LARGER SONORITY DIFFERENCE, LARGER LAG: GESTURAL COORDINATION IN SPEECH PRODUCTION

By

Yunting Gu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Linguistics—Doctor of Philosophy

2025

ABSTRACT

Sonority has been one of the most debated concepts in phonetics and phonology. Constraints involving sonority such as the Sonority Sequencing Principle (SSP), the Sonority Dispersion Principle (SDP), or the Syllable Contact Law have long been used by phonologists to understand syllable structure (Sievers, 1881, 1901; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Parker, 2002, 2011). However, there is no consensus on the phonetic basis of sonority, either in the articulation or the perception of speech (Albert, 2023). This dissertation explores sonority in speech production.

Current speech production theories do not predict the variation of gestural coordination relevant to sonority. However, sonority has been observed to be correlated with systematic variation in gestural coordination, based on CC clusters in Georgian (Crouch, 2022). Also, some observations (Gao, 2008; Shaw and Chen, 2019) suggested that sonority seems to be a factor that systematically correlates to CV gestural coordination variation. In my dissertation, I followed up on these previous studies and explored whether there is a positive correlation between sonority difference and CV lag (the gestural lag between a consonant and a vowel) in English and Mandarin. Based on corpus data of English, as well as Electromagnetic articulography (EMA) experiments participated by English and Mandarin speakers, I found that CV lag positively correlates with CV sonority difference in both languages.

In experiment 1, there were 32 English stimuli from the Wisconsin X-ray Microbeam Database (Westbury et al., 1990) used to test the main claim. Analyzing the corpus data suggested that there is a significant positive correlation between CV lag and sonority difference. To address the limitation of using an existing corpus and to provide a cross-linguistic comparison, EMA data of 24 English stimuli (experiment 2) and 26 Mandarin tone 4 stimuli (experiment 3) were collected and analyzed. Each set of stimuli in the EMA experiments was read 15 times in different randomized lists. When collecting EMA data, sensors were glued to the tongue tip, tongue blade, tongue dorsum, upper lip, and lower lip of each participant. All the kinematic data were annotated in Matlab using the *lp_findgest* algorithm of the *mview* package (Tiede, 2005), where the landmarks were labeled at

20 percent thresholds of peak velocity. The CV lag was computed by subtracting the target onset (onset of gestural plateau) timestamp of the consonant from the target onset timestamp of the vowel (Zhang et al., 2019; Durvasula and Wang, 2023). The sonority difference was quantified by subtracting the C sonority from the V sonority using the sonority scale in Parker (2011). Plots and mixed effects modeling was generated in R (R Core Team, 2017) where CV lag was modeled as a function of the sonority difference, with participant, stimuli, and C duration as random intercepts.

The finding is that for all the data, CV lag positively correlates to sonority difference significantly. Sub-groups of the stimuli controlled for consonant place of articulation or vowel height mostly exhibited the expected correlations. I also used consonant displacement and vowel displacement as estimates for jaw movement, and these findings suggest that jaw movement may not be a valid alternative account to the finding. The dissertation found a positive correlation between sonority and CV gestural coordination in English and Mandarin. If we make an assumption that larger lags are preferred within a syllable, the finding forms a basis to explain universal constraints such as the SSP and the SDP.

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the help and support of many people. First, I would like to express my sincere gratitude to all the professors, staff, and fellow students who have supported me along the way. I would like to express my gratitude to Dr. Karthik Durvasula. Thank you for always being there and supporting me during my PhD career. I appreciate your guidance, encouragement, understanding, and help along the way.

I would also like to thank Dr. Yen-Hwei Lin. I appreciate your patience, guidance, and help throughout my PhD career!

I would like to express my gratitude to Dr. Suzanne Wagner. You made the challenging processes of my PhD journey smoother, and I am grateful for all you have done.

Many thanks to Dr. Silvina Bongiovanni who has always been supportive and encouraging along the way.

I also would like to thank all the linguistic professors such as Dr. Brian Buccola, Dr. Alan Munn, Dr. Betsy Sneller, Dr. Scott Borgeson, Dr. Alan Hezao Ke, and Dr. Cristina Schmitt. You have educated, encouraged, and inspired me over the years.

Second, I would like to say thank you to my parents. They have been understanding and supportive! 谢谢爸妈！ I would also like to thank my family members who have encouraged me and helped me with some of my non-dissertation projects as participants or recruiters.

Third, many thanks to my friends, at MSU linguistic program and beyond. You have shared many moments with me. I especially would like to thank the people in my beginner's tennis club such as Wang, Chen, and Wu. I have shared unforgettable moments with you. Being part of the tennis club made my PhD journey special!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

SSP	Sonority Sequencing Principle
SDP	Sonority Dispersion Principle
C	Consonant
V	Vowel
CV lag	The gestural lag or timing difference between a consonant and a vowel
AP	Articulatory Phonology
EMA	Electromagnetic Articulography
T1	Tongue tip
T2	Tongue blade
T3	Tongue dorsum
T4	Tongue root
TT	Tongue tip
TB	Tongue blade
TD	Tongue dorsum
UL	Upper lip
LL	Lower lip
LA	Lip aperture
Lip	Lip aperture
PVEL	The peak velocity point
MAXC	The maximum constriction point
GON	Gestural onset
TON	Target onset
TOF	Target offset or release
GOF	Gestural offset or release

C-center The midpoint between target onset and target offset

Lag$_{GON}$ CV lag based on gestural onset

Lag$_{TON}$ CV lag based on target onset

CHAPTER 1

INTRODUCTION

1.1 Sonority and sonority-related constraints

The current study explores sonority in speech production. In this section, I first introduce the concept of sonority. Then, I describe phonological constraints that are related to sonority. Sonority has been one of the most debated concepts in phonetics and phonology. It is defined as a unique type of relative, non-binary featurelike phonological concept that potentially categorizes all speech sounds into a hierarchical scale (Parker, 2011). Even though there are different sonority scales in the literature, most phonologists identify the following sonority scale as in (1), where > means *more sonorous than* (Kenstowicz, 1994; Wright, 2004; Pons-Moll, 2008).

(1) vowels > glides > liquids > nasals > obstruents

As an abstract concept, sonority is a primitive that has long been used by phonologists to understand syllable structure (Sievers, 1881, 1901; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Parker, 2002, 2011). There are several generalizations or sonority-related phonological constraints that have been based on, and motivate, the abstract concept of sonority.

First, the Sonority Sequencing Principle (SSP) requires that each syllable should exhibit one peak of sonority in the nucleus, and that, cross-linguistically, a sonority rise (such as [pl]) is preferred in onsets over a sonority plateau (such as [pt]) which in turn is preferred over a sonority fall (such as [lp]) (Sievers, 1881, 1901; Greenberg, 1965; Pike, 1972; Hooper and Bybee, 1976; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Blevins, 1995; Parker, 2002, 2011). A schematic showing the optimal sonority contour can be found in Figure 1.1.

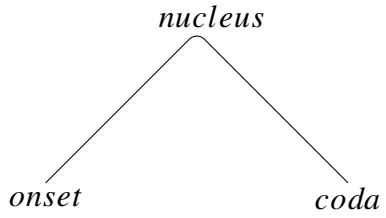

Figure 1.1 Optimal organization of sonority in a syllable. There should be one peak of sonority in the nucleus.

In Table 1.1, there are three sample syllables and their sonority contour at onsets. Sonority increases from bottom to top in this table, and the * symbol denotes the relevant sonority category of each sound. The trajectory of the * symbols shows that [pl] has an increasing sonority, while [pt] has a level sonority or sonority plateau and [lp] has a falling sonority. Note that sonority rise is preferred over sonority plateau over sonority fall. In this case, [pl] is preferred over [pt] over [lp] at onsets.

Vowels						
Glides						
Liquids		*			*	
Nasals						
Obstruents	*		*	*		*
	[p	l]	[p	t]	[l	p]

Table 1.1 Three sample syllables with rising ([pl]), plateau ([pt]), and falling ([lp]) sonority at onsets. Sonority increases from bottom to top, and the * symbol denotes the relevant sonority category of each sound. The trajectory of * shows that [pl] has an increasing sonority, while [pt] has a level sonority or sonority plateau and [lp] has a falling sonority.

Cross-linguistically, there appear to be violations of the SSP, meaning there are sonority fall at what appear to be onsets or sonority rise at what appear to be codas (Yin et al., 2023). For instance, nasal-stop and sibilant-stop onset clusters (such as *skill* [sk], *speak* [sp] in English) have been found in many languages. However, these apparent violations may not be considered as evidence against the SSP since some segment such as [s] can be analyzed as extra-syllabic, which means that [s] is not part of the onset (Cho and King, 2003; Parker, 2011). In Figure 1.2, [sk] in *skill* [skɪl] is an onset cluster, violating the SSP. In Figure 1.3, [s] is not part of the onset but rather an appendix, and the syllable structure is not violating the SSP (Vaux and Wolfe, 2009).

Furthermore, some apparent violations may not be true violations if alternative sonority scales are assumed. Regarding the sonority hierarchy of obstruents, while in Berent et al. (2007), fricatives are assumed to be more sonorous than stops, Parker (2002, 2008, 2011) assumed that *voiced* obstruents are more sonorous than *voiceless* obstruents, as in Table 1.3. In Table 1.3, there is a partial sonority scale where a more sonorous natural class of sounds is indicated by a larger value. Jespersen (1904) also provided a sonority hierarchy as in Table 1.2, where voiceless stops

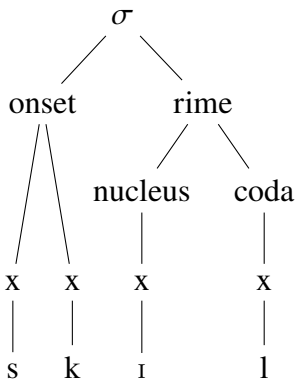

Figure 1.2 Onset cluster [sk] in *skill* [skɪl].

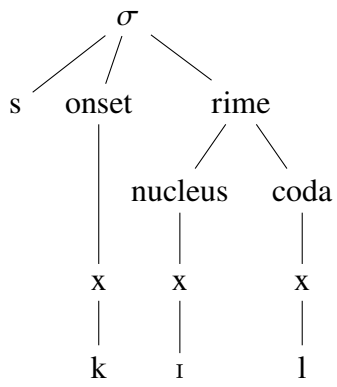

Figure 1.3 Extra-syllabic [s] in *skill* [skɪl].

and fricatives are similar in terms of sonority.

1	voiceless stops
	voiceless fricatives
2	voiced stops
3	voiced fricatives

Table 1.2 Partial sonority hierarchy in Jespersen (1904). The sonority index uses a larger value for more sonorous classes.

Natural class	**Sonority index**
voiced fricatives	6
voiced stops	4
voiceless fricatives (including [h])	3
voiceless stops (including [ʔ])	1

Table 1.3 Partial hierarchy of relative sonority (Parker, 2002, 2008, 2011).

When we analyze the apparent violations of SSP, for example, in the sibilant-stop onset violation cases, if we assume sibilants are more sonorous than stops, then sibilant-stop onset is a violation of the SSP. However, if we assume that voiced fricatives and stops are more sonorous than voiceless fricatives and stops as in Parker (2008), then some sibilant-stop onsets — those with voiceless sibilants and voiced stops — are not true violations of the SSP. Even though voicing assimilation on onset clusters is common, there are languages such as Georgian, Khasi, and Bilaan that have mixed voicing onset clusters (Kreitman, 2010). In Mondern Hebrew, [sd] and [sg] are licit onset clusters according to Kreitman (2010) and they are not violating the SSP if the sonority scale in Parker (2008) is assumed. Indeed, considering alternative assumptions does not account for

all the apparent violations of the SSP. If both extra-syllabicity and alternative sonority scales are considered, the SSP can potentially still be maintained as a cross-linguistic generalization, as long as one is clear that it refers to syllable internal sequences, and not just any sequences. This is similar to the opinion mentioned in Parker (2011).

Second, the Sonority Dispersion Principle (SDP) specifies that in a syllable, from onset to nucleus the sonority difference should be maximized and that from nucleus to coda the sonority difference should be minimized (Steriade, 1982; Clements, 1990; Parker, 2011; Xhaferaj et al., 2022). This predicts that [ta] is better formed than [na] than [la]. Also, the SDP favors syllables that end in a vowel. Clements (2009) observed that at codas, few languages prefer obstruents over sonorants. The languages that allow syllable-final stops and fricatives usually place restrictions on them. Regarding CCV syllables, a preferred syllable has a maximal sum of all the sonority differences for the onset to nucleus part. This study explores the relationship between sonority difference and gestural timing in CV syllables. The main claim can potentially provide a reason about why larger sonority difference is preferred for CV syllables.

Third, the Syllable Contact Law specifies that at syllable boundaries, a larger sonority decrease is preferable (Hooper and Bybee, 1976; Murray and Vennemann, 1983). In other words, at syllable boundaries, the coda A and onset B of the following syllable have a and b as their sonority values. Structure A.B would be more preferable if a-b is larger (Hooper and Bybee, 1976; Murray and Vennemann, 1983). If three consonants A, B and C have sonority values such that A < B < C, then A.C[1], as compared to B.C, is the preferred sequence at the syllabic boundary. For example, the Syllable Contact Law requires that [ɑl.tɑ], with falling sonority, is preferred over [ɑt.lɑ], with rising sonority (Seo, 2011). Another set of specific language examples comes from Korean, where Davis and Shin (1999) used the Syllable Contact Law to analyze Korean phonological processes such as obstruent-nasalization, n-lateralization, l-nasalization, and nasalization of (non-coronal) obstruent-liquid sequences as in Table 1.4.

[1]"." is used to refer to a syllable boundary.

obstruent-nasalization	/sip-nyən/	[sim.nyən]	‘ten years’
n-lateralization	/non-li/	[nol.li]	‘logic’
l-nasalization	/kam-li/	[kam.ni]	‘supervision’
nasalization of obstruent-liquid sequences	/pəp-li/	[pəm.ni]	‘principle of law’

Table 1.4 Korean examples showing the Syllable Contact Law from Davis and Shin (1999).

Lastly, Steriade (1982) and Selkirk (1984) pointed out that there are language-specific requirements for segments of a tautosyllabic consonant cluster to be separated by minimum sonority differences. Parker (2008) interpreted Steriade (1982) and Selkirk (1984) that there is some restriction on the sonority distance between tautomarginal consonant clusters in some languages. For instance, in Spanish /pl/ is possible but not /pn/ since the sonority distance between /p/ and /n/ is not large enough.

As pointed out by Clements (2005), not all principles about sonority are strictly independent. For instance, the Syllable Contact Law is closely related to the Sonority Dispersion Principle (SDP) and may partially derive from it (Clements, 2005). Since the SDP requires that syllable coda prefers high sonority and syllable onset prefers low sonority, a preceding syllable ending in high sonority will form a decrease in sonority when the following syllable onset is of low sonority. Therefore, individual syllables conforming to the Sonority Dispersion Principle are likely to obey the Syllable Contact Law as well (Clements, 2005). Clements (2005) mentioned that it is necessary to have separate principles or laws because some languages obey one sonority principle but not the other.

Examining the SSP, SDP, Syllable Contact Law, and the restriction on sonority distance mentioned above suggests that there is some optimal relative sonority value required or expected for each part of the syllable. The general requirements regarding the sonority of syllable structures focus on the *sonority difference* of adjacent sounds, within one syllable or across syllable boundaries. The four principles suggest an optimal syllable should have peak sonority in the nucleus, with onset sonority rising to the peak and falling coda sonority. Furthermore, the optimal syllable should have a larger sonority difference at its beginning, either CC or CV, and its coda should avoid serving as the beginning of the rising sonority onset of the following syllable.

1.2 Controversy on the basis of sonority

There is no consensus on the phonetic basis of sonority, either in the articulation or the perception of speech (Albert, 2023). I have discussed some phonological generalizations involving sonority. However, there is far less clarity and understanding of the generalizations discussed above that use sonority as a primitive. There have been debates about whether sonority is a primitive or is derivable from other phonetic factors. Some deny the existence of sonority as a primitive, and instead opt to derive it from phonetic properties as (a) a complex function of the acoustics (Ohala, 1990; Ohala and Kawasaki, 1997), (b) a correlate of intensity (Parker, 2008, 2011; Gordon et al., 2012; Ladefoged and Johnson, 2014), (c) a correlate of pitch intelligibility (Albert, 2023), (d) perceptual cue (Henke et al., 2012), (e) articulatory openness (Mattingly, 1981), and (f) articulatory timing (Chitoran, 2016).

Here, I lay out a few arguments about the different ways to derive sonority. For instance, Ohala and Kawasaki (1997) do not believe sonority exists as a primitive. Instead, they argued that the degree of *modulation* should be used to account for phonological universals. This degree of modulation is measured by various acoustic parameters such as amplitude, periodicity, spectral shape, and F0.

Another way of deriving sonority comes from Henke et al. (2012), who argued that a perception cue approach could explain the SSP, the syllable contact law, as well as the unmarked status of the CV syllables. According to Henke et al. (2012), each natural class of sounds has its internal cues of different robustness levels, in terms of manner cues, voicing cues, and place cues. A partial summary can be found in Table 1.5. Moreover, the natural classes also differ in their ability to carry the cues of their adjacent sounds. The internal and carrier cues of each natural class determine the preference for organizing sounds. Henke et al. (2012) argued that this way of accounting for phonological universals has advantages over the SSP in that it can also account for violations of the SSP such as the sibilant-stop cluster at onsets. For instance, they argued that fricatives have internal cues, and therefore, they can bear more gestural overlap than sounds that rely on transitions. Specifically, sibilant fricatives are the least dependent on formant transitions

and therefore are expected to be surrounded by obstruents. They also argued that vowels have robust internal cues in terms of manner cues, voicing cues, and place cues and they are also good carriers of three kinds of cues. Therefore, vowels are optimal as the nuclei of syllables.

class	**manner**	**cues**	**voicing**	**cues**	**place**	**cues**
	internal	carrier	internal	carrier	internal	carrier
vowels	robust	good	robust	good	robust	good
sibilant fricative	robust	medium	medium	poor	robust	poor
stops	poor	poor	poor	poor	poor	poor

Table 1.5 Partial summary of cue robustness in Henke et al. (2012).

Related to the current speech production study, Chitoran (2016) claimed that "the sonority hierarchy can be best understood in its relation to articulatory timing" (p. 46). In the dissertation, I follow up on the primary intuition laid out in Chitoran (2016) that there is a relationship between the sonority hierarchy and articulatory gestural timing.

1.3 Quantifying sonority difference

As mentioned earlier, the phonological constraints related to sonority, including the SSP, SDP, Syllable Contact Law, and the restriction on sonority distance, suggest that there is some optimal relative sonority value expected for each part of the syllable. The requirements regarding the sonority of syllable structures focus on the *sonority difference* of adjacent sounds, within one syllable or across syllable boundaries. In operationalizing the intuition in Chitoran (2016) that there is a relationship between the sonority hierarchy and articulatory gestural timing, it is worth noting that while sonority relates pairs of sound classes in a scale, gestural timing relates two proximal gestures. Therefore, in order for sonority to be reducible to gestural timing, one needs to talk about the sonority of proximal gestures. To put it another way, generalizations regarding sonority and syllable structure can be boiled down to a requirement for sonority difference between adjacent segments. For this reason, I specifically quantified the *sonority difference* between two adjacent segments and explored its relation to gestural timing.

To quantify sonority difference, I considered many proposed sonority scales that are subtly different in the literature (Clements, 1990; Kenstowicz, 1994; Mielke, 2008; Parker, 2008; Kang

et al., 2011). There were controversies on whether rhotics are more sonorous than laterals (Hall, 2002; Parker, 2002), or laterals are more sonorous than rhotics (Hankamer and Aissen, 1974). Also, while in Berent et al. (2007), fricatives are assumed to be more sonorous than stops, Parker (2002, 2008, 2011) argued, based on intensity measurements, that *voiced* obstruents are more sonorous than *voiceless* obstruents. Ultimately, I chose to implement the sonority scale developed by Parker (2002, 2008, 2011) that is shown in Table 1.6.

Natural class	**Sonority index**
low vowels	17
mid peripheral vowels (not [ə])	16
high peripheral vowels (not [ɨ])	15
mid interior vowels([ə])	14
high interior vowels ([ɨ])	13
glides	12
rhotic approximants	11
flaps	10
laterals	9
trills	8
nasals	7
voiced fricatives	6
voiced affricates	5
voiced stops	4
voiceless fricatives (including [h])	3
voiceless affricates	2
voiceless stops (including [ʔ])	1

Table 1.6 The hierarchy of relative sonority (Parker, 2002, 2008, 2011).

The further nuance provided by the sonority scale in Parker (2002, 2008, 2011) is argued to be necessary to account for more complex syllabification patterns observed in some languages. For example, in Imdlawn Tashlhiyt Berber, syllabification is staged and a fine-grained sonority scale separating vowel height, as well as voicelessness in fricatives and stops, is necessary (Dell and Elmedlaoui, 1985). Table 1.7 showed that the distinction between low and high vowel is necessary in a sonority scale to yield the right syllabification in a stage-wise manner (top-to-bottom). Basically, when formulating syllabification, one associates a core onset-nucleus syllable with any sequence (Y)Z, where Z is a low vowel, a high vowel, a liquid, a nasal, a fricative, or a stop. In the example

in Table 1.7, the syllable with a low vowel is syllabified first, then the one with a high vowel. Note that "I" stands for [+son, -cons, +high, -back, -round], and U stands for [+son, -cons, +high, +back, +round]. Without considering the stages and the sonority of different segments, the right syllabification cannot be accounted for.

	[t-IzrUal-In]
low vowel	t-Izr(wa)l-In
high vowel	(t-i)zr(wa)(l-i)n
liquid	(t-i)(z$\underset{\circ}{\text{r}}$)(wa)(l-i)n

Table 1.7 Imdlawn Tashlhiyt Berber example showing staged syllabification (Dell and Elmedlaoui, 1985). The stages of syllabification are presented top-to-bottom. Sensitivity to vowel height differences is observed above. I stands for [+son, -cons, +high, -back, -round], and U stands for [+son, -cons, +high, +back, +round].

Above, I showed that some languages like Imdlawn Tashlhiyt Berber need a more nuanced sonority scale for syllabification. Since ultimately I expect the main claim in the current study to be generalized to all languages, a more nuanced scale is preferred. To sum up, I used the specific scale in Parker (2002, 2008, 2011) because: (a) it is phonetically grounded based on the estimated average intensity of the relevant sound class; (b) it is intended to cover all speech sound categories; (c) the scale is quantified in a clear way; (d) while providing a much more nuanced sonority scale, its relative ranking of major classes is consistent with (1), which accords with other sonority scales (Clements, 1990; Kenstowicz, 1994; Smolensky, 1995; Clements, 2005); (e) it has the potential to be used in cross-linguistic contexts since some languages require more nuanced sonority hierarchies.

Using the sonority index in Table 1.6, I am able to operationalize the independent variable *sonority difference* between a consonant and the subsequent vowel as follows: I subtracted the sonority index of the specific C from that of the following V, according to the scale in Table 1.6. For instance, the sonority difference of [ba] is 13 since $\text{sonority}_{\text{V}}$ - $\text{sonority}_{\text{C}}$ = $\text{sonority}_{\text{low vowel}}$ - $\text{sonority}_{\text{voiced stop}}$ = 17 - 4 = 13. As will be elaborated below, this *sonority difference* forms the independent variable in the dissertation. The dependent variable is articulatory gestural lag, and it will be discussed later in this chapter.

It is important to note while the use of the scale from Parker (2002, 2008, 2011) has the above advantages, it does imply that sonority hierarchy is a linear scale. Furthermore, I recognize that many researchers consider sonority to be a relative notion. I therefore follow up on the main omnibus analyses with a set of more nuanced comparisons specifically meant to address the concerns.

1.4 Theories of gestural coordination in speech production

This dissertation is about the relationship between sonority and gestural timing in CV syllables. In this section, I am going to lay out claims about gestures and different claims related to gestural coordination in speech production. Phonological representations are characterized in terms of gestures and the relations of gestures where the gesture is a basic unit and a relatively abstract concept (Browman and Goldstein, 1989, 1992).[2] Gestures are events that unfold during speech production, and these events consist of the formation and release of constrictions in the vocal tract. The consequences of gestures can be observed in the movement of speech articulators. A schematic illustration of gesture in Gafos (2002) is shown in Figure 1.4, where gestural onset (GON), target onset (TON), C-center (the midpoint between target onset and target offset), target offset or release (TOF), and gestural offset or release (GOF) are denoted from left to right. Many phonologists in speech production do not identify the C-center point as Gafos (2002).

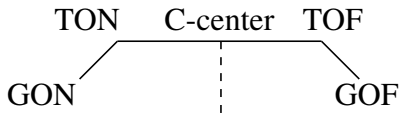

Figure 1.4 A sample gesture according to the view of Gafos (2002).

Within the AP framework, the coupled oscillator model of syllable structure argues that syllabic structure is expressed articulatorily in differential timing relations (Browman and Goldstein, 2000; Hermes et al., 2013; Iskarous and Pouplier, 2022). It is commonly assumed that the gestural coordination pattern is based on the type (consonant or vowel) and position of the gesture. For

[2]This is the claim of Articulatory Phonology (AP). Even though the current study does not necessarily operate under the AP assumptions, the AP framework is reviewed here since a) it is widely assumed in speech production studies and b) it provides a basis for understanding the timing relationships between articulations that are ultimately the focus of this dissertation. The two reasons why AP is not assumed will be mentioned later in this section when relevant.

a consonant-vowel (CV) syllable, AP claims that the consonantal and vowel gestures are timed synchronously (Saltzman and Munhall, 1989; Browman and Goldstein, 2000; Goldstein, 2011; Pouplier, 2020; Krivokapić, 2020; Liu et al., 2022), as in Figure 1.5.[3]

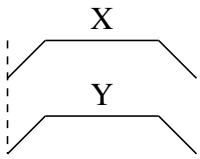

Figure 1.5 Synchronous coordination.

There are other opinions on CV timing as well. For instance, Shaw et al. (2021) viewed the distinction between synchronous and sequential coordination slightly differently from the classic AP view by Browman and Goldstein (1989, 1992). Specifically, they argued that if two gestures start at the same time, they have a synchronous relationship; if two gestures have a sequential relationship, they are coordinated by end-gestural timing as in Figure 1.6. Shaw et al. (2021) pointed out that an observed lag could come from the positive lag of synchronous coordination or a negative lag of end-gestural timing. One interpretation of their claims is that synchronous or sequential relationships are not pre-determined by segmental types of consonants or vowels. Rather, the phasal relationships can be used to describe gestural coordination that fulfills the specific timing requirements. Essentially, if one were to generalize the observation of consonant sequences to all segments, then one would subsequently claim that CV timing could be sequential.

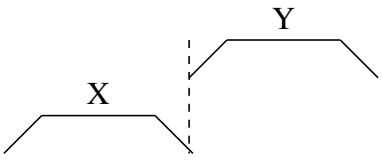

Figure 1.6 Offset-onset timing, sequential timing, or end-gestural timing.

Furthermore, Durvasula and Wang (2023) generalized the findings of Shaw et al. (2021) and argued that if two gestures belong to a pair of adjacent segments, then they have a sequential timing

[3]One reason why the current study does not operate under AP assumptions is that timing differences between C and V in CV syllables were observed. Therefore, C and V are not strictly coordinated synchronously.

relationship, wherein the second gesture is timed to the end of the first gesture. According to them, CV timing should have offset-onset alignment or a sequential timing relationship as in Figure 1.6.

Moreover, Nam (2007) proposed the *split-gesture* hypothesis which suggests that a stop consonant can be split into two sub-gestures[4], one as a closure sub-gesture and another as a release sub-gesture. Each sub-gesture has a synchronous timing relationship with the following vowel while maintaining a *sequential* relationship (as in Figure 1.6) with each other. The sequential timing specifies that a second gesture starts at the extreme displacement point of the abstract oscillatory cycle of the first gesture. The split-gesture hypothesis predicts that the vowel onset is being timed to roughly the 12.5%-16.7% point in the stop articulation (Durvasula and Wang, 2023). This is because gestures typically require 240° - 360° of their internal clock cycle to reach their target configuration (Browman et al., 1990). Nam (2007) assumes that the vowel gesture has a 60° phase difference with the closure sub-gesture and a -60° phase difference with the release sub-gesture. If the closure sub-gesture (CLO in Figure 1.7 and 1.8) and the release sub-gesture (REL in Figure 1.7 and 1.8) both require 240° - 360° of their internal clock cycle to reach their target configuration, to satisfy the vowel gesture coordination condition, the two boundary cases for the CV coordination according to the split-gesture hypothesis can be found in Figure 1.7 and 1.8. Specifically, if we assume that the release gesture (REL) and closure gesture (CLO) requires 240° of their internal clock cycle to reach their target configuration, the vowel gesture is coordinated to the 60°/(60°+60°+240°)=16.7% of the whole C gesture as in Figure 1.7. On the other hand, in another boundary case, if we assume that each gesture requires 360° of their internal clock cycle to reach their target configuration, the vowel gesture is coordinated to the 60°/(60°+60°+360°)=12.5% of the whole C gesture, as in Figure 1.8.

[4]While the term sub-gesture is not a technical term, we use it here for expository convenience to highlight the fact that both sub-gestures are there to model different aspects of a single stop articulation.

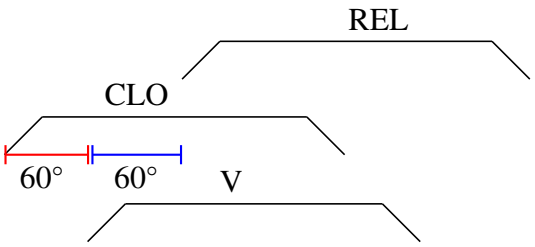

Figure 1.7 Sample CV alignment according to the split-gesture model. Here it is assumed that each gesture requires 240° of their internal clock cycle to reach their target configuration (Browman et al., 1990). The vowel gesture has a 60° phase difference with the closure sub-gesture and a -60° phase difference with the release sub-gesture. Therefore, the vowel gesture is coordinated to the 60°/(60°+60°+240°)=16.7% of the whole C gesture.

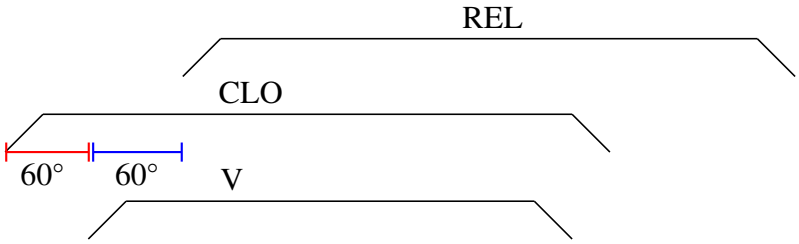

Figure 1.8 Sample CV alignment according to the split-gesture model. In this figure, it is assumed that each gesture requires 360° of their internal clock cycle to reach their target configuration (Browman et al., 1990). The vowel gesture has a 60° phase difference with the closure sub-gesture and a -60° phase difference with the release sub-gesture. Therefore, the vowel gesture is coordinated to the 60°/(60°+60°+360°)=12.5% of the whole C gesture.

In contrast to the above claims, Tilsen (2020) also suggested that CV coordination is eccentric. Specifically, Tilsen (2020) found that "there is category-related information in speech signals well before initiation of the articulatory gestures associated with those categories" (p. 20). Liu et al. (2022) interpreted this to mean that vowel onset initiates before consonant gesture offset and after consonant gesture onset. Also, Öhman (1966) argued that vowels begin during the consonant in CV sequences.

Despite the different claims of CV coordination, the theories mentioned above predict *consistent* CV coordination. Besides CV sequences, for syllables with consonant clusters (CC) at onset, AP specifies the two onset consonants to have a consistent sequential relationship with each other.[5]

[5]Besides the reason mentioned earlier, another reason why AP is not assumed in the current study is that the sonority-driven speech production model proposed in subsection 5.3 assumes that CC onset, CV, and perhaps VC have similar coordination patterns.

In general, even though the theoretical claims mentioned above suggest different specifications regarding CV coordination — synchronous (Browman and Goldstein, 2000; Nam and Saltzman, 2003; Goldstein et al., 2006; Xu et al., 2006; Hermes et al., 2013; Liu et al., 2020; Durvasula et al., 2021; Liu et al., 2022; Iskarous and Pouplier, 2022), sequential (Shaw et al., 2021), and C-center (Nam, 2007), none of the previous theoretical claims have predicted the systematic variation of gestural coordination variation correlates to sonority.

1.5 Gestural coordination variations related to sonority

As discussed in the previous subsection, current theories of gestural coordination do not predict the variation of gestural coordination relevant to sonority. However, some previous studies observed gestural coordination variation, and some of the variation is potentially related to sonority. More specifically, the gestural coordination of CC onset clusters has been argued to be related to sonority (Crouch, 2022; Crouch et al., 2023). In addition, gestural coordination variation has been found to correlate with factors such as a) C voicing (Hoole et al., 2009; Gibson et al., 2019), b) C place (Byrd, 1994, 1996; Gafos et al., 2010; Bombien et al., 2013), c) C manner (Byrd, 1994; Wright, 1996; Byrd, 1996; Hoole et al., 2009; Gibson et al., 2017; Pouplier et al., 2022), d) vowel quality (Fowler and Saltzman, 1993), e) prosodic effects such as stress, domain position (Öhman, 1966; Hardcastle, 1985; Byrd, 1994, 1996; Byrd and Saltzman, 2003; Yanagawa, 2006; Gafos et al., 2010; Gu, 2023), and f) stiffness parameter of the gestures (Du and Gafos, 2023). These factors are crucial for analyzing gestural coordination.

At least some of the observations above can be generalized as a relationship between sonority and gestural timing. As noted, some of the work explicitly pointed out the relevance of sonority to gestural timing (Crouch, 2022; Crouch et al., 2023). Furthermore, the effects of voicing and manner also are potentially related to sonority. In the following subsections, gestural coordination related to sonority will be discussed first. Then, other factors that lead to gestural coordination variation will be briefly discussed. Understanding non-sonority factors relevant to gestural coordination variation is necessary for implementing the present experiment and for interpreting the results of the dissertation.

Crouch (2022) and Crouch et al. (2023) explored the relationship between sonority and CC timing, and observed that there was a correlation between sonority sequencing of consonant onset clusters and their gestural overlap in Georgian. Specifically, they used the 20% threshold algorithm of the `mivew` package (Tiede, 2005).[6] They considered two measurements. The first measurement was termed relative overlap, and it is the same with the onset lag measurement in Pouplier et al. (2022) as in (3a). The second measurement is termed constriction duration overlap as in (2), and it is the opposite of normalized plateau lag in Pouplier et al. (2022).

(2)

$$\text{constriction duration overlap} = \frac{\text{(C1 target offset - C2 target onset)}}{\text{(C2 target offset - C1 target onset)}}$$

Crouch (2022) found that a sequence of two consonants in Georgian with a sonority rise exhibited less overlap than those with a sonority plateau, which in turn were less overlapped than those with a sonority fall. Crouch et al. (2023) speculated that the observed relationship between sonority sequencing and consonant sequences was limited to Georgian consonant onset clusters.

There is also prior work that has observed variation in CC gestural coordination related to C manner and C voicing (Hoole et al., 2009; Gibson et al., 2019; Pouplier et al., 2022). Even though sonority was not used to account for the variation, one could infer from these results that it is really sonority that is related to gestural coordination variation since both voicing and manner relate to sonority. For example, Pouplier et al. (2022) analyzed the CC overlap in 7 languages and argued that manner and voicing can condition CC overlap variation.[7] For each CC onset cluster, they measured onset lag and normalized plateau lag using the formulas in (3). As schematized in Figure 1.9, onset lag was quantified by the difference between C2 gestural onset and C1 target onset (orange line) divides the difference between C1 target offset and C1 target onset (red line). Moreover, normalized plateau lag is quantified by the difference between C2 target onset and C1

[6] All the articulatory studies discussed in this subsection (1.5) used the algorithm.

[7] The 7 languages are: English, French, Russian, Georgian, German, Polish, and Romanian.

target offset (green dashed line) divides the difference between C2 target offset and C1 target onset (blue dashed line).

(3) a.

$$\text{onset lag} = \frac{\text{(C2 gestural onset - C1 target onset)}}{\text{(C1 target offset - C1 target onset)}}$$

b.

$$\text{normalized plateau lag} = \frac{\text{(C2 target onset - C1 target offset)}}{\text{(C2 target offset - C1 target onset)}}$$

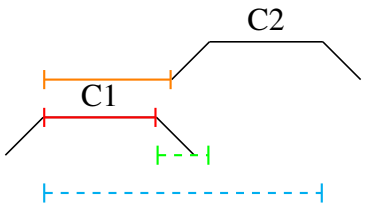

Figure 1.9 Lag measurement in Pouplier et al. (2022). Onset lag is measured by the orange line value divides the red line value. Normalized plateau lag is measured by the green dashed line divides the blue dashed line.

While Pouplier et al. (2022) measured the degree of overlap rather than gestural lags, their results can be reinterpreted in terms of gestural lags. If we assume that the degree of overlap and gestural lag are inversely related, plugging in the numbers of the sonority index in Table 1.6 for the observed sequences shows a positive correlation between sonority difference and gestural lag. For instance, Pouplier et al. (2022) generally observed that consonant clusters that begin with voiceless stops (/pl/, /kl/, and /kn/) had less overlap than those that started with voiced stops (/bl/, /gl/, and /gn/, respectively). Since according to some sonority scales such as Parker (2012) in Table 1.6, voiceless stops are less sonorous than voiced stops, voiceless stops have a larger sonority difference with their subsequent C2. Therefore, the observation in Parker (2012) is consistent with the generalization that there is a positive correlation between sonority difference and gestural lag.

