2-2022

Incomplete Neutralization in Articulatory Phonology

Sejin Oh

Incomplete Neutralization in Articulatory Phonology

by

Sejin Oh

A dissertation submitted to the Graduate Faculty in Linguistics
in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2022

Incomplete Neutralization in Articulatory Phonology

by

Sejin Oh

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Jason Bishop
	Chair of Examining Committee

Date	Cecelia Cutler
	Executive Officer

Supervisory Committee:

Douglas Whalen

Christina Hagedorn

Jason Shaw

Wei-Rong Chen

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Incomplete Neutralization in Articulatory Phonology

by

Sejin Oh

Advisor: Jason Bishop

Previous studies have found small but significant phonetic traces of underlying distinctions for phonologically “neutralized” contrasts. This phenomenon, often called incomplete neutralization, has been found for final devoicing in many languages, (e.g., German; Port, Robert F. & O’Dell, 1985), but has also been reported for other neutralizing phenomena, including flapping in American English (Herd et al., 2010), monomoraic lengthening in Japanese (Braver & Kawahara, 2016), vowel deletion in French (Fougeron & Steriade, 1997), vowel epenthesis in Levantine Arabic (Gouskova & Hall, 2009), among others.

In my dissertation, I explore the (in)completeness of Russian palatalization in the Articulatory Phonology framework, implementing gestural coordination of complex segments and segment sequences. In Russian, the contrast between a palatalized consonant (e.g., /lʲ/) and a plain consonant (e.g., /l/) is reported to be neutralized to the palatal counterpart when a plain consonant is followed by a glide. That is, the palatalization of the plain stop in the environment preceding palatal glides results in neutralization: e.g., /lʲut/ [l̠ʲut] ‘fierce’ (underlyingly palatalization) vs. /ljut/ [l̠ʲjut] ‘pour (3p pl)’(coarticulatory palatalization). However, given that “plain” consonants possibly feature a secondary articulation involving the retraction of the tongue dorsum (velarization/uvularization, see Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963), this

dissertation tests the hypothesis that the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) leads to the incomplete neutralization of underlying and coarticulatory palatalization in Russian.

To this end, this dissertation will explore how complete the neutralization is between underlyingly palatalized consonants and coarticulatorily palatalized consonants (underlyingly plain). In so doing, I will first quantify the extent of palatalization by investigating temporal coordination in both complex segments and segment sequences in Russian and English. I will then present Electromagnetic Articulography (EMA) experiments that examine temporal coordination and spatial positions of articulators involving both underlyingly and coarticulatorily palatalized consonants in Russian. I will also present simulations from computational modeling that can be tested against EMA recordings.

In the first experiment, evidence from articulatory kinematic data collected with EMA on Russian palatalized consonants and English consonant-glide sequences revealed that gestural coordination for complex segments (Russian) differs from segment sequences (English). Specifically, the Russian data is consistent with the hypothesis that the constituent gestures of complex segments are coordinated according to their gesture onsets, showing no correlation between G_1 duration and onset lag. In contrast, the English data exhibits a positive correlation between G_1 duration and onset lag, suggesting that G_2 is timed to some gestural landmark later in the unfolding of G_1.

Results from a second EMA experiment regarding incomplete neutralization of Russian palatalization also reveal that the palatal-plain contrast is neutralized, but more importantly, this neutralization is phonetically incomplete. In particular, both types of palatalizations exhibit the temporal coordination of complex segments, suggesting that plain consonants in the coarticulatory

palatalization context are also palatalized. However, I also find residual evidence of an underlying tongue dorsum retraction for the coarticulatory palatalization. This is in line with previous findings of Russian plain consonants having secondary velarization. The computational simulations show that gestural blending of palatalization and velarization as well as their eccentric timing in coarticulatory palatalization results in incomplete neutralization of underlying and coarticulatory palatalization in Russian.

This dissertation provides new insights for interpreting incomplete neutralization in the AP framework by showing that at least some cases of incomplete neutralization can be accounted for by gestural overlap. The results present substantial potential for the gestural overlap account to be generalized across a wide range of incomplete neutralization, including final devoicing. This dissertation is important both for the analysis of Russian palatalization and for discussion on incomplete neutralization, as well as articulatory phonology more generally.

To Eddie

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my advisor, Jason Bishop, for his invaluable supervision, support, and tutelage during the course of my PhD degree. I still remember our first meeting, where Jason laid out his 5-year plan for my entire PhD education. He spent hours talking with me about the projects we would do together, the training I would receive from him and at the GC, which classes I should take, and so forth. It was in that moment that I could see the kind of advisor he was then, is now, and will continue to be, even after completion of my PhD. My intuition was that choosing this path would make me the luckiest PhD student in the world, in no small part because I had the most amazing advisor in the world - thankfully, I turned out to be right about that! Jason was always readily available to help me regardless of the time of day or night. It is not only his effervescent personality that makes Jason such a joy to work with; his phenomenal insight and expertise, as well as his thorough advice have been fundamental sources for me to grow as a researcher, teacher, and colleague. Without his support and encouragement, I could not have made it this far.

I have another Jason to whom I owe a deep sense of gratitude: Jason Shaw. At first, it was an incredible coincidence that my dissertation topic overlapped with one of his projects. However, he graciously extended me an offer to collaborate with him and his research team, and I decided to move to New Haven to work closely with him. This collaboration with Jason was one of the best things that ever happened during the course of my PhD! My dissertation as well as my personal growth as a researcher have tremendously benefited from his invaluable advice in every single step of the EMA experiments and computational modeling, in addition to benefiting from his incredible expertise in phonology. My dissertation would not have reached this level without his invaluable advice. His amicable personality made this collaboration even more enjoyable. Also,

I would like to express my gratitude to the other two members of the superb research team on complex segments, Alexei Kochetov at Toronto, and Karthik Durvasula at Michigan State. My dissertation has also extensively benefited from Alexei's knowledge of Russian and Karthik's expertise in phonology. I am so thankful that I had the opportunity to collaborate with these outstanding linguists and to learn the importance of teamwork on great research.

I am deeply indebted to Douglas Whalen who introduced me to the fascinating field of speech production and Articulatory Phonology. He generously opened the opportunity for me to work at Haskins, and I feel extremely privileged that I am part of his research group. Also, I would like to express gratitude to Christina Hagedorn for her insightful feedback, which drove me to sharpen my thinking and brought my work to a higher level. I am also thankful to Wei-Rong Chen, who generously and patiently devoted a great deal of time to train me in collecting and analyzing EMA data. It is my great fortune to have these wonderful mentors. Outside of my dissertation committee, I would like to express my gratitude to Juliette Blevins, who opened my eyes to the diachronic perspective of phonological grammar. I am also grateful for her emotional support when I felt lost. I am also thankful to other faculty members at the GC including Janet Fodor, Jason Kandybowicz, and Samer Al Khatib, who encouraged me to explore other fields in linguistics beyond phonetics and phonology.

I am also deeply indebted to Sun-Ah Jun for her support and guidance during my time at UCLA. I have learned a lot from collaborative works with her on intonational phonology of Korean. I would like to express my gratitude to members of the phonology reading group at Yale for their fruitful discussion and constructive feedback. I am especially thankful to Christopher Geissler and Vivian Guo Li for their help with my dissertation data collection as well. I would also like to thank my colleagues from the Haskins Laboratories, as well as the SPAPL Lab at the GC including Mark

Tiede, Kevin Roon, Rion Iwasaki, Jeehyun Yoon, Richard Lissemore, Grace Lambert-Kim, Ghada Shejaeya, Boram Kim, and Jaekoo Kang.

I would like to acknowledge that a portion of Chapter 2 was published in Phonology (See the following citation for the published version: Shaw, J., Oh, S., Durvasula, K., & Kochetov, A. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology, 38*(3), 437-477. doi:10.1017/S0952675721000269). I would like to thank the three anonymous reviewers and the associate editor of Phonology, as well as audiences at AMP 2020 and ISSP 2020 for their comments and discussion. I also feel especially grateful to the speakers who participated in the EMA experiments.

My thanks also go to my friends at the CUNY Graduate Center including Yeonju Lee-Sikka, Chen Zhou, Katherine Howitt, Ekaterina Kistanova, and João Marinotti, as well as my former roommates, Chaeeun Lee and Kyungmin Lee. It is their kind help and support that made my PhD life in New York a wonderful time. I am also thankful to my friends from Chung-Ang University - Jinny Kim, Sangrae Cho, Hankyul Kim, and Emily Kim. I feel especially grateful to my best friends, Yumi Hyun, Soyeong Ahn, Heeseo Song, Goeun Jeong, Junkyu Park, and Minhye Park. Their friendship has been a light to guide me through even the roughest seas.

I am eternally grateful to my mother, Kyungok Kim, and my father, Heungki Oh, for their unwavering support and belief in me. They have been there through my highs and lows, but always made me feel like I could soar. I am also deeply indebted to my grandmother, Inhui Son, who also raised me with her tremendous love. She passed away during the pandemic, but even so, I know that she is looking down at me proudly from heaven. I would like to express my thanks to my brother, Jeryung Oh, my sister-in-law, Yumi Kim, my lovely nephews, Taeseung Oh and Taeyul Oh, and my aunt, Kyongsuk Lynch. Also, I would like to express my gratitude to Dr. Eva Turner and Jacquie Wolf for their endless love and support. I am so blessed to have them in my life as my mothers-in-law and my best friends. I am also thankful to my sweet father-in-law, Dr. Lawrence Forgacs, and my cool sister-in-law, Gina Turner. I also appreciate all the support I received from the rest of my family.

Finally, I would like to express my deepest gratitude to my sweet husband Edward Forgacs, who has tirelessly reviewed countless versions of all of my work, in addition to providing me with moral support throughout the entirety of my doctoral degree. He has always been by my side during difficult times and has traveled across seas and mountains to be with me. Whenever I was in a situation that made me feel like I could cry, it was Eddie that made me smile. Without his tremendous love, understanding, and encouragement, it would have been impossible for me to complete my PhD successfully.

CONTENTS

TABLES

FIGURES

Page

Chapter 1. Introduction

1.0. Introduction to the dissertation

Previous studies have found small but significant phonetic traces of underlying contrasts in phonologically “neutralizing” positions. This phenomenon, often called incomplete neutralization, has been found in final devoicing in many languages (e.g., Bulgarian: Bishop et al., 2019; Russian: Dmitrieva et al., 2010; German: Port & O’Dell, 1985; Polish: Slowiaczek & Dinnsen, 1985), flapping in American English (Herd et al., 2010), vowel epenthesis in Levantine Arabic (Gouskova & Hall, 2009), among other patterns.

Russian contrasts palatalized and plain consonants (so-called “soft” and “hard” consonants, respectively), as shown in (1) (e.g., Avanesov, 1972; Kochetov, 2004; 2006; Padgett, 2001; 2003; Timberlake, 2004). However, even a plain consonant exhibits palatalization when it is followed by a palatal glide, leading to neutralization of the contrast in this context (e.g., Kochetov, 2011; 2013). As shown in (2), for example, the contrast (e.g., /p̲ʲok/ vs. /p̲jot/) is neutralized due to the palatalization of the plain stop in consonant-glide sequences.

