanmu, San (2002). *The Phonology of Standard Chinese*. Oxford University Press.

Ladefoged, Peter, & Ian Maddieson (1996). *The sounds of the world's languages.* Blackwell.

Lin, Yen-Hwei (1989). Autosegmental treatment of segmental processes in Chinese phonology. PhD dissertation, University of Texas at Austin.

Gouskova, Maria, & Juliet Stanton (2021). Learning complex segments. *Language* **97:1**. 151-193.

Purple Culture (2021). Talking Chinese to pinyin/zhuyin converter. URL https://www.purpleculture.net/chinese-pinyin-converter/.

R Core Team (2022). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Shaw, Jason A., Sejin Oh, Karthik Durvasula, & Alexei Kochetov (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology* 38:3. 437-477.

Zehr, Jeremy, & Florian Schwarz (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.