Pouplier et al. (2022) also found that /sk/ and /sp/ clusters are more likely to have a larger overlap than /ʃm/ and /sm/. This observation suggests that when the first consonant is a fricative, clusters where the second C is a stop have a larger overlap than when C2 is a nasal. Since in terms

of sonority, stops < fricatives < nasals, and according to Table 1.6 the difference between stop and fricative is smaller than that between fricative and nasal, /sk/ and /sp/ clusters are likely to have larger overlap than /ʃm/ and /sm/ if we assume a positive relationship between sonority difference and gestural lag. Similarly, Pouplier et al. (2022) found that a) /bl/ and /gl/ have less overlap than /ʃm/, /sm/, and /ʃp/; b) /ʃm/ and /sm/ have less overlap than /sp/ and /sk/; c) for stop-initial data, /gn/ has shorter lag than /bl/, which has larger overlap than /kl/ and /pl/; d) /gl/ has larger overlap than /kl/ and /pl/; e) /sp/ has the lowest onset lag, and other languages' /sp/ extend into the negative range. If we plug in the numbers of the sonority index in Table 1.6 for the relevant sequences, one observes a tendency for a positive correlation between sonority difference and gestural lag can predict those observations of gestural coordination variation, if we also assume degree of overlap and lag are inversely related. Specifically, the sonority values of each onset cluster can seen in Table 1.8 where the sonority difference is calculated by subtracting the C1 sonority index from the C2 sonority index. The relationship between gestural lag and sonority can be found in all observations. For instance, /sp/ has -2 as its sonority difference according to Table 1.6, which is the lowest among the target CC in the study.

(1)	/ʃp/ 1-3=-2	/sm/ 7-3=4	/ʃm/ 7-3=4	<	/gl/ 9-4=5	/bl/ 9-4=5
(2)	/sk/ 1-3=-2	/sp/ 1-3=-2	<	/sm/ 7-3=4	/ʃm/ 7-3=4	
(3)	/gn/ 7-4=3	<	/bl/ 9-4=5	<	/kl/ 9-1=8	/pl/ 9-1=8
(4)	/gl/ 9-4=5	<	/kl/ 9-1=8	/pl/ 9-1=8		
(5)	/sp/ 1-3=-2	<	...			

Table 1.8 Calculating the sonority difference in syllables in Pouplier et al. (2022). The number in the first column refers to the specific observation regarding sonority and gestural timing in the paragraph. The calculation below each cluster shows the CC sonority difference by subtracting the C1 index from the C2 index.

Admittedly, in the above interpretation of Pouplier et al. (2022), the inverse relationship between gestural overlap and lag is assumed but not tested. Additionally, they are patterns that are inferred

from k-means clustering results. Therefore, they should serve as suggestive rather than concrete evidence for a positive correlation between gestural lag and sonority difference. This caveat is also true for other studies on CC that used the measurement of overlap rather than lag.

There is also some acoustic data showing C manner could relate to gestural coordination variation (Wright, 1996). Wright (1996) examined the acoustic data of Tsou, an Austronesian language rich in word-initial consonant clusters. They found that clusters where one or both consonants have internal cues showed a greater degree of overlap. Wright (1996) considered cues to consonant place and manner contrasts. In word-initial stop+stop clusters, for example, the overlap between consonants is minimized to maintain an audible C1 release burst. When C1 is a fricative, more overlap is permitted because fricatives have internal cues to their place and manner. This work is not analyzed beyond as some other observations showing a relationship between C manner and articulatory timing because it used acoustic rather than articulatory data, and I am reluctant to use acoustic results to infer articulatory gestural lags.

Besides Pouplier et al. (2022), some previous studies also observed that the voicing of consonants is relevant to the gestural coordination of consonant cluster at onsets (Hoole et al., 2009; Gibson et al., 2019). Using German EMA data, Hoole et al. (2009) observed that there is less articulatory overlap for the voiceless compared to the voiced C1 for German onset clusters. The overlap of consonant gestures was measured by (2), the same measurement used in Crouch et al. (2023). Since voiceless C is less sonorous than voiced C, the C1C2 sonority difference would be larger for voiceless C1 than voiced C1. Therefore, clusters with voiceless C1 would have a larger lag and less overlap than clusters with voiced C1, given the positive correlation between gestural lag and sonority difference. Furthermore, Gibson et al. (2019) examined the gestural coordination in onset clusters in Spanish using EMA data. Gestural overlap was defined by subtracting the C1 target offset from the C2 target onset as in (4). Gibson et al. (2019) found that two consonants that are both voiced show more articulatory overlap than when C1 is voiceless and C2 is voiced. Again, voiceless obstruents are less sonorous than voiced ones, and therefore, they would have a larger sonority difference as C1. This predicts that $C1_{voiceless}C2_{voiced}$ would have a larger gestural lag —

which can be interpreted as less overlap — than $C1_{voiced}C2_{voiced}$.

(4)

gestural overlap = C2 target onset - C1 target offset

If interpreted in terms of gestural lag, the above results are consistent with the claim that there is a positive correlation between sonority difference and gestural lag. Given that voiceless obstruents are less sonorous than voiced obstruents in the sonority scale in Parker (2002, 2008, 2011).

Some studies showed the relationship between manner and CC gestural timing (Gibson et al., 2017), and again this relationship can be viewed as a link between sonority and gestural timing. Gibson et al. (2017) analyzed EMA data of Spanish onset clusters. C1C2 overlap was quantified by the timing difference between C1 and C2 targets or plateaus. They found that clusters where C2 is a rhotic had significantly larger lag than clusters where C2 is a lateral. Rhotics are more sonorous than laterals. Therefore, $C1C2_{rhotic}$ would have larger sonority difference and larger gestural lag than $C1C2_{lateral}$.

The above findings, suggesting a positive correlation between sonority and gestural coordination in onset CC, are unexpected according to current speech production theories. According to AP, we also do not expect CV coordination to share a similar pattern with CC coordination. However, the following observations show that the positive correlation between sonority and gestural coordination is also likely applicable to CV sequences. First, Gao (2008) observed variation in gestural lag between the onset of the consonant and vowel gestures (CV lag) when exploring Mandarin tone-to-segment alignment using kinematic data collected from Electromagnetic Articulography (EMA) experiments. The CV lag measurement, namely, CV lag based on gestural onset (Lag_{GON}), is schematically shown in Figure 1.10 by the black dashed line. Specifically, Gao (2008) observed that the CV lags of [t]-onset syllables were slightly longer than those of [n]-onset syllables, for words with Tone 4.[8] [t] is less sonorous than [n], so [t]-onset syllables have a larger sonority difference with a same following vowel than [n]-onset syllables.

[8]Tone 4 refers to a falling tone as per the notation introduced by Chao (1930).

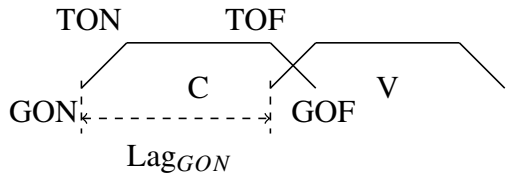

Figure 1.10 Schematic for CV lag computation. For each C or V gesture, the landmarks gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF) are labeled for clarity from left to right. The black dashed line indicates CV lag based on gestural onset (Lag$_{GON}$).

A second related observation was made by Shaw and Chen (2019), who conducted an EMA study of Mandarin speakers producing CV monosyllables, consisting of labial consonants and back vowels, in isolation. In their study, CV lag is the interval between the onset of the consonant gesture and the onset of the vowel gesture — CV lag based on gestural onset (Lag$_{GON}$) as indicated by the black dashed line in Figure 1.10. When testing whether the spatial position of the tongue influences CV coordination, Shaw and Chen (2019) found that the CV lag was significantly shorter in syllables beginning with the nasal stop than in syllables beginning with the oral stop. They observed that syllables beginning with [m] had a shorter CV lag than those that begin with [p]. Again, the nasal [m] is more sonorous than the oral stop [p], so there is a shorter sonority difference and therefore shorter lag for the nasal [m].

Relatedly, some study have claimed that vowel quality relates to gestural coordination variation. Fowler and Saltzman (1993) noted that for /bV/ syllables, as the jaw closes for /b/, the following vowel will oppose this motion; consequently, a following higher vowel such as /i/ will oppose the jaw closing movement less and a following lower vowel /a/ will oppose it more. Even though this claim does not involve gestural timing, one might perhaps extrapolate it to suggest that /ba/ will have a larger lag than /bi/ since there is more opposing force in /ba/. Even though one could also have extrapolated it to suggest that /ba/ will have a smaller lag to counteract the opposition between the /b/ and the /a/, the variation of gestural coordination due to vowel quality difference was hinted at. Since vowel quality is related to sonority, a relationship between sonority and gestural timing is suggested.

To sum up, the results observed in a variety of previous studies are consistent with the claim

that gestural timing variability relates to sonority. Specifically, sonority seems to have a positive correlation with gestural lag on CV and CC sequences. This serves as a promising entry point for figuring out the sonority correlate in speech production. The current study tests this generalization of a potential positive correlation between sonority and gestural lag on CV syllables.

1.6 Other observations of gestural coordination variations

In this subsection, I briefly discuss non-sonority or non-sonority-related factors that were argued to be related to articulatory gestural timing. The factors to be discussed are: a) C place (Byrd, 1994, 1996; Gafos et al., 2010; Bombien et al., 2013), b) prosodic effect (Hardcastle, 1985; Byrd, 1994, 1996; Byrd and Saltzman, 2003; Yanagawa, 2006; Gafos et al., 2010; Gu, 2023), and c) stiffness parameter of the gestures (Du and Gafos, 2023).

First, consonant place leads to gestural coordination variation in German CC onset clusters (Bombien et al., 2013). Specifically, Bombien et al. (2013) found that /kl/ exhibited the highest degree of overlap, followed by /pl/, /ps/, /ks/, and finally /kn/. Liu et al. (2022) interpreted the results of variability as an effect of place — a CC cluster beginning with labials generally has a higher degree of overlap than a CC cluster that begins with velars. Gafos et al. (2010) also found that speaker-specific place order of Moroccan Arabic clusters is related to gestural overlap.

Second, gestural lags are larger at prosodic boundaries (Byrd and Saltzman, 2003) and when stressed (Katsika, 2012; Gu, 2023). Additionally, there are different findings on speech rate's impact on gestural timing, though more seems to find that articulatory overlap increases with speech rate (Hardcastle, 1985; Byrd, 1994; Luo, 2017).[9] Moreover, cluster position in a syllable or word affects gestural timing (Byrd, 1994, 1996; Yanagawa, 2006; Gafos et al., 2010). Byrd (1994, 1996) showed that an onset cluster is less overlapped than coda clusters. Gafos et al. (2010) also found a speaker-specific word position effect in the gestural overlap of Moroccan Arabic clusters. Furthermore, Öhman (1966) showed that in VCV sequences, the first V is affected by the second V. This shows that segment position or syllable position affects gestural coordination. Considering the

[9]Hardcastle (1985) observed that there is more co-articulation during faster speech rate conditions. Similarly, Byrd (1994) observed that articulatory overlap increases with speech rate in English consonant sequences. In contrast, Luo (2017) found no speech rate effect on gestural overlap.

above analyses, when designing experiments and selecting stimuli for the dissertation, monosyllabic citation words are preferred.

Third, Du and Gafos (2023) argued that in onset clusters, C2 stiffness contributes to gestural overlap. They examined articulatory data from German, English, and Spanish participants. The relevance of stiffness to articulatory timing will be discussed later when interpreting results. To sum up, there are various factors that are not captured under the concept of sonority that can contribute to gestural coordination variation. Therefore, when analyzing the effect of sonority on gestural coordination, we need to consider these factors in experimental design and in the interpretation of results.

To sum up, even though many researchers in speech production assume no gestural coordination variation based on segment makeup, some studies showed gestural timing variability relates to various possible factors. Among these factors, sonority's potential positive correlation with gestural lag on CV sequences was not seriously explored, even though there are many pieces of suggestive evidence that can support hypothesizing this positive correlation.

1.7 Claims to be tested

This dissertation tested whether there is a positive correlation between sonority difference and gestural lag. This positive correlation is likely to hold true in CC onset and CV syllables, and I evaluate it on CV syllables in the dissertation. The claim to be tested is that for a CV sequence within a syllable, the sonority difference between C and V positively correlates with the CV lag. This claim regarding CV coordination suggests two sub-claims to be tested related to varying the C for the same V as in (5a) and varying the V for the same C as in (5b). Specifically, claim (5a) predicts that [ba] should have a larger CV lag than [ma] because the stop [b] is less sonorous than the nasal [m]. Claim (5b) predicts that, for instance, [ba] should have a larger CV lag than [bi] because the low vowel [a] is more sonorous than the high vowel [i].

(5) Claims to be tested related to CV timing:

   a. For CV syllables with the same V, a less sonorous C leads to a larger CV lag.

b. For CV syllables with the same C, a more sonorous V leads to a larger CV lag.

In order to quantify the dependent variable, namely, CV lag, I measured the timing difference between the consonant gesture and the following vowel gesture (CV lag = V timestamp - C timestamp). This way of calculating lag by subtracting corresponding timestamps was used in Zhang et al. (2019). Specifically, the CV lag in the current study was computed by subtracting the target onset (onset of gestural plateau) timestamp of the consonant from the target onset timestamp of the vowel. Target onset instead of gestural onset is used since target onset alignment has been argued to be more consistent than gestural onset alignment (Zhang et al., 2019; Durvasula and Wang, 2023). The visual illustration of the lag calculation can be found in Figure 1.11, where the timestamp of *target onset* of a C is subtracted from that of *target onset* of a V to get the CV lag based on target onset, as in the blue dashed line.[10] There are three other measurements based on the method of subtracting corresponding timestamps from another timestamp — CV lag based on gestural onset (as in the black dashed line), CV lag based on gestural offset, and CV lag based on target offset.

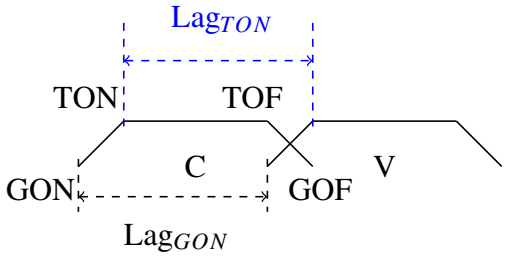

Figure 1.11 Schematic for CV lag computation. For each C or V gesture, the landmarks gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF) are labeled for clarity from left to right. The black dashed line indicates CV lag based on gestural onset (Lag$_{GON}$), and the blue dashed line indicates CV lag based on target onset (Lag$_{TON}$).

In the following chapters, I show the procedure and results of evaluating the claims on CV syllables of an English corpus (Chapter 2), English EMA study (Chapter 3), and Mandarin EMA study (Chapter 4). In the following section, before I detail the experimental procedures, I discuss one major methodological concern in speech production studies, which is how to parse articulatory

[10]This figure is a repetition of Figure 1.10. I am presenting it here for the readers' convenience.

gestures. In general, I discuss the *default* method, the threshold method, and its alternatives such as the comparative method and the ensemble method. I argue that the threshold algorithm is used in the dissertation since it has obvious advantages over its alternatives.

1.8 Measuring articulatory gestures

There are various existing methods of parsing articulatory gestures in speech production. In this section, I discuss some methods of parsing articulatory gestures. Specifically, the threshold technique will be presented in subsection 1.8.1, the minimal contrast technique will be discussed in subsection 1.8.2, and the ensemble technique will be briefly mentioned in subsection 1.8.3. I argue that the threshold technique has advantages over other methods, so it is used in the current dissertation.

1.8.1 The threshold technique

The threshold technique to identify articulatory gestures is widely used in the field since it is the underlying algorithm of the widely used *lp_findgest* in mview (Tiede, 2005). In this section, I am going to briefly discuss the earlier uses of the technique (Hoole et al., 1994; Kroos et al., 1996). Then, I am going to detail the algorithm *lp_findgest*, which is the *default* method of parsing articulatory gestures.

1.8.1.1 The earlier usages of the threshold technique

One earlier usage of the threshold technique was in Hoole et al. (1994), where the study compared the articulation of tense and lax vowels in German. The authors used the threshold technique to identify the CV, nucleus, and VC of CVC syllable as in Figure 1 from Hoole et al. (1994) where the segmentation procedure for CVC utterance /pi:p/ was shown. When pronouncing the utterance, there are opening and closing of lips, which is crucial for identifying the gestural timing patterns. The first step of the procedure is identifying the maximum vertical velocity point of the lower lip from the C1 target to the vowel target. Then, two points that are 20% of this maximum velocity were identified as the CV onset and offset, one when moving upwards and another when moving downwards from this maximum point respectively. Similarly, the VC segment was identified. In

Figure 1 from Hoole et al. (1994), the CV syllables were labeled where the left edge of the label is the onset and the right edge the offset.

According to Hoole et al. (1994), 20% is a high-velocity threshold, and it is chosen to avoid problems with identifying the nucleus stage. They suggested that they want to avoid overlapping CV and VC by using 20% as the threshold. Also, Hoole et al. (1994) claimed that they decided that the nucleus should be a stage rather than a single point because it reflects the observation that tense vowel is longer in duration than lax vowels. They used tangential velocity, which is the velocity signal that incorporates movement in all three available dimensions (Shaw et al., 2023).

The advantage of this measurement criterion is that it yields more stable results (Mooshammer and Fuchs, 2002). In a similar study, Kroos et al. (1996) also used the threshold method in German in similar cases for CV and VC, when they also looked at CVC stimuli. In Kroos et al. (1996), the threshold of 20% is also used after a bunch of experiments of other thresholds. Note that these original studies used the threshold method for syllable identification, which is different from some later uses of the threshold technique which is on a single articulatory gesture.

1.8.1.2 The *lp_findgest* algorithm of mview

The *lp_findgest* algorithm of mview package (Tiede, 2005) is *the* default tool used to identify gestures in speech production studies, and the algorithm assumes the threshold technique, though not strictly the same as the technique as in Hoole et al. (1994). The following indicates the procedure for identifying gestures of the *lp_findgest* algorithm.

To use the algorithm, the researcher first needs to click on a point in the relevant articulatory pellet's information. The mouse click point is usually identified by checking the synchronous acoustic information, and it will roughly be the point of the gestural plateau. There is no clear requirement on where to make this click. After manually identifying the point, the algorithm finds the maximum constriction point (MAXC), which is the closest velocity minimum to the clicked point. Then there are two peak velocity points identified before and after the maximum constriction point, which are called PVEL and PVEL2 respectively. After that, the gestural onset point is marked by identifying the 20% peak velocity between the minimum velocity point before PVEL and the

peak velocity point (PVEL) itself. The nucleus onset is the 20% peak velocity point between PVEL and the maximum constriction point (MAXC), and the nucleus offset is identified by the 20% peak velocity point of the range between MAXC and the following peak velocity PVEL2. Similarly, gestural offset is the 20% peak velocity point between the range between PVEL2 and the following velocity minimum. The velocity in *lp_findgest* algorithm of mview is computed either as tangential velocity (if multiple components are displayed) or as absolute magnitude (if one component is displayed). For instance, in Figure 6 from Shaw et al. (2023), the *lp_findgest* algorithm was used to identify the articulatory trajectory of a bilabial fricative (Shaw et al., 2023). In this case, they used tangential velocities, incorporating movements in the vertical, longitudinal, and lateral dimensions. We can see that even though the vertical dimension as in the second row from above has the largest degree of displacement, there is also movement in the other two dimensions. In Figure 6 from Shaw et al. (2023), the terms start, target, release, and end refer to gestural onset, target onset, target offset, and gestural offset used in this dissertation.

1.8.2 The minimal contrast technique

In this section, I lay out some studies that used the minimal contrast technique. First, Benguerel and Cowan (1974) analyzed French upper lip protrusion patterns. They found that French speakers started lip protrusion for the vowels as early as 6 consonants before the rounded vowel. This finding shows that considering the interaction of surrounding articulations is necessary.

Second, Gelfer et al. (1989) advocated the minimal contrast technique in speech production. Specifically, they argued that in American English a migration of lip rounding back to the beginning of the consonant string in /iCu/ utterances may not support the look-ahead model since a similar lip rounding pattern can also be observed in /iCi/ utterances. They also observed comparable correlation coefficients between electromyographic (EMG) onset time and consonant string duration for utterances with or without lip rounding. The study shows that some speakers produce alveolar consonants with significant lip rounding activity in both rounded and unrounded vowel environments. Crucially, the study advocates that studies of co-articulation should employ the minimal contrast technique.

Third, Liu et al. (2022) is a recent study that promotes the minimal contrast technique. They used a minimal triplet paradigm to analyze Mandarin coarticulation data and argued that syllables are coordinated synchronically. For instance, if the target syllable is C1V1, the triplets will be C2V1, C1V2, and C1V1. Participants produced the targeted utterances by embedding them into the carried phrase *bi* ___ *wei shan* [bi ___ weɪ ʂan] 'more hypocritical than'. Note that this production with carrier phrase is not a complete sentence. And since 'wei shan' [weɪ ʂan] is a compound word meaning hypocritical, there is unlikely to be a pause between wei [weɪ] and shan [ʂan]. This means that there is significant coarticulation between wei [weɪ] and shan [ʂan], and it may not be easy to parse a boundary after wei [weɪ].

To analyze each minimal pair in each triplet (i.e., vowel minimal pair C1V1-C1V2, consonant minimal pair C1V1-C2V1), the time points where two trajectories diverge significantly were identified by generalized additive mixed models (GAMMs). The onset times were determined by when the model indicated a statistically significant difference in the trajectories relevant to either C or V. An issue here is that the point of statistical significance in terms of difference does not necessarily indicate the point of start of non-trivial co-articulation.

Liu et al. (2022) contributed to the discussion of methods by showing the comparison of two triplets as in Figure 14 and Figure 15 in Liu et al. (2022). While the onsets identified by the two methods are more different in Figure 14, the identifications are similar in Figure 15. As pointed out by Liu et al. (2022), the pair <maoliwei> /maʊluweɪ/ and <maoluwei> /maʊliweɪ/ shows more difference because <mao> /maʊ/ has a rounding gesture at its later part. They also mentioned that studies could avoid obvious gestural confounds as such in the stimulus design process. Gelfer et al. (1989) and Liu et al. (2022) argued that the real advantage of the minimal contrast technique is that it can avoid covert confounds in cases where there are no predictably similar gestures in the previous syllable as in Figure 15. However, just because there is a difference in the measurement outcome does not mean that it is a significant confound. In other words, there is no evidence that the inference based on the different techniques are different, even if different techniques result in slightly different values. It seems that a larger absolute difference can be avoided by stimulus design

in cases such as 'maolu' /maʊlu/ vs. 'maoli' /maʊli/ (Figure 14 in Liu et al. (2022)), and yet the covert "confound" does not result in much absolute difference as in 'laili' /laɪli/ vs. 'lailu' /laɪlu/ (Figure 15 in Liu et al. (2022)).

As mentioned by Durvasula and Wang (2023), an issue with the comparative technique is that it requires phonetic minimal pairs rather than phonological minimal pairs. There are logically infinite number of phonetic parameters available, so it is difficult to determine phonetic minimal pairs. Another issue with the minimal contrast technique is inherent phonetics (Durvasula, 2024). For instance, even oral vowels have inherent nasality and low vowels have more inherent nasality than mid or high vowels. Therefore, if looking at nasality, the results would be different if different baselines were chosen.

1.8.3 The ensemble technique

There are other techniques being used such as the minimum velocity technique (Blackwood Ximenes et al., 2017) or the zero velocity technique (Mücke et al., 2012). Since there are pros and cons to each method, implementing several possible methods and adopting the advantages of each method seem to be a solution to the issue of lack of methodology consensus. One recent study, Svensson Lundmark et al. (2021), identified the problem that previous studies on tone gesture used a variety of methods, and they addressed the inconsistency in the varieties of methods by including the comparison of 13 measurements. Specifically, they choose different measurements for different articulators such as lip or tongue. For lip aperture, they used zero velocity and maximum acceleration/deceleration. The temporal landmarks on lip aperture were automatically extracted in R (Team et al., 2013) when the lips had minimal movement, 20% threshold from zero to peak velocity, and when the movement accelerated or decelerated the most. For the tongue body, they used minimal tangential velocity, maximum tangential velocity, zero vertical velocity, and 20% zero vertical velocity.

The conclusion in Svensson Lundmark et al. (2021) is that since consonantal gestures are more often characterized by larger changes in velocity, consonant gestures should be identified by peaks in the acceleration curve. Vowel gestures are made with constantly and relatively slowly

moving tongue body, and Svensson Lundmark et al. (2021) claimed that vowels may use different measurements. It is a challenge to justify those arguments since we need certain assumptions to know about the characteristics of gestures in the first place. While the suggestion regarding specific techniques is inconclusive, Svensson Lundmark et al. (2021) attempted to argue for an ensemble technique that builds on the advantage of different techniques. The potential issue with the ensemble technique is that it makes cross-stimuli comparison difficult. Also, since we are not sure what the right single method is, ensembling several techniques multiplied the scope of the problem.

1.8.4 Summary of measuring techniques

Articulatory trajectory is a complex fact. The position of an articulator is affected not only by the intended movement for the abstract phonological features of a sound, but also by co-articulation and inherent phonetics of those segments. Parsing articulatory gestures involves various decision-making that may not always be straightforward.

The section presents various available methods of identifying the onset and offset of articulatory gestures. There are pros and cons of each method. The threshold method is the *default* one. One of its issues is that many decisions such as the 20% threshold are arbitrary. Another major issue is that it is unclear whether the articulatory movement in question comes from the intended production or some unintended/inherent phonetics. The threshold technique only looked at the articulatory trajectory of one gesture, lacking the ability to parse out the overlapping or confounding gestures (Liu et al., 2022). The comparative technique, also called the minimal contrast technique, has an advantage over the threshold technique in this regard. Using the comparative method, sets of stimuli are used as baselines for the target measure. However, this method is unlikely to solve the inherent phonetics completely because there are an infinite number of dimensions available and it is practically difficult and logically impossible to find real phonetic minimal pairs. I also mentioned the ensemble technique, where the intention is to take advantage of various measurements. The major issues with the ensemble technique are: a) that it is time-consuming for human annotators; b) that cross-stimuli comparison is difficult since different measures are used. The discussions in

this section show that the comparative and ensemble techniques do not have significant advantages over the threshold technique. They also seem to have problems such as the need for more stimuli or measurements. To make the study's result comparable to most available research in the field, the *lp_findgest* algorithm was used. The nuances of choosing the *right* measure will be discussed in Chapter 5.

1.9 Recap of the introduction

In this chapter, I introduced sonority and phonological constraints related to sonority. Even though sonority's function in syllabification has been widely recognized, the phonetic correlation of sonority is controversial. If we take a closer look at sonority and its relationship with gestural timing, there is some previous empirical evidence showing that there is likely to be a positive correlation between sonority difference and gestural timing. Specifically, I showed that this positive correlation has been found in CV and CC sequences. Since Crouch (2022) has specifically tested the relation between CC timing and sonority, I propose to test the claim that there is a positive correlation between CV lag and sonority difference. In the following three chapters, I test the claim on a English corpus (Chapter 2), on English EMA study (Chapter 3), and on Mandarin EMA study (Chapter 4). The methods and results of each experiment will be discussed in each chapter. There will be discussions in Chapter 5 and a brief conclusion in Chapter 6.

This chapter also involves brief review of various methods of parsing articulatory data. Since the threshold method *lp_findgest* has advantages over alternative methods, it will be used to annotate the articulatory data of the study.

CHAPTER 2

EXPERIMENT 1: ENGLISH CORPUS STUDY

2.1 Methods

2.1.1 The corpus

To test the claim that there is a positive correlation between CV lag and sonority difference, I analyzed kinematic data from the Wisconsin X-ray Microbeam Database (Westbury et al., 1990). The data was originally collected using the X-ray Microbeam method to digitally track the movements of the gold pellets on each speaker's mouth. The X-ray Microbeam method relied on X-rays produced by an accelerated electron beam. Then, a narrow beam of the incident X-rays passed through a pinhole aperture where its path is determined by the location of the electron beam. The X-ray path was adjusted during the original data collection to make sure it produced X-ray scans that surrounded the expected pellet location so that the system recorded the coordinates of the movements. The frequency of operation for each pellet was specified separately, at rates ranging between 20 and 180 Hz.

The schematic positions of the pellets can be found in Figure 2.1. To obtain reference points indicated by Ref in Figure 2.1, three pellets were attached to the speaker's head: one on the bridge of the nose, the second on the buccal surface of the maxillary incisors, and the third either on the nosebridge lower than the first or an arm projecting from a snug-fitting pair of eyeglass frames. To extract information about tongue movement, four pellets, which are denoted by T1 to T4 in Figure 2.1, were attached along the longitudinal sulcus of each speaker's tongue. T1 was placed 10 mm posterior to the tongue tip, and T4 was placed about 60 mm posterior to the tongue tip, depending on each speaker's tolerance. Positions of T2 and T3 were chosen so that the four tongue pellets were equally distanced. As for labial articulation, one pellet each was attached to the upper lip (UL) and lower lip (LL).

The Wisconsin X-ray Microbeam Database has 118 speech production tasks of various types such as paragraph reading, sentence reading, citation word production, and number sequence

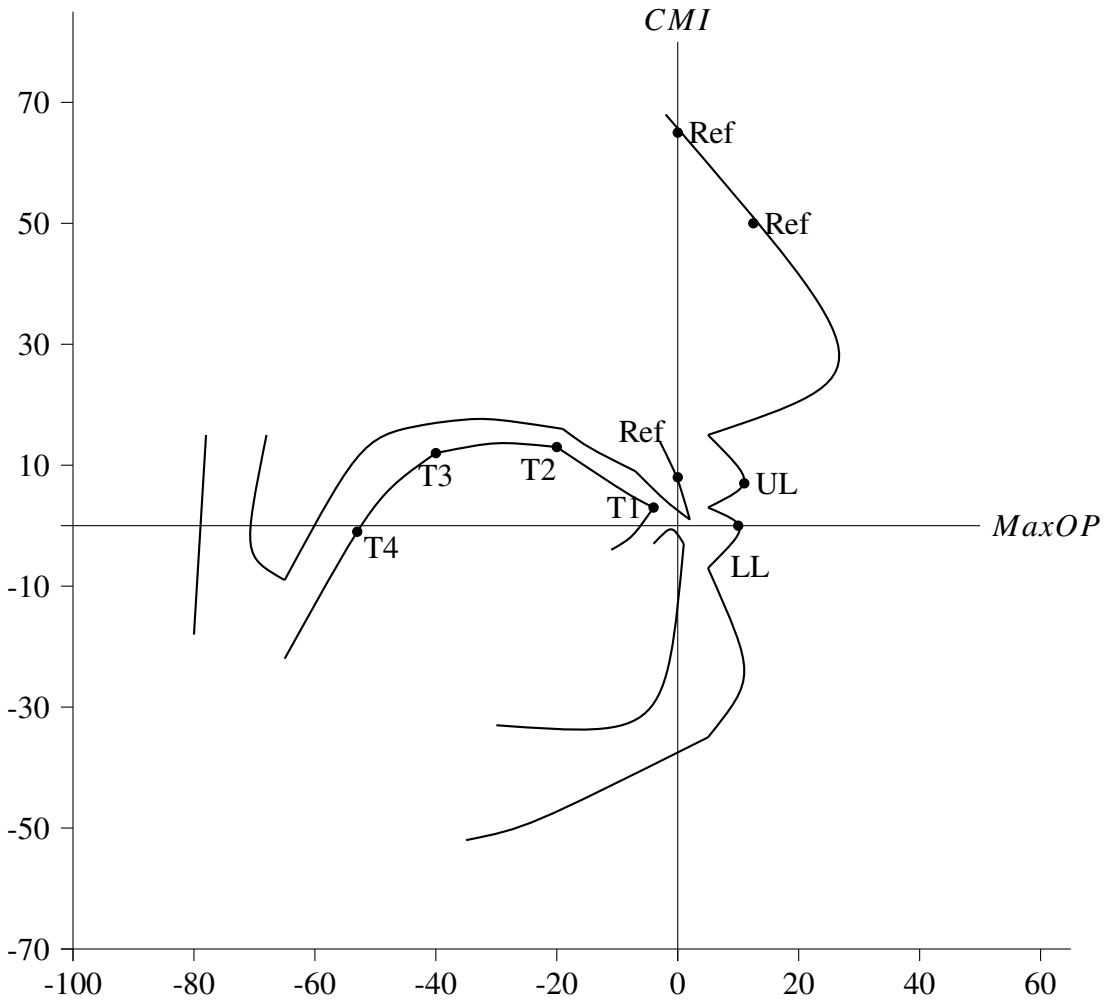

Figure 2.1 Approximate pellet placement locations. The x-axis represents the position with respect to the central mandibular incisor (CMI), and the y-axis shows the position with respect to the maxillary occlusal plane (MaxOP). All numbers are in millimeters (mm). The figure is recreated based on Figure 5.2 of the Wisconsin X-ray Microbeam Database manual (Westbury et al., 1990).

production. Speakers represented in the database were recruited from the University of Wisconsin-Madison as well as the surrounding city, and a majority of speakers spoke an Upper Midwest dialect of American English. Altogether, speech production data of 57 different speakers (32 females and 25 males) were included in the database. The median age for the speaker sample was 21.1 years old (female 21.3 years; male 20.8 years).

2.1.2 Stimuli

All stimuli in the experiment came from the citation word reading tasks of the Wisconsin X-ray Microbeam Database, where the speakers were instructed to "Read each item once, slowly and

clearly, with a brief pause between items. Read in column order." The citation word list reading tasks had monosyllabic words, which ensured that factors like prosody or stress — which may induce gestural variation (Byrd and Saltzman, 2003; Katsika, 2012; Byrd and Krivokapić, 2021; Gu, 2023) — were controlled for. I considered two sets of stimuli, one for each claim in (6).[1]

(6) Claims to be tested related to CV timing:

   a. For CV syllables with the same V, a less sonorous C leads to a larger CV lag.

   b. For CV syllables with the same C, a more sonorous V leads to a larger CV lag.

The list of nonce word stimuli to test claim (6a) is shown in Table 2.1. These words are from task 16 of the corpus and have a template *uhCa*. The varying consonants and the same vowel [ɑ] of the *uhCa* stimuli make them suitable to test the claim (6a) that a less sonorous C leads to a larger CV lag for CV syllables with the same V. All words in task 16 were included in the dissertation except for *uhga* and *uhka* (marked by ⋆), since the velar stop and low vowel both use tongue dorsum as the measurement sensor and it is difficult if not impossible to tease apart whether tongue dorsum movement comes from the consonant or the vowel. Each table of stimuli also includes columns labeled *C-V Pellets*, *Sonority*, and *Difference*; I return to elaborating on these columns later in this section.

[1] This repeats the claim in (5) for the readers' convenience.

	Stimuli	IPA	C-V Pellets	Sonority	Difference
(1)	uhyA	[ə**jɑ**]	T1 - T3	12-17	5
(2)	uhwA	[ə**wɑ**]	Lip - T3	12-17	5
(3)	uhlA	[ə**lɑ**]	T1 - T3	9-17	8
(4)	uhmA	[ə**mɑ**]	Lip - T3	7-17	10
(5)	uhnA	[ə**nɑ**]	T1 - T3	7-17	10
(6)	uhvA	[ə**vɑ**]	LL - T3	6-17	11
(7)	uhzA	[ə**zɑ**]	T1 - T3	6-17	11
(8)	uhzhA	[ə**ʒɑ**]	T1 - T3	6-17	11
(9)	uhdA	[ə**dɑ**]	T1 - T3	4-17	13
(10)	uhbA	[ə**bɑ**]	Lip - T3	4-17	13
(11)	uhfA	[ə**fɑ**]	LL - T3	3-17	14
(12)	uhshA	[ə**ʃɑ**]	T1 - T3	3-17	14
(13)	uhsA	[ə**sɑ**]	T1 - T3	3-17	14
(14)	uhtA	[ə**tɑ**]	T1 - T3	1-17	16
(15)	uhpA	[ə**pɑ**]	Lip - T3	1-17	16
⋆	uhgA	[ə**gɑ**]	T4 - T2	4-17	13
⋆	uhkA	[ə**kɑ**]	T4 - T2	1-17	16

Table 2.1 List of stimuli for the claim (6a). The stimuli are from task 16 of the Wisconsin X-ray Microbeam Database (Westbury et al., 1990). The target CV sequence is in boldface and underlined. The *C-V Pellets* column shows the pellet positions for C and V. The *Sonority* column indicates the C and V sonority respectively, and the *Difference* column lists the sonority difference.

The list of stimuli to test claim (6b) is listed in Tables 2.2 and 2.3. They are all real words except for the ones with an asterisk before them. For the real words, speakers saw the orthography in the column *Stimuli*; but for each nonce word, speakers saw a real word that exemplified the vowel pronunciation of the nonce words as in *sud (dud); *soid (Lloyd); *sowd (loud); *sood (wood); *sayed (bayed). The stimuli in Table 2.2 all have the template *sVd*, where the V varied, and the stimuli are from task 13 of the corpus. I included the second set of stimuli in Table 2.3 to address a potential issue with *sVd* stimuli that the consonant articulations involve an articulator related to that of the following vowel. In contrast, in *bV* words, the consonant and vowel articulations use different articulators, namely, the lips and the tongue. The stimulus *been* is from task 9 and *back* is from task 100.[2] All the stimuli in Table 2.2 or Table 2.3 have the same consonant with varying vowels within each set, which makes them suitable for testing the claim (6b) that a more sonorous

[2]There are two repetitions of *back* for each speaker.

V leads to a larger CV lag for CV syllables with the same C. Note that all CV syllables used to test the claim (6b) occur in a controlled immediate phonological environment (e.g. #s_d#) except for *back* and *been*, which have different coda consonants. The different coda consonants probably do not affect CV timing due to previous observation — Gao (2008) observed that the CV lag of [ma] for Mandarin speakers was not significantly different from that of [man].