(1)	Palatalized consonants	Plain consonants
	/pʲok/ [p̲ʲok] ‘bake (3ps past)’	/pot/ [p̲ot] ‘sweat’
	/bʲust/ [bʲust] ‘bust’	/but/ [b̲ut] ‘booth’

(2)	Palatalized consonants	Plain consonants preceding a palatal glide
	/pʲok/ [p̲ʲok] ‘bake (3ps past)’	/p̲jot/ [p̲ʲjot] ‘drink (3ps pres)’
	/bʲust/ [bʲust] ‘bust’	/bjut/ [bʲjut] ‘beat (3ps pl)’

Interestingly, however, previous studies have sometimes reported that the “plain” stops may actually feature a secondary articulation involving retraction of the tongue dorsum (velarization/uvularization, see Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963). However, while the palatal gesture is understood to exist underlyingly in palatalized consonants (/Cʲ/), its presence in consonant-glide sequences (/Cj/) is derived from the upcoming glide. A question that arises from consideration of these patterns is whether the neutralization between plain and palatalized segments in Russian is phonetically (i.e., acoustically and/or articulatorily) complete.

However, when it comes to assessing the completeness of this neutralizing process, there is the question of how to quantify whether a “plain” consonant preceding a palatal glide is palatalized or not. At first glance, it might seem straightforward − if it is palatalized, the realization would be palatalization of a “plain” consonant e.g., [pʲj] for /pj/, otherwise just a plain consonant without palatalization e.g., [pj]. However, how do we distinguish one from the other, when both involve multiple articulatory gestures, namely a closure of the lips (for [p]) and a movement of the tongue body (for [j])? Moreover, how do we distinguish complex segments such as palatalized segments (e.g., [pʲ]) and segment sequences, such as plain-glide sequences (e.g., [pj]), where both also involve the same gestures?

Even though this distinction has been the focus of much work in phonology and phonetics, there is still no consensus regarding the phonetic properties or the phonological representation that distinguish complex segments from segment sequences. For example, Herbert (1986) and Riehl (2008) have argued that duration is the key factor distinguishing complex segments from segment sequences (namely, that there are longer acoustic durations for segment sequences), while Maddieson and Ladefoged (1993) have argued that any such durational differences are too inconsistent to serve this purpose.

An alternative approach is to focus on the actual articulatory movements of complex segments and segment sequences. Shaw, Durvasula, and Kochetov (2019) recently proposed that complex segments and segment sequences are different in terms of the temporal coordination of the gestures involved. Specifically, they have shown that complex segments feature temporally coordinated onsets, while the gestures of segment sequences show that the offset of G_1 is temporally coordinated with the onset of G_2 (See Section 2.3 for more discussion). Strikingly, however, previous work has only examined complex segments and segment sequences consisting of different components, such as [p^j] vs. [br], which makes it difficult to draw conclusions about fine distinctions in temporal coordination.

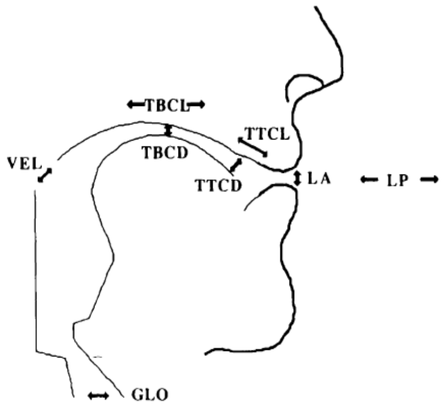
To this end, this dissertation first investigates temporal coordination in complex segments and segment sequences in Russian and English. As described above, differences in temporal patterns are believed to be an important part of the puzzle to how these sound patterns are represented, and yet they remain poorly understood. This dissertation then explores the incompleteness of Russian palatalization in the Articulatory Phonology framework, implementing gestural coordination of complex segments and segment sequences. The current dissertation, therefore, addresses pressing issues in laboratory phonology generally, as well as proper analysis of the temporal organization and coordination of speech units. The remainder of this chapter is structured as follows: Section 1.1 – 1.2 provides a theoretical background of Articulatory Phonology and of incomplete neutralization. In Section 1.3, I discuss Russian palatalization as a putative case of incomplete neutralization. Section 1.4 provides questions and the outline of the chapters of this dissertation.

1.1. Articulatory phonology

In the Articulatory Phonology (henceforth, AP) framework, gestures, which are abstract representations of movement of articulators in the vocal tract, serve as the primitive phonological units (e.g., Browman & Goldstein, 1986; 1989; 1992; 1995a; Pouplier, 2020). Speech can be organized into constellations of gestures, and each gesture is defined as an event which forms and releases a constriction in the vocal tract. Crucially, gestures are specified spatially as well as temporally. As discussed more below, this framework modeled by a dynamical system is different from classical phonological representations, in that it bridges abstract phonological representations and continuous physical movement.

1.1.1. The representation of gestures

Gestures are discrete and abstract in the sense that they are specifically defined by a set of dynamical parameters which characterize each gesture distinctively (e.g., Browman & Goldstein, 1986; 1989; 1992; 1995a; Pouplier, 2020). As shown in Figure 1, gestures are specified with respect to vocal tract variables. AP utilizes a set of gestural descriptors which distinguish contrastive gestures: Constriction degree (CD), constriction location (CL), and stiffness (k). Tract variable goals (input values for CD and CL) determine the inherent spatial aspect, while the stiffness specifies the intrinsic temporal aspect of each gesture. Values for the possible descriptor are shown in (3) (adopted from Browman & Goldstein, 1989, p. 207). For example, /s/ and /ʃ/ differ in their values for CL (alveolar vs. postalveolar, respectively), and /s/ and /t/ differ in their values for CD (critical vs. closed).



Tract variable		**Articulators involved**
LP	Lip protrusion	Upper & lower lips, jaw
LA	Lip aperture	Upper & lower lips, jaw
TTCL	TT constrict location	Tongue tip, body, jaw
TTCD	TT constrict degree	Tongue tip, body, jaw
TBCL	TB constrict location	Tongue body, jaw
TBCD	TB constrict degree	Tongue body, jaw
VEL	Velic aperture	Velum
GLO	Glottal aperture	Glottis

Figure 1: Tract variables (adopted from Browman & Goldstein, 1989, p. 207)

(3) CD descriptors: closed, critical, narrow, mid, wide
CL descriptors: protruded, labial, dental, alveolar, postalveolar, palatal, velar, uvular, pharyngeal

1.1.2. Gestural score

The spatiotemporal activation of gestures can be displayed in a gestural score with spatial information (specifications for tract variables) on the vertical axis and temporal information on the horizontal axis (e.g., Browman & Goldstein, 1989; 1992). For example, a gestural score for the word ‘pen’ shows the input values for CD and CL of each gesture as well as their intergestural timing. For example, as shown in Figure 2, there are gestures associated with /p/ at the beginning: a closure gesture of the lips and a wide glottal constriction. The TB gesture for /ɛ/ also starts at the beginning of the utterance overlapping with the gestures associated with [p]. The final consonant /n/ also has two gestures: a tongue tip closure and a velic opening, which also overlap with the preceding vowel gesture. The overlap between the velic opening and the vowel gesture leads to partial nasalization of the vowel.

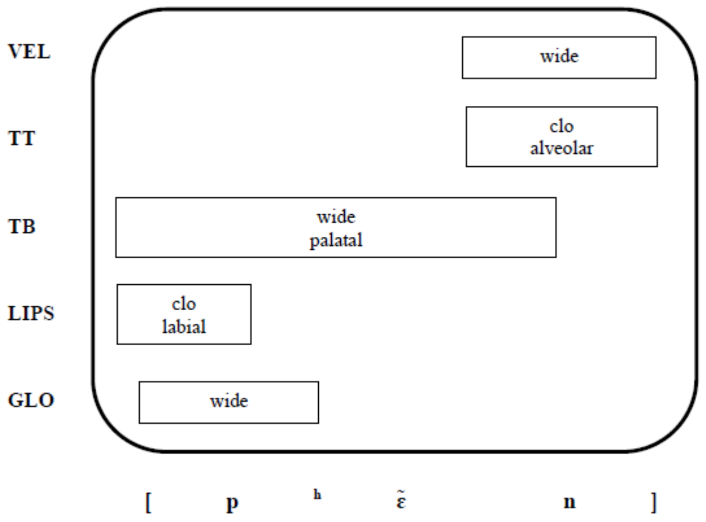


Figure 2: A gestural score for ‘pen’

Browman and Goldstein (1989) proposed that phonological phenomena such as deletion, insertion, assimilation, and weakening can be captured by two general processes: ‘hiding’ and ‘blending’ of gestures. When gestures significantly overlap on the different articulatory tiers, one gesture may hide the other acoustically, despite both gestures still being present articulatorily. For example, the deletion of /t/ in ‘perfect memory’ at a fast speech rate is better described as gestural hiding (Tiede et al., 2001). As shown in Figure 3, the alveolar gesture for /t/ completely overlaps with the preceding velar gesture for /k/ and the following labial gesture for /m/, resulting in the hidden acoustic consequence of /t/.

On the other hand, when two gestures overlap on the same articulatory tier, they compete with each other to achieve their own goals. This kind of overlap may lead to ‘blending’ of the dynamical parameters of these gestures. The gestural outcome of blending is different from that of either of the individual gestures. Instead, the outcome falls somewhere in between the two gestures, the extent of which depends on the strength of each gesture. For example, the place assimilation of /n/ in ‘ten themes’ at a fast speech rate is better described as a gestural blending. As shown in

Figure 4, the gestures for /n/ and /θ/ overlap on the same TT tract variables, resulting in gestural blending between /n/ and /θ/.

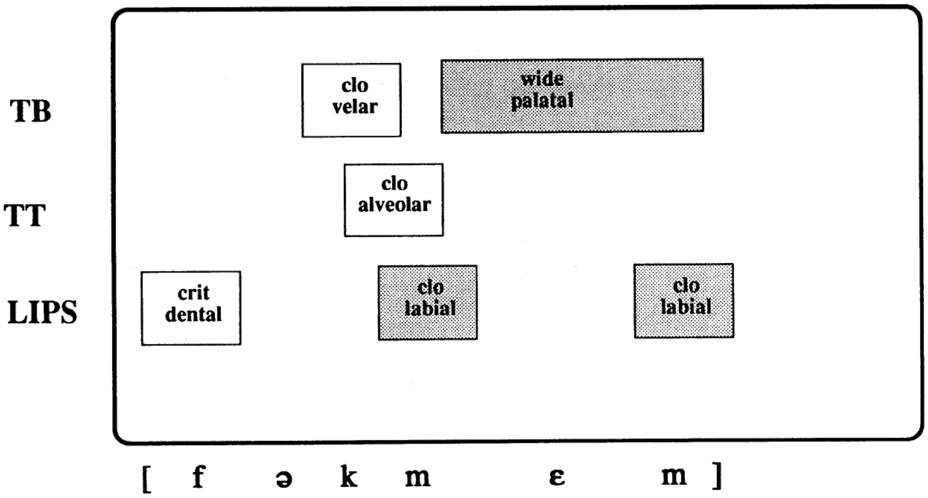


Figure 3: The partial gestural score for ‘perfect memory.’ The last syllable of ‘perfect’ and the first syllable of ‘memory’ are shown in unshaded and shaded boxes, respectively (adopted from Browman & Goldstein, 1989, p. 216)[1]