	Stimuli	IPA	C-V Pellets	Sonority	Difference
(1)	seed	[**si**d]	T1 - T3	3-15	12
(2)	sid	[**sɪ**d]	T1 - T3	3-15	12
(3)	sued	[**su**d]	T1 - T3	3-15	12
(4)	*sood	[**sʊ**d]	T1 - T3	3-15	12
(5)	*sayed	[**sɛ**ɪd]	T1 - T3	3-16	13
(6)	surd	[**sɜ˞**d]	T1 - T3	3-16	13
(7)	said	[**sɛ**d]	T1 - T3	3-16	13
(8)	*sud	[**sʌ**d]	T1 - T3	3-16	13
(9)	sewed	[**so**d]	T1 - T3	3-16	13
(10)	sawed	[**sɔ**d]	T1 - T3	3-16	13
(11)	*sowd	[**sa**ud]	T1 - T3	3-17	14
(12)	side	[**sa**ɪd]	T1 - T3	3-17	14
(13)	sod	[**sɔ**d]	T1 - T3	3-16	13
(14)	*soid	[**sɔ**ɪd]	T1 - T3	3-16	13
(15)	sad	[**sæ**d]	T1 - T3	3-17	14

Table 2.2 sVd stimuli for claim (6b). The stimuli came from task 13 of the Wisconsin X-ray Microbeam Database (Westbury et al., 1990). The target CV sequences are in boldface and underlined. The *C-V Pellets* column shows the pellet positions for C and V. The *Sonority* column indicates the C and V sonority respectively, and the *Difference* column lists the sonority difference.

	Stimuli	IPA	C-V Pellets	Sonority	Difference
(16)	been	[**bi**n]	Lip - T3	4-15	11
(17)	back	[**bæ**k]	Lip - T3	4-17	13

Table 2.3 bV stimuli for claim (6b). The stimuli *been* is from task 9, and *back* is from task 100 of the Wisconsin X-ray Microbeam Database (Westbury et al., 1990). The target CV sequences are in boldface and underlined.

The pellet positions corresponding to the consonant and vowel gestures of each stimulus are shown in the *C-V Pellets* column of the above tables, where the C and V measurements were separated by a hyphen. The legend for the pellet positions is in Table 2.4. The stimulus tables

document that different pellet positions were used for vowels. During the data annotation process, the pellets were selected to ensure that the articulatory movement correctly reflected the acoustics of the relevant segment, and this selection was done before any analysis was performed on the data. While lip aperture was automatically computed by the *mdp_LipAperture* algorithm of the `mivew` package (Tiede, 2005), other pellets' information came directly from data collection. For the current study, the relevant measurements used for C and V were chosen based on previous literature (Gao, 2008; Hall, 2010; Zhang et al., 2019) and an understanding of the articulatory events involved. For example, /n/ involves tongue tip alveolar closure gestures, so T1 (tongue tip) was measured for the consonant closure of /n/. Since consonants such as /j/, /z/, and /s/ also involve tongue tip articulation, T1 was also measured for them. Furthermore, the feature [labial] corresponds to the use of the lip tract variables, so lip aperture was measured for syllables with [w], [m], [b], or [p] (Gao, 2008; Hall, 2010; Zhang et al., 2019). Similarly, the gesture for labiodental fricatives [f] and [v] was measured by lower lip. As for the vowels, I evaluated the potential pros and cons of different measurements. We could use the same pellet to measure vowels with all qualities, leading to consistent measurement but less precise estimate of each vowel. Alternatively, we could use different tongue pellets for each type of vowel — for instance, T2 for the front vowel and T4 for the back vowel. While our original analysis had pellets varying based on what the annotator thought best represented the acoustics of the vowel, on the recommendation of some anonymous experts, I chose to use a single pellet to represent the vowels. Therefore, in order to test the claims stated earlier, I chose to use one sensor T3 consistently for the *sVd* stimuli. In general, using the above pellet choices did allow us to identify gestures that were consistent with the acoustic waveforms or spectrographic information for the relevant consonants and vowels.

Index	Gesture
T1	tongue tip
T2	tongue blade
T3	tongue dorsum
T4	tongue root
Lip	lip aperture
LL	lower lip

Table 2.4 Measure indexes and their correlated gestures.

The specific sonority difference for each stimulus is based on the sonority indexes of C and V, and the information can be found in the *Difference* column (abbreviated from Sonority Difference) of the stimulus tables (Tables 2.1-2.3). The sonority differences were calculated based on the C and V sonority indexes indicated in the *Sonority* column, where the first number and second number indicate the sonority index of C and V respectively based on the hierarchy in Parker (2012) shown previously in Table 1.6.[3] Note that English has passively or weakly voicing in its voiced stops (Iverson and Salmons, 1995), but they are still labeled as voiced stops.

2.1.3 Data annotation and analysis

The kinematic data were annotated in Matlab using the default settings of the *lp_findgest* algorithm of the `mview` package where gestural onset, gestural offset, nucleus onset, and nucleus offset used the 20% threshold of the velocity profile (Tiede, 2005). The tangential velocity of x and y axes was considered by the algorithm.

The procedure for identifying gestures of the *lp_findgest* algorithm can be found in Section 1.8.1 of Chapter 1.1. Basically, to use the algorithm, the annotator first clicked on a point in the relevant articulatory pellet's information. The mouse click point was usually identified by checking the synchronous acoustic information, and it would roughly be the point of the gestural plateau. After manually identifying the point, the algorithm found the maximum constriction point (MAXC),

[3]In Table 2.2, if a syllable has a diphthong with two vowels of different sonority indexes, the sonority index of the first vowel of the diphthong is recorded as the sonority index for V, since Hsieh (2017) suggests that vowels in English diphthongs are coordinated sequentially. A second option to index diphthong sonority would have been to use the average sonority index of the two targets in the diphthong, which implies that two vowels in diphthong are coupled synchronously as in Dutch and Romanian (Collier et al., 1982; Marin and Goldstein, 2012). In the dissertation, I did not consider the diphthongal realizations of tense vowels. Future research is necessary to probe the nuances of diphthong articulation.

which was the closest velocity minimum to the clicked point. Then, there were two peak velocity points identified before and after the maximum constriction point, which were called PVEL and PVEL2 respectively. After that, the gestural onset point was marked by identifying the 20% peak velocity between the minimum velocity point before PVEL and the peak velocity point (PVEL) itself. The nucleus onset was the 20% peak velocity point between PVEL and the maximum constriction point (MAXC), and the nucleus offset was identified by the 20% peak velocity point of the range between MAXC and the following peak velocity PVEL2. Similarly, gestural offset was the 20% peak velocity point between the range between PVEL2 and the following velocity minimum. In the current study, the velocity in *lp_findgest* algorithm of mview was computed as tangential velocity since multiple components are displayed. The lip aperture was automatically calculated by the *mdp_LipAperture_old* algorithm of the mview package. The *mdp_LipAperture_old* algorithm computed the Euclidean distance between the Lower Lip sensor and Upper Lip sensor at each time point. The formula can be found in (7).

(7)
$$LA = \sqrt{(UL_x - LL_x)^2 + (UL_y - LL_y)^2 + (UL_z - LL_z)^2}$$

Based on the information on the acoustics as well as the articulatory movement trajectories, the consonant gesture and the following vowel gesture of each token were annotated by the first author. For instance, if it is a low vowel, we would expect the T3 gesture to be lower in the vertical dimension. Figure 2.2 shows a sample gestural annotation for [mɑ] — where only relevant rows LA (in white) and T3 (in red) are displayed here for clarity. The white text was added to denote gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF) for the gesture of the consonant, and red texts were added for that of the vowel. These landmarks of the articulation were provided automatically by the algorithm. At the bottom of Figure 2.2, there is an axis indicating time, and the timestamps for gestural onsets, gestural offsets, target onsets, and target offsets were recorded for the consonant gesture and vowel gesture of each token. Using the

timestamp information, CV lags were computed for each CV sequence.

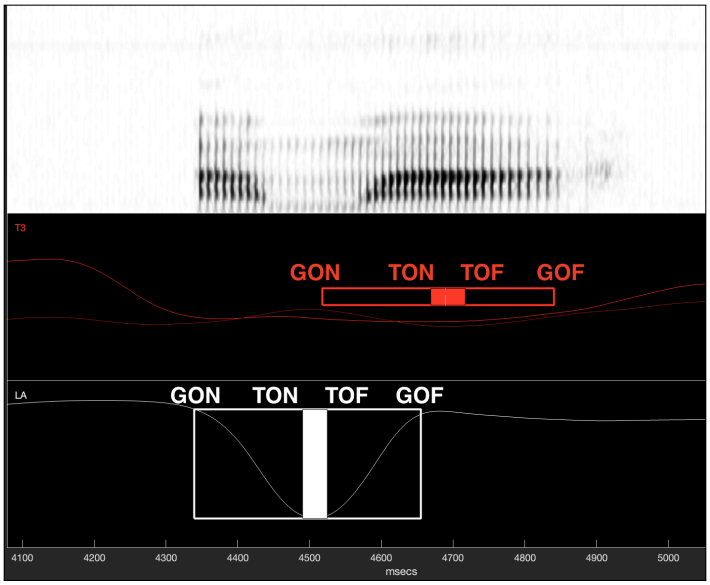

Figure 2.2 Sample annotation of [mɑ] in uhmA of Speaker JW11, Task 16. The white labels refer to the LA (lip aperture) gesture, and the red labels refer to the T3 (tongue dorsum) gesture. The curves show the displacement of sensors at the x and y axes for T3 and other non-LA rows. The y-axis has a lighter color in each row. For both LA and T3, the labels were added by the author to denote gestural onset (GON), target onset (TON), target offset (TOF), and gestural offset (GOF).

After collecting data from the corpus and computing CV lags, the relationship between the CV lags and the sonority difference was analyzed and plotted using the `tidyverse` package (Wickham et al., 2019). Subsequent mixed-effects modeling was done using the `lme4` (Bates et al., 2014) and `lmerTest` (Kuznetsova et al., 2017) packages in R (R Core Team, 2017), where each CV LAG was modeled as a function of SONORITY DIFFERENCE, with PARTICIPANT, CONSONANT DURATION, and sometimes WORD as random intercepts. CONSONANT DURATION was included in the model to address the alternative explanation that longer consonant duration is related to a larger CV lag. Also, WORD is not used as a random intercept if in the subset there is a one-to-one mapping between sonority difference and word.

2.2 Results

2.2.1 Overall analysis

All 57 different speaker datasets in the Wisconsin X-ray Microbeam corpus have been included in the current analysis, and altogether 3399 tokens were measured. Excluding [gɑ] and [kɑ], there were 3214 tokens measured in the overall analysis of all the data, and the claim I proposed was supported. Specifically, I claimed that the CV lag positively correlates to the sonority difference between the C and V. As can be observed in Figure 2.3, there is indeed such a positive correlation between CV lag and the sonority difference.

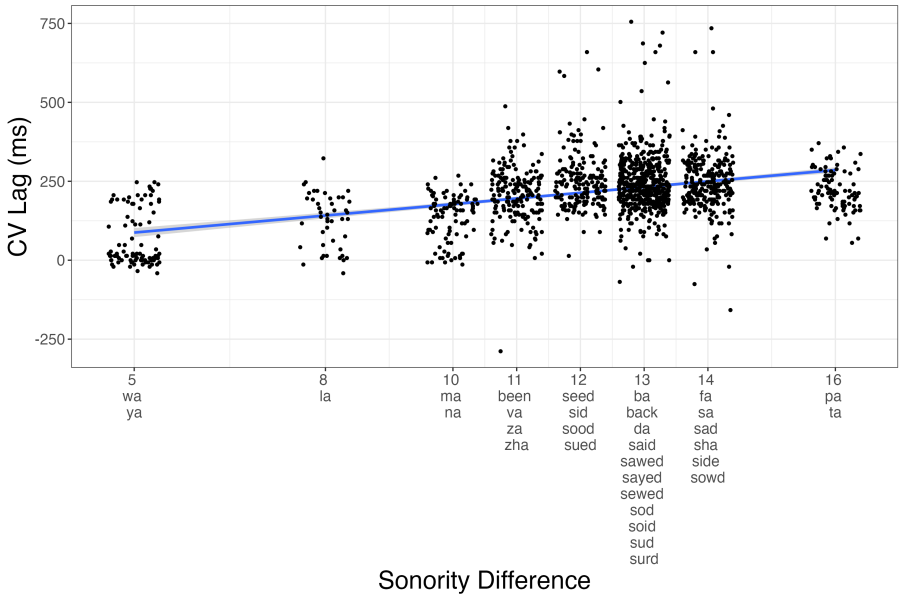

Figure 2.3 CV lag increases with sonority difference for all data based on target onset. The *geom_jitter* option (size = 0.9) is used to spread out the overlapping dots for clarity.

Furthermore, the mixed effects model results, presented in Table 2.5, are also consistent with the visual inspection of the data above. Though I have not presented the data here in the interest of concision, CV lag based on other landmarks such as gestural onset, gestural offset, and target offset all exhibited the same expected pattern.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-1.34	35.39	30.39	-0.04	0.97
Sonority difference	17.91	2.84	29.97	6.30	<0.0001

Table 2.5 Mixed effects model results for all data.

One possible explanation of the observation may be that there is a positive correlation between sonority difference and consonant duration, and longer consonant duration is related to a larger lag. To evaluate this alternative explanation, I measured the gesture duration of the consonant and included it in the statistical model. Adding consonant duration as one more random intercept in the mixed effect model still shows a positive correlation between sonority difference and CV lag based on target onset as in Table 2.6. Henceforth, I will include consonant duration as a random intercept in all comparisons to control for this confound.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	25.95	27.84	30.33	0.93	0.34
Sonority difference	16.49	2.21	28.60	7.45	<0.0001

Table 2.6 Mixed effects model results for all data. Random intercepts: C duration, word, participant.

While an analysis with *all* the data and the use of a sonority difference score as a predictor is straightforward to statistically model and has higher statistical power, it does have some issues. First, the analysis assumes that the sonority scale is linear and not just relative, which is contrary to most phonologists' beliefs. Furthermore, it collapses across different articulators or gestures. For these reasons, I also analyzed more nuanced sub-groups of the data. Specifically, I looked at sets of stimuli that control for the place of articulation or gesture of the consonant. I also looked at comparisons where there is agreement on the predicted lag variation even if different sonority scales are considered. The results for the subsets for claim (5a) are presented in Section 2.2.2, and those for claim (5b) in Section 2.2.3. Analyzing the subgroups also allows us to address potential alternative explanations of the observation such as jaw movement, or place of articulation.

2.2.2 Claim 5a: the same vowel with different consonants

2.2.2.1 Different consonants using lips as the primary articulator

To eliminate gesture as a potential confounding variable, the results for stimuli with different consonants using lips as the primary articulators were separated into the lip aperture group (target CV sequences [wɑ], [mɑ], [bɑ], and [pɑ]) and lower lip group (target CV sequences [fɑ] and [vɑ]).

Figure 2.4 shows the results for the lip aperture group ([wɑ], [mɑ], [bɑ], and [pɑ]), where the CV lag based on target onset clearly shows the expected pattern that CV lag increases with sonority difference. Additionally, the mixed effects model indicated a statistically significant positive slope as in Table 2.7.[4]

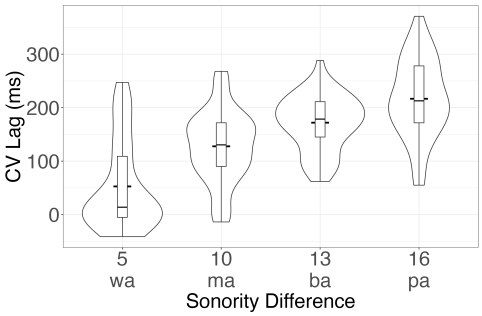

Figure 2.4 CV lag based on target onset for lip aperture consonants.

This sub-analysis of the data only involves vowels of one quality, and the variation of gestural timing is found. Therefore, vowel quality alone cannot be used to account for the observation.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-20.17	14.54	154.23	-1.39	0.17
Sonority difference	14.71	1.20	130.12	12.26	<0.0001

Table 2.7 Mixed effects model results for lip aperture consonants. Random intercepts: C duration, participant.

Even though stimuli from the lip aperture group (target CV sequences [wɑ], [mɑ], [bɑ], and [pɑ]) all involve the same oral gesture for the consonant, they do not share the same manner of articulation or voicing; however, manner or voicing may affect gestural overlap (Du and Gafos, 2023). To control

[4]In fact, all other CV lag landmarks — CV lag based on target onset, target offset, and gestural offset — exhibited significant results in the expected direction. This is mostly true for other sub-analyses of the study, too.

for voicing, [mɑ] and [bɑ] were compared. Note, the comparison also controls for jaw movement. The descriptive plot for the CV lag comparison (Figure 2.5) and the corresponding mixed-effects model (Table 2.8) suggest that CV sequences with oral bilabial stops generally induced a larger CV lag than their counterparts with nasal consonants. This replicates the observation in Shaw and Chen (2019) mentioned above for Mandarin that CV lag for nasal stop is shorter than CV lag for oral stop. Furthermore, the major difference between the [m] and [b] articulations is the lowering or raising of velum, and there is no obvious articulatory reason that velum movement by itself should cause gestural lag variation between the lips and the tongue. The finding suggests that jaw movement may not be a valid alternative account of the observed gestural lag variation, since in this pair the two segments [m] and [b] involve similar degrees of jaw movement. Therefore, by comparing CV lag for [mɑ] and [bɑ], our claim is more strongly supported.

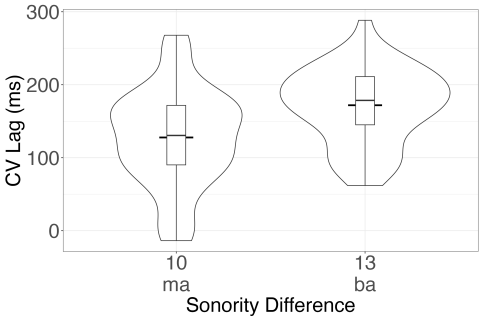

Figure 2.5 CV lag based on target onset comparison for [mɑ], [bɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-19.57	42.86	47.43	-0.46	0.65
Sonority difference	14.73	3.68	45.00	4.01	0.0002

Table 2.8 Mixed effects model results for [mɑ], [bɑ]. Random intercepts: C duration, participant. Adding or not adding C duration in the model as a random effect does not change the model results.

Previous studies have found that consonant manner and place could lead to gestural coordination variation (Bombien et al., 2013; Wright, 1996; Pouplier et al., 2022). Comparing the CV lag for [pɑ] and [bɑ] can control for the potential confounding factors since the pair has the same manner and place of articulation, as well as jaw movement. The results for the [pɑ] and [bɑ] comparison can

be found in Figure 2.6 and Table 2.9. There is a significant positive correlation between sonority difference and CV lag for stimuli with voiced and voiceless bilabial stops and the same vowel. This shows that manner and place of articulation cannot account for the observation, which strengthens the claim supporting the link between sonority and gestural timing.

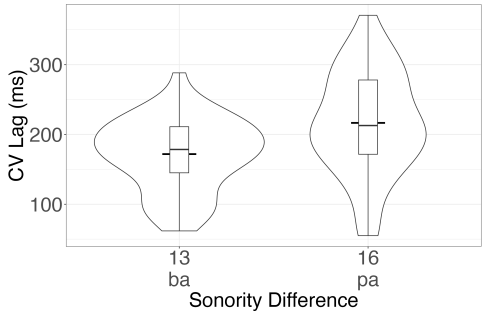

Figure 2.6 CV lag based on target onset comparison for [pɑ], [bɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-15.23	55.06	47.22	-0.28	0.78
Sonority difference	14.40	3.77	45.51	3.82	0.0004

Table 2.9 Mixed effects model results for [pɑ], [bɑ]. Random intercepts: C duration, participant.

Figure 2.7 and Table 2.10 show the results for stimuli involving lower lip as the primary consonant articulator. The visual inspection of the plot suggested that CV lags based on target onsets increase with the rise in sonority difference, and correspondingly a positive correlation was observed in the statistical modeling though the effect is not statistically significant. Given that the estimate is in the same direction, I suggest that this might be a power issue, related to a limited amount of data.

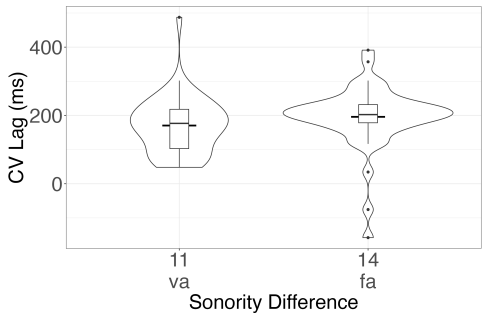

Figure 2.7 CV lag based on target onset comparison for [fɑ], [vɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	92.13	57.05	44.08	1.62	0.11
Sonority difference	7.11	4.54	42.41	1.57	0.13

Table 2.10 Mixed effects model results for [fɑ], [vɑ]. Random intercepts: C duration, participant.

2.2.2.2 Different consonants using tongue tip as the primary articulator

In this section, I present the results of my analysis for the stimuli where tongue tip was used as the primary articulator for the consonant. Within this group, there are nine *uhCa* nonce words with the target CV sequences [jɑ], [lɑ], [nɑ], [zɑ], [ʒɑ], [dɑ], [ʃɑ], [sɑ], and [tɑ]. The results for stimuli involving the T1 pellet for the consonant can be seen in Figure 2.8 and Table 2.11. Both the visual inspection and the mixed effects model again suggest a positive correlation between CV lag and sonority difference.

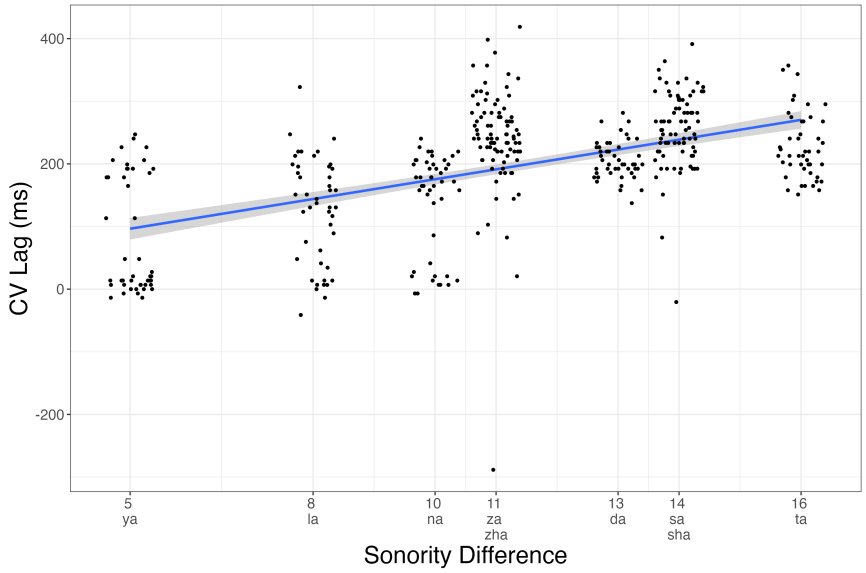

Figure 2.8 CV lag based on target onset for consonants using T1 gesture. The *geom_jitter* option (size = 0.9) is used to spread out the overlapping dots for clarity.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	21.72	44.31	6.77	0.49	0.64
Sonority difference	15.53	3.75	6.70	4.14	0.005

Table 2.11 Mixed effects model results for consonants using T1 gesture. Random intercepts: C duration, participant.

However, despite the clear positive relationship, the T1 group analyzed above involves different places of articulation for the consonant. Namely, six of them ([lɑ], [nɑ], [zɑ], [dɑ], [sɑ], and [tɑ]) are alveolar consonants, while [ʒ] and [ʃ] are postalveolar and [j] is palatal. To make sure stimuli with the same place of consonant articulation are compared to each other, the stimuli with an alveolar consonant are analyzed as a whole, as in Figure 2.9 and Table 2.12. We still see a significant positive slope with roughly the same magnitude of difference.

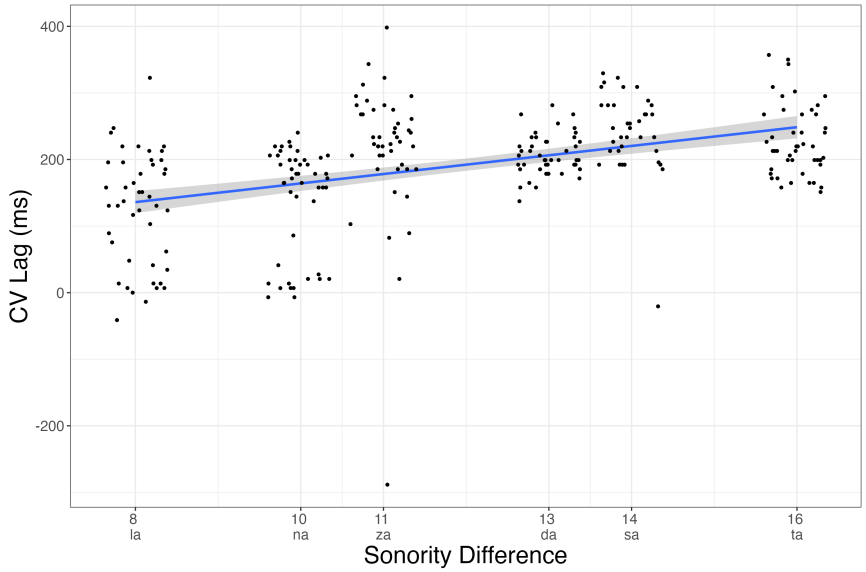

Figure 2.9 CV lag based on target onset for alveolar consonants. The *geom_jitter* option (size = 0.9) is used to spread out the overlapping dots for clarity.

Target Onset	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	23.36	45.29	3.71	0.52	0.64
Sonority difference	14.24	3.68	3.68	3.87	0.02

Table 2.12 Mixed effects model results for alveolar consonants. Random intercepts: C duration, participant.

Note that the group of stimuli with alveolar consonants ([lɑ], [nɑ], [zɑ], [dɑ], [sɑ], and [tɑ]) involve different manners of articulation and voicing for the consonant articulation. To exclude the account that jaw movement is the cause of the gestural lag variation, and to control for voicing, the voiced alveolar consonants [nɑ] and [dɑ] are compared in Figure 2.10 and Table 2.13. Again,

the comparison clearly shows that the larger sonority difference between C and V significantly correlates with larger CV lags for [nɑ] and [dɑ].

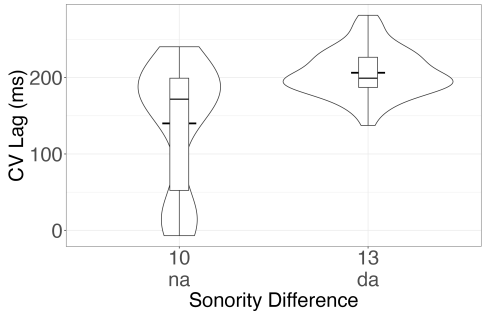

Figure 2.10 CV lag based on target onset comparison for [nɑ], [dɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-80.88	49.13	90.00	-1.65	0.10
Sonority difference	22.09	4.24	90.00	5.21	<0.0001

Table 2.13 Mixed effects model results for [nɑ], [dɑ]. Random intercepts: C duration, participant. Adding or not adding C duration as a random effect in the model yielded the same results.

To control for orality and jaw movement, I compared [tɑ] and [dɑ] as in Figure 2.11 and Table 2.14. The results show that there is a significant positive correlation between gestural lag and sonority difference for the two stimuli that differ in voicing. The above result shows that manner or place of articulation, or jaw movement, cannot account for the observation, since the stimuli are the same in the two aspects but still differ in gestural timing.

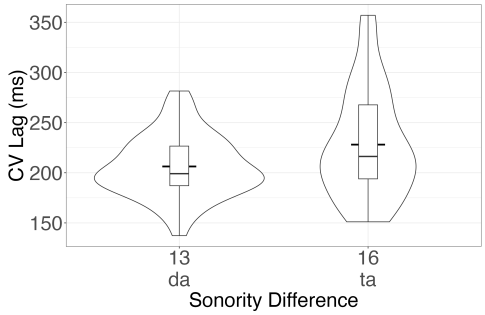

Figure 2.11 CV lag based on target onset comparison for [tɑ], [dɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	134.72	37.42	43.12	3.60	0.00
Sonority difference	5.84	2.53	40.82	2.31	0.03

Table 2.14 Mixed effects model results for [tɑ], [dɑ]. Random intercepts: C duration, participant.

For a similar reason, I also looked at [zɑ] and [sɑ], which are different in voicing. The results are in Figure 2.12 and Table 2.15. The visual inspection and the mixed effects modeling generally show a positive correlation, but it is not significant. Given the expected direction of the estimate, the insignificant effect is likely due to the lack of statistical power.[5]

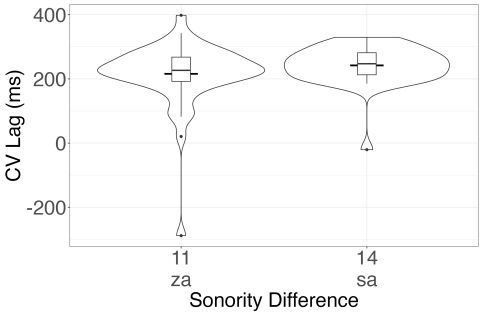

Figure 2.12 CV lag based on target onset comparison for [zɑ], [sɑ].

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	105.92	72.41	50.04	1.46	0.15
Sonority difference	9.74	5.77	47.96	1.69	0.10

Table 2.15 Mixed effects model results for [zɑ], [sɑ]. Random intercepts: C duration, participant.

2.2.3 Claim 5b: the same consonants with different vowels

In general, I conclude that claim 5a — for CV syllables with the same V, a less sonorous C leads to a larger CV lag — has been supported by the Wisconsin Microbeam corpus data. I now turn to probing the second claim by keeping the consonant constant and varying the vowel. As mentioned before, fifteen nonce words with the template *sVd*, with the crucial vowel in between, were measured along with two real words *back* and *been*. I first present the results for the *sVd* words (Section 2.2.3.1) and then present the results for the bV real words (Section 2.2.3.2).

[5]I did not compare the postalveolar fricative [ʒ] and [ʃ] since T1 is not a precise pellet position to measure post-alveolar consonants.

2.2.3.1 sVd words

In this section, I looked at the 15 sVd stimuli in our stimulus set from the corpus. The results for the 15 sVd stimuli that are shown in Figure 2.13 do not suggest a clear positive relationship, though the estimate is in the expected direction. Moreover, the mixed effects models for target onsets indicate a positive correlation as in Table 2.16, though the positive correlation is not statistically significant.

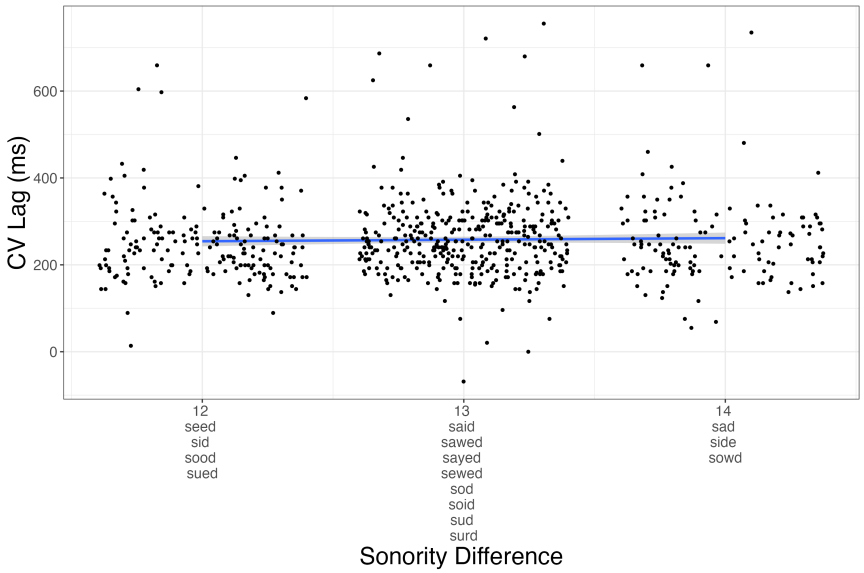

Figure 2.13 CV lag based on target onset for sVd words. The *geom_jitter* option (size = 0.9) is used to spread out the overlapping dots for clarity.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	187.61	69.42	13.53	2.70	0.02
Sonority difference	6.25	5.35	13.47	1.17	0.26

Table 2.16 Mixed effects model results for sVd words. Random intercepts: C duration, word, participant.

The null result, however, should be interpreted with caution since there is a tradeoff with using the same sensor to measure all vowels, which are of different quality, as mentioned before. Given that the estimate was in the expected direction, the null result is potentially due to additional variance in the measurements. It is possible that the use of the same pellet for all vowels resulted in imprecise vowel measurements, and therefore led to more noise. Note, the issue is further

exarcerbated because the consonants are alveolar. To address the issue of compatibility and the issue of varying vowel quality, I present the following analysis which involves a subset of the original dataset. Originally, for each stimulus, either T2 or T3 was used for vowel measurement, depending on what I thought was most indicative of the vowel in the acoustics — this was done prior to any data analysis and was based on the judgment of the annotator. However, to address the worry of consistency in the use of pellets, I only analyzed the measurements with the pellet that was used for the majority of the tokens of a stimulus. For instance, for *seed*, there are 50 measurements using T2 and 6 measurements using T3, so the 50 CV syllables with T2 measurement were included and analyzed in the subset. Table 2.17 showed vowel measurement for each stimulus. It seems that overall T2 matches more with the acoustic information.

	Stimuli	**IPA**	**V Pellet**
(1)	seed	[**si**d]	T2
(2)	sid	[**sɪ**d]	T2
(3)	sued	[**su**d]	T2
(4)	*sood	[**sʊ**d]	T2
(5)	*sayed	[**sɛ**ɪd]	T3
(6)	surd	[**sɜ˞**d]	T2
(7)	said	[**sɛ**d]	T2
(8)	*sud	[**sʌ**d]	T2
(9)	sewed	[**so**d]	T2
(10)	sawed	[**sɔ**d]	T2
(11)	*sowd	[**sa**ud]	T2
(12)	side	[**sa**ɪd]	T2
(13)	sod	[**sɔ**d]	T2
(14)	*soid	[**sɔ**ɪd]	T2
(15)	sad	[**sæ**d]	T2

Table 2.17 sVd stimuli for claim (5b). The stimuli came from task 13 of the Wisconsin X-ray Microbeam Database. The target CV sequences are in boldface and underlined. The *C-V Pellets* column shows the pellet positions for C and V.

The results for the subset of *sVd* stimuli can be found in Figure 2.14 and Table 2.18. The estimate is much larger. Furthermore, if word is removed as a random intercept, I also have more confidence to reject the null hypothesis (estimate = 12.1; Pr(>|t|)=0.07). These are just speculative thoughts at this point. However, given the direction of the estimate, I believe the insignificant result

is due to insufficient statistical power and the complexity of vowel measurement. Analyzing more data may bring out the positive correlation more significantly.

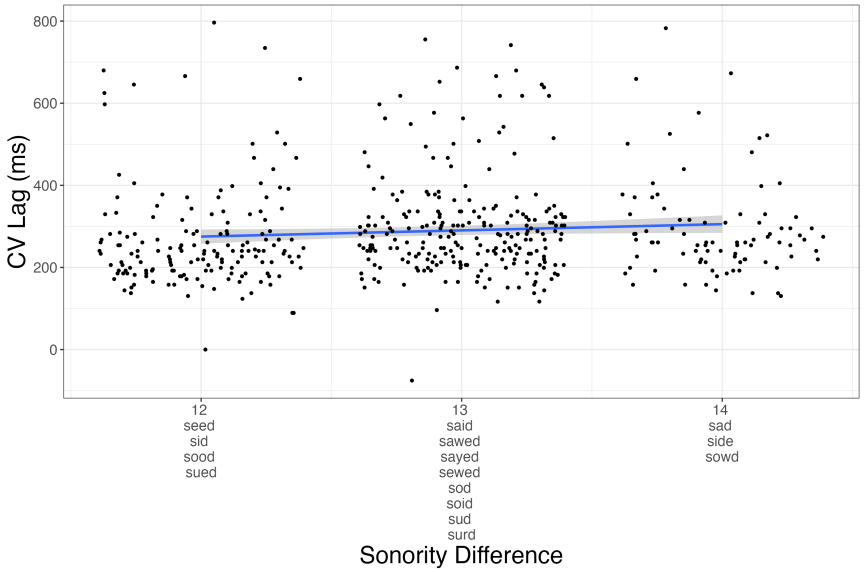

Figure 2.14 CV lag based on target onset for a subset of sVd words. The *geom_jitter* option (size = 0.9) is used to spread out the overlapping dots for clarity.

Target Onset	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	142.47	161.43	11.70	0.88	0.40
Sonority difference	12.09	12.49	11.80	0.97	0.35

Table 2.18 Mixed effects model results for a subset of sVd words. Random intercepts: C duration, word, participant.

2.2.3.2 bV real words

While there is some separability between the tongue tip gesture and the tongue body gesture to parse out the initial consonant and vowels in sVd stimuli, there is still a possibility of interference between the gestures. To resolve this issue, I also looked at bV words, which have different C and V articulators. The two bV real words in question have a bilabial consonant measured by lip aperture and a vowel measured by T3. The expected pattern cannot be clearly seen in target onsets as exemplified by Figure 2.15 and Table 2.19.

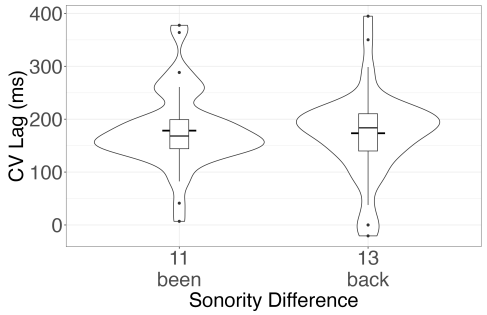

Figure 2.15 CV lag based on target onset for bV real words.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	193.07	83.07	32.38	2.32	0.03
Sonority difference	-1.49	6.88	31.25	-0.22	0.83

Table 2.19 Mixed effects model results for bV real words. Random intercepts: C duration, participant.

Given the issue of the use of a uniform pellet discussed earlier, I followed up with analysis similar to the *sVd* stimuli, I also subset the *bV* stimuli. Namely, I annotated according to the acoustics, and only included the majority pellet in the analysis. Again, I would like to remind the reader that this annotation was prior to any analysis. For most speakers, the T3 pellet best matches the vowel acoustics in the *back* and *been* stimuli.

	Stimuli	IPA	V Pellet
(16)	been	[**bi**n]	T3
(17)	back	[**bæ**k]	T3

Table 2.20 bV stimuli for claim (5b). The stimuli *been* is from task 9 and *back* is from task 100 of the Wisconsin X-ray Microbeam Database. The target CV sequences are in boldface and underlined.

The results for the subset can be found in Figure 2.16 and Table 2.21. For *bV* stimuli, there is a significant positive correlation between sonority difference and lag in CV syllables. I believe that there are some significant results for *bV* but not *sVd* due to the separate lip and tongue measurements for C and V respectively for bilabial stimuli.

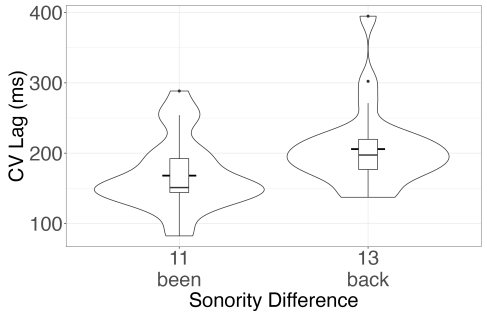

Figure 2.16 CV lag based on target onset for a subset of bV real words.

Target Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-63.13	72.48	34.26	-0.87	0.39
Sonority difference	21.06	6.03	33.59	3.49	0.001

Table 2.21 Mixed effects model results for a subset of bV real words. Random intercepts: C duration, participant.

2.2.4 Could the significant positive correlation be an artifact of vowel displacement?

Shaw and Chen (2019) observed that CV lag based on gestural onsets is negatively correlated with the displacement of the vowel from gesture onset to the achievement of the target. Essentially, if the tongue body has to move more to achieve the vowel target, then the movement starts earlier, and consequently, the CV lag based on gestural onsets is smaller. It is therefore logically possible that my results are somehow artifactual and based on the relation observed by Shaw and Chen (2019). To check for this possibility, I ran another analysis wherein I added another fixed effect, namely, the horizontal distance from vowel gesture onset to target achievement.[6] This additional analysis involved all the stimuli. I chose the whole dataset as it was the largest stimulus set and therefore the analysis would suffer the least in terms of statistical power from the addition of a post-hoc variable. Note that vowel displacement could be an estimate of jaw movement. Therefore, the post-hoc analysis also serves as another exploration of the potential effect of jaw movement on gestural coordination.

[6]Shaw and Chen (2019) observed that this was a better predictor than the Euclidean distance traversed. Therefore, I employ this measure.

2.2.4.1 Post-hoc analyses with vowel displacement using all the stimuli

The model with both sonority difference and vowel displacement as independent variables is in Table 2.22. As can be seen from the table, the model shows that there is a negative relationship between CV lag based on target onsets and vowel displacement. This replicates the findings of Shaw and Chen (2019). Crucially, for our purposes, the effect of sonority difference is still clearly present and to almost the same degree as in the original model presented before (Estimate in original model = 16.49 vs. estimate in the current model = 15.76). Again, I interpret the result as showing that the main finding in this article, that there is a positive relationship between sonority difference and CV lag, once there is an adjustment for the contributory effect of vowel displacement.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	29.20	25.93	30.83	1.13	0.27
Sonority difference	15.76	2.06	29.00	7.65	< 0.00001
Vowel displacement	-3.97	0.67	901.04	-5.95	< 0.00001

Table 2.22 Mixed effect model for all stimuli with sonority difference and vowel displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

2.2.5 Could the significant positive correlation be a confound of jaw movement?

It is obvious that jaw movement can vary by consonant (Gracco and Lofqvist, 1994). Furthermore, it has been observed that jaw movement correlates with variation in gestural coordination (Gracco, 1994; Gracco and Lofqvist, 1994; Mooshammer et al., 2003; Redford, 1999; MacNeilage and Davis, 2000). Could the significant positive correlation in the current study be actually due to a confound of jaw movement? I observed both the voiced C-voiceless C comparison, as well as the nasal C-oral C comparison exhibited a positive correlation between sonority and gestural timing, despite having putatively similar jaw movements. However, it is still worth confirming for those comparisons that jaw movement is not the (unique) source CV lag variation observed here. In the following subsections, I tested for this possibility by adjusting for any effect of jaw movement in our data. Consonant displacement was chosen to be an approximation of jaw movement. Inspired by Shaw and Chen (2019), consonant displacement means the horizontal distance from consonant gesture onset to target achievement. I evaluated

consonant displacement on the whole dataset, and all pairs that I claimed are controlled for jaw movement. These are pairs of stimuli that differ in voicing or nasality.

2.2.5.1 Post-hoc analyses with consonant displacement using all the stimuli

The model with both sonority difference and consonant displacement as fixed effects is in Table 2.23. I can see that there is a significant positive correlation between consonant displacement and CV lag (estimate = 2.19). However, this does not show that the correlation with sonority difference was confounded since the sonority difference effect remains effectively unaltered (estimate in original model = 16.49 vs. estimate in the current model = 16.76). In the following subsections, I am going to confirm that the pairwise comparison which I believed controlled for jaw movement indeed exhibits little effect of consonant displacement.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	24.78	24.31	30.07	1.02	0.32
Sonority difference	16.76	1.93	28.02	8.70	< 0.00001
C displacement	2.19	0.54	266.73	4.03	0.0001

Table 2.23 Mixed effect model for all stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

2.2.5.2 Post-hoc analyses with consonant displacement using the voicing pairs

There were two pairs of stimuli which differ in consonant voicing used to control for jaw movement — [pɑ, bɑ] and [tɑ, dɑ]. Here, I test whether the pairs truly control for consonant displacement, which is an approximation of jaw movement.

I first look at the [pɑ, bɑ] pair. The model with both sonority difference and consonant displacement as fixed effects is in Table 2.24. There is an insignificant negative correlation between consonant displacement and CV lag variation. Since the effect size for sonority difference still remains similar when one considers consonant displacement (estimate in original model = 14.40 vs. estimate in the current model = 15.03), I conclude that for [pɑ, bɑ] comparison, the observed positive correlation is not a confound of jaw movement.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-53.17	58.65	53.91	-0.91	0.37
Sonority difference	15.03	3.76	46.18	4.00	0.0002
C displacement	-2.31	1.28	80.64	-1.81	0.07

Table 2.24 Mixed effect model for [pɑ], [bɑ] stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

I then looked at another pair where the stimuli differ in consonant voicing — [tɑ, dɑ]. The model with both sonority difference and consonant displacement as fixed effects is in Table 2.25. When considering consonant displacement, there is still a significant positive correlation between CV lag and sonority difference (Estimate in original model = estimate in the current model = 5.84). The effect size, though statistically significant, is smaller than other pairs. This is probably due to the fact that coronal consonants share the same tongue articulator with vowels. Including consonant displacement in the model shows that when controlled for jaw movement, the [tɑ, dɑ] pair exhibited a positive correlation between sonority difference and CV lag.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	136.09	39.03	44.78	3.49	0.00
Sonority difference	5.84	2.63	41.90	2.22	0.03
C displacement	-1.34	1.47	64.39	-0.91	0.36

Table 2.25 Mixed effect model for [dɑ], [tɑ] stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

2.2.5.3 Post-hoc analyses with consonant displacement using the nasality pairs

In the previous subsection, I looked at two pairs that differ in consonant voicing. I confirmed that those two pairs showed significant positive correlations between CV lag and sonority, even when controlled for jaw movement. In this subsection, I conduct similar analyses for pairs of stimuli that differ in nasality of the consonant — [bɑ, mɑ] and [nɑ, dɑ]. I claimed that the pairs should have controlled jaw movement, but I am going to confirm it here.

The model for [bɑ, mɑ] is shown in Table 2.26. Even though there is an insignificant negative correlation between consonant displacement and CV lag, there is still a significant positive

correlation between sonority and CV lag (estimate in the original model = 14.73 vs. estimate in the current model = 13.73).

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-35.18	44.39	52.30	-0.79	0.43
Sonority difference	13.73	3.76	47.69	3.65	0.001
C displacement	-2.24	1.53	51.63	-1.47	0.15

Table 2.26 Mixed effect model for [bɑ], [mɑ] stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

The model for [nɑ, dɑ] is shown in Table 2.27. Again, there is a negative correlation between consonant displacement and CV lag. However, the effect of the relationship between sonority and CV lag appears to be unchanged (estimate in the original model = 22.09, and estimate in the current model = 22.68).

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-81.43	48.35	89.00	-1.68	0.10
Sonority difference	22.68	4.18	89.00	5.43	< 0.00001
C displacement	-5.08	2.57	89.00	-1.98	0.05

Table 2.27 Mixed effect model for [nɑ], [dɑ] stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

2.2.6 Summary

In general, experiment 1 using English corpus data showed that there is a positive correlation between CV lag and sonority difference. The overall results, including results for subgroups, are summarized in Table 2.28, and the pairwise comparison results can be found in Table 2.29. In Table 2.28, all the data as well as subgroups of the whole dataset showed the expected positive correlation except for the *sVd* stimulus group. The expected positive correlation can still be observed when considering vowel or consonant displacement. In Table 2.29, all the pairs that control voicing, nasality, or vowel height exhibited a significant positive correlation between CV lag and sonority difference. However, the comparison between voiced and voiceless fricatives did not show the expected pattern. The non-significant results for *sVd* stimuli may be due to the interaction between C and V measures since both use tongue sensors. Also, the non-significant of the fricative pairs

may be due to that the pair does not differ in voicing in their realization. Future research may consider coding sonority index according to the voicing realizations of obstruents.

Dataset (English corpus data)	**Estimate (sonority diff)**	**Estimate (displace)**
All English corpus data	16.49 ***	
All English corpus data, V displacement	15.76 ***	-3.97 ***
All English corpus data, C displacement	16.76 ***	2.19 ***
sVd stimuli (subset)	12.09	
T1 C stimuli	15.53 **	
Alveolar C stimuli (la, na, za, da, sa, ta)	14.24 *	
Lip aperture (wa, ma, ba, pa)	14.71 ***	

Table 2.28 Summarizing the results of experiment 1. *** means that p $\leqslant$ 0.001; ** means that p $\leqslant$ 0.01; * means that p $\leqslant$ 0.05.

Pairwise comparison	**Stimulus pair**	**Estimate (sonority difference)**
Nasality differ	ma, ba	14.73 ***
	na, da	22.09 ***
Voicing differ, stop	pa, ba	14.40 ***
	da, ta	5.84 *
Voicing differ, fricative	fa, va	7.11
	sa, za	9.74
Vowel height (subset)	been, back	21.06 ***

Table 2.29 Summarizing pairwise comparison of experiment 1. *** means that p $\leqslant$ 0.001; ** means that p $\leqslant$ 0.01; * means that p $\leqslant$ 0.05.

2.3 Conclusion

Even though the main claim was supported, this experiment on corpus data was not ideal for the following reasons. First, the stimuli in the corpus study were restricted because of using an existing corpus. For claim (5a) that for CV syllables with the same V, a less sonorous C leads to a larger CV lag, only nonce words with low vowel [ɑ] were involved. It is not clear whether real words with other vowels, and even [ɑ] will support the claim. Moreover, for claim (5b) that for CV syllables with the same C, a more sonorous V leads to a larger CV lag, only [s] and [b] were tested as the onsets, and analyzing a variety of consonants is necessary. Second, experiment 1 was based on English corpus data, lacking a cross-linguistic validation. To address the above issues, EMA

experiments using carefully designed English and Mandarin real words are proposed to evaluate the main claim.

CHAPTER 3

EXPERIMENT 2: ENGLISH EMA STUDY

The study of English corpus data showed that there was a positive correlation between CV lag and sonority difference. The same pattern has been consistently observed for stimuli with controlled C and varying V, as well as for stimuli with controlled V and varying C. The positive correlation has also been observed when considering alternative factors such as jaw movement, voicing, or nasality. However, it has limited stimulus selection. To address this issue, I conducted an EMA experiment in English, as described in the following sections of this chapter.

3.1 Methods

3.1.1 Data collection

The NDI Vox-EMA System (VOX) manufactured by Northern Digital Inc. (NDI) was used to collect articulatory data of the study in the Phonetics Lab of Michigan State University. Articulatory data was collected at a sampling rate of 400 Hz, and acoustic data was recorded simultaneously at a sampling frequency of 16 kHz. 8 sensors were attached to the participants' articulators and other reference points using PeriAcryl Oral Tissue Adhesive. Specifically, 3 sensors were glued to the tongue — one on the tongue tip (TT), about 1 cm from the anatomical tip; one on the tongue dorsum (TD), as far back as comfortable; one between tongue tip (TT) and tongue dorsum (TD) so that there was equal distance between two sensors — tongue blade (TB). Two sensors, used to track lip movements, were glued to the vermillion border of the upper and lower lips respectively. Reference sensors were glued to the left and right mastoid, and nasion to correct head movement. Some rough positions of the sensors can be found in Figure 3.1.

3.1.2 Participant recruitment

English participants were recruited through email. Recruitment emails were sent to sections of *LIN 401 Introduction to Linguistics* and *IAH 231C Roles of Language in Society* of the 2023 Fall semester. See the recruitment email in Appendix A. Potential participants first completed a pre-screening survey via Google Forms. The pre-screening survey can be found in Appendix B.

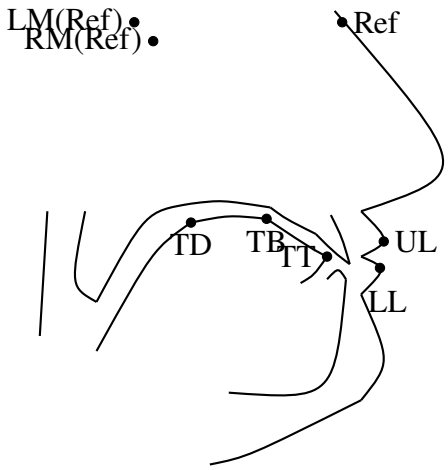

Figure 3.1 Approximate pellet placement locations. Ref: reference sensor. LM: left mastoid; RM: right mastoid. UL: upper lip. LL: lower lip. TT: tongue tip. TB: tongue blade. TD: tongue dorsum.

Then, they were contacted by the experimenter to schedule the actual experiment.

3.1.3 Experimental set-up and procedure

The participants were given time to read and sign the consent forms when they first greeted the experimenters. Before or after the participant read the consent form, one experimenter briefly introduced EMA, its setup, and the participants' task in layman terms. Participants were then instructed to use a disposable toothbrush to brush the midline of their tongue, before they rinsed their mouth with water. Participants were provided with water to drink during the experiment, and they were advised to use the restroom before the actual experiment, which lasted 45 minutes to 1.5 hours. When ready, participants sat facing a computer screen. The field generator was placed towards the left side of the participant, and its position was adjusted after the sensor application so that the sensors were roughly at the center of the field. See Figure 3.2 for a photo of the lab setup. During the experiment, one experimenter sat to the right of the participant to monitor the recording process. Another experimenter was sitting behind the participant to fill in the protocol file, which was a document about the experimental procedure, containing information about mispronounced words or falling sensors. After the attachment of sensors on mastoids and the nasion, participants were instructed to bite a bite plane in a still position, and the still position of the sensors was recorded 3 times where each recording was 2 seconds.

Then the lingual sensors were glued to the participants' mouths. First, sensors were glued to

Figure 3.2 The set-up of the EMA lab from the view of the participant.

the tongue, from the back to the front. Then, sensors were glued to the upper and lower lips of the participants. Dental edible pigments were used to denote 3 points for tongue sensors. The first mark was the 1cm point to the tongue tip, and this was the Tongue Tip sensor. The second mark was about 5-6 cm from the tongue tip (not Tongue Tip sensor point), or in some cases, the furthest back that the participant could tolerate without discomfort. This was the Tongue Dorsum sensor. Lastly, the third mark denoting the position of the Tongue Blade sensor was on the midpoint of the first two marks.

At the end of the experiment session, 30 dollars in cash was given to the participant, and participants signed the receipt to indicate that they had received the money.

Altogether data from 18 English participants were collected, and those from 10 were annotated and analyzed in the current study. For the 10 participants, there were already 7268 manual annotations of gestures for the EMA English data. I annotated the data according to the reverse order of data collection, and the rest of the data were not annotated due to time constraints. Among the 10 participants, 9 were female and 1 was male, and the average age was 19.9 years old.

3.1.4 Data processing and annotation

The data collected were head-corrected and annotated in Matlab. I used the *findgest* algorithm, where gestural onset, gestural offset, nucleus onset, and nucleus offset used the 20% threshold of the velocity profile (Tiede, 2005). The tangential velocity rather than absolute velocity was used since multiple components are displayed.

I annotated the data according to the following assumptions. The gesture was selected based on previous literature, the understanding of articulatory movement, and by considering the consistency of the overall analysis. For instance, alveolar consonants like [t,d,n,s,l] were measured by tongue tip, and [p, b, m, w] were measured by lip aperture. The vowels were annotated by tongue dorsum, which is to ensure the consistency of the comparison across all stimuli. The disadvantage of the decision is that some sensors may not be the *exact* tongue gesture used, but some tradeoff has to be made and arguably tongue gestures are not independent of each other. In each annotation, the click was on the mid-point of the gestural movement area. In annotating a lip aperture, for instance, the closure was located roughly by the acoustics and then automatically by the algorithm.

During the annotation process, sometimes I was not confident about the annotation. In such cases, the annotation was marked "Questionable", "MultipleMeasure", "NoneDefault", "SensorUnavailable", "Mispronounced". The label names are intuitive, but their meanings can be found in Appendix E. For instance, if a word (A) was pronounced incorrectly (as B), then the actual pronunciation (B) as well as the label "Mispronounced" were coded in the datasheet. Out of 3528 syllables in question, 3241 syllables (92%) did not have any labels, and these unambiguous annotations were analyzed in the current study.[1]

3.1.5 Data analysis

The data analysis process for the English study is similar to that of experiment 1. After collecting data from the corpus and computing CV lags, the relationship between the CV LAG and the SONORITY DIFFERENCE was analyzed and visualized using the `tidyverse` package (Wickham et al., 2019). Subsequent mixed-effects modeling was done using the `lme4` (Bates et al., 2014) and

[1] When syllables regardless of label or certainty level were analyzed, similar results were shown but not presented here.

lmerTest (Kuznetsova et al., 2017) packages in R (R Core Team, 2017), where the CV LAG was modeled as a function of the SONORITY DIFFERENCE, with PARTICIPANT and CONSONANT DURATION as random intercepts. WORD was chosen as another random effect in the comparison involving different types of stimuli.[2] For smaller subgroups where stimuli were controlled for C or V and only had one varying V or C, WORD was not considered as a random effect since the stimuli in the subset perfectly correlate with SONORITY DIFFERENCE.[3]. For instance, in the pairwise comparison of *peak*, *pack*, WORD was not considered as a random effect since the potential random effect of WORD is perfectly correlated with the fixed effect of vowel height. Since the subgroups have 5 or less than 5 stimuli, the decision also follows the "convention" in mixed-effect modeling that there should be at least 5 levels of a variable to be considered as a random effect (Gelman, 2007; Kéry and Royle, 2020; Harrison et al., 2018; Arnqvist, 2020; Harrison, 2015).

3.2 Stimuli

There were 24 English stimuli, and each participant repeated them in 15 randomized lists with filler words between the blocks. This means that there were 15 repetitions of each stimulus. When organized in different ways, the subgroups of the stimuli can be used to test the two sub-hypotheses of the dissertation. I will first present subgroups of stimuli used to test the claim that for CV syllables with the same V, a less sonorous C leads to a larger CV lag. Then, I will present subgroups of the stimuli used to test the claim that for CV syllables with the same C, a more sonorous V leads to a larger CV lag. A summary of the English stimuli can be found at the end of this section.

3.2.1 Same V different C

The groups of stimuli in this subsection were used to test the claim that for CV syllables with the same V, a less sonorous C leads to a larger CV lag. There were two types of consonants: bilabial (labial) or coronal. In each of the subsections, from the top to the bottom of each table of stimuli, CV gestural lag is expected to decrease since sonority difference decreases due to C sonority increases.

[2]The code is lmer(CV lag based on target onset~Sonority difference+(1|Participant)+(1|Stimuli)+(1|C duration)
[3]The code is lmer(CV lag based on target onset~Sonority difference+(1|Participant)+(1|C duration)

3.2.1.1 Same V different bilabial C

Presented in this subsection are stimuli with the same vowel and different bilabial consonants. Since the vowel uses the tongue gesture and bilabial (labial) consonants use the lip as the primary articulator, words with bilabial consonants have been the top choice for speech production studies. As in Table 3.1, 3.2, and 3.3, high, mid, and low vowels are considered and the coda environment is controlled in each subgroup. The first column in each table has the index for each stimulus. As the reader can see in subsection 3.2.2, the stimuli are organized in a different way to test another sub-claim of the study. The stimulus tables for the Mandarin EMA study also have a similar index column.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
1	peak	p	i	bilabial	high	1	15	14
2	beak	b	i	bilabial	high	4	15	11
3	meek	m	i	bilabial	high	7	15	8
4	week	w	i	bilabial	high	12	15	3

Table 3.1 Same high V different bilabial C.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
5	pain	p	e	bilabial	mid	1	16	15
6	bane	b	e	bilabial	mid	4	16	12
7	main	m	e	bilabial	mid	7	16	9
8	wane	w	e	bilabial	mid	12	16	4

Table 3.2 Same mid V different bilabial C.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
9	back	b	æ	bilabial	low	4	17	13
10	pack	p	æ	bilabial	low	1	17	16
11	Mac	m	æ	bilabial	low	7	17	10
12	whack	w	æ	bilabial	low	12	17	5

Table 3.3 Same low V different bilabial C.

3.2.1.2 Same V different coronal C

In the following Tables 3.4, 3.5, and 3.6, there were stimuli with the same vowel and varying coronal consonants in each subgroup of stimuli, and the vowels were high, mid, and low vowels

respectively.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
13	two	t	u	coronal	high	1	15	14
14	sue	s	u	coronal	high	3	15	12
15	do	d	u	coronal	high	4	15	11
16	new	n	u	coronal	high	7	15	8

Table 3.4 Same high V different coronal C.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
17	toe	t	o	coronal	mid	1	16	15
18	so	s	o	coronal	mid	3	16	13
19	doe	d	o	coronal	mid	4	16	12
20	know	n	o	coronal	mid	7	16	9

Table 3.5 Same mid V different coronal C.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
21	talk	t	ɑ	coronal	low	1	17	16
22	sock	s	ɑ	coronal	low	3	17	14
23	dock	d	ɑ	coronal	low	4	17	13
24	knock	n	ɑ	coronal	low	7	17	10

Table 3.6 Same low V different coronal C.

3.2.2 Same C different V

The following stimuli groups have the same C and different V. The stimuli in the subsection are a rearrangement of the stimuli above for the purpose of testing another sub-claim. To control for the coda environment, only high and low vowel words are considered for the bilabial group, and only high and mid-vowel words are considered for the coronal group. For the bilabial stimuli in Table 3.7, in each subgroup, the high vowel stimuli should have a smaller lag than low vowel stimuli — since a high vowel is less sonorous than a low vowel, a high vowel also has a smaller sonority difference than a low vowel. Similarly, for the coronal subgroup in Table 3.8, the high vowel stimuli should have a smaller lag than the mid-vowel stimuli.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
1	peak	p	i	bilabial	high	1	15	14
10	pack	p	æ	bilabial	low	1	17	16
2	beak	b	i	bilabial	high	4	15	11
9	back	b	æ	bilabial	low	4	17	13
3	meek	m	i	bilabial	high	7	15	8
11	Mac	m	æ	bilabial	low	7	17	10
4	week	w	i	bilabial	high	12	15	3
12	whack	w	æ	bilabial	low	12	17	5

Table 3.7 Same bilabial C different V.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
13	two	t	u	coronal	high	1	15	14
17	toe	t	o	coronal	mid	1	16	15
14	sue	s	u	coronal	high	3	15	12
18	so	s	o	coronal	mid	3	16	13
15	do	d	u	coronal	high	4	15	11
19	doe	d	o	coronal	mid	4	16	12
16	new	n	u	coronal	high	7	15	8
20	know	n	o	coronal	mid	7	16	9

Table 3.8 Same coronal C different V.

3.2.3 Summary of English experiment stimuli

A summary of all English stimuli can be found in Table 3.9. The sonority index for the consonant and vowel can be found in the *C sonority* and *V sonority* columns. The sonority difference of each stimulus can be found in the last column.

Index	Stimuli	C	V	C category	V category	C sonority	V sonority	Sonority diff
1	peak	p	i	bilabial	high	1	15	14
2	beak	b	i	bilabial	high	4	15	11
3	meek	m	i	bilabial	high	7	15	8
4	week	w	i	bilabial	high	12	15	3
5	pain	p	e	bilabial	mid	1	16	15
6	bane	b	e	bilabial	mid	4	16	12
7	main	m	e	bilabial	mid	7	16	9
8	wane	w	e	bilabial	mid	12	16	4
9	back	b	æ	bilabial	low	4	17	13
10	pack	p	æ	bilabial	low	1	17	16
11	Mac	m	æ	bilabial	low	7	17	10
12	whack	w	æ	bilabial	low	12	17	5
13	two	t	u	coronal	high	1	15	14
14	sue	s	u	coronal	high	3	15	12
15	do	d	u	coronal	high	4	15	11
16	new	n	u	coronal	high	7	15	8
17	toe	t	o	coronal	mid	1	16	15
18	so	s	o	coronal	mid	3	16	13
19	doe	d	o	coronal	mid	4	16	12
20	know	n	o	coronal	mid	7	16	9
21	talk	t	ɑ	coronal	low	1	17	16
22	sock	s	ɑ	coronal	low	3	17	14
23	dock	d	ɑ	coronal	low	4	17	13
24	knock	n	ɑ	coronal	low	7	17	10

Table 3.9 English stimuli summary.

Target Sounds	**Articulatory Sensor**	**Gesture**
Bilabial [p, b, m, w]	lower and upper lip	lip aperture
Alveolar [t, d, n, s]	tongue tip	tongue tip
Vowel	tongue dorsum	tongue dorsum

Table 3.10 Articulatory sensors and gestures for each type of sounds – English.

3.3 Results

3.3.1 Overall analysis

The analysis of all the data based on target onset can be found in Figure 3.3 and the mixed effect model can be found in Table 3.11. There was a significant positive correlation as predicted. As mentioned in the previous chapter, while analysis with *all* the data and the use of a sonority

difference score as a predictor is easier to statistically model and has higher statistical power, it does have some issues. First, the analysis assumes that the sonority scale is linear and not just relative, which is contrary to most phonologists' beliefs. Furthermore, it collapses across different articulators or gestures. For these reasons, I analyzed more nuanced sub-groups of the data. The results for the subsets can be found in the following sections.

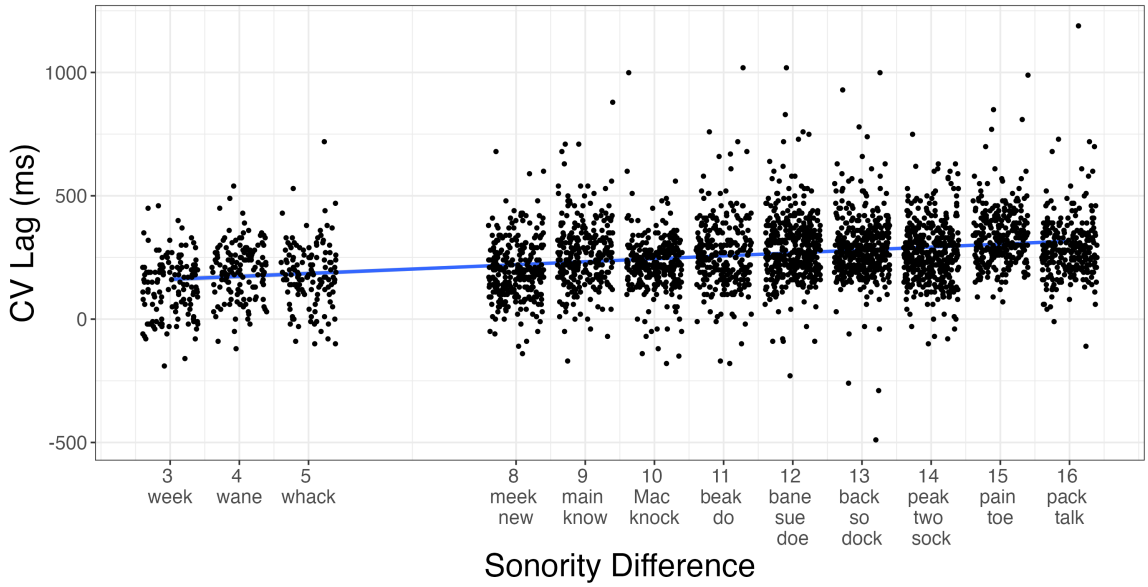

Figure 3.3 CV lag based on target onset for English participants.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	156.70	22.73	33.01	6.89	<0.00001
Sonority difference	11.24	1.64	21.92	6.86	<0.00001

Table 3.11 Mixed effects model results for English participants.

3.3.2 Results for claim 5a: the same vowel with different consonants

I first tested the claim 5a that for the CV syllables with the same vowel and different consonants, a less sonorant consonant leads to larger CV lag. Results for stimuli with the same vowel and varying bilabial consonant will be presented before results for stimuli with the same vowel and varying coronal consonant.

3.3.2.1 Same V and varying bilabial C

This subsection shows the result for subgroups of stimuli that have the same V and varying bilabial C. The results for bilabial consonants and high vowels are shown in Figure 3.4 and Table 3.12. There was a significant positive correlation between sonority difference and gestural lag for stimuli with different bilabial consonants and the same high vowel.

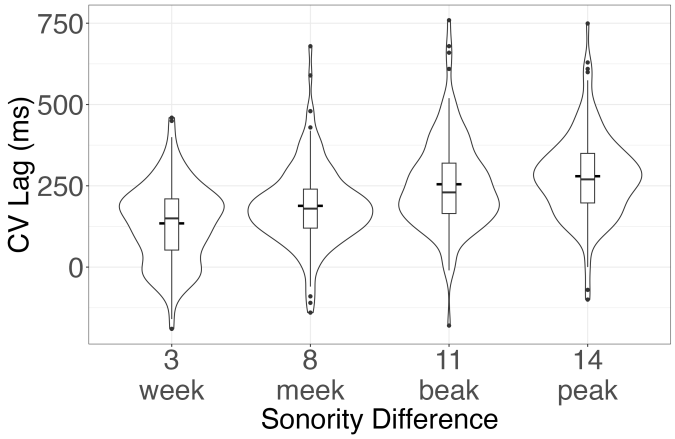

Figure 3.4 CV lag based on target onset for English participants: bilabial C and high V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	95.46	19.47	22.19	4.90	0.00
Sonority difference	14.00	1.20	439.70	11.71	< 0.00001

Table 3.12 Mixed effects model results for English participants: bilabial C and high V.

As for stimuli with different bilabial C and the same mid V, there was also a significant positive correlation between gestural lag and sonority difference, as in Figure 3.5 and Table 3.13.

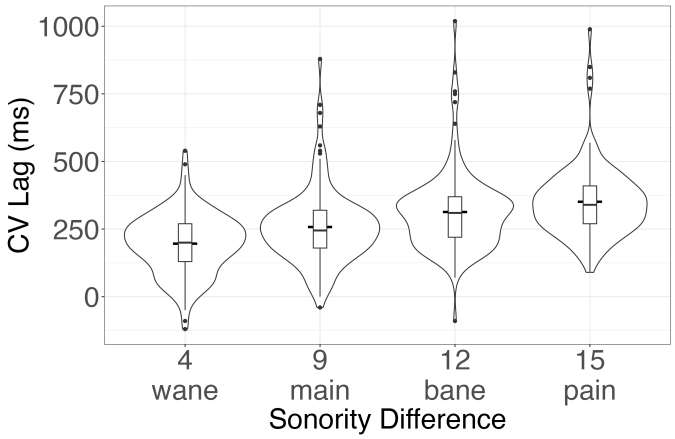

Figure 3.5 CV lag based on target onset for English participants: bilabial C and mid V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	145.14	20.87	28.11	6.96	<0.00001
Sonority difference	14.35	1.26	453.34	11.42	< 0.00001

Table 3.13 Mixed effects model results for English participants: bilabial C and mid V.

When there were the same low vowel and different bilabial consonants, there also was a significant positive correlation as in Figure 3.6 and Table 3.14. The effect size was slightly lower for the stimuli with low vowel, as compared to the stimuli with high and mid vowels.

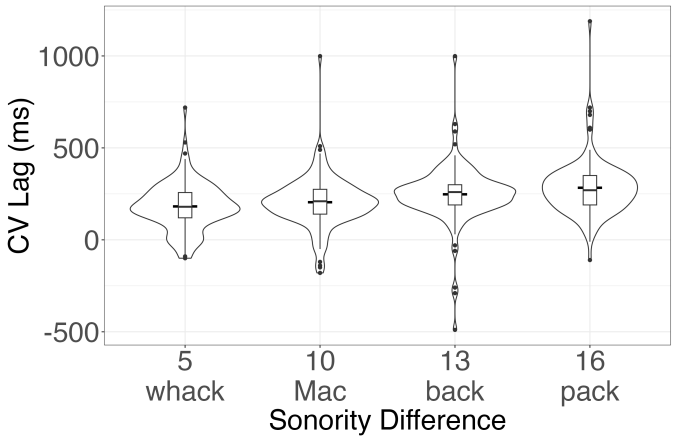

Figure 3.6 CV lag based on target onset for English participants: bilabial C and low V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	124.19	23.04	42.61	5.39	0.000003
Sonority difference	10.47	1.47	363.78	7.11	< 0.00001

Table 3.14 Mixed effects model results for English participants: bilabial C and low V.

Previous studies have found that consonant manner and place could lead to gestural coordination variation (Bombien et al., 2013; Wright, 1996; Pouplier et al., 2022). Comparing the CV lag for bilabial stimuli differing in the voicing of C can control for the potential confounding factors such as manner, place, and jaw movement, since the pair has the same manner and place of articulation, as well as jaw movement. The results for the *peak, beak* comparison can be found in Figure 3.7 and Table 3.15. There was a positive correlation between CV lag and sonority difference. As in Appendix H, I also added C DISPLACEMENT as a fixed effect. There was no effect of C displacement, confirming that the jaw movement was controlled.

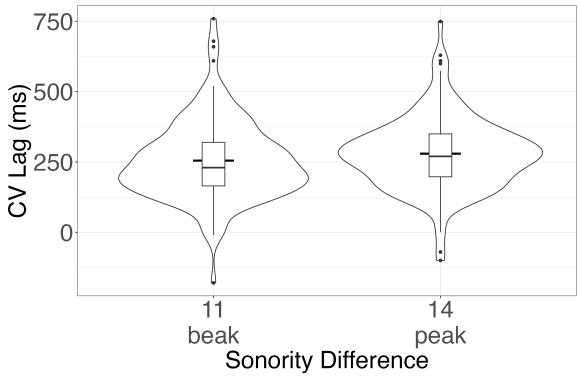

Figure 3.7 CV lag based on target onset for English participants: *peak, beak*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	162.74	64.22	203.01	2.53	0.01
Sonority difference	8.92	4.87	200.26	1.83	0.07

Table 3.15 Mixed effects model results for English participants: *peak, beak*.

The results for *pain, bane* can be found in Table 3.16 and Figure 3.8. There was a significant positive correlation between CV lag and sonority difference.

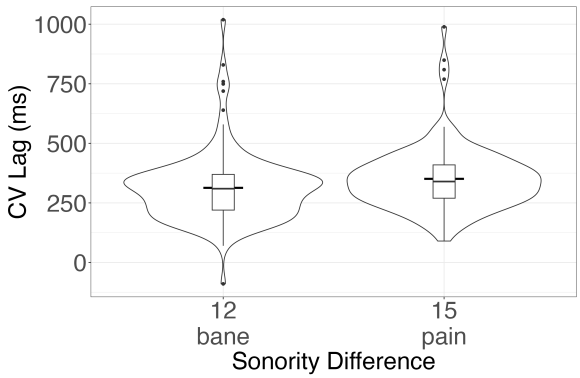

Figure 3.8 CV lag based on target onset for English participants: *pain, bane*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	171.58	69.45	207.31	2.47	0.01
Sonority difference	12.81	5.01	186.32	2.56	0.01

Table 3.16 Mixed effects model results for English participants: *pain, bane*.

The results for the low vowel group *pack, back* can be found in Figure 3.9 and Table 3.17. There was a significant positive correlation between sonority difference and CV lag. Adding C displacement as a fixed effect showed that there was no clear relationship between CV lag and C displacement. This showed that when jaw movement or manner of articulation was controlled, there was still a significant positive correlation between CV lag and sonority difference observed.

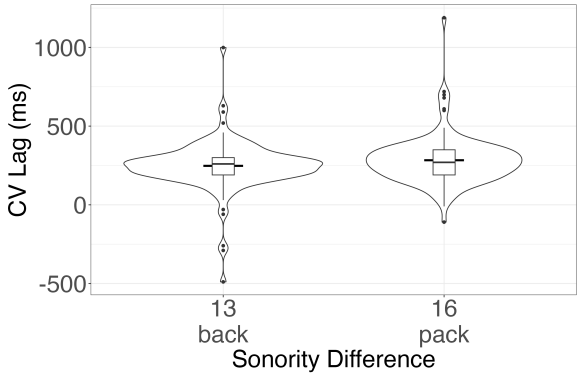

Figure 3.9 CV lag based on target onset for English participants: *pack, back*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	92.79	82.98	166.39	1.12	0.27
Sonority difference	12.66	5.52	149.64	2.29	0.02

Table 3.17 Mixed effects model results for English participants: *pack, back*.