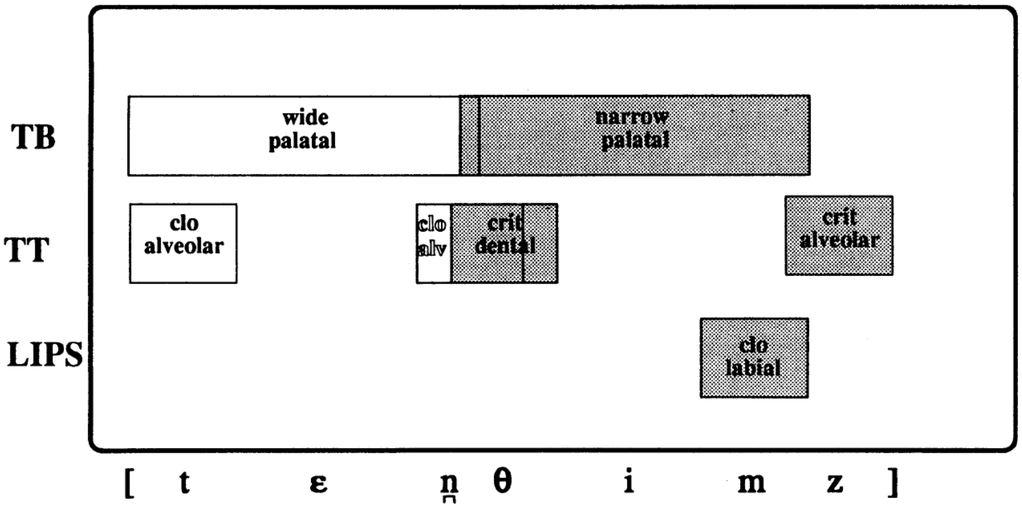


Figure 4: A gestural score for ‘ten themes.’ ‘ten’ and ‘themes’ are shown in unshaded and shaded boxes, respectively (adopted from Browman & Goldstein, 1989, p. 218)

---

[1] Schwa is not shown in the gestural score, as it is modeled as a targetless vowel.

An important advantage of AP is that it allows phonologists to directly test their hypotheses using articulatory data. Another advantage is that it eliminates the need for “rule” implementation, which other theories rely upon. Later in the dissertation, these advantages are shown to be particularly crucial to understanding the nature of Russian palatalization patterns. I will also present simulations from computational modeling that can be tested against Electromagnetic Articulography (EMA) recordings.

1.2. Incomplete neutralization

1.2.1. Final devoicing

Previous studies have found small but significant phonetic traces of underlying contrasts for phonologically “neutralized” contrasts. This phenomenon, often called incomplete neutralization, has been found for final devoicing in many languages. In the case of final devoicing, the voicing contrast is preserved in word-initial and word-medial positions. However, in the word-final position, both underlying voiced and underlying voiceless obstruents surface as voiceless. In German, for example, the voicing contrast of the alveolar stops in (4) are neutralized in word-final positions, while the contrast is preserved in word-medial positions as shown in (4).

(4) Examples of final devoicing in German

Rat [ʁa:**t**] *(‘council’)*	Räte [ʁæ:**tə**] *(‘councils’)*
Rad [ʁa:**t**] (*‘wheel’*)	Räder [ʁæ:**d**ɐ] (*‘wheels’*)

However, previous studies have provided considerable evidence that such phonological neutralization is phonetically incomplete in German (e.g., O'Dell & Port, 1983; Port, Robert & Crawford, 1989; Roettger et al., 2014), as well many other languages, such as Catalan (e.g., Charles-Luce & Dinnsen, 1987), Dutch (e.g., Warner et al., 2004), Polish (e.g., Slowiaczek & Dinnsen, 1985), and Russian (e.g., Dmitrieva, 2005; Dmitrieva et al., 2010; Kharlamov, 2012; 2014). Previous studies have shown that there are small acoustic and articulatory differences between underlying voiced and voiceless obstruents, and such phonetic differences surface in the direction expected for the underlying form. More specifically, the underlyingly voiced obstruents tend to have shorter final stop closure durations, a shorter release burst, a longer preceding vowel, and/or more extensive voicing into closure than the underlying voiceless obstruents (Charles-Luce & Dinnsen, 1987; Dinnsen & Charles-Luce, 1984; Mascaró, 1987 for Catalan; Ernestus & Baayen, 2007; Warner et al., 2004 for Dutch; O'Dell & Port, 1983; Port & Crawford, 1989; Roettger et al., 2014 for German; Port & Crawford, 1989; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985 for Polish; Dmitrieva et al., 2010; Kharlamov, 2012; 2014 for Russian). In German, for example, Port and O'Dell (1985) observed significant differences on all four of these parameters, while, in Russian, Kharlamov (2014) reported shorter consonantal duration and more extensive voicing into closure in voiced obstruents. These differences, though statistically significant, tend to be very small in magnitude, however – generally on the order of 10–20 milliseconds at most. Perception studies have shown that listeners can perceive even the small phonetic differences between underlying voiced and voiceless obstruents with above-chance accuracy (Ernestus & Baayen, 2009; Kleber et al., 2010; Port & Crawford, 1989; Port & O'Dell, 1985; Roettger et al., 2014; Warner et al., 2004).

1.2.2. Phonological accounts

Incomplete neutralization has been the focus of much work in phonology as well, since it is difficult to incorporate incomplete neutralization into models of the grammar (e.g., Braver, 2019;

Brockhaus, 1995; Dinnsen & Charles-Luce, 1984; Ernestus & Baayen, 2009; Piroth & Janker, 2004; Port & O'Dell, 1985; Roettger et al., 2014; Van Oostendorp, 2008). For example, in serial/rule-based models, a phonological rule has to be applied before low-level phonetic implementation rules. However, if final devoicing (the phonological rule) is applied first, it is impossible to apply the phonetic implementation rules, since the voicing contrast has already been neutralized. To solve this issue, previous studies have proposed a number of solutions.

For example, it has been claimed that phonetic implementation rules may apply before or simultaneously with phonological rules (Dinnsen & Charles-Luce, 1984; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985). However, previous studies have also questioned whether it is even necessary to posit different abstract representations, or different rule orderings, at all, given the small size of the phonetic effects in question. For example, two opposing views that have been proposed to account for incomplete neutralization both do so without posting changes to rule ordering. On the one hand, functionalists have argued that incomplete neutralization is driven by hypercorrection of orthographic cues in the service of communicative goals (e.g., Fourakis & Iverson, 1984; Iverson & Salmons, 2011; Jassem & Richter, 1989; Mascaró, 1987; Piroth & Janker, 2004; Warner et al., 2004). That is, due to the existence of minimal pairs and their different graphemes for underlying voiced and voiceless obstruents, speakers often modulate their speech to signal the underlying voicing contrast for functional utility. However, recent studies have shown that incomplete neutralization can be induced without any clear functional motivation and without the influence of orthography, suggesting that incomplete neutralization is not derived, at least not solely, from such communication-related pressures (e.g., Ernestus & Baayen, 2009; Roettger et al., 2014). Previous studies proposed that incomplete neutralization reflects fine-grained phonetic information in the mental lexicon (Alegre & Gordon, 1999; Baayen et al., 1997; Bybee, 1994;

1995; 2001; Ernestus & Baayen, 2009; Roettger et al., 2014). Ernestus and Baayen (2009) proposed that even though both underlying voiced and voiceless obstruents have the same exemplar representation in the mental lexicon as voiceless[2], the co-activation of paradigmatically related words induces the subtle phonetic differences between underlying voiced and voiceless obstruents. In addition, previous studies have found incomplete neutralization for nonce words, suggesting that existing words with similar spellings are co-activated, leading to incomplete neutralization for nonce words (e.g., Ernestus & Baayen, 2009; Roettger et al., 2014).

My dissertation aims to provide a solution to this long-standing issue by showing that at least some cases of incomplete neutralization can be modeled as gestural overlap in the AP framework.

1.2.3. Other cases of incomplete neutralization

Even though the majority of work on incomplete neutralization heavily focuses on final devoicing, the phenomenon is not restricted to final devoicing. Other patterns that have long been described as neutralization have also turned out to be cases of incomplete neutralization: flapping in American English (e.g., Herd et al., 2010), coda aspiration in Eastern Andalusian Spanish (e.g., Gerfen, 2002), monomoraic lengthening in Japanese (e.g., Braver & Kawahara, 2016), vowel epenthesis in Levantine Arabic (e.g., Gouskova & Hall, 2009), vowel deletion in French (e.g., Fougeron & Steriade, 1997), blended vowels in Romanian (e.g., Marin, 2012), etc.

---

[2] Ernestus and Baayen (2009)'s analysis is fundamentally different from one that assumes abstract underlying forms, abstract rules and derived surface forms, as they assume exemplar representations. In exemplar representations, the representation of final stops, e.g., German, can always be voiceless, since exemplar representations are based on episodic memory of surface forms.

As with final devoicing, there are small but significant acoustic differences between underlying voiced and voiceless obstruents in the case of flapping in American English. Specifically, the durations of preceding vowels for underlying /d/ have been found to be significantly longer than those for underlying /t/, but the difference in duration is generally very small, consistent with the many cases of final devoicing that have been reported (4 ms for Braver & Kawahara, 2016; 6 ms for Herd et al., 2010; 16 ms for Patterson & Connine, 2001; 9 ms for Sharf, 1962).

Another case of incomplete neutralization is coda aspiration in Eastern Andalusian Spanish (EAS). In EAS, obstruents are realized as aspiration/breathy voicing in the syllable-final position, resulting in neutralization of coda obstruents. For example, /as.ta/ "until", /ap.ta/ "apt", and /ak.ta/ "certificate" are all neutralized into [a^{h}t:a]. However, previous studies reported that the underlying /s/ showed a longer aspiration duration than the other underlying obstruents /p/ and /k/, supporting the incomplete neutralization of the coda obstruents in EAS (Bishop, 2007; Gerfen & Hall, 2001; Gerfen, 2002).

Previous discussions on incomplete neutralization have also been expanded to vowels and suprasegmental contrasts. In Lebanese Arabic, the epenthetic vowel /i/ is inserted into final CC clusters, and it is assumed to be identical with the lexical vowel /i/ on the surface in Lebanese phonology (e.g., Abdul-Karim, 1980). However, Gouskova and Hall (2009) have shown that there are small acoustic differences between the epenthetic and lexical /i/ in Lebanese Arabic. In particular, the results from eight Lebanese speakers revealed that the epenthetic /i/ showed a significantly shorter duration and lower F2 than lexical /i/, with some inter-speaker variation.

Braver and Kawahara (2016) and Braver (2013) examined incomplete neutralization of monomoraic vowel lengthening in Japanese. Vowels in Japanese exhibit a length contrast (e.g.,

/t͡ɕi/ ‘blood’ vs. /t͡ɕi:/ ‘social status’). However, a vowel in a monomoraic noun undergoes lengthening when it is not followed by a particle, but no lengthening occurs when it is (e.g. [t͡ɕi:] ‘blood’ vs. [t͡ɕi mo] ‘blood-also’). However, the results of these studies show lengthened monomoraic vowels to still be significantly shorter than underlyingly long vowels, with durational differences ranging from 26.55 ms to 32.47 ms. Such differences are relatively larger than the differences reported in most other incomplete neutralization cases, which, as described above, are typically 10–20 ms at most. Braver (2013:163) argues that “incomplete neutralization is not a homogenous process, but, rather, consists of a continuum from almost completely neutralized (and imperceptible) to relatively less neutralized (and perceptible).”