Below I compared the nasal and stop with the same vowel and coda environment. A significant difference found in the comparison can strengthen the main claim tested because [m] and [b] differ in nasality, and they have controlled jaw movement, frontness, and voicing. As we can see in this section, a significant positive correlation was found in bilabial nasal and stop with different vowel heights. As in Figure 3.10 and Table 3.18, for the same rime environment with high vowels, the bilabial stop had a significantly larger lag than bilabial nasal. The major difference between the [m] and [b] articulations is the lowering or raising of velum, and there is no obvious articulatory reason that velum movement by itself should cause gestural lag variation between the lips and the tongue. The finding suggests that jaw movement may not be a valid alternative account of the observed gestural lag variation, since in this pair the two segments [m] and [b] differ in nasality and involve similar degrees of jaw movement. Therefore, by comparing CV lag for the pair *meek, beak*, our claim is more strongly supported.

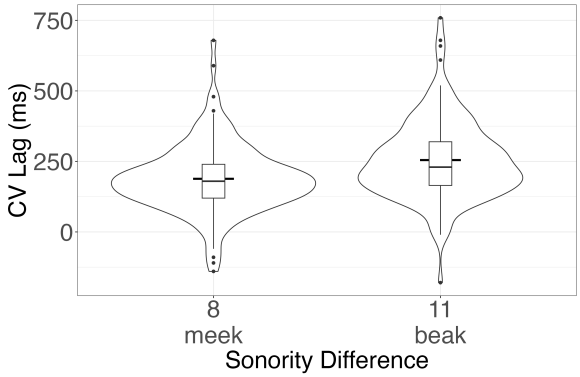

Figure 3.10 CV lag based on target onset for English participants: *meek, beak*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	19.11	49.92	231.70	0.38	0.70
Sonority difference	21.84	4.98	233.55	4.38	0.00002

Table 3.18 Mixed effects model results for English participants: *meek, beak*.

As in Figure 3.11 and Table 3.19, for the stimulus pairs with mid vowels, the bilabial stop had a larger lag than the bilabial nasal. This brought out the main claim tested in the dissertation by confirming that the results are likely to be confounded by voicing and jaw movement — since the pair had controlled jaw movement and voicing.

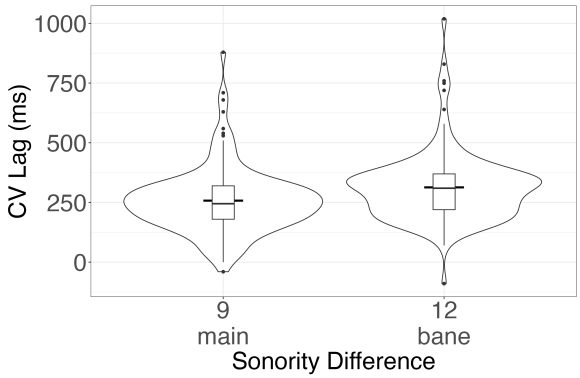

Figure 3.11 CV lag based on target onset for English participants: *main, bane*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	62.26	57.34	220.29	1.09	0.28
Sonority difference	21.59	5.21	204.31	4.14	0.0001

Table 3.19 Mixed effects model results for English participants: *main, bane*.

The results for low vowels with bilabial onsets can be found in Figure 3.12 and Table 3.20. A significant correlation between sonority difference and gestural lag can be found in the subgroup as well, showing the result was not confounded by jaw movement and voicing of consonants.

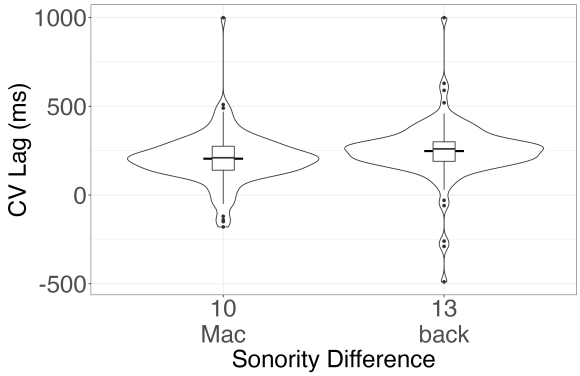

Figure 3.12 CV lag based on target onset for English participants: *Mac, back*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	106.32	67.63	216.79	1.57	0.12
Sonority difference	11.20	5.65	213.27	1.98	0.05

Table 3.20 Mixed effects model results for English participants: *Mac, back*.

3.3.2.2 Same V and varying coronal C

In this subsection, stimulus subgroups with varying coronal C and the same V were analyzed. Figure 3.13 and Table 3.21 showed the results for stimuli with varying coronal consonants and high vowel, that there was a significant positive correlation between sonority difference and CV gestural lag.

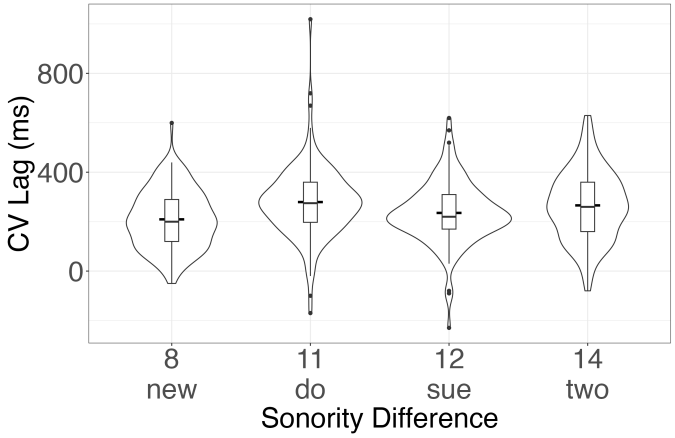

Figure 3.13 CV lag based on target onset for English participants: coronal C and high V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	155.68	32.88	144.16	4.74	0.00001
Sonority difference	8.02	2.57	506.29	3.12	0.002

Table 3.21 Mixed effects model results for English participants: coronal C and high V.

For the varying coronal consonants and mid vowel group, the expected positive correlation was found as in the plot in Figure 3.14 and Table 3.22.

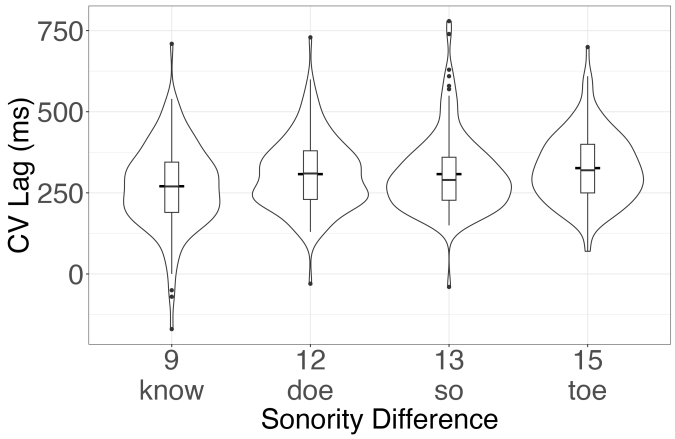

Figure 3.14 CV lag based on target onset for English participants: coronal C and mid V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	191.71	29.69	142.13	6.46	0.00
Sonority difference	9.40	2.11	446.85	4.46	< 0.0001

Table 3.22 Mixed effects model results for English participants: coronal C and mid V.

The varying coronal consonants and low vowel group also exhibited significant positive correlation as in Figure 3.15 and Table 3.23. Note that the effect size for the coronal C subgroups was generally smaller than that of the bilabial C subgroups. This could be due to the fact that coronal consonants and vowels all used tongue as the main articulator, and therefore separating the gestures became less straightforward.

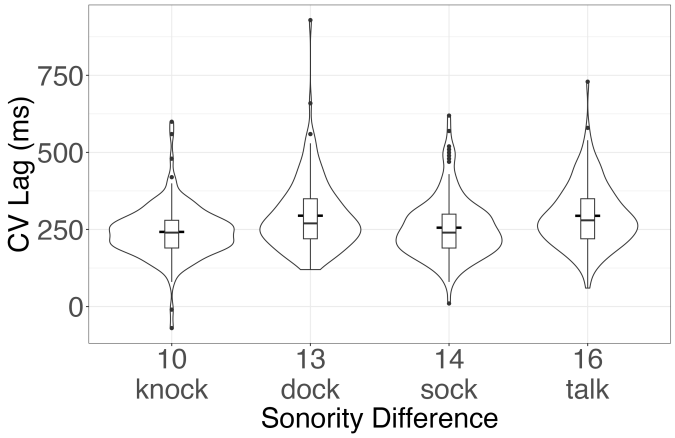

Figure 3.15 CV lag based on target onset for English participants: coronal C and low V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	174.68	27.62	146.50	6.33	0.00
Sonority difference	7.43	1.82	524.96	4.08	< 0.0001

Table 3.23 Mixed effects model results for English participants: coronal C and low V.

I also compared the stimuli with voiceless and voiced coronal C, to control for jaw movement and manner of articulation. C duration was not used as a random intercept since voiced C and voiceless C differ in duration (Denes, 1955). In other words, since the durational difference is correlated to C voicing and I am testing the effect of voicing, adding C duration as a random intercept would counteract the pattern related to the target effect. There was no significant positive correlation found for the *two, do* pair, as in Figure 3.16 and Table 3.24.

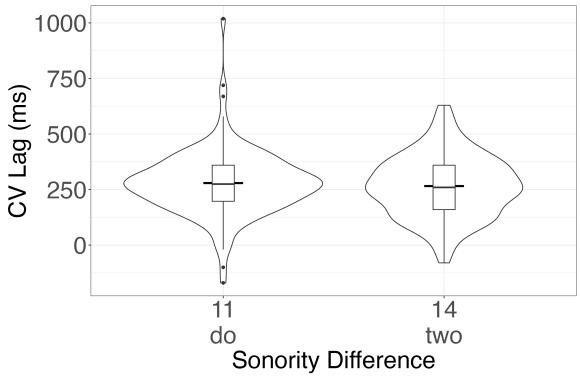

Figure 3.16 CV lag based on target onset for English participants: *two, do*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	341.11	72.43	249.13	4.71	<0.00001
Sonority difference	-5.64	5.64	246.90	-1.00	0.32

Table 3.24 Mixed effects model results for English participants: *two, do*.

For the *toe, doe* comparison, there was an insignificant positive correlation between sonority difference and CV lag as in Table 3.25 and Figure 3.17.

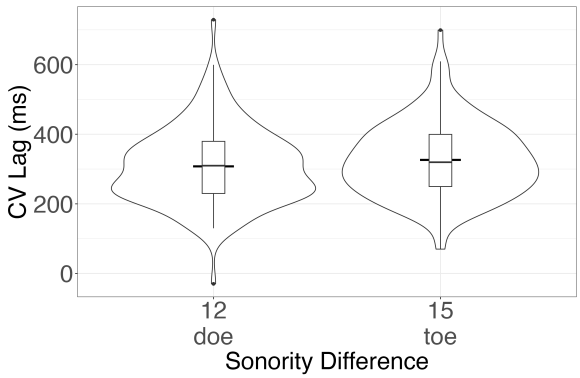

Figure 3.17 CV lag based on target onset for English participants: *toe, doe*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	228.02	55.96	268.06	4.08	0.0001
Sonority difference	6.47	3.99	272.97	1.62	0.11

Table 3.25 Mixed effects model results for English participants: *toe, doe*.

Also, there was no clear relationship observed between sonority and lag found for *talk, dock*, as in Figure 3.18 and Table 3.26. In general, there was no clear relationship observed for coronal C stimulus pairs that differ in voicing.

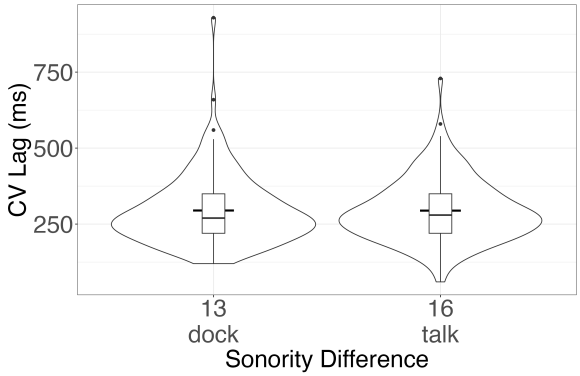

Figure 3.18 CV lag based on target onset for English participants: *talk, dock*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	296.47	57.11	233.89	5.19	<0.00001
Sonority difference	-0.23	3.72	275.03	-0.06	0.95

Table 3.26 Mixed effects model results for English participants: *talk, dock*.

Stimulus pairs of coronal nasal and stop with the same vowel were analyzed. As mentioned before, this is to control the jaw movement and voicing. The results show that when the nasal and stop pair is with a high, mid, or low vowel, significant correlations were found. The high vowel pair comparison can be found in Figure 3.19 and Table 3.27, and the high vowel with coronal stop had a significantly larger lag than that with coronal nasal.

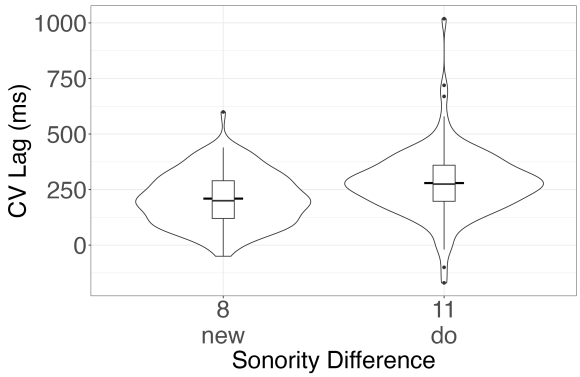

Figure 3.19 CV lag based on target onset for English participants: *do, new*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	18.06	51.89	248.12	0.35	0.73
Sonority difference	23.74	5.26	240.84	4.52	0.00001

Table 3.27 Mixed effects model results for English participants: *do, new*.

Figure 3.19 and Table 3.28 show that the coronal stop with mid vowel had a larger lag than the coronal nasal with the same mid vowel.

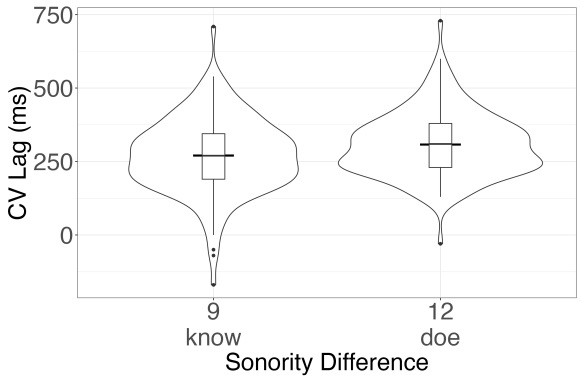

Figure 3.20 CV lag based on target onset for English participants: *doe, know*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	165.83	50.11	230.31	3.31	0.001
Sonority difference	11.83	4.49	244.86	2.64	0.01

Table 3.28 Mixed effects model results for English participants: *doe, know*.

The syllable that starts with a coronal stop and ends with a low vowel had a significantly larger lag than that with a coronal nasal as in Figure 3.21 and Table 3.29.

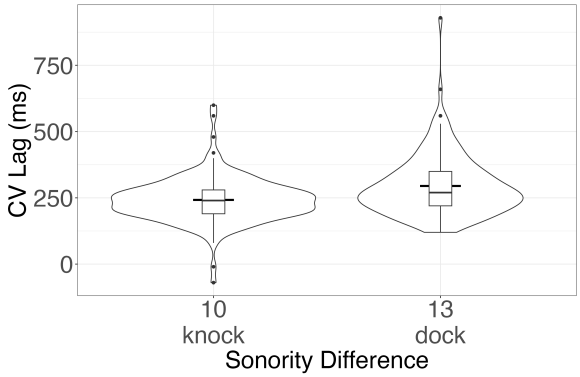

Figure 3.21 CV lag based on target onset for English participants: *dock, knock*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	68.00	44.00	226.27	1.55	0.12
Sonority difference	17.58	3.59	269.80	4.90	0.000002

Table 3.29 Mixed effects model results for English participants: *dock, knock*.

3.3.3 Results for claim 5b: the same C and different V

3.3.3.1 Same bilabial C and different V

For the syllables with the same bilabial C [p] and different vowel, there was a positive correlation between gestural lag and sonority difference as in Figure 3.22. However, the positive correlation was not statistically significant as in Table 3.30.

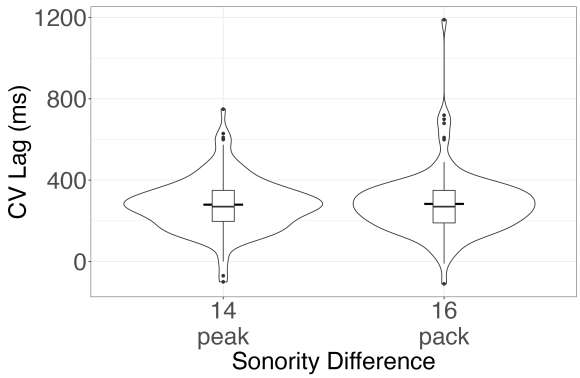

Figure 3.22 CV lag based on target onset for English participants: *peak, pack*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	239.06	119.20	154.58	2.01	0.05
Sonority difference	3.20	7.86	143.32	0.41	0.69

Table 3.30 Mixed effects model results for English participants: *peak, pack*.

The same bilabial consonant [b] with a low vowel had a shorter lag than those with a high vowel, but this difference was not statistically significant, as in Table 3.31 and Figure 3.23.

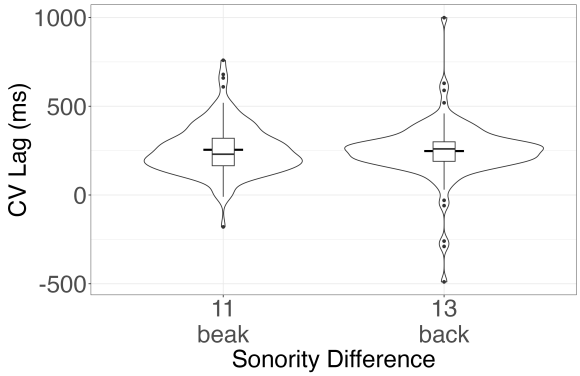

Figure 3.23 CV lag based on target onset for English participants: b and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	267.69	92.36	171.80	2.90	0.00
Sonority difference	-0.22	7.46	156.97	-0.03	0.98

Table 3.31 Mixed effects model results for English participants: b and different V.

Words that have onset [m] and a rime of a low vowel had a larger lag than that with a high vowel. However, the difference was not significant.

Figure 3.24 CV lag based on target onset for English participants: *meek, Mac*.

Gestural Onset	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	134.53	70.24	260.07	1.92	0.06
Sonority difference	7.61	7.65	246.81	1.00	0.32

Table 3.32 Mixed effects model results for English participants: m and different V.

Words with bilabial consonant [w] and a low vowel had a significantly larger lag than those with a high vowel, as in Figure 3.25 and Table 3.33. The reason why [w] and different vowels exhibited a difference but other bilabial consonants did not show the expected pattern is unclear. It is possible that the vowel and the coda consonant in each word all involved tongue movement, so the tongue movement from the following coda consonant may affect the preceding vowel. Therefore, there were no significant patterns in most bilabial consonants. Syllables with bilabial consonant [w] and vowels did show that sonority difference is positively correlated to gestural lag. It may be that the observation was confounded by the fact that [w] also involved tongue movement. The tongue movement in [w] also affected the tongue movement of vowels, which surfaced as a significant observation.

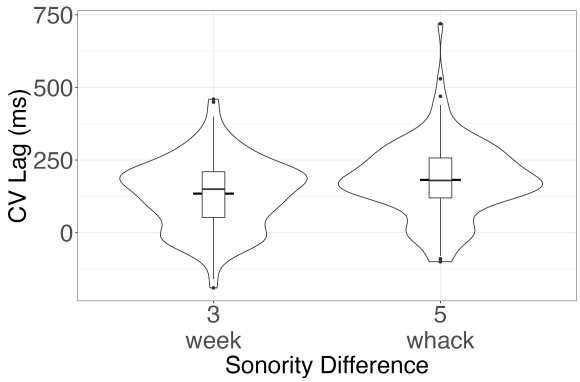

Figure 3.25 CV lag based on target onset for English participants: w and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	80.66	33.96	57.56	2.38	0.02
Sonority difference	18.23	6.80	250.31	2.68	0.01

Table 3.33 Mixed effects model results for English participants: w and different V.

3.3.3.2 Same coronal C and different V

I look at results for the same coronal C and different V. For stimuli with [t] and different V, there was a positive correlation between CV lag and sonority difference as in Figure 3.26 and Table 3.34. The effect sizes in this subsection are bigger than those in other analyses. The bigger effect size may not indicate that coronal C stimuli had a more significant correlation. Rather, it may come from some measuring confounds since C and V used the same articulator tongue.

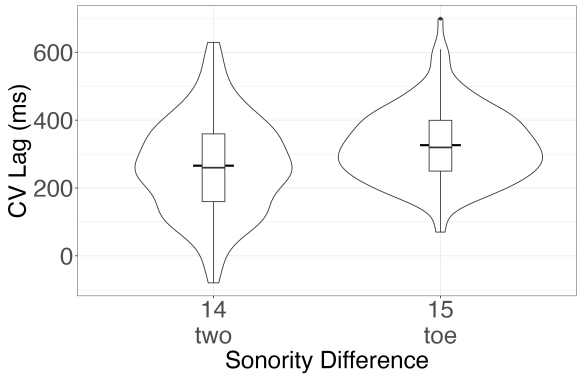

Figure 3.26 CV lag based on target onset for English participants: t and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-604.05	202.26	259.52	-2.99	0.003
Sonority difference	61.90	13.86	256.03	4.47	0.00001

Table 3.34 Mixed effects model results for English participants: t and different V.

For stimuli with [d] and different vowels, there was a positive correlation between CV lag and sonority difference, as in Figure 3.27 and Table 3.35.

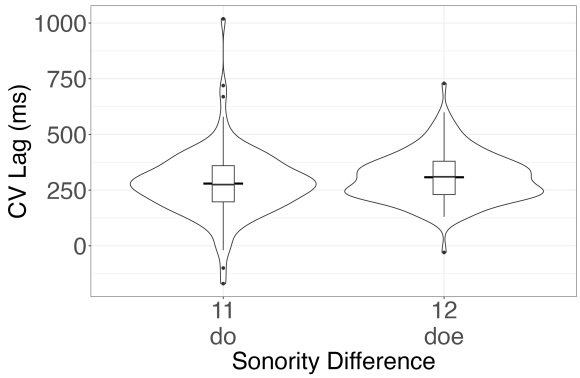

Figure 3.27 CV lag based on target onset for English participants: d and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-23.95	175.16	264.96	-0.14	0.89
Sonority difference	27.51	15.17	262.21	1.81	0.07

Table 3.35 Mixed effects model results for English participants: d and different V.

For stimuli with [s] and different vowels, there was a significant positive correlation between CV lag and sonority difference, as in Figure 3.28 and Table 3.36.

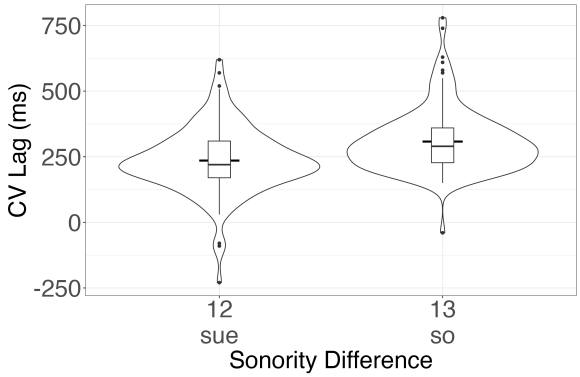

Figure 3.28 CV lag based on target onset for English participants: s and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-678.84	177.35	257.04	-3.83	0.0002
Sonority difference	76.03	14.13	253.25	5.38	0.0000002

Table 3.36 Mixed effects model results for English participants: s and different V.

There was also a significant positive correlation found for stimuli with [n] and different vowels, as in Figure 3.29 and Table 3.37.

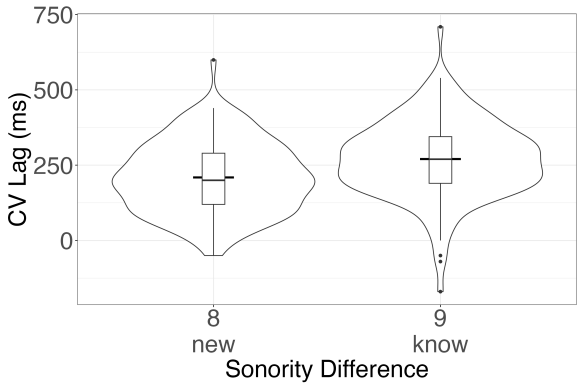

Figure 3.29 CV lag based on target onset for English participants: n and different V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-278.23	120.45	251.69	-2.31	0.02
Sonority difference	60.78	14.00	242.75	4.34	0.00002

Table 3.37 Mixed effects model results for English participants: n and different V.

3.3.4 The vowel displacement and C displacement analysis

As justified by experiment 1, it is logically possible that our results are somehow artifactual and based on the relation observed by Shaw and Chen (2019) — that CV lag based on gestural onsets is negatively correlated with the displacement of the vowel from gesture onset to the achievement of the target. Furthermore, as mentioned before, jaw movement correlates with variation in gestural coordination (Gracco, 1994; Gracco and Lofqvist, 1994; Mooshammer et al., 2003; Redford, 1999; MacNeilage and Davis, 2000).

To check for this possibility that the significant correlation is not due to consonant or vowel displacement, I ran another analysis wherein I added random intercepts which are vowel

displacement and consonant displacement. The vowel displacement is the horizontal distance from vowel gesture onset to target achievement. For measuring consonant displacements, I subtracted the gesture onset value from the target onset value. Bilabial consonant displacement was measured by lip aperture displacement difference between gesture onset and target onset. Coronal consonant displacement was measured by the y-axis (vertical) distance between gesture onset and target onset.

This additional analysis involved all the stimuli. I chose the whole dataset as it was the largest stimulus set and therefore the analysis would suffer the least in terms of statistical power from the addition of a post-hoc variable. Note that both vowel displacement and consonant displacement could be an estimate of jaw movement. Therefore, the post-hoc analysis also serves as another exploration of the potential effect of jaw movement on gestural coordination. Besides the previous random intercepts of PARTICIPANT, WORD, CONSONANT DURATION, there are also two more random intercepts of VOWEL DISPLACEMENT and CONSONANT DISPLACEMENT.[4] Since the C displacement is measured differently for stimuli with bilabial C and stimuli with coronal C, there were separate analyses conducted, one for the bilabial C stimuli as in Table 3.38, one for the coronal C stimuli as in Table 3.39. In both cases, there were significant positive correlations between sonority difference and CV gestural lag. I also tested the C displacement and V displacement as fixed effects. As in Appendix G, vowel displacement in fact contributes to the CV lag variation in both positive and negative directions, and the effect of sonority difference was still there. Specifically, for bilabial C stimuli, the effect size for sonority difference was 14.07 with V displacement as a fixed effect. For coronal C stimuli, the effect size for sonority difference was 6.74 with V displacement as a fixed effect. It is likely that C displacement does not have a significant effect on CV lag in the dataset tested here.

[4]The R formula is shown here: CV lag based on target onset~Sonority difference+(1|Participant)+(1|Word)+(1|C duration)+(1|V displacement)+(1|C displacement).

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	120.97	29.37	13.98	4.12	0.001
Sonority difference	12.74	2.46	10.11	5.17	0.0004

Table 3.38 Mixed effects model results for all bilabial C stimuli. Random intercepts: participant, word, consonant duration, vowel displacement, consonant displacement.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	182.46	44.14	11.80	4.13	0.00
Sonority difference	8.31	3.39	10.00	2.45	0.03

Table 3.39 Mixed effects model results for all coronal C stimuli. Random intercepts: participant, word, consonant duration, vowel displacement, consonant displacement.

3.3.5 Summary

The summary of the results can be found in the following Table 3.40 and Table 3.41. In Table 3.40, all the subgroups showed that there is a significant positive correlation between CV lag and sonority difference. The bilabial group had a slightly larger effect size than the coronal group.

	Dataset (English EMA data)	Estimate (sonority difference)
	All English EMA data	11.24 ***
Bilabial C	All bilabial C data, C and V displacement	12.74 ***
	High V (week, meek, beak, peak)	14.00 ***
	Mid V (wane, main, bane, pain)	14.35 ***
	Low V (whack, Mac, back, pack)	10.47 ***
Coronal C	All coronal C data, C and V displacement	8.31 *
	High V (new, do, sue, two)	8.02 **
	Mid V (know, doe, so, toe)	9.40 ***
	Low V (knock, dock, sock, talk)	7.43 ***

Table 3.40 Summary of English EMA results. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$.

In Table 3.41, the pairs differ in nasality for both bilabial C and coronal C stimuli exhibited significant positive correlations between CV lag and sonority difference. The coronal C stimuli with differences in C voicing did not show the expected pattern. Also, the bilabial C stimuli with differences in vowel height did not consistently show the expected pattern. The non-significant result is potentially due to the same V measure for different vowel heights, as well as the imprecise C voicing coding.

	Pairwise comparison	Stimulus pair	Estimate (sonority difference)
Bilabial C	Nasality differ	Mac, back	11.20 *
		meek, beak	21.84 ***
		main, bane	21.59 ***
	Voicing differ	beak, peak	8.92
		bane, pain	12.81 **
		back, pack	12.66 *
	Vowel height differ	peak, pack	3.2
		beak, back	-0.22
		meek, Mac	7.61
		week, whack	18.23 **
Coronal C	Nasality differ	new, do	23.74 ***
		know, doe	11.83 **
		knock, dock	17.58 ***
	Voicing differ	two, do	-5.64
		toe, doe	6.47
		talk, dock	-0.23
	Vowel height differ	two, toe	61.90 ***
		do, doe	27.51
		sue, so	76.03 ***
		new, know	60.78 ***

Table 3.41 Summary of English EMA pairwise comparison. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$.

3.4 Conclusion

Overall the English experiment showed the expected pattern, on both bilabial and coronal consonants combined with different vowel heights. I argued that the observed correlation cannot simply be attributed to vowel quality or the effect of vowel displacement. Groups of stimuli with the same vowel quality showed a significant positive correlation. Therefore, vowel quality is unlikely to be the driver of observed CV lag variation. Additionally, in a post-hoc analysis, I showed there were still positive correlation between sonority difference and CV lag observed for both stimuli with bilabial C and stimuli with coronal C.

My results also suggest that jaw movement is unlikely to be an important factor driving the observed correlation between sonority and gestural lag variation. First, using consonant displacement as an approximation of jaw movement, I showed that the observed positive correlation is not confounded by jaw movement. Second, for the pairwise comparisons controlled

for jaw movement, I used post-hoc analyses to confirm that consonant displacement was indeed controlled. Relatedly, the comparison using vowel displacement (which is also an approximation of jaw movement) makes a similar point. The positive correlation between sonority difference and CV lag is still exhibited when including vowel displacement. Based on the above analysis, jaw movement is unlikely to be the factor leading to the observed relationship between CV lag and sonority difference.

However, for some pairs the observation was not significant, though still in the expected direction. Here are some potential reasons why the expected pattern is not shown – that the lag could be affected by the coda environment and that the vowel sensor is not precise. First, the coda consonant [k] may affect the articulation of the preceding vowel, and this may be the reason why the three pairs – *peak, pack*; *beak, back*; *meek, Mac* – did not show the expected significant correlation. The question remain why *week, whack* exhibited the expected pattern. It may be that the significant effect is confounded by the velar or tongue movement in [w]. Similarly, in the pair *main, bane*, the vowel may be nasalized due to the following coda nasal. Since [n] uses the tongue and nasality is more sonorous, the CV timing may be affected. Furthermore, the consistent tongue sensor is used to measure vowels of different heights, so maybe for this reason the pair *do, doe* did not exhibit expected patterns. It might be that some characteristic of the [d] articulation made the pair more sensitive to the consistent vowel sensor measure.

CHAPTER 4

EXPERIMENT 3: MANDARIN EMA STUDY

Both the English corpus study and the English EMA study showed a significant positive correlation between sonority difference and CV lag. The relationship between sonority sequence and CV lag has been found for stimuli with coronal C or bilabial C, combined with high, mid, or low vowels in English.

As mentioned earlier, Crouch (2022) and Crouch et al. (2023) observed that sonority difference positively correlates to CC onset lag in Georgian, and they argued it is due to language-specific mechanism in Georgian. However, as shown in the previous two chapters, a positive correlation has also been found in English. Since the main claim of the study is intended to be language-independent, a further question would be whether CV lag in languages other than English still exhibits this correlation between sonority and CV lag. To provide a cross-linguistic perspective, I conducted an EMA study on Mandarin.

4.1 Methods

The Mandarin experiments shared the same experimental procedure despite the differences discussed below. The first difference in terms of method is the language used when communicating with participants. For the Mandarin experiment, the communications with participants include recruitment messages (see Appendix C), pre-screening surveys (see Appendix D), and communications during experimental sessions. All the spoken communications were in Mandarin Chinese and all the written communications were in simplified Chinese. The second difference is that Mandarin participants were recruited through WeChat, the primary social media platform among Chinese people. Thirdly, the Mandarin experiment had a carrier phrase. Since the English experiments without carrier phrases sometimes had unreasonably large gestures extended into the pause, in the later conducted Mandarin experiments, the carrier phrase *zhe4 ge4* __ *mo* [ʈʂə kə __ mə] 'this __' was used. The carrier phrase was chosen because before the target word, there is a schwa, which is the neutral position. Another reason for choosing the carrier phrase is that after the vowel of the target word CV, there is a bilabial consonant, which uses a

different articulator (i.e., lips) than the vowel (i.e., tongue). The carrier phrase did seem to serve this purpose because there were fewer gestures that were annotated with uncertainty — English 8% uncertain labels, Mandarin 4% uncertain labels. Lastly, the uncertainty labels in Mandarin were slightly different than those in the English experiment.[1] Even though most labels were the same, in Mandarin there were new labels such as "NaLamispron", which means that either [l] is pronounced as [n], or [n] is pronounced as [l]. Some participants also told the experimenters during the sessions that they could not distinguish between [n] and [l]. This is evidence of the merger-in-progress of word-initial lateral [l] and [n], which occurred in many Chinese languages such as Nanjing Mandarin, Chengdu, Southwestern Mandarin, languages in Southern China (Shi, 2015; Johnson and Song, 2016; Zhang and Levis, 2021; Cheng et al., 2023). This merger of [n] and [l] has been observed in both production and perception of Chinese languages (Cheng et al., 2023). Of the 10 participants analyzed, 2 from Jiangsu, which belongs to Southern China, had the [n-l] merger.

Altogether data from 20 Mandarin participants were collected, and those from 10 were annotated and analyzed in the current study. The data were annotated in the reverse order of data collection, which means that the last 10 participants' data were analyzed. The data of the first 10 participants is not considered in the current dissertation due to time constraints. Altogether there were 4004 annotated Mandarin syllables, and 3849 of them (96.1%) were not marked with any uncertainty labels. The results of these unambiguous annotations can be found in the Results section. Among the 10 participants of the EMA Mandarin experiments, 9 participants were female and 1 was male. The 10 participants aged from 23 to 56 years old, with an average age of 33.8 years old.

4.2 Stimuli

There were 27 Mandarin stimuli, and each participant repeated them in 15 randomized lists with filler words between the blocks. This means that there were 15 repetitions of each stimulus. Just like in the English experiment, when organized in different ways, the subgroups of the stimuli can be used to test the two sub-hypotheses of the dissertation. I will first present subgroups of stimuli

[1] See the Mandarin annotation labels in Appendix F.

used to test the claim that for CV syllables with the same V, a less sonorous C leads to a larger CV lag. Then, I will present subgroups of the stimuli used to test the claim that for CV syllables with the same C, a more sonorous V leads to a larger CV lag. Similarly, a summary of the Mandarin stimuli can be found at the end of this section. Similar to the English EMA experiment, the first column in each stimulus table below has the index for each stimulus.[2]

4.2.1 Same V different C

The stimuli in this subsection were used to test the claim that for a CV syllable, the larger the sonority difference, the larger the CV lag. In each of the subsections, from the top to the bottom of each table of stimuli, CV gestural lag decreases since sonority difference decreases due to C sonority increases.

4.2.1.1 Same V different bilabial C

There are two sets of stimuli for bilabial consonants in order to involve more variation of bilabial consonants. For the low vowel stimuli, for instance, there were low vowel stimuli with nasalized vowels and non-nasalized vowels.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
1	僻	pi	4	p	i	bilabial	high	distant	1	15	14
2	臂	bi	4	b	i	bilabial	high	arm	4	15	11
3	秘	mi	4	m	i	bilabial	high	secret	7	15	8

Table 4.1 Same high V different bilabial C. *C cat* means C category, and *V cat* means V category. *T* stands for tone. This is also true for other Mandarin stimuli tables.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
4	帕	pa	4	p	a	bilabial	low	handkerchief	1	17	16
5	坝	ba	4	b	a	bilabial	low	dam	4	17	13
6	骂	ma	4	m	a	bilabial	low	scold	7	17	10
7	袜	wa	4	w	a	bilabial	low	sock	12	17	5

Table 4.2 Same low V different bilabial C.

[2]Similar to the English experiment, I annotate the obstruents with indexes assuming true voicing distinction. A more careful study in the future may consider the realization of voicing and code accordingly.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
8	配	pei	4	p	ə	bilabial	mid	match	1	14	13
9	贝	bei	4	b	ə	bilabial	mid	shell	4	14	10
10	妹	mei	4	m	ə	bilabial	mid	sister	7	14	7
11	味	wei	4	w	ə	bilabial	mid	flavor	12	14	2

Table 4.3 Same mid-V different bilabial C.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
13	盼	pan	4	p	æ	bilabial	low	hope	1	17	16
14	半	ban	4	b	æ	bilabial	low	half	4	17	13
15	曼	man	4	m	æ	bilabial	low	grace	7	17	10
16	万	wan	4	w	æ	bilabial	low	ten thousand	12	17	5

Table 4.4 Same low nasalized V different bilabial C.

4.2.1.2 Same V different coronal C

This subsection has stimuli of the same V and different coronal C.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
18	兔	tu	4	t	u	coronal	high	rabbit	1	15	14
19	素	su	4	s	u	coronal	high	plain	3	15	12
20	度	du	4	d	u	coronal	high	degree	4	15	11
21	怒	nu	4	n	u	coronal	high	anger	7	15	8
22	路	lu	4	l	u	coronal	high	road	9	15	6

Table 4.5 Same high V different coronal C.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
23	踏	ta	4	t	a	coronal	low	step	1	17	16
24	飒	sa	4	s	a	coronal	low	cool	3	17	14
25	大	da	4	d	a	coronal	low	big	4	17	13
26	那	na	4	n	a	coronal	low	that	7	17	10
27	腊	la	4	l	a	coronal	low	wax	9	17	8

Table 4.6 Same low V different coronal C.

4.2.2 Same C different V

For this subsection, the stimuli are organized in another way to test another sub-claim. In each of the subsections, the high vowel group should have a smaller lag than the mid vowel group, and the mid vowel group should have a smaller lag than the low vowel group - since high vowel is less

sonorous than mid vowel than low vowel, high vowel also has smaller sonority difference than mid vowel than low vowel. For a same labial C, the syllable with lower vowel is predicted to have larger CV gestural lag since its sonority difference is larger. All the syllable pairs in question share the same coda environment.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
1	僻	pi	4	p	i	bilabial	high	distant	1	15	14
4	帕	pa	4	p	a	bilabial	low	handkerchief	1	17	16
2	臂	bi	4	b	i	bilabial	high	arm	4	15	11
5	坝	ba	4	b	a	bilabial	low	dam	4	17	13
3	秘	mi	4	m	i	bilabial	high	secret	7	15	8
6	骂	ma	4	m	a	bilabial	low	scold	7	17	10

Table 4.7 Same labial C different V.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
18	兔	tu	4	t	u	coronal	high	rabbit	1	15	14
23	踏	ta	4	t	a	coronal	low	step	1	17	16
19	素	su	4	s	u	coronal	high	plain	3	15	12
24	飒	sa	4	s	a	coronal	low	cool	3	17	14
20	度	du	4	d	u	coronal	high	degree	4	15	11
25	大	da	4	d	a	coronal	low	big	4	17	13
21	怒	nu	4	n	u	coronal	high	anger	7	15	8
26	那	na	4	n	a	coronal	low	that	7	17	10
22	路	lu	4	l	u	coronal	high	road	9	15	6
27	腊	la	4	l	a	coronal	low	wax	9	17	8

Table 4.8 Same coronal C different V.

4.2.3 Summary of Mandarin experiment stimuli

A summary of all Mandarin stimuli can be found in Table 4.9. The consonant and vowel categories can be found in *C cat* and *V cat* columns. The sonority index for the consonant and vowel can be found in the *C son* and *V son* columns. The sonority difference of each stimulus can be found in the last column *S diff*.

Index	Word	Pinyin	T	C	V	C cat	V cat	Gloss	C son	V son	S dif
1	僻	pi	4	p	i	bilabial	high	distant	1	15	14
2	臂	bi	4	b	i	bilabial	high	arm	4	15	11
3	秘	mi	4	m	i	bilabial	high	secret	7	15	8
4	帕	pa	4	p	a	bilabial	low	napkin	1	17	16
5	坝	ba	4	b	a	bilabial	low	dam	4	17	13
6	骂	ma	4	m	a	bilabial	low	scold	7	17	10
7	袜	wa	4	w	a	bilabial	low	sock	12	17	5
8	配	pei	4	p	ə	bilabial	mid	match	1	14	13
9	贝	bei	4	b	ə	bilabial	mid	shell	4	14	10
10	妹	mei	4	m	ə	bilabial	mid	sister	7	14	7
11	味	wei	4	w	ə	bilabial	mid	flavor	12	14	2
12	肺	fei	4	f	ə	labial	mid	lung	3	14	11
13	盼	pan	4	p	æ	bilabial	low	hope	1	17	16
14	半	ban	4	b	æ	bilabial	low	half	4	17	13
15	曼	man	4	m	æ	bilabial	low	grace	7	17	10
16	万	wan	4	w	æ	bilabial	low	ten thousand	12	17	5
17	饭	fan	4	f	æ	labial	low	meal	3	17	14
18	兔	tu	4	t	u	coronal	high	rabbit	1	15	14
19	素	su	4	s	u	coronal	high	plain	3	15	12
20	度	du	4	d	u	coronal	high	degree	4	15	11
21	怒	nu	4	n	u	coronal	high	anger	7	15	8
22	路	lu	4	l	u	coronal	high	road	9	15	6
23	踏	ta	4	t	a	coronal	low	step	1	17	16
24	飒	sa	4	s	a	coronal	low	cool	3	17	14
25	大	da	4	d	a	coronal	low	big	4	17	13
26	那	na	4	n	a	coronal	low	that	7	17	10
27	腊	la	4	l	a	coronal	low	wax	9	17	8

Table 4.9 Summary of Mandarin stimuli. The consonant and vowel categories can be found in *C cat* and *V cat* columns. The sonority index for the consonant and vowel can be found in the *C son* and *V son* columns. The sonority difference of each stimuli can be found in the last column *S diff*.

Target Sounds	**Articulatory Sensor**	**Gesture**
Bilabial [p, b, m, w]	lower and upper lip	lip aperture
Labial-dental [f]	lower lip	lower lip
Alveolar [t, d, n, s, l]	tongue	tongue tip
Vowel	tongue	tongue dorsum

Table 4.10 Articulatory sensors and gestures for each type of sounds – Mandarin.

4.3 Results

4.3.1 Overall analysis

For all the Mandarin data in the study, there was a positive correlation between sonority difference and gestural lag, as in the plot in Figure 4.1 and Table 4.1. Following the reasoning in the English experiment, results for analyzing the subgroups of the stimuli can be found in the following subsections.

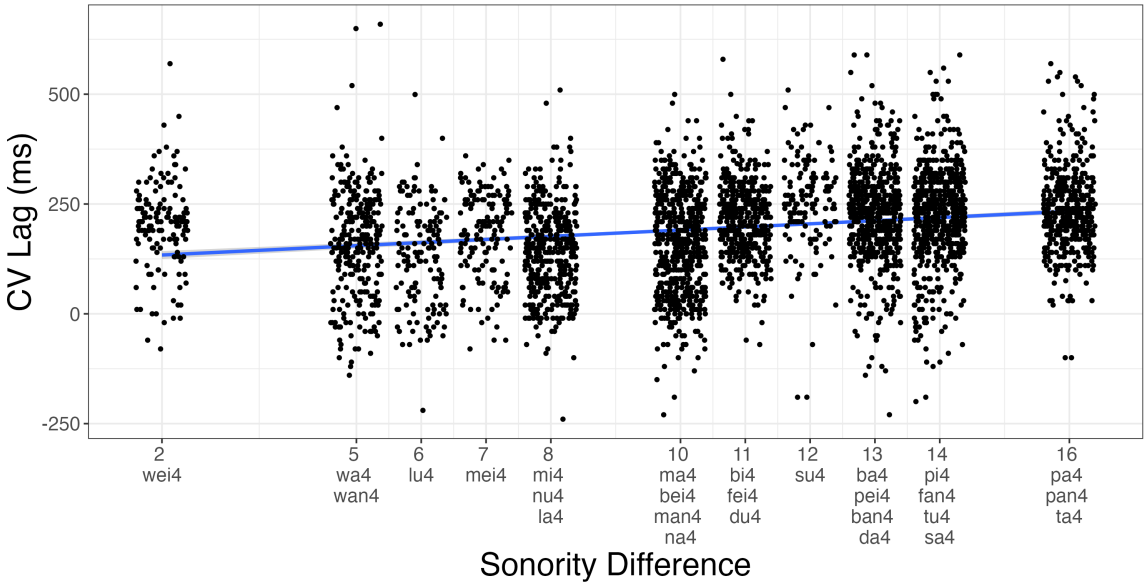

Figure 4.1 CV lag based on target onset for Mandarin participants.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	120.24	24.51	30.74	4.91	0.00003
Sonority difference	7.11	2.03	25.00	3.51	0.002

Table 4.11 Mixed effects model results for Mandarin participants.

4.3.2 Results for claim 5a: the same V and different C

I first show results for the claim 5a that for the same V and different C, a less sonorous C leads to a larger CV lag. I will first show results for the bilabial C subgroup, then I will present the results for the coronal C subgroup.

4.3.2.1 Same V and varying bilabial C

The analyses in this subgroup have the same vowel and varying bilabial consonants. Figure 4.2 and Table 4.12 showed that different bilabial consonants with the same high vowel in Mandarin had a significant positive correlation between CV lag and sonority difference.

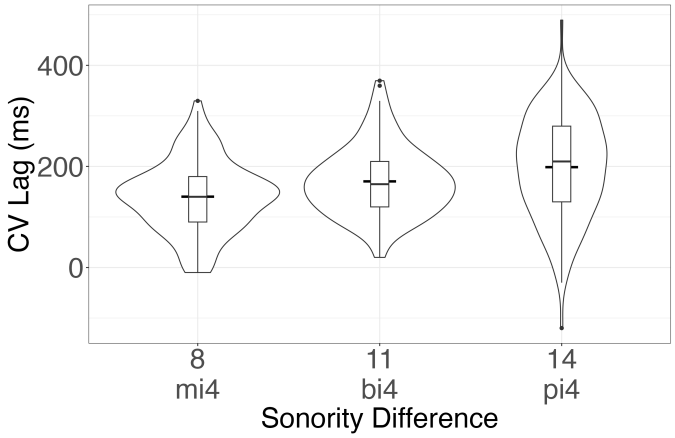

Figure 4.2 CV lag based on target onset for Mandarin participants: bilabial C and high V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	57.43	19.19	103.69	2.99	0.00
Sonority difference	10.23	1.48	419.31	6.90	< 0.00001

Table 4.12 Mixed effects model results for Mandarin participants: bilabial C and high V.

For Mandarin tone 4 stimuli with bilabial C and mid V, there was a significant positive correlation between sonority difference and gestural lag as in Figure 4.3 and Table 4.13.

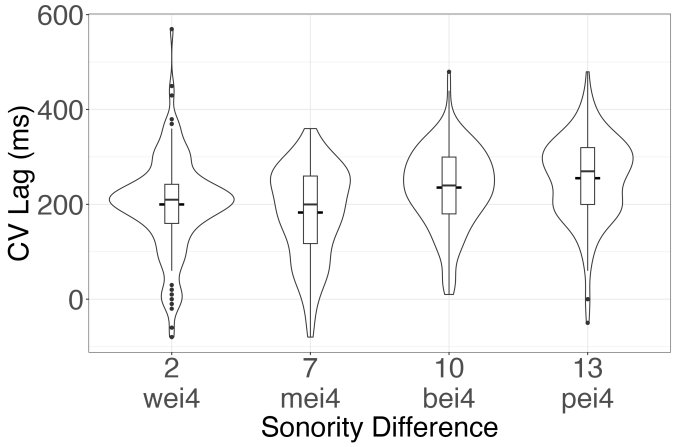

Figure 4.3 CV lag based on target onset for Mandarin participants: bilabial C and mid V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	174.17	14.32	16.50	12.17	0.00
Sonority difference	5.55	0.88	575.69	6.28	< 0.00001

Table 4.13 Mixed effects model results for Mandarin participants: bilabial C and mid V.

When only low vowel stimuli without coda nasal were analyzed, there was a significant positive correlation as in Figure 4.4 and Table 4.14.

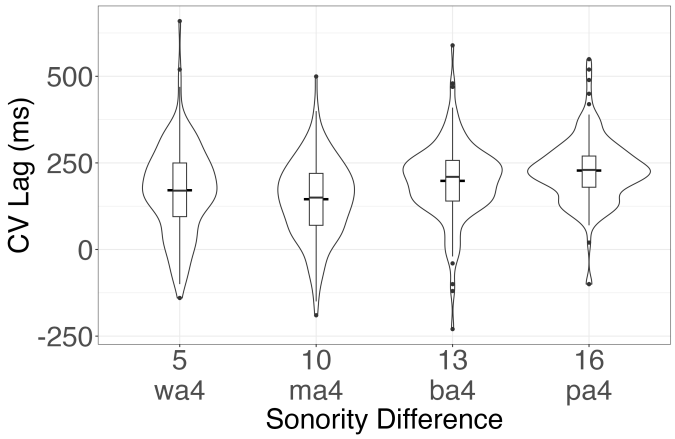

Figure 4.4 CV lag based on target onset for Mandarin participants: bilabial C and low V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	123.98	15.89	71.23	7.80	<0.00001
Sonority difference	5.73	1.17	542.16	4.91	<0.00001

Table 4.14 Mixed effects model results for Mandarin participants: bilabial C and low V.

When only low vowel stimuli with coda nasal were analyzed, there was a significant positive correlation as in Figure 4.5 and Table 4.15.

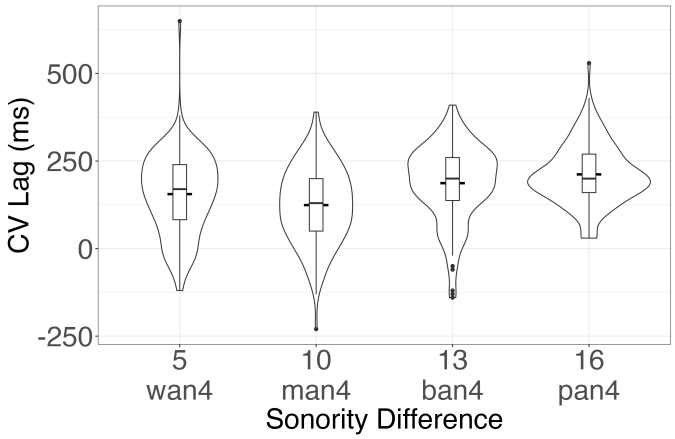

Figure 4.5 CV lag based on target onset for Mandarin participants: bilabial C and low V, with coda nasal.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	107.92	15.26	61.94	7.07	0.00
Sonority difference	5.61	1.10	543.82	5.11	< 0.00001

Table 4.15 Mixed effects model results for Mandarin participants: Bilabial C and low V, with coda nasal.

As mentioned in the corpus English experiment which is experiment 1, there is no consensus on whether voiceless stops are more sonorous than voiced stops. To resolve this potential ambiguity and to control for jaw movement, the Mandarin bilabial nasals and stops were compared.[3] If for the same vowel, the syllable beginning with a stop has a larger gestural lag, then the main claim of the dissertation will be supported. For the same high vowel, the syllable with the bilabial stop had a significant larger lag than the syllable with the bilabial nasal, as in Figure 4.6 and Table 4.16.

[3]The pairwise comparison of two stimuli differ in voicing can be found in Appendix I.

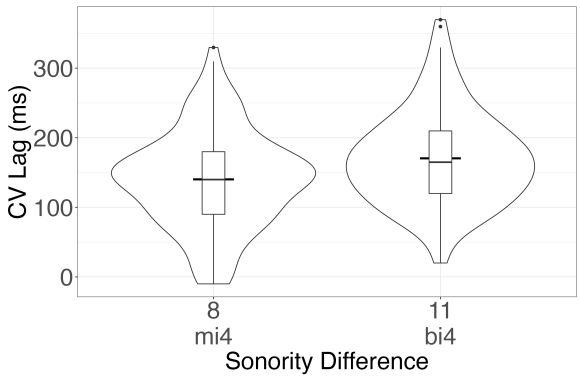

Figure 4.6 CV lag based on target onset for Mandarin participants: *bi4, mi4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	58.84	25.53	176.97	2.31	0.02
Sonority difference	10.18	2.46	276.23	4.14	0.00005

Table 4.16 Mixed effects model results for Mandarin participants: *bi4, mi4*.

For the same mid vowel, there was also a significant positive correlation between gestural lag and sonority difference as in Figure 4.7 and Table 4.17.

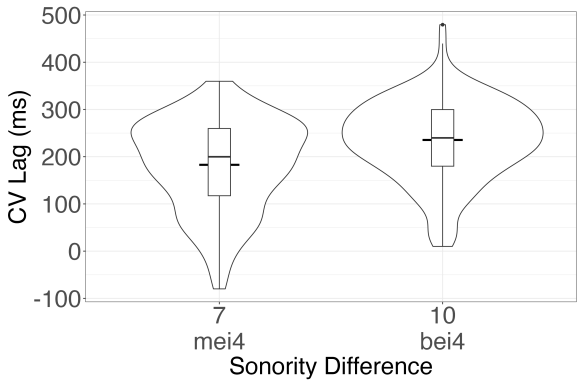

Figure 4.7 CV lag based on target onset for Mandarin participants: *bei4, mei4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	58.96	30.26	146.55	1.95	0.05
Sonority difference	17.62	3.18	286.01	5.55	< 0.00001

Table 4.17 Mixed effects model results for Mandarin participants: *bei4, mei4*.

For the same low vowel, syllables with bilabial stops had a significantly larger lag than the one with bilabial nasals as in Figure 4.8 and Table 4.18.

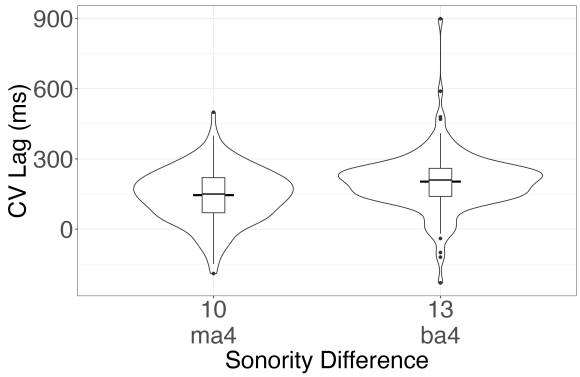

Figure 4.8 CV lag based on target onset for Mandarin participants: *ba4, ma4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-32.56	50.42	284.01	-0.65	0.52
Sonority difference	17.93	4.29	277.47	4.18	0.00004

Table 4.18 Mixed effects model results for Mandarin participants: *ba4, ma4*.

For the syllable with low vowel and a coda nasal, syllables with bilabial nasals had larger lags than those with bilabial stops as in Figure 4.9 and Table 4.19.

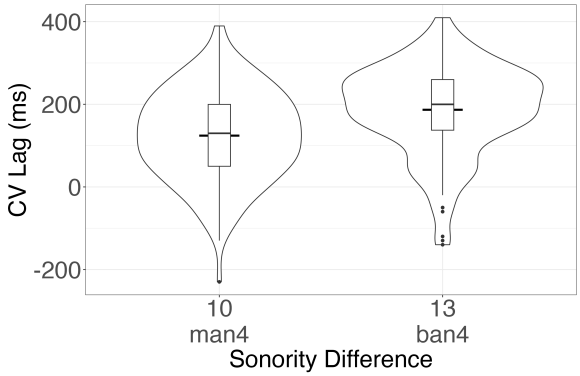

Figure 4.9 CV lag based on target onset for Mandarin participants: *ban4, man4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-80.27	49.06	244.41	-1.64	0.10
Sonority difference	20.54	4.09	264.72	5.02	< 0.00001

Table 4.19 Mixed effects model results for Mandarin participants: *ban4, man4*.

4.3.2.2 Same V and varying coronal C

Below are the results of analyzing coronal C and the same V. For Mandarin tone 4 stimuli with coronal C and high V, there was a significant positive correlation between gestural lag and sonority difference as in Figure 4.10 and Table 4.20.

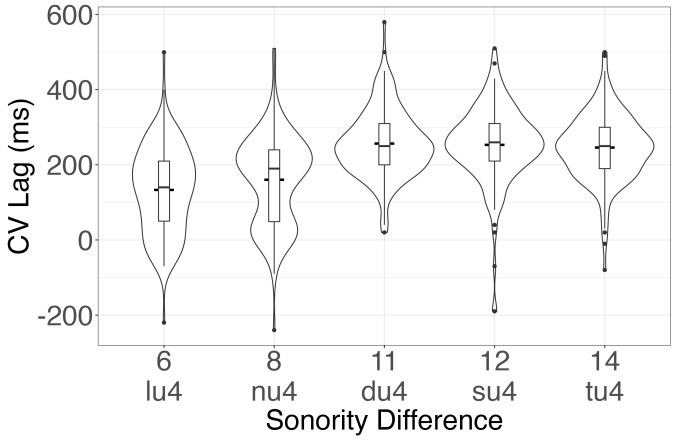

Figure 4.10 CV lag based on target onset for Mandarin participants: coronal C and high V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	41.60	17.76	46.64	2.34	0.02
Sonority difference	16.54	1.30	698.27	12.71	< 0.00001

Table 4.20 Mixed effects model results for Mandarin participants: coronal C and high V.

For Mandarin stimuli with coronal C and low V, there was a significant positive correlation between CV lag and sonority difference as in Figure 4.11 and Table 4.21.

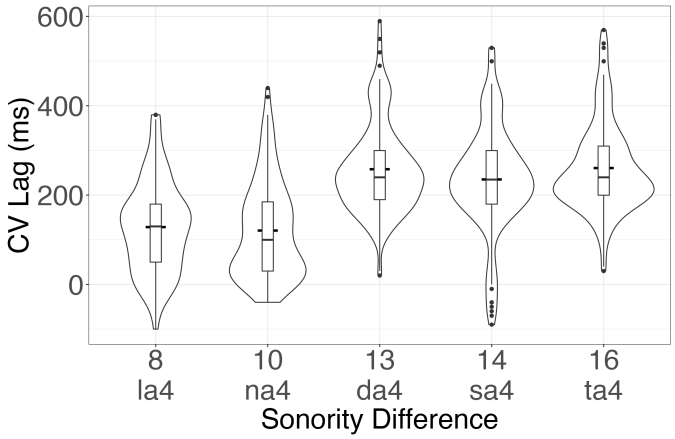

Figure 4.11 CV lag based on target onset for Mandarin participants: coronal C and low V.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-41.02	21.55	59.64	-1.90	0.06
Sonority difference	19.84	1.37	691.61	14.48	< 0.00001

Table 4.21 Mixed effects model results for Mandarin participants: coronal C and low V.

Syllables with coronal nasals and stops were compared in the following subsection. For high vowel [u], syllables with coronal stops have significantly larger lag than those with coronal stops, as in Figure 4.12 and Table 4.22.

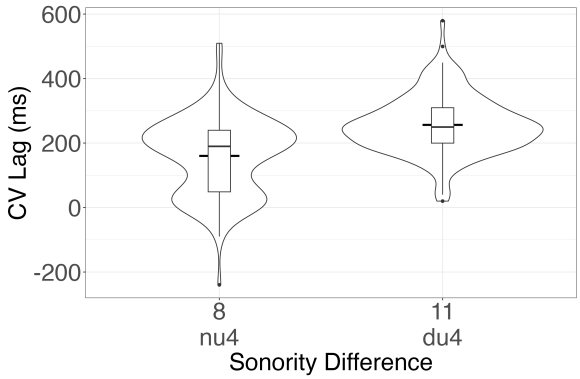

Figure 4.12 CV lag based on target onset for Mandarin participants: *du4, nu4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-83.39	39.46	153.88	-2.11	0.04
Sonority difference	30.89	3.71	273.54	8.33	< 0.00001

Table 4.22 Mixed effects model results for Mandarin participants: *du4, nu4*.

For low vowel Mandarin stimuli, syllables with oral stops had significantly larger lag than those with nasal stops as in Figure 4.13 and Table 4.23.

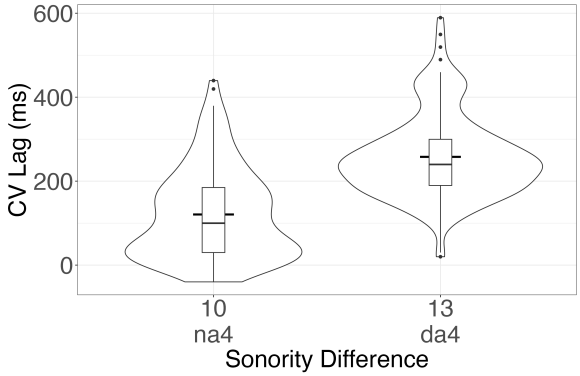

Figure 4.13 CV lag based on target onset for Mandarin participants: *da4, na4*.

Gestural Onset	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	-311.41	46.45	210.16	-6.70	0.00
Sonority difference	43.65	3.75	270.35	11.65	< 0.00001

Table 4.23 Mixed effects model results for Mandarin participants: *na4, da4*.

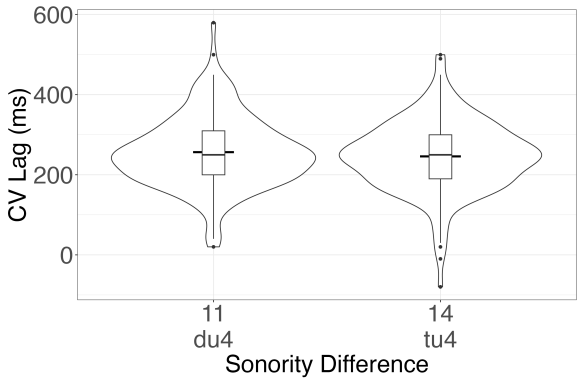

Figure 4.14 CV lag based on target onset for Mandarin participants: *tu4, du4*.

Gestural Onset	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	292.34	40.22	192.30	7.27	0.00
Sonority difference	-3.31	2.96	281.53	-1.12	0.26

Table 4.24 Mixed effects model results for Mandarin participants: *tu4, du4*.

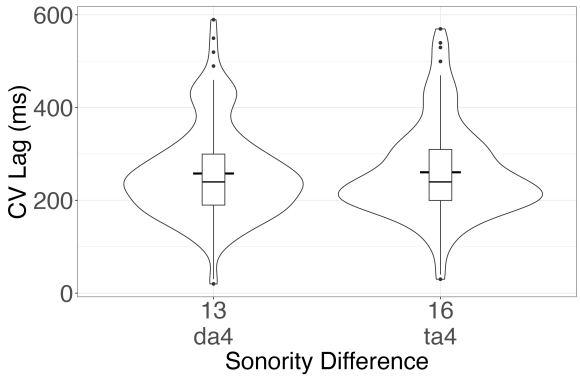

Figure 4.15 CV lag based on target onset for Mandarin participants: *ta4, da4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	241.93	54.23	280.05	4.46	0.00
Sonority difference	1.12	3.62	280.00	0.31	0.76

Table 4.25 Mixed effects model results for Mandarin participants: *ta4, da4*.

4.3.3 Results for claim 5b: the same C and different V

In the previous subsection, I showed that there was a significant positive correlation between CV lag and sonority difference. In the current subsection, I showed the results for claim 5b that for the same C and different V, a more sonorous V leads to a larger CV lag. In the following subsections, I present results for the same bilabial C first, then the same coronal C.

4.3.3.1 Same bilabial C and different V

The analyses below were used to test the claim on the same bilabial C and different V. If syllables with low vowels have larger lags than those with high vowels, the main claim of the current study will be supported. For target Mandarin syllables with [p], lower vowel syllables had a larger lag as in Figure 4.16 and Table 4.26.

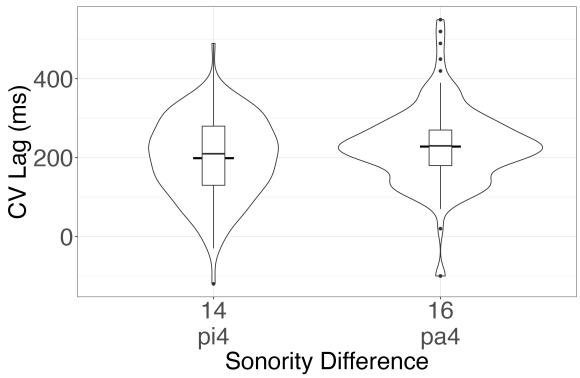

Figure 4.16 CV lag based on target onset for Mandarin participants: *pi4, pa4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	2.42	87.02	274.81	0.03	0.98
Sonority difference	14.09	5.78	270.81	2.44	0.02

Table 4.26 Mixed effects model results for Mandarin participants: *pi4, pa4*.

For syllables starting with [b], the high vowel syllables had significantly larger lags than low vowel syllables as indicated by Figure 4.17 and Table 4.27.

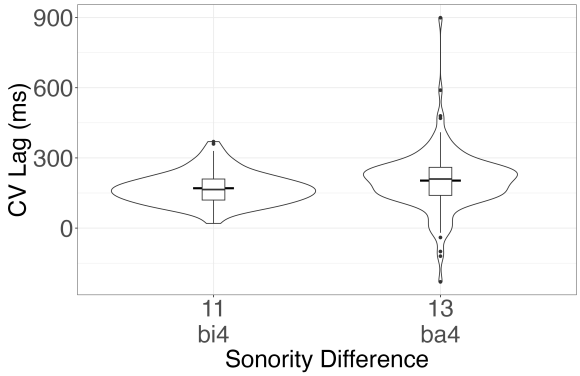

Figure 4.17 CV lag based on target onset for Mandarin participants: *bi4, ba4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-12.33	69.25	282.51	-0.18	0.86
Sonority difference	16.72	5.69	274.16	2.94	0.004

Table 4.27 Mixed effects model results for Mandarin participants: *bi4, ba4*.

For bilabial nasal syllables, those with high vowels had a larger lag than those with low vowels. However, the fitted mixed effect model did not exhibit a significant pattern (as in Table 4.28) and the descriptive plot in Figure 4.18 also did not show any obvious difference.

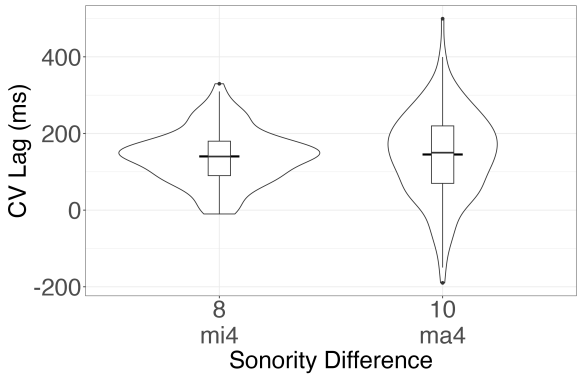

Figure 4.18 CV lag based on target onset for Mandarin participants: *mi4, ma4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	119.54	49.58	288.91	2.41	0.02
Sonority difference	2.58	5.41	280.01	0.48	0.63

Table 4.28 Mixed effects model results for Mandarin participants: *mi4, ma4*.

4.3.3.2 Same coronal C and different V

This subsection shows the results for stimuli with the same coronal consonant and different vowels. If the syllables with low vowels have larger lags than those with higher vowels, the main claim of the dissertation will be supported. Figure 4.19 and Table 4.29 shows the comparison of *tu4* and *ta4 – ta4* had a larger lag than *tu4*.

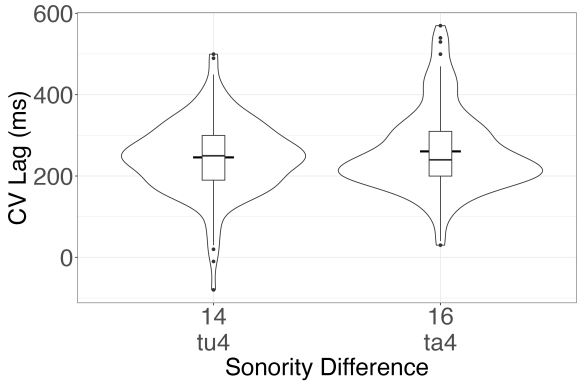

Figure 4.19 CV lag based on target onset for Mandarin participants: *tu4, ta4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	144.45	79.73	291.31	1.81	0.07
Sonority difference	7.23	5.27	284.04	1.37	0.17

Table 4.29 Mixed effects model results for Mandarin participants: *tu4, ta4*.

The syllable pair *su4* and *sa4* had similar CV lags, and high vowel syllables had larger lags than low vowel ones as in Figure 4.20 and Table 4.30.

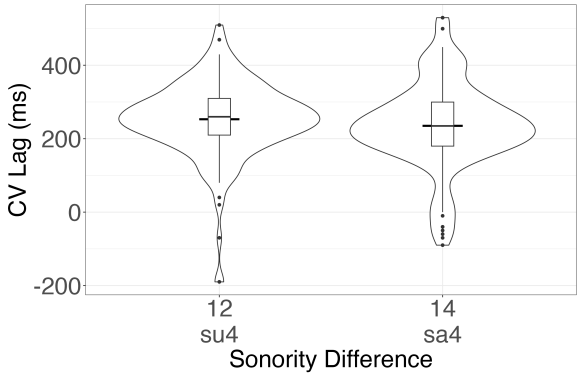

Figure 4.20 CV lag based on target onset for Mandarin participants: *su4, sa4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	354.88	80.82	290.98	4.39	0.00
Sonority difference	-8.47	6.12	282.06	-1.38	0.17

Table 4.30 Mixed effects model results for Mandarin participants: *sa4, su4*.

The coronal consonant syllable *du4* and *da4* also had similar CV lags as in Figure 4.21 and Table 4.31. In other words, the expected pattern was not exhibited for this pair.

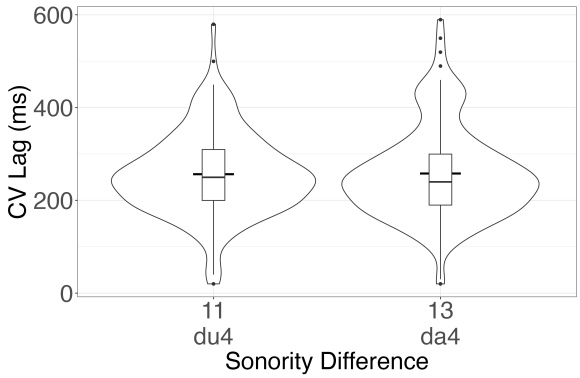

Figure 4.21 CV lag based on target onset for Mandarin participants: *du4, da4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	249.92	60.55	279.19	4.128	0.00
Sonority difference	0.51	4.88	281.01	0.10	0.92

Table 4.31 Mixed effects model results for Mandarin participants: *du4, da4*.

The syllables *nu4* had a significantly larger lag than *na4*, as in Figure 4.22 and Table 4.32. This observation was opposite to the main claim of the dissertation. It could be that the vowels were nasalized, and this changed the sonority difference.

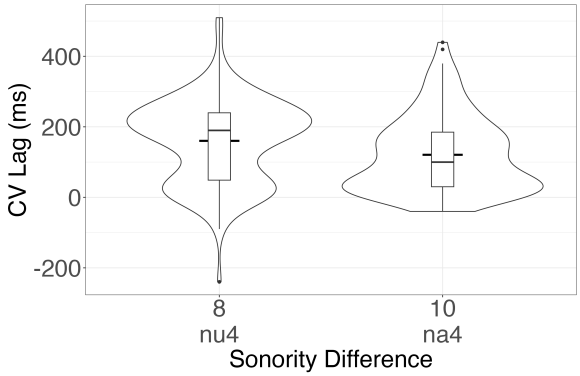

Figure 4.22 CV lag based on target onset for Mandarin participants: *nu4, na4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	308.76	58.87	248.16	5.25	0.00
Sonority difference	-18.67	6.26	260.49	-2.98	0.003

Table 4.32 Mixed effects model results for Mandarin participants: *nu4, na4*.

The coronal consonant syllables *lu4* and *la4* had similar CV lags, which means that the expected positive correlation was not found here. See Figure 4.23 and Table 4.33.

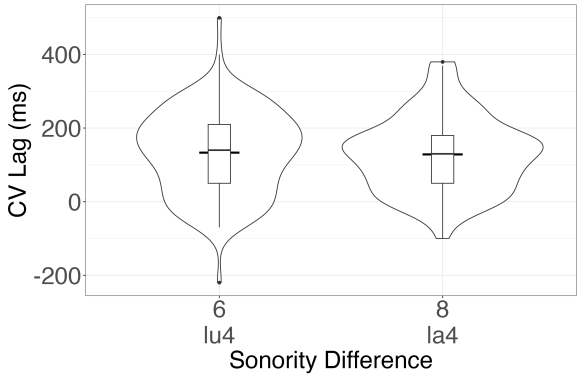

Figure 4.23 CV lag based on target onset for Mandarin participants: *lu4, la4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	154.44	45.42	215.13	3.40	0.001
Sonority difference	-3.70	6.20	266.08	-0.60	0.55

Table 4.33 Mixed effects model results for Mandarin participants: *lu4, la4*.

4.3.4 The vowel displacement and C displacement analysis

As mentioned earlier, Shaw and Chen (2019) found that there was a negative correlation between CV lag and vowel displacement. Also, C displacement was used as an estimate of jaw movement. Just like the English experiment, in the following, I considered V displacement and C displacement as random intercepts.[4] Since C displacement was measured differently for stimuli with coronal C vs. stimuli with bilabial C, the two types of stimuli were analyzed separately. The mixed-effect model for stimuli with bilabial C can be found in Table 4.34, where there was a positive

[4]The R formula is shown here: CV lag based on target onset~Sonority difference+(1|Participant)+(1|Word)+(1|C duration)+(1|V displacement)+(1|C displacement).

correlation between sonority difference and CV lag. However, the effect of sonority difference was not significant.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	136.54	22.71	16.13	6.01	0.00
Sonority difference	3.65	1.96	13.21	1.86	0.09

Table 4.34 Mixed effects model results for Mandarin participants. The bilabial C stimuli. Random intercepts: participant, word, C duration, V displacement, C displacement.