Previous studies have also found that the neutralization of certain phonological contrasts is both articulatorily and acoustically incomplete. In French, the deletion of schwa results in a consonant cluster. For example, the /dr/ cluster in ‘d'rôle’ [drol] is driven by the deletion of schwa in ‘de rôle’ [dərol] "of role." However, French also has the same consonant cluster /dr/, such as in the word ‘drôle’ [drol] "funny." Fougeron and Steriade (1997) examined whether the clusters driven by the deletion of schwa are completely merged with the underlying consonant clusters. They collected electropalatographic and acoustic data from two French speakers and two types of consonant sequences: [dr] (20 repetitions) and [kl] (10 repetitions). The results revealed that [d] in ‘d'rôle’ showed significantly larger linguopalatal contact, a longer lingual occlusion gesture, and less frequency of lenition than the [d] in the underlying consonant cluster (‘drôle’). [kl] clusters showed no significant differences between the two conditions regarding linguopalatal contact or the acoustic duration. However, the results from the inter-gestural timing analysis revealed that the clusters driven by the deletion of schwa showed more overlap between [k] and [l] than the underlying [kl] cluster.

Marin (2012) proposed a production model to examine the incomplete neutralization of Romanian vowels. In Romanian, the vowel /e/ alternates with diphthong /ea/ (derived /e/), and acoustic analysis from Marin (2012) revealed that for Romanian vowels, the derived /e/ is significantly more central than the vowel /e/ that is underlyingly /e/ (underived /e/). She hypothesized that it might be attributable to different production mechanisms between derived and underived /e/, and tested her hypothesis by comparing acoustic data to modeled stimuli. Using an articulatory based synthesizer, TADA (Task Dynamic Application), the underived /e/ was modeled with the gestural specifications of a single gesture /e/, while the derived /e/ was modeled as a 'blending' of two gestures /e/ and /a/, reflecting its underlying diphthong, /ea/. The results revealed that the blending of two gestures /e/ and /a/ showed similar acoustic properties to naturally produced derived /e/, and modeled stimuli for underived /e/ was also similar to naturally produced underived /e/.

These studies suggest that incomplete neutralization is not restricted to final devoicing. An application of phonological rules may result in very similar outcomes of two phonological entities at the surface level. Crucially, however, if there is an underlying contrast between them, the outcomes they produce may not be identical. This leads to the question of whether this can be applied to other phonological phenomena in general. The current study aims to examine this using underlying and coarticulatory palatalization in Russian as a test case, in which the underlying contrast between plain and palatalized consonants is assumed to be neutralized.

1.3. Russian palatalization as incomplete neutralization

As discussed in Section 1.0, the contrast between palatalized and plain consonants in Russian is neutralized when a plain consonant is followed by a glide (e.g., Kochetov, 2011).[3] That is, both a palatalized consonant (underlying palatalization) and a consonant preceding a palatal glide (coarticulatory palatalization) are realized phonetically as a palatalized consonant. For example, a plain consonant preceding a palatal glide in /p̲jot/ is realized as a palatalized consonant [pʲ], resulting in neutralization of the contrast between plain and palatalized consonants in Russian, as shown in (5).

(5)	Palatalized consonants (Underlying palatalization)	Plain consonants preceding a palatal glide (Coarticulatory palatalization)
	/pʲok/ [p̲ʲok] ‘bake (3ps past)’	/p̲jot/ [p̲ʲjot] ‘drink (3ps pres)’
	/bʲust/ [bʲust] ‘bust’	/bjut/ [bʲjut] ‘beat (3p pl)’

In addition to the difference regarding underlying vs. coarticulatory palatalization between palatalized segments and plain consonants that precede a palatal glide, previous studies have also reported that these “plain” stops possibly have a secondary articulation involving retraction of the tongue dorsum (velarization/uvularization, see Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963). For example, a recent ultrasound study by Roon and Whalen (2019) confirmed that plain consonants in Russian are velarized (and/or uvularized) with intra-speaker variation. In particular,

---

[3] Some C+j sequences are morphologically derived (e.g., /pj-a-n-ij/ from /pʲi-tʲ/ ‘to drink’, via /i/-gliding), others are underlying (e.g., /djakon/ and /rjanij/), at least synchronically. Sequences can occur morpheme-internally (as in the above examples above) and across morphemes (prefix + stem and stem + suffix; e.g., /s-jom-k-a/, /brat-ja/) or words (preposition + stem; e.g. /s jamoj/ ‘with a pit’). C1 in tautomorphemic and stem + suffix sequences is nondistinctively palatalized (unless it is unpaired with respect to palatalization, e.g., ʂjot ‘to knit (3rd sg)’). Palatalization is described as applying variably across prefix + stem boundaries and being absent across preposition (clitic) + stem boundaries, as well as across prosodic words (Avanesov, 1972; Timberlake, 2004).

articulatory data from 3 Russian native speakers revealed that there are consistent and discernable dorsal gestures regardless of the manner and syllable positions (initial vs. final) at least within labials [p, f, m], but the location of constriction varied by speakers (velar to uvular). These patterns raise the question of whether the underlying palatalization and coarticulatory palatalization in Russian are phonetically identical. As described further below, the present dissertation will pursue the hypothesis that gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) leads to incomplete neutralization of the underlying palatalization and coarticulatory palatalization in Russian.

Apart from the discussion of whether the contrast between palatalized and plain consonants is neutralized or not, the consonant-glide sequences themselves are not necessarily identical to the palatalized consonants due to the existence of a glide as a separate segment in the coarticulatory palatalization. Indeed, since there is a phonological contrast in Russian between /C^j/ and /Cj/, there must be a perceivable difference between these forms. In the articulatory kinematics, the palatal gesture in /Cj/ is known to be longer than in /C^j/ (Kochetov, 2006), a duration difference that may support the contrast. Acoustic studies of Russian have shown differences that are consistent with this observation about the kinematics. For example, Diehm (1998) reported that consonant-glide sequences (Cj) exhibit significantly higher F2 values at the transition onset and significantly longer F2 steady-state duration than palatalized consonants (C^j). In addition, Suh and Hwang (2016) also found that the vocalic duration comprising the j+V portion of CjV syllables is significantly longer than the j+V portion of C^jV syllables. These acoustic results confirm that there is a salient acoustic cue to the difference between complex segments and corresponding segment sequences. However, the acoustic differences between consonant-glide sequences (Cj) and palatalized consonants (C^j) do not provide any substantive information as to whether a “plain” consonant preceding a palatal glide is palatalized or not. With regard to evaluating the incomplete neutralization, the quantification of palatalization (or lack thereof) of the consonants preceding the palatal glide has yet to be determined.

Such a quantification of palatalization in Russian might be achieved by examining temporal coordination for complex segments and segment sequences proposed by Shaw et al. (2019). They hypothesized that complex segments have a temporal basis—two articulatory gestures, G1 and G2, that belong to the same complex segment if the onset of G2 is temporally coordinated with the onset of G1. In contrast, two gestures belong to sequences of segments if the onset of G2 is temporally coordinated with the offset of G1. These competing coordination relations were explored by investigating how the lag between the onset of G1 and the onset of G2 varied with G1 duration. The key finding involved differences between English consonant-glide sequences, e.g., [bj], [mj], [pj], [vj], and Russian palatalized labials, e.g., [pʲ], and segment sequences, [br]. The Russian segment sequences and the English stop-glide sequences patterned together — as consonant duration increased, so too did the lag between consonant and glide gestures. Russian palatalized consonants were different. For palatalized consonants, variation in duration had no effect on lag, which is consistent with the hypothesized temporal basis for complex segments.

Strikingly, however, these studies have only examined complex segments and corresponding segment sequences in two different languages, or complex segments and segment sequences consisting of different components in the same language, such as [pʲ] vs. [br]. This makes it difficult to draw conclusions about fine distinctions in temporal coordination.

1.4. Questions and organization of the dissertation

As mentioned above, the goal of the dissertation is to explore the (in)completeness of Russian palatalization in the Articulatory Phonology framework, implementing gestural coordination of complex segments and segment sequences. The main two questions of this dissertation are as follows:

1. Is a “plain” consonant preceding /j/ palatalized, leading to neutralization between the underlying palatalization and coarticulatory palatalization in Russian?
2. If so, is the neutralization between underlying and coarticulatory palatalization in Russian incomplete?

In Chapter 2, I quantify Russian palatalization by examining temporal coordination in complex segments versus segment sequences in Russian and English. In Chapter 3, the dissertation investigates the incompleteness of neutralization between underlying and coarticulatory palatalization in Russian. In particular, I conducted an Electromagnetic Articulography (EMA) experiment examining temporal coordination and the spatial position of the tongue body for underlying and coarticulatory palatalization. In Chapter 4, I modeled underlying and coarticulatory palatalization using an articulatory-based synthesizer, Task Dynamic Application (TADA), and compared articulatory data to modeled stimuli. Finally, Chapter 5 presents a discussion and conclusions.

Chapter 2. Temporal basis of complex segments and segment sequences

2.1. Introduction[4]

How the continuous dimensions of speech, i.e., the articulatory kinematics and resulting acoustics, relate to phonological categories presents a major scientific challenge. It is a specific instance of the broader challenge of relating discrete and continuous aspects of a cognitive system. Relating phonological categories to speech is often not straightforward since similar speech signals can have different phonological interpretations across languages. A key example, and the focus of this chapter, is complex segments and segment sequences. For example, the segment sequences in (6a) have complex segment counterparts in (6b).

(6) Examples of complex segments and segment sequence counterparts
  (a) Complex segments: /bʲ/, /kʷ/, /k͡p/, /t͡s/
  (b) Segment sequences: /bj/, /kw/, /kp/, /ts/

Complex segments can consist of the same articulatory gestures as segment sequences. For example, a closure of the lips (for [b]) and a movement of the tongue blade (for [j]) characterize both the complex segment /bʲ/ and the segment sequence /bj/. Phonologically, however, languages differ in whether such articulations constitute sequences of multiple segments (e.g., /bj/ in English [bjuti] 'beauty') or as single segments with complex internal structure (e.g., /bʲ/ in Russian [bʲust] 'bust'). A fundamental question is how these similar articulations map to different phonological structures, single complex segments or segment sequences, in different languages.

---

[4] Some portion of this chapter has been published (See the following citation for the published version: Shaw, J., Oh, S., Durvasula, K., & Kochetov, A. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology, 38*(3), 437-477. doi:10.1017/S0952675721000269).

The characterization of complex segments has been the focus of much work in phonology and phonetics, although there is still no consensus regarding whether it is their phonological representation or their phonetic properties that distinguish them from segment sequences. Some researchers have proposed distinct underlying representations for the two cases (Anderson, 1976; Lombardi, 1990; Riehl, 2008; Sagey, 1986). Among others, Sagey (1986) argued that affricates and prenasalized stops are contour segments in which a single root node projects two ordered features, while Lombardi (1990) argued that affricates can be correctly analyzed as complex segments with unordered features. On the other hand, Feinstein (1979) argued that complex segments and corresponding segment sequences have the same underlying representation, but they differ in terms of syllable structure.

Regarding their phonetic properties, Herbert (1986) and Riehl (2008) have argued that phonetic duration is the key factor distinguishing complex segments from segment sequences. They suggest that segment sequences have a longer phonetic duration than complex segments consisting of the same gestures. However, Maddieson and Ladefoged (1993) noted that any such durational differences are too inconsistent to serve this purpose.