To test whether this insignificance is due to vowel displacement or consonant displacement, I added V displacement or C displacement as fixed effects in the model. The results of considering vowel displacement can be found in Table 4.35.[5] After considering vowel displacement, there was a significant positive correlation between sonority difference and CV lag. Also, there was a significant positive correlation between sonority difference and vowel displacement. The positive correlation between sonority and vowel displacement was surprising as it did not replicate either Shaw and Chen (2019) or experiment 1. Further research needs to be conducted to conclude a reason.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	107.87	21.38	15.42	5.04	0.00
Sonority difference	5.73	1.87	13.12	3.07	0.01
V displacement	11.90	0.64	1860.38	18.61	< 0.00001

Table 4.35 Mixed effect model for bilabial C stimuli with sonority difference and V displacement as fixed effects. Random intercepts: participants, word, C duration.

I also added C displacement as a fixed effect, and there was no significant relationship between C displacement and CV lag as in Table 4.36.[6]

[5]CV lag based on target onset~Sonority difference + V displacement + (1 | Participant) + (1 |Word) + (1 | C duration) + (1 | C displacement).

[6]The code is: CV lag based on target onset~Sonority difference+C displacement+(1|Participant)+(1|Word)+(1|V displacement).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	134.04	23.15	16.41	5.79	0.00
Sonority difference	3.34	2.00	13.38	1.67	0.12
C displacement	-1.07	1.01	758.35	-1.06	0.29

Table 4.36 Mixed effect model for bilabial C stimuli with sonority difference and C displacement as fixed effects. Random intercepts: participant, word, V displacement.

Also, the mixed effect model result for considering C and V displacement for stimuli with coronal C can be found in Table 4.37. When considering C and V displacements, there was a significant positive correlation between CV lag and sonority difference.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	29.54	34.29	9.15	0.86	0.41
Sonority difference	15.02	2.84	7.80	5.30	0.001

Table 4.37 Mixed effects model results for Mandarin participants. The coronal C stimuli. Random intercepts: participant, word, C duration, V displacement, C displacement.

4.3.5 Summary

As in Table 4.38, most groups in Mandarin EMA data exhibited a significant positive correlation between CV lag and sonority difference. Pairwise comparison in Table 4.39 showed that the stimuli with bilabial C generally had significant positive correlations when the stimuli differed in voicing, nasality, and vowel height. In those cases, the stimuli had controlled nasality, voicing, as well as C place and manner.

	Dataset (Mandarin EMA data)	Est (son diff)	Est (displ)
	All Mandarin EMA data	7.11 **	
Bilabial C	C, V displacement as random intercepts	3.65	
	V displacement as fixed effect	5.73 **	11.90 ***
	High V (mi4, bi4, pi4)	10.23 ***	
	Mid V (wei4, mei4, bei4, pei4)	5.55 ***	
	Low V no coda (wa4, ma4, ba4, pa4)	5.73 ***	
	Low V with coda (wan4, man4, ban4, pan4)	5.61 ***	
Coronal C	C, V displacement as random intercepts	15.02	
	High V (lu4, nu4, du4, su4, tu4)	16.54 ***	
	Low V (la4, na4, da4, sa4, ta4)	19.84 ***	

Table 4.38 Summary of Mandarin EMA results. *Son diff* means sonority difference, *est* means estimate, and *displ* means displacement. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$.

	Pairwise comparison	Stimulus pair	Estimate (sonority difference)
Bilabial C	Nasality differ	mi4, bi4	10.18 ***
		mei4, bei4	17.62 ***
		ma4, ba4	17.93 ***
		man4, ban4	20.54 ***
	Voicing differ	bi4, pi4	10.57 ***
		bei4, pei4	6.95 *
		ba4, pa4	9.69 *
		ban4, pan4	8.11 *
	Vowel height differ	mi4, ma4	2.58
		bi4, ba4	16.72 **
		pi4, pa4	14.09 *
Coronal C	Nasality differ	nu4, du4	30.89 ***
		na4, da4	43.65 ***
	Voicing differ	tu4, du4	-3.31
		ta4, da4	1.12
	Vowel height differ	tu4, ta4	7.23
		su4, sa4	-8.47
		du4, da4	0.51
		nu4, na4	-18.67 **
		lu4, la4	-3.7

Table 4.39 Summary of EMA Mandarin pairwise comparisons. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$.

4.4 Conclusion

Overall the Mandarin experiment replicated the observations in the English experiment — in other words, the Mandarin experiment showed that there is a positive correlation between sonority difference and CV gestural lag for tone 4 Mandarin words. This is generally true for both the claim 5a which is on different C and controlled V, and the claim 5b which is about the same C and different V.

However, just like the English EMA experiment, there were a few sub-analyses that did not exhibit the expected pattern. The possible reasons for the non-significant observations may be the following — that the vowel sensor was the same regardless of vowel height, and that the nasality of surrounding sounds affected the sonority level of the vowel. First, the coronal consonant and different vowel groups had some unexpected patterns. This is probably due to the reasons a) that the C and V share the tongue as the articulator and b) that there is one consistent tongue measure used regardless of vowel height. For a similar reason, *do* and *doe* in English also did not exhibit the expected pattern. Second, nasalization may affect the sonorous level or gestural lag of adjacent sounds. This could be the reason why the expected pattern was not found in the *mi4*, *ma4* pair. Also, when including stimuli without a controlled environment (such as the bilabial group), the pattern was not significant.

CHAPTER 5

DISCUSSION

The dissertation showed that there is a significant positive correlation between CV lag and sonority difference for both English monosyllabic words and Mandarin tone 4 words. The positive correlation was found when there is the same C and different V, as well as when there is the same V and different C. Furthermore, the claim was also supported by more controlled comparisons of labial and coronal consonants, as well as vowels of different heights.

For experiment 1, I observed the positive correlation in English corpus data. In experiment 2, a similar positive correlation was observed for English EMA data, which has more variation of vowels and consonants in the stimuli. Furthermore, the correlation was not limited to English since a similar positive correlation has been found in Mandarin EMA data in experiment 3.

In all 3 experiments, I argued that the observed correlation cannot simply be attributed to vowel quality or the effect of vowel displacement. Specifically, I showed that voicing is unlikely to be the sole factor leads to CV lag variation because two stimuli that differ in nasality and share the same voicing — such as [bɑ, mɑ] — exhibited the positive correlation. Moreover, groups of stimuli with the same vowel quality showed a significant positive correlation — such as the bilabial C group and the alveolar C group. Therefore, vowel quality is unlikely to be the driver of observed CV lag variation. Additionally, in a post-hoc analysis, I showed that vowel displacement in fact contributes to the CV lag variation in the direction *opposite* to what I observed for sonority difference, and the effect of sonority difference was almost unchanged.

The results of this dissertation also suggest that jaw movement is unlikely to be an important factor driving the observed correlation between sonority and gestural lag variation. First, using consonant displacement as an approximation of jaw movement, I showed that the observed positive correlation is not confounded by jaw movement. Second, I looked at voicing pairs and nasality pairs which controlled for jaw movement. The voiced and voiceless pairs differ in voicing and involve a similar level of jaw movement showed the expected effect. Furthermore, in nasality pairs [mɑ, bɑ] and [nɑ, dɑ], the stimuli with the nasal segment have a larger lag than the oral

segment of the same place of articulation. Since the two segments are mainly different in nasality and involve similar degrees of jaw movement, the lag difference cannot be attributed to jaw movement. For the pairwise comparisons controlled for jaw movement, I used post-hoc analyses to confirm that consonant displacement was indeed controlled. Relatedly, the comparison using vowel displacement (which is also an approximation of jaw movement) makes a similar point. The positive correlation between sonority difference and CV lag is still exhibited when including vowel displacement. Based on the above analysis, jaw movement is unlikely to be the factor leading to the observed relationship between CV lag and sonority difference.

As noted in Chapter 1.1, previous theoretical claims about speech production have typically predicted a consistent relationship of CV coordination, assuming prosodic factors are held constant (Browman and Goldstein, 1989, 1992; Nam, 2007; Liu et al., 2020; Durvasula and Wang, 2023; Liu et al., 2022). However, the results of the current study suggest that CV lags are correlated with the sonority difference between the consonant and the vowel; therefore, sonority should be a factor in modeling articulatory timing. In addition, the results of the dissertation also suggest that sonority needs to be considered in experiments studying gestural coordination, particularly in making comparisons between segment sequences consisting of different segments.

In my 3 experiments, Mandarin and English both showed significant positive correlations between CV lag and sonority difference. The estimates were around 10 as in the following Table 5.1, 5.2, and 5.3.

Dataset (English corpus data)	**Estimate (sonority diff)**	**Estimate (displace)**
All English corpus data	16.49 ***	
All English corpus data, V displacement	15.76 ***	-3.97 ***
All English corpus data, C displacement	16.76 ***	2.19 ***
sVd stimuli (subset)	12.09	
T1 C stimuli	15.53 **	
Alveolar C stimuli (la, na, za, da, sa, ta)	14.24 *	
Lip aperture (wa, ma, ba, pa)	14.71 ***	

Table 5.1 Summarizing the results of experiment 1. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 2.28, for the readers' convenience.

	Dataset (English EMA data)	Estimate (sonority difference)
	All English EMA data	11.24 ***
Bilabial C	All bilabial C data, C and V displacement	12.74 ***
	High V (week, meek, beak, peak)	14.00 ***
	Mid V (wane, main, bane, pain)	14.35 ***
	Low V (whack, Mac, back, pack)	10.47 ***
Coronal C	All coronal C data, C and V displacement	8.31 *
	High V (new, do, sue, two)	8.02 **
	Mid V (know, doe, so, toe)	9.40 ***
	Low V (knock, dock, sock, talk)	7.43 ***

Table 5.2 Summary of English EMA results. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 3.40, for the readers' convenience.

	Dataset (Mandarin EMA data)	Est (son diff)	Est (displ)
	All Mandarin EMA data	7.11 **	
Bilabial C	C, V displacement as random intercepts	3.65	
	V displacement as fixed effect	5.73 **	11.90 ***
	High V (mi4, bi4, pi4)	10.23 ***	
	Mid V (wei4, mei4, bei4, pei4)	5.55 ***	
	Low V no coda (wa4, ma4, ba4, pa4)	5.73 ***	
	Low V with coda (wan4, man4, ban4, pan4)	5.61 ***	
Coronal C	C, V displacement as random intercepts	15.02	
	High V (lu4, nu4, du4, su4, tu4)	16.54 ***	
	Low V (la4, na4, da4, sa4, ta4)	19.84 ***	

Table 5.3 Summary of Mandarin EMA results. *Son diff* means sonority difference, *est* means estimate, and *displ* means displacement. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 4.38, for the readers' convenience.

For the pairwise comparisons, the stimulus pairs that differ in nasality exhibited the most consistent results. There was less consistency in the differ-in-voicing pairs as well as the differ-in-vowel-height pairs. The non-consistent results could be due to a) the same tongue sensor measurements for different vowels or b) the imprecise coding of C voicing of the stimuli.

Pairwise comparison	Stimulus pair	Estimate (sonority difference)
Nasality differ	ma, ba	14.73 ***
	na, da	22.09 ***
Voicing differ, stop	pa, ba	14.40 ***
	da, ta	5.84 *
Voicing differ, fricative	fa, va	7.11
	sa, za	9.74
Vowel height (subset)	been, back	21.06 ***

Table 5.4 Summarizing pairwise comparison of experiment 1. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 2.29.

	Pairwise comparison	Stimulus pair	Estimate (sonority difference)
Bilabial C	Nasality differ	Mac, back	11.20 *
		meek, beak	21.84 ***
		main, bane	21.59 ***
	Voicing differ	beak, peak	8.92
		bane, pain	12.81 **
		back, pack	12.66 *
	Vowel height differ	peak, pack	3.2
		beak, back	-0.22
		meek, Mac	7.61
		week, whack	18.23 **
Coronal C	Nasality differ	new, do	23.74 ***
		know, doe	11.83 **
		knock, dock	17.58 ***
	Voicing differ	two, do	-5.64
		toe, doe	6.47
		talk, dock	-0.23
	Vowel height differ	two, toe	61.90 ***
		do, doe	27.51
		sue, so	76.03 ***
		new, know	60.78 ***

Table 5.5 Summary of English EMA pairwise comparison. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 3.41.

	Pairwise comparison	Stimulus pair	Estimate (sonority difference)
Bilabial C	Nasality differ	mi4, bi4	10.18 ***
		mei4, bei4	17.62 ***
		ma4, ba4	17.93 ***
		man4, ban4	20.54 ***
	Voicing differ	bi4, pi4	10.57 ***
		bei4, pei4	6.95 *
		ba4, pa4	9.69 *
		ban4, pan4	8.11 *
	Vowel height differ	mi4, ma4	2.58
		bi4, ba4	16.72 **
		pi4, pa4	14.09 *
Coronal C	Nasality differ	nu4, du4	30.89 ***
		na4, da4	43.65 ***
	Voicing differ	tu4, du4	-3.31
		ta4, da4	1.12
	Vowel height differ	tu4, ta4	7.23
		su4, sa4	-8.47
		du4, da4	0.51
		nu4, na4	-18.67 **
		lu4, la4	-3.7

Table 5.6 Summary of EMA Mandarin pairwise comparisons. *** means that $p \leqslant 0.001$; ** means that $p \leqslant 0.01$; * means that $p \leqslant 0.05$. This Table is a repetition of Table 4.39.

In the following, I discuss potential explanations of the current finding in subsection 5.1. Moreover, I discuss the potential impact of the link between sonority and gestural timing. Specifically, the study provides a potential basis to account for several cross-linguistic typological patterns such as the Sonority Sequencing Principle (SSP) (Sievers, 1881, 1901; Greenberg, 1965; Pike, 1972; Hooper and Bybee, 1976; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Blevins, 1995; Parker, 2002, 2011), the Sonority Dispersion Principle (Clements, 1990; Parker, 2011), and the asymmetry between CV and VC syllable frequencies (Ohala, 1990; Tabain et al., 2004; Nam et al., 2009). These general typological patterns of human language may be accounted for based on our results along with another premise of a preference for larger gestural lag over shorter ones. Furthermore, in subsection 5.3, I propose a sonority-driven speech production constraint. Lastly, in subsection 5.4, I discuss some caveats and directions for future studies.

5.1 Potential explanations and theory for the finding

In this section, I intend to evaluate some claims that could potentially explain the finding of the study that there is a positive correlation between CV lag and sonority difference. Firstly, the language-specific mechanism in Georgian mentioned by Crouch (2022) and Crouch et al. (2023) seems the least likely since similar patterns have been found in other languages such as English and Mandarin (Gao, 2008; Shaw and Chen, 2019). For the rest of the section, to explain the findings of the dissertation, I consider principles in speech production such as parallel transmission (Mattingly, 1981), coarticulatory resistance (Bladon and Al-Bamerni, 1976), and perceptual recoverability (Chitoran et al., 2002). I also consider *the prosodic gesture model* to model the results (Byrd and Saltzman, 2003).

The first claim I evaluate here comes from Mattingly (1981), who argued that segments are grouped into syllables because listeners have certain expectations regarding the *parallel transmission* of information. Specifically, segments from different classes of articulatory manners should be ordered in a way that a more closed constriction must occur in the process of being released to a more open constriction. They suggest that this kind of ordering or organization of articulatory gestures ensures the parallel transmission of a syllable, which could be argued to be more efficient than decoding isolated consonants or vowels. More specifically, they suggest that a sonority rise is more likely to transmit in parallel over a sonority plateau (and a sonority plateau over a sonority fall). Furthermore, they suggest that this view can be used to explain the Sonority Sequencing Principle (SSP), which requires that each syllable should exhibit one peak of sonority in the nucleus, and that, cross-linguistically, a sonority rise (such as [pl]) is preferred in onsets over a sonority plateau (such as [pt]) which in turn is preferred over a sonority fall (such as [lp]) (Sievers, 1881, 1901; Greenberg, 1965; Pike, 1972; Hooper and Bybee, 1976; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Blevins, 1995; Parker, 2002, 2011). Following their reasoning, one would expect a sonority rise to have the least lag and a sonority fall to have the longest lag — but this prediction is contrary to the observations, both in the current study and those in Crouch (2022) and Crouch et al. (2023). Therefore, parallel transmission could not

account for the primary finding of the paper.

Another concept relevant to the interpretation of the finding is *coarticulatory resistance*, which was originally used to account for the coarticulatory variation in English /l/ (Bladon and Al-Bamerni, 1976). Coarticulatory resistance is likely not to play a role here because it is non-directional. On the contrary, the main claim of the dissertation requires a directional calculation of sonority difference and gestural lag. Furthermore, Kent and Minifie (1977) argue that though the concept of coarticulatory resistance can be used to model the observed variation, it seems unable to predict or explain the general link between gestural timing and sonority difference, because it is expected to vary by language, segment, and even potentially context. Rather, it serves as the numerical redescription of the articulatory variation. Therefore, even though we could say that CV syllables of larger sonority difference have higher coarticulatory resistance, the claim does not explain the observations.

A third possible way to account for the claim is based on *the prosodic gesture model* (Byrd and Saltzman, 2003), which suggests gestural lag variation. The prosodic or π-gesture model suggests that prosodic gestures "temporally stretch gestural activation trajectories" (p. 149) and prosodic gestures make the gestures in their activation domain longer, larger, and further apart (Byrd and Saltzman, 2003). In Figure 5.1, prosodic gesture occurs in the prosodic tier, and it slows down the gestural coordination of gesture 1 and gesture 2 between the two dashed lines. Cho (2006) suggested that there are various strengths of prosodic gestures. Therefore, it is worth considering sonority as a prosodic gesture. Note that if sonority is a prosodic gesture, the C and V should also be lengthened as the lags are lengthened. To evaluate this, I test whether C duration positively correlates to sonority difference and whether V duration positively correlates to sonority difference. Specifically, C or V duration was modeled as a function of sonority difference, where participants and words were modeled as random intercepts. For English corpus data from experiment 1, C duration (estimate = 4.66, p = 0.15) or V duration (estimate = 4.42, p = 0.35) did not have a relationship with sonority difference. For English EMA data from experiment 2, C duration (estimate = 4.53, p = 0.05) had a positive correlation with sonority difference, but V duration (estimate = 2.55, p = 0.20) did not

have a positive correlation with sonority difference. For Mandarin EMA data from experiment 3, C duration (estimate = -3.31, p = 0.12) did not have a relationship with sonority difference, but V duration had a positive correlation with sonority difference (estimate = 7.024, p= 0.01). These results show that C and V gestures did not lengthen consistently according to sonority. Therefore, sonority may not be modeled as a prosodic gesture.

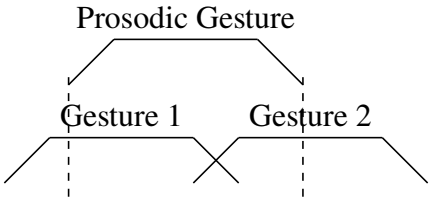

Figure 5.1 Prosodic gesture.

A fourth relevant claim regarding the modulation of gestural timing comes from Chitoran et al. (2002), who suggests that *perceptual recoverability* could be the underlying reason for the gestural coordination they observed in Georgian. Specifically, Chitoran et al. (2002) proposed that, cross-linguistically, a syllable should have structures that allow maximum gestural overlap with minimal loss of information. Even though the hypothesis was rejected by Crouch (2022) and Crouch et al. (2023) based on their results in Georgian CC timing, perceptual recoverability can provide a valid explanation of the finding if we assume sonority is essentially an abstraction of *intensity*. Intensity is considered to be the phonetic correlate of sonority in Parker (2002), where the sonority scale of the dissertation comes from.

The speculative explanation for the link between gestural timing and sonority that I would like to suggest is that a larger intensity difference requires a larger gestural lag to ensure perceptual recoverability. *Intensity* is the acoustic correlate of sonority in Parker (2008), and the dissertation assumes the sonority scale in Parker (2002). Therefore, the observed correlation between sonority difference and CV lag could actually be due to the correlation between *intensity difference* and CV lag. It could be that if two adjacent sounds are very different in intensity, it is likely that a large degree of overlap will result in the masking of the lower-intensity sound. On the other hand, if two sounds are similar in terms of intensity, they may be more likely to withstand a large degree of

overlap. Since the assumption is that a consonant and a vowel in a given CV syllable should have maximum overlap with minimal loss of information, a C and a V that differ more in intensity should have a larger lag (less overlap) than syllables with smaller intensity differences between C and V. The remaining question is about the directional sensitivity of the intensity difference. It is possible that the human perceptual system is sensitive to the sound intensity direction — the lower intensity sound is more likely to be masked by a *following* higher intensity sound than a *preceding* higher intensity sound. The two predictions of the explanation — 1) two adjacent sounds in a syllable with higher intensity difference need to have larger lags (less overlapping degree) to be perceived; 2) a lower intensity sound followed by a higher intensity sound is more challenging to be perceived than the exact sound preceded by a higher intensity sound — can be tested by independently designed speech perception experiments.

The perceptual experiments discussed here can be used to test the relationship between perceptual recoverability and intensity difference. Synthesized stimuli made up of two segments could be used to test the claim. The two segments in a stimulus should have different intensity difference values. For instance, there could be stimuli such as AB where $B_{intensity}$ - $A_{intensity}$ = 20 dB or CD where $D_{intensity}$ - $C_{intensity}$ = 40 dB. For each pair of segments of the same intensity difference, different degrees of overlap of the two segments were created. For example, for AB where $B_{intensity}$ - $A_{intensity}$ = 20 dB, A and B should have various overlapping degrees as follows: B aligned to the 0%, 20%, 40%, 60%, 80%, and 100% of A. The recordings will be played to the participants to test whether they can identify the stimuli. It is expected that with the increase in intensity difference, participants would need more lag (less overlap) to correctly identify the segment pairs. To test directionality, the perception of stimuli AB will be compared with that of stimuli BA, and it is predicted that a lower-intensity segment followed by a higher-intensity segment (e.g., AB) should require more lag than a higher-intensity segment followed by a lower-intensity segment (e.g., BA).

This hypothesis of perceptual recoverability and how it connects to gestural timing aligns with the argument in Wright (1996), who argued that the perceptual demands of the listener contribute

to the production strategy of the speaker. In fact, scholars have explored the perception-production link and observed altered speech production patterns when the auditory feedback changes (Katseff et al., 2012). After collecting perceptual evidence, we will be more equipped to evaluate the explanation of the observation in the future.

Note that the preference for larger intensity difference over smaller difference was also suggested by Henke et al. (2012). They claimed that greater amplitude change results in more robust information. As mentioned in Chapter 1, Henke et al. (2012) argued that different natural classes of sounds have internal and transitional perception cues of various robust levels. Therefore, it is expected that there are different overlapping degrees of adjacent sounds for different natural classes to ensure the transmission of information in the perception.

5.2 Providing a basis for some phonological universals

The findings of the dissertation can provide a basis for several phonological universals if we add a premise that humans prefer larger gestural lags in articulation. Notably, the implications do not depend on sonority as a primitive. In other words, even if sonority were ultimately derived from some other set of factors, and the gestural timing of CV sequences was really correlated with those other factors, the implications of the gestural timing differences for the various typological observations that I discuss below will still hold true.

5.2.1 Providing a basis for the Sonority Dispersion Principle and the SSP

The main finding of the current study has the potential to help explain phonological universals related to the syllable structure — such as the Sonority Dispersion Principle and the Sonority Sequencing Principle (SSP). Firstly, the Sonority Dispersion Principle states that in a syllable CV, the onset and nucleus differ from each other in sonority as much as possible (Clements, 1990; Parker, 2011). In other words, this principle requires that a CV syllable should have larger *sonority difference*. The Sonority Dispersion Principle is potentially derivable from the finding of the current dissertation along with another premise that a larger gestural lag is preferred, potentially for reasons of perceptual recoverability (Chitoran et al., 2002). For instance, it may be that a larger lag is correlated with more perceptual salience so it is preferred. It is also possible that a larger

lag is preferred because it is easier to articulate. Ohala (1990) argued that some sequences may be disfavored due to their being difficult to articulate — it is possible that having a shorter gestural lag for gestures serves as the physical manifestation of articulatory difficulty to implement a certain ordered sequence. More work needs to be done to assess articulatory ease, which has been elusive to define or study (Shariatmadari, 2006).

Secondly, the findings of the dissertation could be used to explain the Sonority Sequencing Principle (SSP) if we generalize the link to CC onset clusters. The SSP requires that a sonority rise (such as [pl]) is preferred in onsets over a sonority plateau (such as [pt]) which in turn is preferred over a sonority fall (such as [lp]) cross-linguistically (Sievers, 1881, 1901; Greenberg, 1965; Pike, 1972; Hooper and Bybee, 1976; Steriade, 1982; Selkirk, 1984; Clements, 1990; Kenstowicz, 1994; Blevins, 1995; Parker, 2002, 2011). Based on the observations of the current study, the SSP can be accounted for as follows. If the sonority index of the clusters C_1C_2 is coded according to Table 1.6, and if the sonority difference of clusters is calculated by subtracting the sonority index of the first consonant from that of the second consonant (i.e., C_2 - C_1), then the sonority difference of the clusters [pl], [pt], and [lp] are 8, 0, -8 respectively (Table 5.7). In this current dissertation, I found that gestural lag positively correlates to sonority difference. If one generalizes the finding on CV sequences to CC sequences, one would predict that sonority rise has a larger lag than sonority plateau, which has a larger lag than sonority fall. This prediction was already been supported by Georgian (Crouch, 2022; Crouch et al., 2023). If the premise that humans prefer larger gestural lag within a syllable is true, I would predict the phonological constraint that sonority rise is preferred over plateau over fall.

Onset cluster	**Sonority difference**
Sonority rise [pl]	9-1=8
Sonority plateau [pt]	1-1=0
Sonority fall [lp]	1-9=-8

Table 5.7 Providing a basis for the SSP. In the Sonority difference column, the first two numbers refer to the sonority indexes of the two consonants in the cluster, and the last number represents the sonority difference.

5.2.2 Relevance to the Syllable Contact Law

A related question would be how the current finding can inform the explanation of the Syllable Contact Law, which specifies that the structure A.B would be more preferable if a-b is larger (Hooper and Bybee, 1976; Murray and Vennemann, 1983). I would argue that a general preference toward larger gestural lag within the syllable could derive the Syllable Contact Law. Consider the syllable $CV_1A.BV_2$, where there are two syllables CV_1A and BV_2, with A.B at the syllable boundary. If both syllables need to satisfy the requirement that larger lags are preferred in a syllable, then it comes as a consequence that a-b is larger (a and b refer to the sonority of A and B respectively). Since in each syllable, every segment sequence should satisfy the large lag requirement, V_2-b and a-V_1 should have larger sonority differences. This means that V_2 and a should be larger, and b and V_1 should be less sonorous. If a is large and b is small, and a-b would be larger. See the stepwise derivation in (8).

(8) Preference in question: larger lags (larger sonority difference) are preferred within a syllable, where sonority difference is calculated by $\text{Sonority}_{\text{later}}$ - $\text{Sonority}_{\text{former}}$.

a. Sample syllables: $CV_1A.BV_2$

b. First syllable: CV_1A

c. First syllable satisfying preference: a-V_1 larger → **a large**; V_1 small

d. Second syllable: BV_2

e. Second syllable satisfying preference: V_2-b larger → V_2 large, **b small**

f. **a large**, **b small** → a-b large

The derivation shows that the Syllable Contact Law may be the consequence of adjacent syllables satisfying the larger lag requirement within each syllable. Admittedly, the derivation is not an explanation of the Syllable Contact Law, but rather a prediction about tendencies.

5.2.3 Providing a basis for the syllable frequency asymmetry: CV versus VC

The premises used above can also be used to provide a basis to explain why cross-linguistically, CV syllables are much more common than VC syllables (Ohala, 1990; Tabain et al., 2004; Nam

et al., 2009). Specifically, VC will have a shorter gestural lag than CV, which is disfavored if the premise that human prefers larger lag is true.

Of course, we also need to account for the fact that VC sequences do exist in many languages. It is likely that the observed relationship between sonority and gestural timing should be seen akin to a force that has a certain effect, keeping all other things constant (Kröger et al., 1995); however, if there is a sufficiently strong antagonistic force that requires VC as a sequence (perhaps as a language-specific segmental sequence), then VC sequences could still surface. Such an analysis makes the prediction that VC is dispreferred, but is still possible under the right circumstances — this correctly predicts the asymmetry in CV and VC sequences within syllables across languages.

5.2.4 Providing a basis for the link among sonority, stress, and vowel height

There are several observations or correlations involving sonority, stress, and vowel height, and the findings of the dissertation may provide explanations for those too. First, the finding of the dissertation allows us to partially understand why lower vowels are acoustically longer and have higher intensity than higher vowels (Lehiste, 1970; Gordon et al., 2012). Following the main claim of the paper, since lower vowels are more sonorous, they will have a larger lag with the preceding consonant. Therefore, there is less overlap with the preceding consonant, and thus there is less acoustic "hiding" of the vowel. Consequently, a low vowel is likely to be acoustically longer and louder than a high vowel.

Second, our proposal leads to the prediction that lower vowels are perceptually and acoustically longer, making them better tone and stress holders than higher vowels. Such cases can be found in many languages. For example, Zuraw (2003) found that in Palauan, more sonorous vowels are dispreferred in unstressed syllables. Similarly, Gordon et al. (2012) found that in Armenian, Javanese, and Kwak'wala, the reduced phonological sonority of schwa relative to peripheral vowels is manifested in the rejection of stress on schwa. This indicates a positive correlation between sonority and stress — reduced sonority correlates to the absence of stress. Moreover, avoidance of stressed high vowels has been observed in Takia (Ross, 2002, 2003; De Lacy, 2007) — while the final syllable is stressed by default, if the final vowel is a high vowel, stress falls on a non-high

vowel elsewhere in the word. Finally, it is also argued that languages where stress seeks out vowels of lower sonority and disregards higher sonority ones are unattested (De Lacy, 2007).

The relationship between sonority difference and lag could be one of the reasons why stress favors high sonority vowels. As observed by Gu (2023), Katsika (2016) and Katsika (2012), stressed vowels are correlated with larger gestural lags than unstressed syllables. Since given the same consonant, vowels high in sonority are correlated with larger gestural lags than vowels low in sonority, this in turn leads to an acoustically and perceptually longer lower vowel, which therefore provides a better holder for stress.[1]

But why are larger lags preferred? Chitoran (2016) extended the argument in Pouplier and Beňuš (2011) and claimed that a larger lag provides a favorable environment for energy peak to emerge. So maybe humans prefer this energy peak environment between two consonants.

5.3 A sonority-driven speech production model

This current study has the potential to support a new speech production model that assumes sonority determines gestural coordination patterns. The assumption of this sonority-based speech production model is that all gestures coordinate according to the sonority differences with the gestures of the adjacent segment within a syllable. Consequently, the findings of the current study should be generalized to CC, CV, and VC sequences in a syllable in all languages, where a positive correlation between sonority difference and gestural lag is predicted for cases beyond CV syllables. It is also possible that this claim even holds true across syllable boundaries or even larger prosodic boundaries. This claim of a sonority-driven speech production model is consistent with the results in Aziz (2024), who interpreted the model as a Sonority-Driven Gestural Timing constraint ranked high in languages such as Malagasy. To argue for this model, admittedly, one needs to also test gestural coordination in coda positions, the CV coordination in syllables with consonant clusters, and various other conditions, which I plan to undertake in future work.

The intent of extending to all languages was not intended to be immodest but to make the claim

[1]There are languages where stress systems are insensitive to sonority (De Lacy, 2007), and the detailed discussion of those phenomena requires substantial amount of careful literature review and experiment. Therefore, they are left for future studies.

testable. If I were to say that such a generalization holds only for the dialect of English studied here, it is not clear how someone else tests our claim beyond simply replicating our results for the relevant dialect, and it would not be clear how it would inform phonetic theory more generally. There is a section 5.4 about caveats that discusses relevant information.

Indeed, there are a lot of exceptions to the SSP, as mentioned in Chapter 1. I want to reiterate that the generalization is about syllables and needs to be judged on segment sequences that are in the same syllable structure context. For example, in English sC sequences form putative exceptions to SSP; however, it has been argued that the [s] is not part of the onset in such cases, and instead forms a foot-level appendix (Vaux and Wolfe, 2009). Similarly, Moroccan Arabic and Jazani Arabic have word-initial consonant sequences; however, they do not violate the SSP as it has been argued that such sequences do not form complex onsets and all but the last consonant are not in the same syllable as the last consonant in such sequences (Goldstein et al., 2007; Shaw et al., 2009, 2011; Hermes et al., 2013, 2017). Claims about the SSP and the SDP are strictly speaking about syllable-internal sequencing, so it is likely that the apparent cross-linguistic exceptions are just an analysis along the lines of the superficial exceptions discussed above.

The sonority-driven speech production model assumes that sonority *causes* gestural coordination variation, and I have discussed how this is likely. It is logically possible that a third factor is causing the sonority differences and gestural lag variation. There is a question about whether abstract sonority or the phonetic correlate of sonority is related to CV lag. I would argue that primarily it is the abstract sonority that leads to a certain gestural coordination pattern. Also, it would appear that the phonetic correlate of sonority systematically relates to gestural lag variation.

5.4 Caveats and directions for future studies

Even though the main claim of the current study has been largely supported by a series of experiments, there are some caveats and future directions that are discussed in this section.

One caveat is that it is unclear if the underlying concept is sonority or a strength hierarchy (Honeybone, 2008) since they are mirror images of each other. At this point, I am unsure of how

to distinguish between the two concepts, given they are inverses of each other.

In the following, I am going to discuss the necessity of more comprehensive cross-linguistic analysis, as well as the nuances of lag measurement and gestural parsing techniques.

5.4.1 Cross-linguistic analysis

Future studies on other languages are necessary to test the cross-linguistic impact of the current finding. For instance, it may be worth replicating the study in German since German onset clusters exhibited gestural coordination variation, but not in the same direction of the dissertation (Bombien et al., 2013). Specifically, Bombien et al. (2013) found that /kn/ has a larger target onset lag than /kl/ and that /ks, ps/ has a larger lag than /kl, pl/. One potential reason for the discrepant results between the current study and that of Bombien et al. (2013) is that their study also explored prosodic effects but the segmental and syllabic material in the context carrier phrases was not controlled across the stimuli, so it is possible that the effect they found was driven by these other characteristics of the stimuli. Another reason for the result difference could be that in the case of /kn/ and /kl/ (and for that matter /ks/), the measurements for both consonants were both based on tongue movement, so it is possible that this led to misparsing of the gestures (as did in our own /kɑ/ and /gɑ/ stimuli). Therefore, it would be optimal to re-run the study on German with these considerations in mind. If the original pattern observed in Bombien et al. (2013) sustains despite the change to the stimuli that I am suggesting, then the sonority hypothesis will be difficult to maintain as is — it will either have to be relativized to just CV sequences or some other independent factor would have to be added to the theory.

5.4.2 The lag measurement

A disclaimer I would like to note is that the current results, without considering C duration in the model, are about *observed* CV lags — that is, CV lag duration was treated as informing us about the lag relationship between CV. However, as pointed out before, there are at least two different sources of observed CV lags: an absolute/constant lag and a proportional lag (Solé, 1992; Mücke et al., 2020; Durvasula and Wang, 2023). The observed lag due to the former is not expected to vary with different durations of the first gesture, but it is expected to vary with different durations

of the first gesture for the latter. To address the potential confound due to measurement choice, consonant duration is considered in the mixed effect models of all three experiments. Therefore, arguably the significant positive correlation between CV lag and sonority difference can be found, regardless we consider absolute lag or proportional lag. However, it is worth mentioning that I did not calculate proportional lag according to the method in Durvasula and Wang (2023), so future work may need to verify whether similar results can be found for proportional lag.

5.4.3 The gestural parsing method

In speech production studies, kinematic data need to be analyzed and gestures need to be parsed in order to answer the relevant question in articulation. There are logically infinite ways of measuring and parsing an articulatory gesture, as discussed in section 1.8 of Chapter 1. However, it is not clear what is the *right* way to parse gestures. The decision is arbitrary, and the decision may directly affect the answers to the research questions. One potential way to evaluate different methods is that the right measurement will bring out consistent observations. But since we are not sure about the expectations of observations in the first place, the process of figuring out the right method involves some circular argumentation, and there is no obvious way around it.

One method future studies could use is to use more than one measurement technique to answer each research question. If several methods lead to the same conclusion, then the conclusion is likely to be true. However, if the conclusion is dependent on methods, then one needs to try to make sense of the differences if possible. Some remaining relevant questions include: whether the observed outcome difference is significant and how to make sense of the difference — is it in the grammar, in the phonetics, or is it trivial?

In the current dissertation, the threshold algorithm *lp_findgest* was used, but there are some issues or ambiguities related to it. First, Shaw et al. (2023) did mention that *lp_findgest* can sometimes yield unrealistic parsing of gestures. For instance, sometimes peak velocity is not large enough to parse out gestures. In some cases, tangential velocities result in inaccurate sums of distinct gestures, where component velocities should be used instead. The question remains about what is a realistic gesture. Should it be a one-to-one mapping to acoustic signals? Should the

gesture be marked based on acoustic output? There is no straightforward answer to these questions, but there should be more attempts to address the questions to push forward our understanding of speech production studies.

Second, the gesture boundaries determined by the threshold technique are sensitive to articulatory stiffness. Liu et al. (2022) argued that the reason why some research concluded that a consonant and a vowel in a CV syllable are coordinated sequentially is that consonants are articulated with higher stiffness than vowels.

Third, there are actually variations in the assumptions of the threshold technique and how the method is applied. Comparing the algorithm of *lp_findgest* and the threshold technique in Hoole et al. (1994) suggests that the gesture identification procedures are different, regardless of where the methods are used. While *lp_findgest* first identifies the nucleus and expands the gesture from the center, the original technique first identifies the maximum velocity points at edges and then the nucleus stage serves as the "bi-product". Therefore, it is incorrect to assume that any threshold technique was operated under the same assumption underlyingly. Furthermore, observing the uses of the threshold technique suggests that there is a variety of scenarios — single articulatory gesture or syllable — to use the method, though the scope of the technique was not explicit.