Moreover, this type of duration-based diagnostic is only possible when there is a within-language contrast between complex segments and phonetically matched segment sequences or through cross-linguistic comparison. However, such within-language comparison is highly difficult, as few languages provide the requisite evidence for contrast. The duration-based diagnostic can also be complicated by numerous factors that affect segment duration, including the prosodic boundaries (e.g., Cho, 2016; Fougeron & Keating, 1997), the information density of syllables (e.g., Coupé et al., 2019), the local predictability of a segment (e.g., Shaw & Kawahara, 2019), or even a segment's average predictability (e.g., Cohen-Priva, 2017). Moreover, each of these factors may potentially interact with the analysis of gestures as a complex segment or a segment sequence.

An alternative approach is to focus on the articulatory movements of complex segments and segment sequences based on the concept of coordination (Bernstein, 1967; Browman & Goldstein, 1995; Kugler et al., 1982; Turvey, 1990). As discussed in Section 1.0, the goal of this chapter is to establish the temporal diagnosis for complex segments and segment sequences to assess the phonetic realization of two types of Russian palatalizations: underlying and coarticulatory palatalizations. In particular, I propose a specific instantiation of the coordination hypothesis and to test it using kinematic recordings of complex segments with closely matched segment sequences, collected using Electromagnetic Articulography (EMA). The complex segment case refers to the palatalized consonants (underlying palatalization) in Russian and the case of segment sequences is consonant-glide sequences in English. I selected this pair for comparison because they offer a clear case of similar gestures that show phonologically different behavior across languages. As the main aim of this chapter is to test whether different phonological entities, complex segments vs. segment sequences, are also differentiated by virtue of how the component articulatory gestures are coordinated in time, it is crucial that I establish independent phonological evidence for the distinction in question. I, therefore, discuss the phonological arguments in Russian and consonant-glide sequences in English in section 2.2. In particular, I review phonological evidence for treating palatalized consonants in Russian as complex segments (2.2.1) and corresponding gestures in English as segment sequences (2.2.2). Past kinematic studies on these languages are also briefly summarized in Section 2.3. I then lay out my hypotheses and predictions in Section 2.4. Through computational simulations, explicit predictions are made for how the distinct complex segments and segment sequence coordination patterns structure distinct patterns of variation in the kinematic signal. In section 2.5, I transition to an empirical test of the hypotheses. This sets the stage for a new experiment, described in section 2.5, and reported in section 2.6. The discussion and the summary are presented in sections 2.7 and 2.8, respectively.

2.2. Phonological evidence

Complex segments and segment sequences differ in their phonological behavior, and these differences have formed the primary basis for arguments supporting the structural distinction. The basic form of the argumentation is as follows: a pair of gestures[5] is parsed as a single (complex) segment, as opposed to a segment sequence, if it shows the same phonological behavior as other (simplex) segments, otherwise, they are assumed to be parsed as a segment sequence. The phonological behavior supporting this type of argument can be classified into at least four types: (i) phonological contrast, (ii) phonological distribution, (iii) morpho-phonological patterning, and (iv) language games. I discuss each type of argument for palatalized consonants in Russian as an example of complex segments. Then, I provide phonological arguments for consonant-glide sequences in English as an example of a segment sequence.

2.2.1. Russian palatalized consonants as complex segments

2.2.1.1. Phonological contrast

In Standard Russian, there is a phonological contrast between C^j, palatalized consonants, and corresponding segment sequences, i.e., C+j sequences, both word-initially (7a) and word-medially

---

[5] For simplicity in exposition, I focus on whether a pair of gestures constitutes a complex segment or a segment sequence, but the basic idea generalizes as well to the *n*-gesture case. That is, three (or more) gestures also constitute a complex segment if they, together, show the same behavior as a single segment.

(7b) (Avanesov, 1972; Timberlake, 2004).[6] That is, palatalized consonants and corresponding segment sequences pattern differently, phonologically. In the following sections, I provide further evidence that palatalized consonants in Russian are unambiguously complex segments.

(7) Contrast between complex segments, Cʲ, and segment sequences, C+j[7]

(a) Word-initial position		
/pʲatij/ [pʲatɨj] ‘fifth’	vs.	/pjanij/ [pʲjanɨj] ‘drunk’
/bʲust/ [bʲust] ‘bust’	vs.	/bjut/ [bʲjut] ‘beat (3p pl)’
/dʲatel/ [dʲatʲel] 'woodpecker'	vs.	/djakon/ [dʲjakon] 'deacon'
/lʲut/ [lʲut] ‘fierce’	vs.	/ljut/ [lʲjut] ‘pour (3p pl)’
/rʲadom/ [rʲadom] ‘near’	vs.	/rjanij/ [rʲjanɨj] ‘zealous’
(b) Word-medial position		
/kopʲa/ [kopʲa] ‘save (part.)’	vs.	/kopja/ [kopʲja] ‘spear (gen. sg.)’
/xamʲa/ [xamʲa] ‘to be rude (part.)’	vs.	/skamja/ [skamʲja] ‘bench’
/batʲa/ [batʲa] ‘dad’	vs.	/bratja/ [bratʲja] ‘brothers’
/sudʲa/ [sudʲa] ‘judge (part.)’	vs.	/sudja/ [sudʲja] ‘judge (noun)’

---

[6] For simplicity of presentation, I do not indicate morpheme boundaries in phonemic forms (unless these are crucial for the phonetic realization of C+j), and I do not indicate stress or vowel reduction in phonetic forms. Phonetic transcriptions indicate the following processes: palatalization of non-palatalized consonants before /e/ and /j/ (see below); backing of /i/ to [ɨ] after non-palatalized consonants; devoicing of voiced obstruents word-finally; regressive voicing assimilation of obstruents in clusters; regressive palatality assimilation in certain clusters (see Timberlake, 2004 for a description of these patterns).

[7] Some C+j sequences are morphologically derived, e.g., /pj-a-n-ij/ from /pʲi-tʲ/ ‘to drink’ via /i/-gliding, others are underlying, e.g., /djakon/ and /rjanij/, at least synchronically. Consonant-glide sequences can occur morpheme-internally (as in the examples above) and across morphemes (prefix + stem and stem + suffix: e.g., /s-jom-k-a/, /brat-ja/) or words (preposition + stem; e.g., /s jamoj/ ‘with a pit’). C_1 before a palatal glide in tautomorphemic and stem + suffix sequences is pronounced as non-contrastively palatalized (e.g., /djakon/ [dʲjakon] ‘deacon’), with the exception of prefix-stem boundaries (e.g., /pod-jom/ [podjom] ‘rise, lift’) and variably if it is labial, e.g., /pjanij/ [pʲjanɨj] ~ [pjanɨj] ‘drunk’ (Avanesov, 1972, pp. 348-377).

2.2.1.2. Phonological distribution

In Russian, palatalized segments can occur in the same environments as non-palatalized (simplex) segments, but C+j sequences are more restricted. For example, C+j sequences do not occur word-finally or preconsonantally, while palatalized consonants are common in these positions as shown in (8a). Moreover, palatalized consonants occur in consonant clusters, both prevocalically and preconsonantally, as well as both in onset and coda positions. In these positions, palatalized consonants pattern together with non-palatalized counterparts of the same manner. For example, as shown in (8b), both palatalized and non-palatalized laterals occur as C_1 in two-consonant onset clusters, while comparable l+j sequences cannot. The occurrence of the glide /j/ in clusters is limited to immediately prevocalic onset and immediately postvocalic coda positions only. Lastly, palatalized and non-palatalized liquids occur as C_4 in 4-consonant onset clusters, which are the maximally permitted onsets in the language as shown in (8c).

(8) Distributional evidence for complex segmenthood of palatalized consonants

(a) /golu**bʲ**/ [golu**pʲ**] ‘pigeon’ vs. */…bj/
/se**mʲ**/ [sʲe**mʲ**] ‘seven’ vs. */…mj/
/ma**tʲ**/ [ma**tʲ**] ‘mother’ vs. */…tj/
/pro**sʲ**ba/ [pro**zʲ**ba] ‘request’ vs. */…sjb…/
/vo**lʲ**nij/ [vo**lʲ**nɨj] ‘free’ vs. */…ljn…/
/go**rʲ**ko/ [go**rʲ**ko] ‘bitter (adv.)’ vs. */…rjk…/

(b) /**lʲ**gota/ [**lʲ**gota] ‘benefit’ vs.*/ljCV…/ cf. /**l**gatʲ/ ‘to lie’
/**lʲ**dʲina/ [**lʲ**dʲina] ‘ice floe’ vs.*/ljCV…/ cf. /**l**be/ [lbʲe] ‘forehead (gen.sg.)

(c) /vzglʲad/ [vzg**lʲ**at] ‘glance’ vs.*/CCCCjV…/ cf. /vzplaknutʲ/ [fsp**l**aknutʲ] ‘to cry a bit’
/vstrʲatʲ/ [fst**rʲ**atʲ] ‘to stick in’ vs.*/CCCCjV…/ cf. /vzgrustnutʲ/ [vzg**r**usnutʲ] ‘to feel sad a bit’

2.2.1.3. Morpho-Phonological patterning

Russian word formation and morpho-phonology provide some evidence that the C+j sequence is separable in ways that palatalized segments are not. As shown in (9a-b), C+j sequences are broken up by a vowel in alternating forms, resulting in C+V+j sequences, while this does not apply to palatalized segments.

(9) Morpho-phonological patterning for complex segmenthood of palatalized consonants

(a) /se**mj**a/ [sʲemʲja] ‘family’ vs. /se**mej**nij/ [sʲemʲejnɨj] ‘legal’
(b) /vre**m**ʲa/ [vrʲemʲa] ‘time’ vs. /vre**m**ʲ**e**nnij/ [vrʲemʲennɨj] ‘temporary’
c.f., */vre**mej**nij/

2.2.1.4. Language games

To round out the phonological arguments for Russian, there is also some evidence from language games, in which palatalized consonants are treated as single segments, not segment sequences. This is, for example, the case in a children’s secret language described in Vinogradov et al. (2005). The language game has the following rules:

- Rule 1: In words beginning with a single consonant or a cluster, the first consonant is substituted with the fricative /ʂ/ (e.g., ja → ʂa).
- Rule 2: The original (C)(C)V moves to the end of the word (e.g., ja → ʂa.ja).
- Rule 3: Another syllable, /ʦi/, is added right after it (e.g., ja → ʂa.ja.ʦi).

The line /ja # nʲi.ʧe.vo # ne.po.nʲi.ma.ju # po # kra.je.ve.de.nʲju/ ‘I don’t understand anything about local history (school subject)’ is realized in the language game as /ʂa.ja.ʦi # ʂi.ʧe.vo.ne.ʦi ʂe.po.nʲi.ma.ju.nʲi.ʦi # ʂo.po.ʦi # ʂra.je.ve.de.nju.kra.ʦi/. Each word and the corresponding language game transformation form are shown in (10). The portion of each original

Russian word that is substituted by [ʂ] in the language game is boldfaced. The key evidence provided by the language game comes from the fact that palatalized consonants, [nʲ] in (10b-c), pattern with single (simplex) segments, [j] in (10a), [p] in (10d), in being substituted with the single segment [ʂ]. When a Russian word starts with a segment sequence, [kr] in (10e), only the first of two segments is substituted. The language game, as illustrated by these transformations, provides additional evidence for the complex segment status of palatalized consonants in Russian.

(10) The Russian forms and the corresponding language game transformation forms for /ja # nʲi.ʧe.vo # ne.po.nʲi.ma.ju # po # kra.je.ve.de.nʲju/ ‘*I don’t understand anything about local history (school subject)*’

	Original Russian	→	**Language Game transformation**
(a)	**j**a	→	ʂa.**j**a.ʦi
(b)	**n**ʲi.ʧe.vo	→	ʂi.ʧe.vo.**n**ʲi.ʦi
(c)	**n**e [nʲe] po.nʲi.ma.ju	→	ʂe.po.nʲi.ma.ju.**n**e[nʲe].ʦi
(d)	**p**o	→	ʂo.**p**o.ʦi
(e)	**k**ra.je.ve.de.nju [kra.je.vʲe.dʲe.nʲju]	→	ʂra.je.ve.de.nju **k**ra.ʦi

In sum, Russian palatalized consonants present a clear case of complex segments. Phonological evidence supporting this analysis includes contrast, distributional facts, morpho-phonological alternations, as well as language games.

2.2.2. English labial-glide consonants as segment sequences

As a control case for Russian complex segments, I opted for segment sequences in English consisting of a consonant and a palatal glide: C+j. As mentioned earlier, phonological contrast sometimes distinguishes complex consonants from consonant sequences. However, English does not contrast [Cj] and [Cʲ]. Furthermore, the absence of contrast by itself does not inform us as to the segmental structure of the observed sequence; C+j could in principle be [Cj] or [C^{j}]. Therefore, in what follows I provide evidence from morpho-phonology and language games to establish that the gestures composing these sequences are organized phonologically as two segments, i.e.: [Cj].

2.2.2.1. Morpho-phonological patterning

One piece of evidence for C+j as a [Cj] sequence in English comes from an affixation pattern. The pattern, adopted from Yiddish and termed "*Shm*-fixed segmentism" involves reduplication and segment substitution to denote a sort of dismissal of the targeted word (Feinsilver, 1961; McCarthy & Prince, 1986; Nevins & Vaux, 2003). In this morpho-phonological pattern, when there is a single word-initial consonant, the initial consonant is typically replaced by [ʃm-], as can be seen in (11a-b). When there is an initial consonant sequence, either the initial consonant or the whole syllable onset can be replaced by [ʃm-] (11d). Most relevant here is the fact that, in words that begin with [Cj] sequences, the first consonant can be replaced by [ʃm] to the exclusion of the glide (11e-f, left), which suggests that the two are independent segments. Note, as with other pre-vocalic consonant sequences, such as [br] in (11d), the whole [Cj] glide can also be replaced by [ʃm] (11d-e, right). In this respect as well, the behavior of [Cj] parallels other segment sequences in its morpho-phonological patterning.

(11) *Shm*-fixed segmentism in English (crucial segments are boldfaced; d-e from Nevins & Vaux, 2003)

(a)	[**b**eɪgl̩ **ʃm**eɪgl̩]			"bagel (dismissively)"
(b)	[**t**$^{\mathbf{h}}$eɪk **ʃm**eɪk]		*[**t**$^{\mathbf{h}}$eɪk **ʃmh**eɪk]	"take (dismissively)"
(c)	[**tʃ**æd **ʃm**æd]		*[**tʃ**æd **ʃmʃ**æd]	"chad (dismissively)"
(d)	[**b**rɛkfəst **ʃm**rɛkfəst]	(or)	[**br**ɛkfəst **ʃm**ɛkfəst]	"breakfast (dismissively)"
(e)	[**k**jut **ʃm**jut]	(or)	[**kj**ut **ʃm**ut]	"cute (dismissively)"
(f)	[**p**juk **ʃm**juk]	(or)	[**pj**uk **ʃm**uk]	"puke (dismissively)"

2.2.2.2. Language games

Another piece of evidence for the bi-segmentality of [Cj] sequences in English comes from the language game Pig Latin (Barlow, 2001; Davis & Hammond, 1995; Idsardi & Raimy, 2005; Nevins & Vaux, 2003). In Pig Latin, a word-initial consonant or syllable onset is moved to the end of the word, and [eɪ] is then added to the dislocated segment. Most relevant to current interests is the behavior of word-initial phonetic sequences of [Cj] in the variant of the game that Davis and Hammond (1995) call *Dialect A*.[8] In this variety, the initial consonant in words with an initial [Cj] sequence can be separated from the glide as shown in (12a-b). This suggests that the consonant and the glide are separate segments in the language. In fact, similar arguments for the separability of phonetic [Cj] sequences can also be made on the basis of other language games: "The Name Game" (Davis & Hammond, 1995), Ibenglish (Idsardi & Raimy, 2005), Ubbi Dubbi (Vaux, 2011).

(12) Pig Latin and palatal glides in English (Davis & Hammond, 1995)

	English	*Pig Latin (Dialect A)*	
(a)	[kjut]	[jutkeɪ]	"cute"
(b)	[pjuk]	[jukpeɪ]	"puke"

2.2.3. Summary

What is notable about the phonological arguments described above is that they refer only to the "behavior'' of segments within phonological systems to illustrate instances in which single complex segments behave differently from corresponding segment sequences. The phonological arguments in question rarely address the issue of how these patterns are realized phonetically. Temporal properties of speech have often been raised as a promising place to look for phonetic

---

[8] Davis and Hammond (1995) document a second "dialect" of Pig Latin, where the palatal glide is simply deleted, e.g., [utke] for "cute"; this dialect is not informative as to the segmental nature of the consonant-glide sequences and is therefore not presented here.

differences, at least for some classes of complex segments. For example, Ladefoged and Maddieson (1998) proposed that total gesture duration may serve to differentiate the class of complex segments they describe as “secondary articulations” from segment sequences consisting of a consonant and an approximant. However, this only works in the presence of contrast within a language or with a suitable cross-linguistic comparison, which introduces a number of complications in interpreting segment durations. For other cases, such as prenasalized stops, total gestural duration may fail to differentiate complex segments from sequences (Browman & Goldstein, 1986; Gouskova & Stanton, 2021; c.f., Maddieson, 1989, who also notes the importance of converging phonological evidence).

I, therefore, pursue an alternative basis for the phonological distinction, one that is rooted in the concept of gestural coordination (Bernštejn, 1967; Browman & Goldstein, 1995; Kugler et al., 1982; Turvey, 1990). As discussed in Chapter 1.0, the goal of this chapter is to establish the temporal diagnosis for complex segments and segment sequences to assess the phonetic realization of two types of Russian palatalization. However, evaluating coordination is not as straightforward as measuring phonetic duration, as differences in coordination are not necessarily detectable in phonetic duration. In the following section, I reviewed past results on coordination structures of complex segments and segment sequences.

2.3. Past results on English and Russian timing

English and Russian are relatively well-studied languages, including their phonetic aspects. There are detailed phonetic accounts of segment sequence timing in both languages (e.g., Russian: Davidson & Roon, 2008; Pouplier et al., 2017; English: Umeda, 1977) as well as phonetic descriptions of palatalization (e.g., Russian: Diehm, 1998; Kochetov, 2009; Kochetov, 2013; Suh

& Hwang, 2016; English: Zsiga, 1995) and direct comparisons of the languages (Zsiga, 2000)

The most directly relevant research comparing Russian and English is reported by Shaw and colleagues (2019), who test temporal coordination of complex segments and segment sequences, using previously collected data, including a reanalysis of Russian data first reported in Kochetov (2006) and an analysis of English data from the Wisconsin X-Ray Microbeam Speech Production corpus (Westbury et al., 1994). The Russian data compared the palatalized labial /p^j/ with the consonant sequence /br/. They hypothesized that complex segments have a temporal basis—two articulatory gestures, G1 and G2, belong to the same complex segment if the onset of G2 is temporally coordinated with the onset of G1. In contrast, two gestures belong to sequences of segments, if the onset of G2 is temporally coordinated with the offset of G1. These competing coordination relations were explored by investigating how the lag between the onset of G1 and the onset of G2 varied with G1 duration. Variation in onset-to-onset lag, defined as the interval from the onset of G_1 to the onset of G_2, as a function of stop-consonant duration (/p/ for /p^j/ and /b/ for /br/) is plotted in Figure 5. EMA data from 3 Russian native speakers revealed that gesture lag increased with stop-consonant duration for /br/ (Fig 5: left) but not for the complex segment /p^j/ (Fig. 5: right). The English data from Shaw et al. (2019) address the /bj/ sequence at the onset of the word *beautiful* from 20 speakers. As plotted in Figure 6, for English, as C_1 duration increases, the lag between gestures also increases.

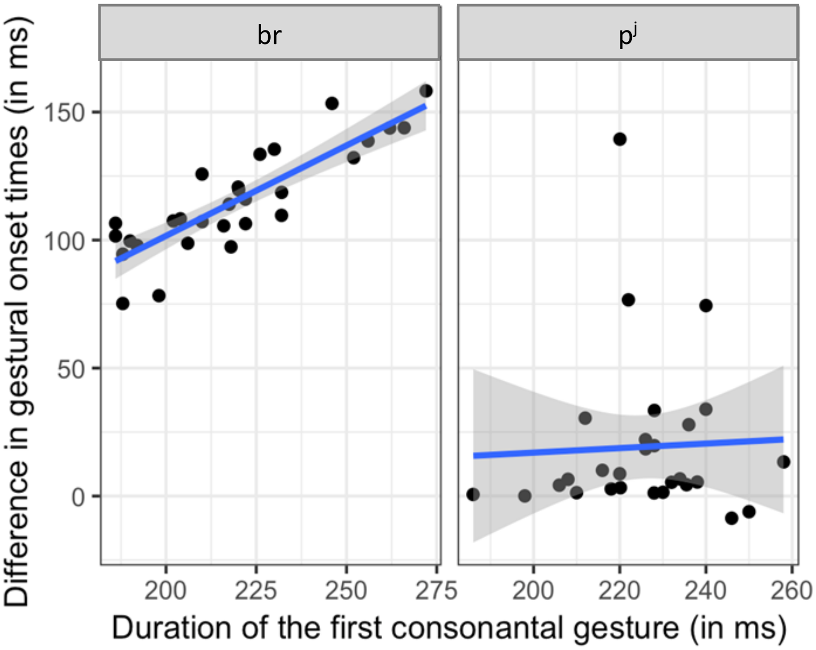


Figure 5: Russian data showing the gestural lag, y-axis, as a function of first consonant duration, x-axis, for /br/ (left) and /pj/ (right). Figure reproduced from Shaw et al. (2019)

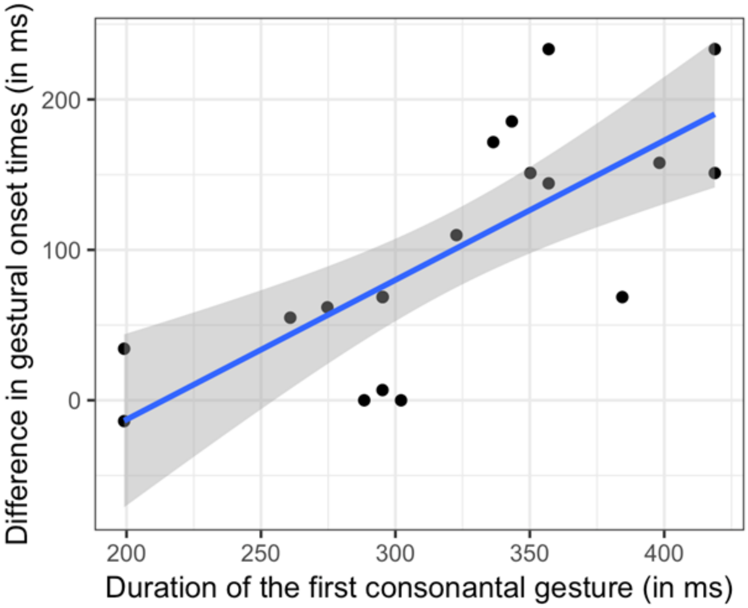


Figure 6: English data showing the gestural lag, y-axis, as a function of first consonant duration, x-axis, for the /bj/ sequence in ‘beautiful’. Figure reproduced from Shaw et al. (2019)