Moreover, there are variations on the exact threshold used. In previous studies, the 20% threshold was frequently used, probably because it is the default threshold in the *lp_findgest* algorithm. However, the justification for the 20% threshold is missing in many studies. It is possible this threshold was chosen in the *lp_findgest* algorithm based on the previous study in Hoole et al. (1994). It is worth noting though, as mentioned earlier, that the two threshold techniques are not entirely the same, and the 20% was chosen in Hoole et al. (1994) because it generates the expected prediction in terms of German tense-lax vowel durational difference. Note that even though changing the threshold from 10% to 30% does not affect the conclusion in Durvasula and Wang (2023), this does not mean that all thresholds do not affect results in all studies. For instance, Kuberski and Gafos (2023) does show that increasing threshold values result in better performance in linear regression models based on the analysis of thresholds such as 0%, 5%, 10%, 15%, 20%,

and 25%.[2] This suggests that ideally thresholds should be justified.

Fourth, regarding the specific threshold algorithm *lp_findgest*, there are many decisions that are arbitrary in this algorithm, but the arbitrariness of the decision was sometimes missing. One specific decision to make is that when the algorithm identifies "the closest velocity minimum", it needs to specify the range of this search. However, it is unclear what the appropriate range is, and it is not clear how to balance the two constraints — closest and minimum. The challenge of determining an appropriate range also occurs when identifying the "minimal velocity point before PVEL" and "the velocity minimum following PVEL2".

Another decision to make is that — when talking about 20% peak (PVEL) velocity between minimal velocity point and peak velocity point, it is also not clear whether it is the 20% when the velocity is increasing or the 20% when the velocity is decreasing. Also, when talking about the 20% point, it seems that we need to find the velocity point that is *exactly* 20%. A remaining question is how to make the decision when there is not a point being recorded that is exactly 20% of the peak velocity. One option is to use the existing real data to make a prediction of the velocity contour and locate the point that is exactly the 20% point. Alternatively, one could locate a velocity point only according to the existing velocity data collected.

Additionally, in general, when talking about velocity, there is a question about whether to look at the velocity of one axis, multiple axes, or some calculation based on existing information. Here are some possible velocity calculations: a) taking the maximum velocity among the three axes' velocity; b) using the sum of three absolute velocities. Besides velocity, in some EMA systems such as the NDI Vox, there is a 6D data frame output for data collection. Namely, for each timestamp, the system has information for not only x, y, and z positions, but also four data points in terms of quaternion rotation. Algorithms could potentially include quaternion rotation in its calculation as well.

These challenges are for the field. While I have chosen to use the threshold technique in order to have results comparable to most others in the field, and because it has fewer issues than the

[2]Note that this is based on observed lag and does not distinguish between proportional lag and observed lag. One may argue that using proportional lag can lead to consistent results.

comparative technique as mentioned in section 1.8, I leave a more detailed comparison of possible techniques for future work.

CHAPTER 6

CONCLUSION

The dissertation found that there is a positive correlation between gestural lag and sonority difference for Mandarin and English CV syllables. There are a few questions that the current study aims to address. The first question is about gestural coordination in speech production. The current study found that there is a systematic variation in gestural coordination relevant to sonority. The second question is the articulatory correlation of sonority — the current study suggests that sonority may be a fundamental factor in speech production. The third question is the source of phonological constraints — the dissertation suggests that a few phonological universals could come from human beings' general preference toward larger gestural lag.

The current study probed the link between gestural timing and sonority. Based on corpus data and newly collected EMA data, I found a positive correlation between gestural lag and sonority difference in English CV syllables. This finding provides a basis for typological universals — SSP, SDP, and CV-VC syllable frequency asymmetry — if one adds another premise that larger gestural lags are preferred. It also can account for the correlation among sonority, stress, and vowel height. A potential explanation of the finding is available if we consider intensity as the phonetic correlate of sonority — larger directional sonority differences of adjacent sounds require larger gestural lag to ensure perceptual recoverability. The current study provides empirical evidence that suggests the necessity of revamping the current speech production model. In general, the dissertation provides evidence and claims that can help us form a more nuanced understanding of speech production.

BIBLIOGRAPHY

Albert, A. (2023). *A model of sonority based on pitch intelligibility*. BoD–Books on Demand.

Arnqvist, G. (2020). Mixed models offer no freedom from degrees of freedom. *Trends in ecology & evolution*, 35(4):329–335.

Aziz, J. (2024). *The Phonetics and Phonology of So-Called Vowel Devoicing in Malagasy*. PhD thesis, UCLA.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Benguerel, A.-P. and Cowan, H. A. (1974). Coarticulation of upper lip protrusion in french. *Phonetica*, 30(1):41–55.

Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3):591–630.

Blackwood Ximenes, A., Shaw, J. A., and Carignan, C. (2017). A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from american and australian english. *The Journal of the Acoustical Society of America*, 142(1):363–377.

Bladon, R. A. W. and Al-Bamerni, A. (1976). Coarticulation resistance in english/l. *Journal of Phonetics*, 4(2):137–150.

Blevins, J. (1995). The syllable in phonological theory. In *Handbook of phonological theory*, pages 206–244. Blackwell.

Bombien, L., Mooshammer, C., and Hoole, P. (2013). Articulatory coordination in word-initial clusters of german. *Journal of Phonetics*, 41(6):546–561.

Browman, C. P. and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2):201–251.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Browman, C. P. and Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*, (5):25–34.

Browman, C. P., Goldstein, L., et al. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 1:341–397.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of phonetics*, 24(2):209–244.

Byrd, D. and Krivokapić, J. (2021). Cracking prosody in articulatory phonology. *Annual Review of Linguistics*, 7:31–53.

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2):149–180.

Byrd, D. M. (1994). *Articulatory timing in English consonant sequences*. University of California, Los Angeles.

Chao, Y.-R. (1930). A system of tone letters. *Le maître phonétique*, 30:24–30.

Cheng, R., Jongman, A., and Sereno, J. A. (2023). Production and perception evidence of a merger:[l] and [n] in fuzhou min. *Language and Speech*, 66(3):533–563.

Chitoran, I. (2016). Relating the sonority hierarchy to articulatory timing patterns: A cross-linguistic perspective. *Challenging sonority: Cross-linguistic evidence*, pages 45–62.

Chitoran, I., Goldstein, L., and Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from georgian. *Laboratory phonology*, 7(4-1):419–447.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in english. *Laboratory phonology*, 8:519–548.

Cho, Y.-m. Y. and King, T. H. (2003). Semisyllables and universal syllabification. *The syllable in optimality theory*, pages 183–212.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology: Volume 1, Between the Grammar and Physics of Speech*, 1:283.

Clements, G. N. (2005). Does sonority have a phonetic basis? comments on the chapter by vaux. 14 pp. In *In Raimy, E. & Cairns, C.(Eds.), Contemporary Views on Architecture and Representations in Phonological Theory*. Citeseer.

Clements, G. N. (2009). Does sonority have a phonetic basis. *Contemporary views on architecture and representations in phonology*, 48:165.

Collier, R., Bell-Berti, F., and Raphael, L. J. (1982). Some acoustic and physiological observations on diphthongs. *Language and Speech*, 25(4):305–323.

Crouch, C. (2022). *Postcards from the syllable edge: sonority and articulatory timing in complex onsets in Georgian*. PhD thesis, UC Santa Barbara.

Crouch, C., Katsika, A., and Chitoran, I. (2023). Sonority sequencing and its relationship to articulatory timing in georgian. *Journal of the International Phonetic Association*, pages 1–24.

Davis, S. and Shin, S.-H. (1999). The syllable contact constraint in korean: An optimality-theoretic analysis. *Journal of East Asian Linguistics*, 8(4):285–312.

De Lacy, P. (2007). The interaction of tone, sonority, and prosodic structure. *The Cambridge handbook of phonology*, 281:281–308.

Dell, F. and Elmedlaoui, M. (1985). Syllabic consonants and syllabification in imdlawn tashlhiyt berber.

Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, 27(4):761–764.

Du, S. and Gafos, A. I. (2023). Articulatory overlap as a function of stiffness in german, english and spanish word-initial stop-lateral clusters. *Laboratory Phonology*, 14(1).

Durvasula, K. (2024). Lecture 8 handout of laboratory phonology. Unpublished lecture notes.

Durvasula, K., Ruthan, M. Q., Heidenreich, S., and Lin, Y.-H. (2021). Probing syllable structure through acoustic measurements: case studies on american english and jazani arabic. *Phonology*, 38(2):173–202.

Durvasula, K. and Wang, Y. (2023). Revisiting cv timing with a new technique to identify inter-gestural proportional timing. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, pages 2284–2288.

Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and speech*, 36(2-3):171–195.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural language & linguistic theory*, 20(2):269–337.

Gafos, A. I., Hoole, P., Roon, K., Zeroual, C., Fougeron, C., Kühnert, B., D'Imperio, M., and Vallée, N. (2010). Variation in overlap and phonological grammar in moroccan arabic clusters. *Laboratory phonology*, 10:657–698.

Gao, M. (2008). *Mandarin tones: An articulatory phonology account*. PhD thesis, Yale University.

Gelfer, C. E., Bell-Berti, F., and Harris, K. S. (1989). Determining the extent of coarticulation: Effects of experimental design. *The Journal of the Acoustical Society of America*, 86(6):2443–2445.

Gelman, A. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge

university press.

Gibson, M., Sotiropoulou, S., Tobin, S., and Gafos, A. (2017). On some temporal properties of spanish consonant-liquid and consonant-rhotic clusters. *Proceedings of the 13th Tagung Phonetik und Phonologie im deutschsprachigen Raum (PP13)*, pages 73–76.

Gibson, M., Sotiropoulou, S., Tobin, S., and Gafos, A. I. (2019). Temporal aspects of word initial single consonants and consonants in clusters in spanish. *Phonetica*, 76(6):448–478.

Goldstein, L. (2011). Back to the past tense in english. *Representing language: Essays in honor of Judith Aissen*, pages 69–88.

Goldstein, L., Byrd, D., and Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, pages 215–249.

Goldstein, L., Chitoran, I., and Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from georgian vs. tashlhiyt berber. In *Proceedings of the XVIth international congress of phonetic sciences*, pages 241–244. Saarbrücken Univ. des Saarlandes Saarbrücken, Germany.

Gordon, M., Ghushchyan, E., McDonnell, B., Rosenblum, D., and Shaw, P. A. (2012). Sonority and central vowels: A cross-linguistic phonetic study. *The sonority controversy*, pages 219–256.

Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech, Language, and Hearing Research*, 37(1):4–27.

Gracco, V. L. and Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*, 14(11):6585–6597.

Greenberg, J. H. (1965). Some generalizations concerning initial and final consonant sequences. *Linguistics*, 3(18):5–34.

Gu, Y. (2023). Exploring the effect of stress on gestural coordination. *Proceedings of the Linguistic Society of America*, 8(1):5539.

Hall, N. (2010). Articulatory phonology. *Language and Linguistics Compass*, 4(9):818–830.

Hall, T. A. (2002). Against extrasyllabic consonants in german and english. *Phonology*, 19(1):33–75.

Hankamer, J. and Aissen, J. (1974). The sonority hierarchy. *Papers from the parasession on natural phonology. Chicago: Chicago Linguistic Society*, 11.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during/kl/sequences. *Speech Communication*, 4(1-3):247–263.

Harrison, X. A. (2015). A comparison of observation-level random effect and beta-binomial models for modelling overdispersion in binomial data in ecology & evolution. *PeerJ*, 3:e1114.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., Robinson, B. S., Hodgson, D. J., and Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6:e4794.

Henke, E., Kaisse, E. M., and Wright, R. (2012). Is the sonority sequencing principle an epiphenomenon. *The sonority controversy*, 18:65–100.

Hermes, A., Mücke, D., and Auris, B. (2017). The variability of syllable patterns in tashlhiyt berber and polish. *Journal of Phonetics*, 64:127–144.

Hermes, A., Mücke, D., and Grice, M. (2013). Gestural coordination of italian word-initial clusters: the case of 'impure s'. *Phonology*, 30(1):1–25.

Honeybone, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. *Lenition and fortition*, pages 9–93.

Hoole, P., Bombien, L., Kühnert, B., and Mooshammer, C. (2009). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters.

Hoole, P., Mooshammer, C., and Tillmann, H. G. (1994). Kinematic analysis of vowel production in german. In *Third international conference on spoken language processing*.

Hooper, J. B. and Bybee, J. L. (1976). *An introduction to natural generative phonology*. Academic Press.

Hsieh, F.-Y. (2017). *A gestural approach to the phonological representation of English diphthongs*. PhD thesis, University of Southern California.

Iskarous, K. and Pouplier, M. (2022). As time goes by: A critical appraisal of space and time in articulatory phonology in the 21st century. *USC online articles*.

Iverson, G. K. and Salmons, J. C. (1995). Aspiration and laryngeal representation in germanic. *Phonology*, 12(3):369–396.

Jespersen, O. (1904). Lehrbuch der phonetik (leipzig and berlin). *G. Teubner*.

Johnson, K. and Song, Y. (2016). Gradient phonemic contrast in nanjing mandarin. *UC Berkeley Phonlab Annual Report*, 12(1).

Kang, Y., van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K. (2011). *The Blackwell companion to phonology*.

Katseff, S., Houde, J., and Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and speech*, 55(2):295–308.

Katsika, A. (2012). *Coordination of prosodic gestures at boundaries in Greek*. Yale University.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in greek. *Journal of phonetics*, 55:149–181.

Kenstowicz, M. J. (1994). *Phonology in generative grammar*, volume 7. Blackwell Cambridge, MA.

Kent, R. D. and Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of phonetics*, 5(2):115–133.

Kéry, M. and Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Kreitman, R. (2010). Mixed voicing word-initial onset clusters. *Laboratory phonology*, 10(4):4.

Krivokapić, J. (2020). Prosody in articulatory phonology. *Prosodic theory and practice*.

Kröger, B. J., Schröder, G., and Opgen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *The Journal of the Acoustical Society of America*, 98(4):1878–1889.

Kroos, C., Hoole, P., Kühnert, B., and Tillmann, H. G. (1996). Phonetic evidence for the phonological status of the tense-lax distinction in german. *Journal of the Acoustical Society of America*, 100:2691.

Kuberski, S. R. and Gafos, A. I. (2023). How thresholding in segmentation affects the regression performance of the linear model. *JASA Express Letters*, 3(9).

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, 82:1–26.

Ladefoged, P. and Johnson, K. (2014). *A course in phonetics*. Cengage learning.

Lehiste, I. (1970). Suprasegmentals, cambridge, massachusetts & london, uk.

Liu, Z., Xu, Y., and Hsieh, F. (2020). Coarticulation as synchronised sequential target approximation: An ema study. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, volume 2020, pages 1381–1385. International Speech Communication Association (ISCA).

Liu, Z., Xu, Y., and Hsieh, F.-f. (2022). Coarticulation as synchronised cv co-onset–parallel

evidence from articulation and acoustics. *Journal of Phonetics*, 90:101116.

Luo, S. (2017). Gestural overlap across word boundaries: Evidence from english and mandarin speakers. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 62(1):56–83.

MacNeilage, P. F. and Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465):527–531.

Marin, S. and Goldstein, L. (2012). A gestural model of the temporal organization of vowel clusters in romanian. *Consonant Clusters and Structural Complexity. Berlin/Boston, De Gruyter*, pages 177–203.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In *Advances in Psychology*, volume 7, pages 415–420. Elsevier.

Mielke, J. (2008). *The emergence of distinctive features*. Oxford University Press.

Mooshammer, C. and Fuchs, S. (2002). Stress distinction in german: simulating kinematic parameters of tongue-tip gestures. *Journal of Phonetics*, 30(3):337–355.

Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P. H., and Fuchs, S. (2003). Coordination of lingual and mandibular gestures for different manners of articulation. In *Proceedings of the 15th International Congress of Phonetic Sciences, August 3-9, 2003, Barcelona, Spain*, number 1, pages 81–84. International Phonetic Association.

Mücke, D., Hermes, A., and Tilsen, S. (2020). Incongruencies between phonological theory and phonetic measurement. *Phonology*, 37(1):133–170.

Mücke, D., Nam, H., Hermes, A., and Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. *Consonant clusters and structural complexity*, 26:205.

Murray, R. W. and Vennemann, T. (1983). Sound change and syllable structure in germanic phonology. *Language*, pages 514–528.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. *Laboratory phonology*, 9:483–506.

Nam, H., Goldstein, L., and Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to phonological complexity*, 16:299–328.

Nam, H. and Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences*, volume 1.

Ohala, J. J. (1990). Alternatives to the sonority heirarchy. In *Papers from the 26th Regional Meeting of the Chicago Linguistics Society*, volume 2.

Ohala, J. J. and Kawasaki, H. (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Language and its ecology: Essays in memory of Einar Haugen*, 100:343.

Öhman, S. E. (1966). Coarticulation in vcv utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1):151–168.

Parker, S. (2002). *Quantifying the sonority hierarchy*. PhD thesis, University of Massachusetts at Amherst.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of phonetics*, 36(1):55–90.

Parker, S. (2011). Sonority. *The Blackwell companion to phonology*, pages 1–25.

Parker, S. (2012). *The sonority controversy*, volume 18. Walter de Gruyter.

Pike, K. L. (1972). Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds.

Pons-Moll, C. (2008). The sonority scale: categorical or gradient. In *Poster presented at the CUNY Conference on the Syllable*.

Pouplier, M. (2020). Articulatory phonology. In *Oxford research encyclopedia of linguistics*.

Pouplier, M. and Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from slovak. *Laboratory phonology*, 2(2):243–273.

Pouplier, M., Pastätter, M., Hoole, P., Marin, S., Chitoran, I., Lentz, T. O., and Kochetov, A. (2022). Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics*, 93:101153.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Redford, M. A. (1999). The mandibular cycle and reversed-sonority onset clusters in russian. In *Proceedings from the 14th International Congress of Phonetic Sciences*, pages 1893–1896.

Ross, M. (2002). Takia. the oceanic languages ed. by john lynch, malcolm ross & terry crowley, 216-248.

Ross, M. (2003). Seminar on takia, a papuanised austronesian language of papua new guinea.

Saltzman, E. L. and Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4):333–382.

Selkirk, E. (1984). On the major class features and syllable theory. *Language sound structure*.

Seo, M. (2011). Syllable contact. *The Blackwell companion to phonology*, pages 1–18.

Shariatmadari, D. (2006). Sounds difficult? why phonological theory needs 'ease of articulation'. *School of Oriental and African Studies Working Papers in Linguistics*, 14:207–226.

Shaw, J. and Chen, W.-r. (2019). Spatially conditioned speech timing: evidence and implications. *Frontiers in Psychology*, 10:2726.

Shaw, J., Gafos, A. I., Hoole, P., and Zeroual, C. (2009). Syllabification in moroccan arabic: evidence from patterns of temporal stability in articulation. *Phonology*, 26(1):187–215.

Shaw, J., Kawahara, S., and Shaw, J. A. (2023). Limits on gestural reorganization following vowel deletion: The case of tokyo japanese. *Laboratory Phonology*, 14(1).

Shaw, J., Oh, S., Durvasula, K., and Kochetov, A. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology*, 38(3):437–477.

Shaw, J. A., Gafos, A. I., Hoole, P., and Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of moroccan arabic consonant clusters. *Phonology*, 28(3):455–490.

Shi, X. (2015). 成都话响音的鼻化度——兼论其/n, l/不分的实质及类型 [the nasality degree of sonorants in chengdu dialect]. 中国语音学报. *Chinese Journal of Phonetics*, 10:92–100.

Sievers, E. (1881). Grundzüge der phonetik: Breitkopf und hartel.

Sievers, E. (1901). *Grundzüge der Phonetik: zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*, volume 1. Breitkopf & Härtel.

Smolensky, P. (1995). On the structure of the constraint component con of ug. *Handout of the talk presented at UCLA*.

Solé, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and speech*, 35(1-2):29–43.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. PhD thesis, Massachusetts Institute of Technology.

Svensson Lundmark, M., Frid, J., Ambrazaitis, G., and Schötz, S. (2021). Word-initial consonant–vowel coordination in a lexical pitch-accent language. *Phonetica*, 78(5-6):515–569.

Tabain, M., Breen, G., and Butcher, A. (2004). Vc vs. cv syllables: a comparison of aboriginal languages with english. *Journal of the International Phonetic Association*, 34(2):175–200.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Tiede, M. (2005). Mview: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

Tilsen, S. (2020). Detecting anticipatory information in speech with signal chopping. *Journal of Phonetics*, 82:100996.

Vaux, B. and Wolfe, A. (2009). 5 the appendix. *Contemporary views on architecture and representations in phonology*, 48:101.

Westbury, J., Milenkovic, P., Weismer, G., and Kent, R. (1990). X-ray microbeam speech production database. *The Journal of the Acoustical Society of America*, 88(S1):S56–S56.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of open source software*, 4(43):1686.

Wright, R. (2004). A review of perceptual cues and cue robustness. *Phonetically based phonology*, 34:57.

Wright, R. A. (1996). *Consonant clusters and cue preservation in Tsou*. University of California, Los Angeles.

Xhaferaj, A. et al. (2022). The sonority dispersion principle in albanian. *European Journal of Social Science Education and Research*, 9(1):40–47.

Xu, Y., Liu, F., et al. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, 18(1):125.

Yanagawa, M. (2006). *Articulatory timing in first and second language: a cross-linguistic study*. Yale University.

Yin, R., van de Weijer, J., and Round, E. R. (2023). Frequent violation of the sonority sequencing principle in hundreds of languages: how often and by which sequences? *Linguistic Typology*, 27(2):381–403.

Zhang, M., Geissler, C., and Shaw, J. (2019). Gestural representations of tone in mandarin: evidence from timing alternations. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, pages 1803–1807. Australasian Speech Science and Technology Association Inc Canberra, ACT.

Zhang, W. and Levis, J. M. (2021). The southwestern mandarin/n/-/l/merger: effects on production in standard mandarin and english. *Frontiers in Communication*, 6:639390.

Zuraw, K. (2003). Vowel reduction in palauan reduplicants. In *Proceedings of the 8th Annual Meeting of the Austronesian Formal Linguistics Association [AFLA 8]*, pages 385–398.

APPENDIX A

ENGLISH RECRUITMENT EMAIL

Subject: Interested in being part of a speech production study?

Hello,

We are conducting a speech production experiment on native speakers of English/Mandarin/Japanese/Spanish. The experiment will involve the participant producing sentences in their native language while they are connected to an Electromagnetic Articulograph machine that will track their speech articulations through sensors. From the project, we hope to learn how speakers coordinate articulatory events (gestures) during speech production.

An Electromagnetic Articulograph allows researchers to measure the positions of parts of the mouth as they are moved during speech articulation. The machine is connected to sensors that will be placed both on the face/lips and on the tongue of the participant for the duration of the experiment. The sensors will be affixed in place with a standard dental adhesive. Note, long-term exposure to the electromagnetic fields of the Electromagnetic Articulograph machine has not been shown to be harmful to human health, but it is recommended to avoid subjects who are pregnant or who utilize pacemakers. Guidelines place the limit for safe continuous exposure between 100μT and 200μT. The interested participant can read more about the technology at this link: https://en.wikipedia.org/wiki/Electromagnetic_articulography.

A prospective participant will be a healthy individual with no history of hearing or speech deficiencies who is at least 18 years old, is not pregnant and does not use a pacemaker.

The experiment will last at most 2 hours and you will be paid $30 for your participation.

If you are interested, fill out this pre-screening survey, and we will contact you if you meet our criteria.

Sincerely,

Yunting

APPENDIX B

ENGLISH PRE-SCREENING SURVEY

Google form title: Linguistics experiment (speech production)

1. A prospective participant will be a healthy individual with no history of hearing or speech deficiencies who is at least 18 years old, is not pregnant and does not use a pacemaker.

    - Yes, I have read the above text.
    - No, I haven't read the above text.

2. Do you have a history of hearing or speech deficiencies?

    - Yes.
    - No.
    - Prefer not to answer.

3. Are you at least 18 years old?

    - Yes.
    - No.
    - Prefer not to answer.

4. What's your preferred email address? (We will contact you if you pass the pre-screening.)

5. What is your first/primary language? (Multiple-choice question)

    - English
    - Mandarin
    - Spanish
    - Japanese
    - Other ___

6. What state did you grow up in if you grew up in the US? (Answer NA if you did not grow up in the US.)

APPENDIX C

MANDARIN RECRUITMENT MESSAGE

有兴趣参加语言学实验吗?

您好，我们正在进行一项关于英语/普通话/日语/西班牙语母语者的语言学实验。在该实验中，参与者将用他们的母语说出一些句子，同时他们将与一台电磁发音仪(Electromagnetic Articulography；EMA) 相连，该仪器可通过传感器跟踪他们的发音。我们希望通过这个项目了解说话者在发音过程中如何协调发声。研究人员通过电磁发音仪测量口部各部分在发音过程中的位置。该机器连接有传感器。整场实验中，这些传感器都将通过标准牙科粘合剂固定在参与者的脸部/嘴唇和舌头上。需要注意的是，尽管持续暴露于电磁发音仪的电磁场对人类健康未显示出有害影响，我们仍建议孕妇或使用心脏起搏器者避免参与实验。实验规范显示：连续暴露于磁场强度100 μT至200 μT之间是安全的。受试者如有兴趣，可以通过此链接了解更多该技术的相关信息：https://en.wikipedia.org/wiki/Electromagnetic_articulography。

符合要求的实验参与者应是健康的，没有听力或言语缺陷的记录，年龄至少为18岁，不在孕期，并且不使用心脏起搏器。实验最长将进行2小时，您将获得30美元的参与费用。

如果您感兴趣，请填写此表: https://forms.gle/NPSEwv9V9xj9N3V46。如果符合我们的条件，我们将与您联系。

谢谢!

密歇根州立大学语音实验室

APPENDIX D

MANDARIN PRE-SCREENING SURVEY

语言学实验调查

1. 参与试验者必须是健康的，没有听力或言语缺陷的，18岁或18岁以上的，没有怀孕的，不使用心脏起搏器的成年人。
    - 是的，我已阅读以上信息。
    - 不，我没阅读以上信息。
2. 您是18岁或18岁以上吗?
    - 是
    - 否
    - 无法提供信息
3. 您有听力或言语缺陷记录吗?
    - 有
    - 没有
    - 无法提供信息
4. 您的电子邮箱地址是？我们会在初筛结束之后联系您。
5. 您的母语是？（如有多种母语，请全部填写。）
6. 您会说哪些语言/方言?
7. 您是否在中国出生和长大?
    - 是
    - 否

8. 您在哪个省份/城市出生？哪个省份/城市长大?

9. 您在中国生活过几年?

APPENDIX E

ANNOTATION LABELS AND THEIR MEANINGS IN EXPERIMENT 2

	Label	Meaning
1	"Questionable"	I am unsure about the annotation.
2	"MultipleMeasure"	More than one gesture is measured for a sound. Example: TB and TD were measured.
3	"NoneDefault"	The non-default gesture is used. Example: TT was measured for vowel.
4	"SensorUnavailable"	The target sensor is unavailable.
5	"Mispronounced"	

Table E.1 Annotation labels and their meanings in experiment 2: the English EMA experiment.

APPENDIX F

ANNOTATION LABELS AND THEIR MEANINGS IN EXPERIMENT 3

	Label	Meaning
1	"Questionable"	I am unsure about the annotation.
2	"MultipleMeasure"	More than one gesture is measured.
3	"NoneDefault"	The non-default gesture is used.
4	"SensorUnavailable"	The target sensor is unavailable.
5	"UnclearPro"	Unclear pronunciation.
6	"NaLamispron"	There is a [n, l] merger. Example: 腊 (la) is pronounced as 那 (na).

Table F.1 Annotation labels and their meanings in experiment 3: the Mandarin EMA experiment.

APPENDIX G

ENGLISH EXPERIMENTS RESULTS WITH VOWEL DISPLACEMENT AS FIXED EFFECT

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	159.69	23.19	34.58	6.89	<0.00001
Sonority difference	10.96	1.70	23.88	6.46	<0.00001
Vowel displacement	-0.45	0.70	1710.92	-0.65	0.52

Table G.1 Mixed effect model for all stimuli with sonority difference and vowel displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	111.58	26.61	21.28	4.19	0.0004
Sonority difference	14.07	1.92	10.57	7.34	0.00002
Vowel displacement	4.05	0.99	1233.65	4.11	0.00004

Table G.2 Mixed effect model for bilabial C stimuli with sonority difference and vowel displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	415.92	132.53	187.41	3.14	0.00
Sonority difference	-8.29	8.73	180.30	-0.95	0.34
Vowel displacement	-8.17	2.87	158.34	-2.85	0.005

Table G.3 Mixed effect model for *peak, pack* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	173.50	108.30	171.41	1.60	0.11
Sonority difference	7.34	8.70	161.65	0.84	0.40
Vowel displacement	5.29	2.90	171.37	1.83	0.07

Table G.4 Mixed effect model for *beak, back* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	72.02	80.61	260.30	0.89	0.37
Sonority difference	13.98	8.620	255.34	1.62	0.11
Vowel displacement	4.75	2.86	229.61	1.66	0.10

Table G.5 Mixed effect model for *meek, Mac* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-23.66	38.13	68.16	-0.62	0.54
Sonority difference	34.53	6.98	224.40	4.95	0.000001
Vowel displacement	10.70	1.76	217.54	6.09	<0.00001

Table G.6 Mixed effect model for *week, whack* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	183.60	41.69	12.13	4.40	0.001
Sonority difference	6.74	3.19	10.15	2.11	0.06
Vowel displacement	-6.70	1.12	1177.57	-6.01	<0.00001

Table G.7 Mixed effect model for coronal C stimuli with sonority difference and vowel displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-527.73	240.84	264.38	-2.19	0.03
Sonority difference	56.36	16.79	264.45	3.36	0.001
Vowel displacement	-1.93	3.30	205.60	-0.59	0.56

Table G.8 Mixed effect model for *two, toe* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-185.42	205.99	258.89	-0.90	0.37
Sonority difference	33.80	16.78	256.49	2.02	0.05
Vowel displacement	-17.18	4.02	256.98	-4.28	0.00003

Table G.9 Mixed effect model for *sue, so* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	321.94	198.28	205.77	1.62	0.11
Sonority difference	-4.24	17.47	202.99	-0.24	0.81
Vowel displacement	-9.05	3.20	197.04	-2.83	0.01

Table G.10 Mixed effect model for *do, doe* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-66.85	132.19	261.62	-0.51	0.61
Sonority difference	34.19	15.63	257.73	2.19	0.03
Vowel displacement	-8.22	2.30	251.65	-3.57	0.0004

Table G.11 Mixed effect model for *new, know* with sonority difference and vowel displacement as fixed effects. Random intercepts: participants and consonant duration.

APPENDIX H

ENGLISH EXPERIMENT RESULTS WITH CONSONANT DISPLACEMENT AS FIXED EFFECT

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	137.07	28.00	16.97	4.90	0.0001
Sonority difference	12.22	2.23	10.47	5.49	0.0002
C displacement	-0.76	0.89	1434.78	-0.86	0.39

Table H.1 Mixed effect model for bilabial C stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	178.57	65.18	202.80	2.74	0.01
Sonority difference	8.50	4.87	204.82	1.75	0.08
C displacement	2.88	2.11	172.16	1.37	0.17

Table H.2 Mixed effect model for *peak, beak* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	208.80	71.22	198.13	2.93	0.003
Sonority difference	11.68	5.00	186.08	2.33	0.02
C displacement	6.24	2.57	170.60	2.43	0.02

Table H.3 Mixed effect model for *pain, bane* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	108.33	83.25	162.65	1.30	0.20
Sonority difference	12.87	5.50	149.96	2.34	0.02
C displacement	4.62	2.81	159.42	1.65	0.10

Table H.4 Mixed effect model for *pack, back* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	181.35	43.99	11.76	4.12	0.001
Sonority difference	8.33	3.38	10.01	2.46	0.03
C displacement	-1.63	2.33	1431.03	-0.70	0.48

Table H.5 Mixed effect model for coronal C stimuli with sonority difference and consonant displacement as fixed effects. Random intercepts: stimuli, participants, and consonant duration.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	318.42	72.36	248.90	4.40	0.00002
Sonority difference	-3.97	5.65	244.01	-0.70	0.48
C displacement	-15.75	8.02	251.99	-1.97	0.05

Table H.6 Mixed effect model for *two, do* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	214.86	56.22	265.08	3.82	0.0002
Sonority difference	7.20	3.99	271.23	1.80	0.07
C displacement	-13.63	6.30	278.76	-2.16	0.03

Table H.7 Mixed effect model for *toe, doe* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	295.35	57.40	234.82	5.15	<0.00001
Sonority difference	-0.17	3.73	274.04	-0.05	0.96
C displacement	-1.12	4.95	277.74	-0.23	0.82

Table H.8 Mixed effect model for *talk, dock* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	11.21	52.02	251.54	0.22	0.83
Sonority difference	23.95	5.29	249.89	4.53	<0.00001
C displacement	-13.15	6.58	259.00	-2.00	0.05

Table H.9 Mixed effect model for *do, new* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	158.88	50.78	234.77	3.13	0.002
Sonority difference	12.27	4.54	263.26	2.70	0.01
C displacement	0.54	6.40	271.89	0.09	0.93

Table H.10 Mixed effect model for *doe, know* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	67.29	44.39	223.92	1.52	0.13
Sonority difference	17.52	3.61	273.96	4.86	<0.00001
C displacement	2.79	4.91	279.36	0.57	0.57

Table H.11 Mixed effect model for *dock, knock* with sonority difference and consonant displacement as fixed effects. Random intercepts: participants.

APPENDIX I

MANDARIN RESULTS FOR PAIRWISE COMPARISON DIFFER IN C VOICING

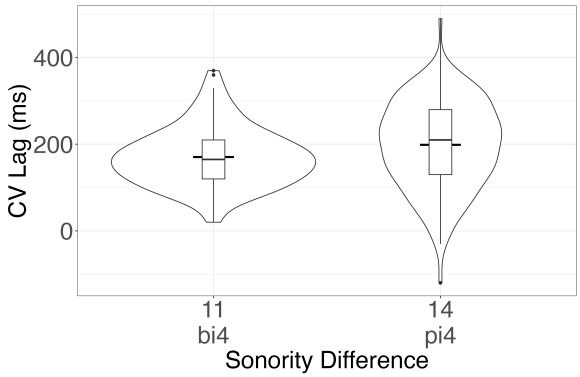

Figure I.1 CV lag based on target onset for Mandarin participants: *bi4, pi4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	54.04	40.86	268.60	1.32	0.19
Sonority difference	10.57	3.15	270.81	3.35	0.001

Table I.1 Mixed effects model results for Mandarin participants: *bi4, pi4*.

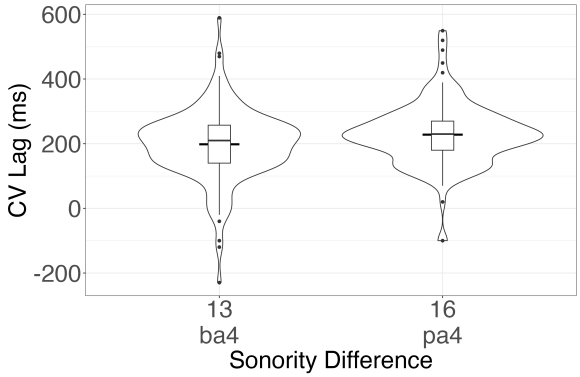

Figure I.2 CV lag based on target onset for Mandarin participants: *ba4, pa4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	73.45	58.80	269.06	1.25	0.21
Sonority difference	9.69	3.99	259.96	2.43	0.02

Table I.2 Mixed effects model results for Mandarin participants: *ba4, pa4*.

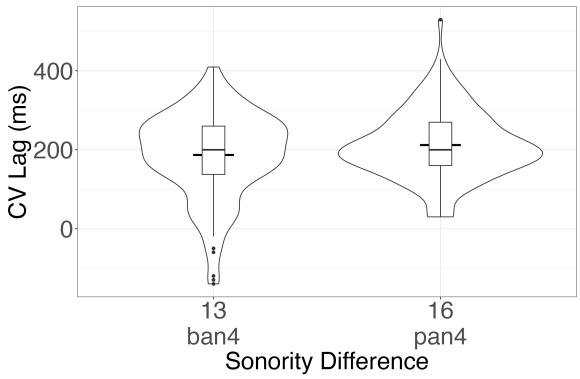

Figure I.3 CV lag based on target onset for Mandarin participants: *ban4, pan4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	82.40	55.82	269.80	1.48	0.14
Sonority difference	8.11	3.75	262.96	2.16	0.03

Table I.3 Mixed effects model results for Mandarin participants: *ban4, pan4*.

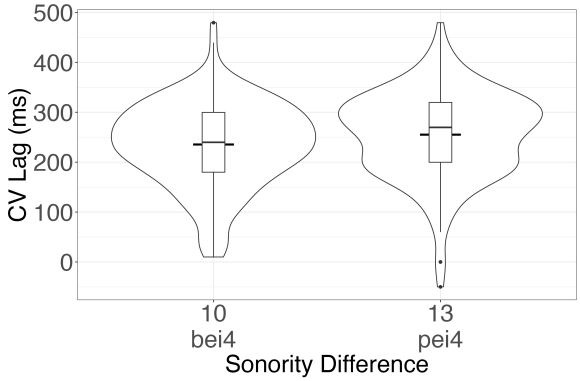

Figure I.4 CV lag based on target onset for Mandarin participants: *bei4, pei4*.

Gestural Onset	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	165.91	36.18	189.85	4.59	0.00001
Sonority difference	6.95	2.90	282.52	2.40	0.02

Table I.4 Mixed effects model results for Mandarin participants: *bei4, pei4*.