Taken together, the results in Figure 5 and Figure 6 suggest that the gestures of complex segments are coordinated based on gesture onsets, while the gestures of segment sequences are timed sequentially. However, the data provide only an imperfect test of the hypothesis, for a number of reasons. In the Russian data, /br/ and /p^j/ differ in numerous ways: /br/ was extracted from a real word while /p^j/ was extracted from a nonsense word; /br/ was phrase-initial while /p^j/ was phrase medial. More fundamentally, the voicing of the labial stop differed, and the gestures involved in the production of /r/, an apical trill, are distinct from those involved in the production of a palatal glide. For the trill, the tongue body is positioned to support tongue tip raising towards the alveolar ridge; for the palatal glide, the tongue body rises towards the palate. It is, of course, possible that abstract timing relations generalize across end-effectors (tongue tip, tongue blade, lips, etc.) such that it is perfectly appropriate to compare the relative timing of the lips and tongue tip in /br/ with the lips and tongue body for /p^j/. After all, quite different articulators enter into qualitatively similar coordination patterns in numerous cases. For example, in Moroccan Arabic, rising sonority consonant clusters, e.g., /kfl/, show qualitatively similar patterns of coordination as falling sonority clusters, e.g., /msk/ (Shaw et al., 2011); see also Jazani Arabic (Durvasula et al., 2021; Ruthan et al., 2019). Similarly, in Romanian, stop-initial clusters show qualitatively similar patterns of timing, regardless of the place of articulation of C1, e.g., /ksenofob/ ‘xenophobe’ vs. /psalm/ ‘psalm’ (Marin, 2013). However, there are, of course, other cases in which the timing of gestures varies systematically across contexts, with differences possibly conditioned by the magnitude of movements (e.g., Brunner et al., 2014).

For these reasons, the ideal test to examine the temporal coordination of complex segments and segment sequences would better control for segmental/prosodic context, as well as the articulators involved in the gestures. The cross-language comparison between English /bj/ and

Russian /p^j/ involves similar places of articulation, but the stops differ in voicing, which is known to influence timing, at least in some languages (Bombien et al., 2010). Additionally, the source of consonant duration variation differs in the two datasets. The Russian data comes from three speakers producing two items 4-5 times each—variation in consonant duration comes from item, speaker, and repetition. In contrast, the English data comes from many more speakers producing just one repetition of one item. Thus, all of the variation in consonant duration comes from inter-speaker variation. Greater control over the experimental materials, including the segments involved in coordination, the prosodic position of the target items, and the sources of variability would provide additional clarity.

In what follows, I propose hypotheses and predictions regarding the temporal coordination of complex segments and segment sequences. Then, I report on a new experiment designed to improve on past work, eliciting closely matched gestures in Russian, where they constitute complex segments, and in English, where they constitute segment sequences.

2.4. Hypotheses and Predictions

The fundamental question is whether the gestures of complex segments are coordinated differently than gestures of segment sequences, i.e., it is a difference in coordination that provides the basis for the phonological distinction. Specifically, I propose that the gestures of complex segments are coordinated with reference only to gesture onsets, while segment sequences are coordinated with reference to the offset of the first gesture and the onset of the second. This distinction is schematized in Figure 7. Panel (a) shows complex segment timing, while panel (b) shows segment sequences. Before elaborating on this proposal and the predictions it makes for the phonetic signal, I lay out a few foundational assumptions on which the proposal rests.

First, I assume that gestures are forces that drive articulators to phonologically relevant task goals over time; this is a foundational assumption of Articulatory Phonology (e.g., Browman & Goldstein, 1986) and one that, to the best of my knowledge, is uncontroversial, at least within Articulatory Phonology. Even as the theory of the gesture has undergone development in its dynamic formulation—e.g., from an autonomous linear dynamical system with step activation (Saltzman, Elliot L. & Munhall, 1989) to a linear dynamical system with continuous activation (Kröger et al., 1995) to a non-linear dynamical system (Sorensen & Gafos, 2016) to hybrid interacting dynamical systems (Parrell & Lammert, 2019) —the assumption that speech movements are under the control of phonological goals remains a constant working assumption.

The second assumption is that gestures can be decomposed into a series of states or "gestural landmarks" which are available for coordination. That is, coordination relations are expressed in terms of gestural landmarks. For the purposes of this chapter, I reference only two such landmarks, the gesture *onset* landmark, which corresponds to the start of gesturally-controlled movement, and the gesture *offset* landmark, which corresponds to the end of controlled movement. How many additional gestural landmarks are in principle available and what additional landmarks besides these two may also be required to describe the range of coordination patterns in a language or across languages is beyond the scope of this chapter, but see, e.g., Browman and Goldstein (1990; 2000), Gafos (2002), Borroff (2007), Goldstein (2011), Shaw and Chen (2019), for further discussion.

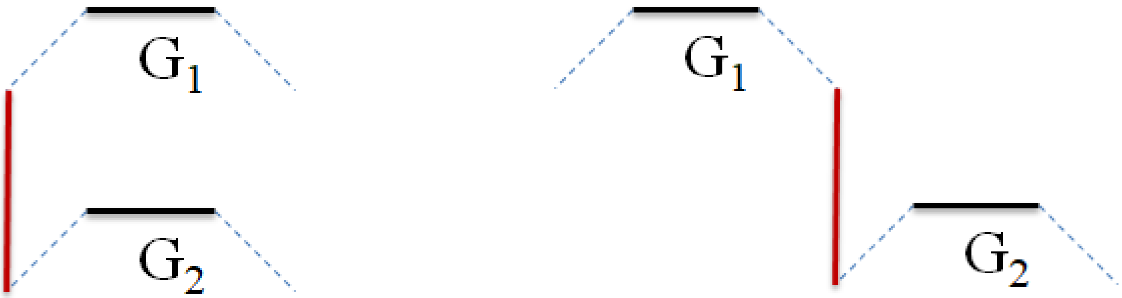
The gestural coordination patterns central to the main hypothesis have antecedents in the literature; they are roughly (caveat below) equivalent to "in-phase" and "anti-phase" coupling (Goldstein et al., 2009; Nam et al., 2009). Two gestures coordinated in-phase will start at the same time. For gestures coordinated anti-phase, the gestures will be sequential, such that the second gesture starts when the first ends. The approach of coupling gestures according to phase angle enables the specification of a continuous range of coordination relations (Browman & Goldstein, 1990), which can be restricted by other principles, including (i) recoverability—coordination relations that do not allow gestures to be perceived will be dispreferred (Browman & Goldstein, 2000; Silverman, 1997)—or (ii) stability (Nam et al., 2009). Drawing on a theory of coordination developed from observations of manual movement data (Haken et al., 1985), Nam and colleagues (2009) proposed that in-phase and anti-phase modes of coordination are available without learning and therefore intrinsically stable.

My hypothesis for complex segments is consistent with in-phase coupling with the following caveat: I assume that landmark-based coordination relations can be stated with consistent lag, as per the phonetic constants in the models of Shaw and Gafos (2015). For example, two gestures can be coordinated such that the onset of movement control is synchronized with a consistent +/- lag. Possible instantiations are shown in panels (c) and (d) of Figure 7. Panel (c) shows complex segment timing with a positive lag; panel (d) shows gestures timed as segment sequences with negative lag. Notably, owing to the influence of the +/- lag, the surface timing of (c) and (d) is identical despite being coordinated based on different articulatory landmarks.

Allowing for the theoretical possibility that gesture landmarks are coordinated with a consistent +/- lag introduces a possible disassociation between the notion of coordination, which is central to the hypothesis, and observations of the relative timing of articulatory movements in the kinematics. Accordingly, this also influences my approach to hypothesis testing. From this theoretical perspective, measures of gestural overlap alone may under-determine temporal control structures, as illustrated in Figure 7 (c) and (d). The same surface timing could be derived from different combinations of coordination relations and lag values: (1) in-phase timing with a positive lag (c), anti-phase timing with a negative lag (d), or even an intermediate timing relation, e.g., “c-center” timing, however derived,[9] with no lag. Crucially, however, these competing hypotheses about temporal control structure can be differentiated by considering relations between temporal intervals, defined on the basis of articulatory landmarks observable in the kinematic signal.

(a) Complex segment, no lag

(b) Segment sequence, no lag



(c) Complex segment, positive lag

(d) Segment sequence, negative lag

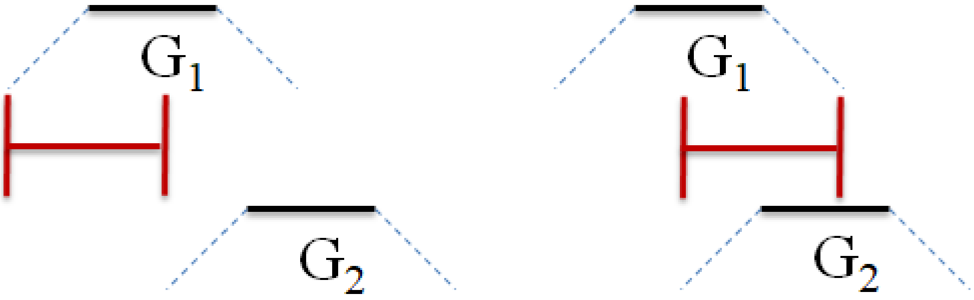


Figure 7: Hypothesized gestural coordination patterns for complex segments (a) and (c), and segment sequences (b) and (d). The upper two panels show surface timing patterns with no +/- lag so that the surface timing faithfully reflects the hypothesized coordination relations. The lower two panels show surface timing patterns that deviate systematically from the hypothesized coordination relation due to a +/- lag.

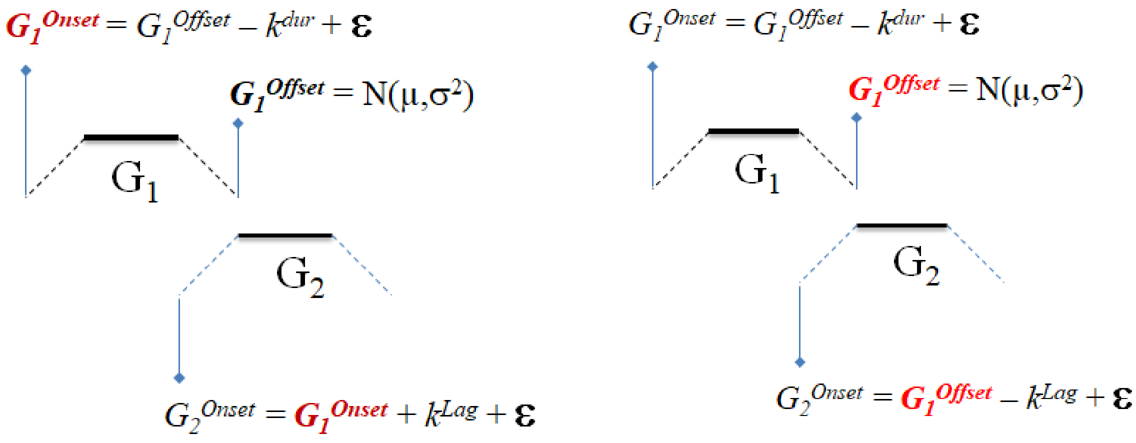
---

[9] “C-center timing” refers to a pattern whereby the vowel starts around the midpoint of preceding consonant gestures (Browman & Goldstein, 1988) and can be derived from the interaction of a network of in-phase and anti-phase coordination relations in a number of ways, including least squares minimization (Browman & Goldstein, 2000), violable constraints in Optimality Theory (Gafos, 2002) or coupled oscillators (Goldstein et al., 2009).

The strategy for differentiating hypotheses is to consider how the temporal interval between gesture onsets varies with gesture duration. The basic strategy follows Shaw et al. (2011) in that I evaluate how temporal coordination conditions covariation between phonologically relevant intervals. The competing hypotheses schematized above make different predictions about how the interval between gesture onsets will covary with gesture duration. For complex segments, variation in first gesture duration will have no effect on the interval between gesture onsets. This is because the onset of G_2 is only dependent on the onset of G_1. For segment sequences, however, any increase in G_1 duration will delay the onset of G_2 since the onset of G_2 is dependent on the offset of G_1.

Notably, the patterns of structure-specific covariation are independent of any constant +/- timing lag that may mediate between the hypothesized coordination relations and the observed timing in the kinematics. Covariation between G_1 duration and the inter-gestural onset interval is predicted only for *segment sequences* (b,d) and not for *complex segments* (a,c). The reasoning is as follows: if the gesture onsets are timed to each other, even if there is a positive lag, then variation in G_1 duration will be entirely independent of the interval between G_1 onset and G_2 onset. Therefore, a longer G_1 duration will not delay G_2 onset, since, in this case, G_2 onset is dependent only on G_1 onset. In contrast, if G_2 is timed to some gestural landmark later in the unfolding of G_1—e.g., gesture offset, as in (d)—then increases in G_1 duration will delay the onset of G_2, increasing the temporal lag between gesture onsets.

To make the above reasoning concrete, simple mathematical models were coded, which illustrate the hypothesized timing relations and simulated patterns of covariation between G_1 duration and the interval between gesture onsets. The simulation algorithm for each model is summarized in Figure 8. The algorithms first sample the G_1^{Offset} landmark from a normal distribution defined by a mean, μ, and a variance, σ^2. The particular parameters of this distribution have no bearing on the simulation results. For the simulation below the mean was 500 and the variance was 400. The G_1^{Onset} landmark was defined as preceding the G_1^{Offset} landmark by a constant, k^{dur}, and an error term, ε. The error term is normally distributed error. Together, the constant and the error term define a normal distribution that characterizes the duration of G_1. For the simulations below, k^{dur} ranged from 200 to 250 and the associated error term was 50. These parameters are identical for both models. The key difference is in how the onset of G_2 is determined. For the complex segment model, G_2^{Onset} is timed to G_1^{Onset}, plus a constant k^{Lag} and associated error term, ε. For the segment sequence model, G_2^{Onset} is instead timed to G_1^{Offset}.



(a) Complex segment

(b) Segment sequence

Figure 8: Simulation algorithm for complex segments (a) and segment sequences (b)

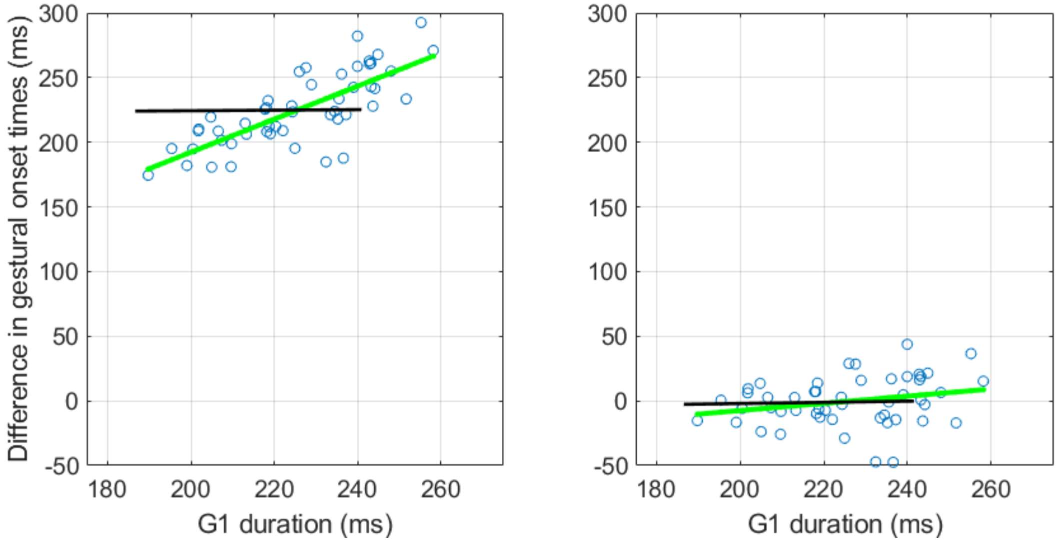
Figure 9 illustrates two sets of simulations based on the models. In both sets of simulations, k^{dur}, the constant that determines G_1 duration, was gradually varied to evaluate how variation in G_1 duration impacts the interval between gesture onsets. In the first set of simulation results, (a) and (b) (shown on the top row), the models were implemented with no lag by setting the k^{lag}

parameter to 0. The associated error term was 100. In the second set of simulations, (c) and (d) (shown on the bottom row), k^{lag} was set to 100 and the error term was maintained at 100. A key illustration is that the pattern of covariation is the same across coordination patterns regardless of lag. For segment sequences there is a positive correlation; for complex segments, there is no strong association between G_1 duration and difference in gestural onset times. Note, however, that even though the pattern of covariation remains constant across different lag values, there are other measures that change. For example, there is a clear difference in the interval between gestural onsets in (a) and (b). If there is no lag, i.e., $k^{lag} = 0$, then complex segments have a greater overlap between gestures than segment sequences. However, in the bottom panels, the difference in onset-to-onset lag between complex segments and sequences goes away. Thus, on the set of theoretical assumptions I have adopted, gesture overlap can successfully diagnose the difference between complex segments and segment sequences only under certain conditions. In contrast, the variation between temporal intervals is structured consistently regardless of variation in gesture overlap. Covariation between G_1 duration and onset-to-onset lag provides a reliable diagnostic of coordination regardless of gesture overlap, i.e., whether there is a +/- lag between gestures.

As the simulations illustrate, the differences in coordination that I have hypothesized as a basis for the phonological distinction between complex segments and segment sequences can be differentiated in the kinematic signal because of how they structure variation in temporal intervals, defined on gestural landmarks. I now turn to empirical tests of the hypothesis.

(a) Segment sequence, no lag

(b) Complex segment, no lag



(c) Segment sequence, negative lag

(d) Complex segment, positive lag

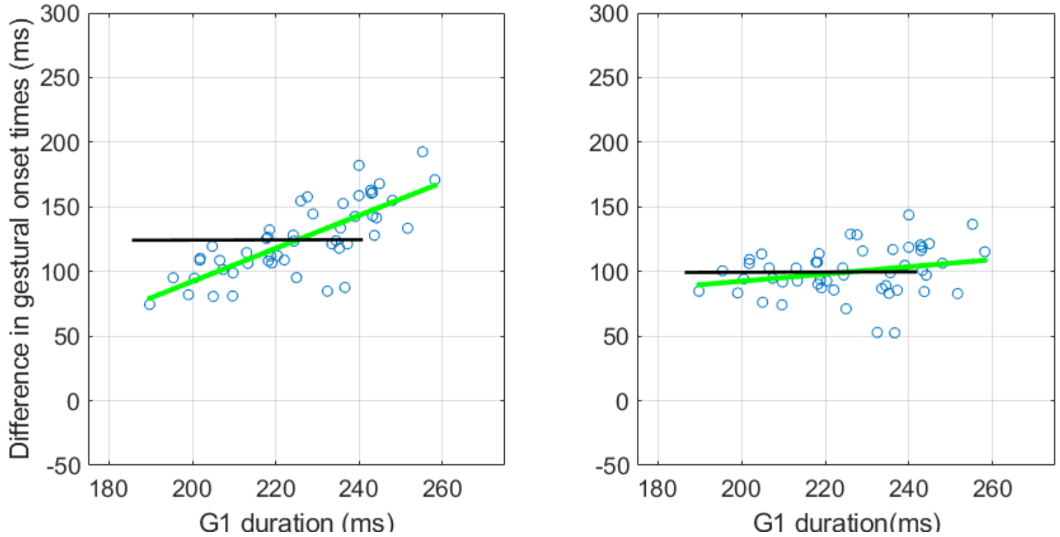


Figure 9: Simulation results showing the gestural lag (y-axis) for complex segments (left) and segment sequences (right) as G1 duration (x-axis) varies. The green line represents the least squares linear fit to the data; the black line shows the mean lag.

2.5. Methods

2.5.1. Participants

Four native speakers of Russian (3 male; 1 female) and four native speakers of English (2 male; 2 female) participated in the study. All speakers were in their 20s at the time of recording and living in the United States. The Russian speakers were born in Russia and moved to the United States as adults.

2.5.2. Materials

The target Russian materials consisted of the six words shown in Table 1 (left). All words begin with palatalized labial consonants followed by a back vowel, either /u/ or /o/. The English items begin with a labial consonant and a palatal glide and are followed by the vowel /u/. The Russian words were read in the carrier phrase: /ʌˈna ____ pəftʌˈrʲilʌ/. ‘She ____ repeated.’ In this phrase, the target word is preceded by the vowel /a/, which is typically reduced, and followed by /p/. The English words were read in the carrier phrase ‘It’s a ______ perhaps’. In this phrase, the target word was also preceded by a reduced vowel and followed by /p/.

Table 1: Stimulus items

Russian			**English**	
word	**IPA**	**gloss**	**word**	**IPA**
пёк	/pʲok/	bake (3ps past)	pew	/pju/
бюст	/bʲust/	bust (breast/sculpture)	butte	/bjut/
мю	/mʲu/	Greek letter	muse	/mjuz/
Фёдор	/fʲodor/	Fyodor (name)	musical	/mjuzikəl/
вёз	/vʲoz/	carry (3ps past)	view	/vju/
вёдра	/vʲodra/	bucket (pl)		

2.5.3. Procedure

Data collection was executed in the Phonetics Lab at Yale University. Articulatory movements were recorded using the NDI Wave Speech Production system, which uses Electromagnetic Articulography to track small sensors, approximately 3 mm in diameter. The sensors were attached to the tongue, lips, and jaw using high viscosity periacryl. Three sensors were attached along the sagittal midline of the tongue. The most posterior of these three lingual sensors was attached to the tongue body, approximately 5 cm behind the tongue tip. The most anterior lingual sensor was placed approximately 1 cm behind the tongue tip. A third sensor was placed on the tongue blade, halfway between the sensors on the tongue tip and tongue body, approximately 3 cm behind the tip. I refer to this sensor as the "tongue blade" (TB) sensor. Sensors were also attached to the upper and lower lips, just above and below the vermillion border. To track jaw movement, another sensor was placed on the gum line just below the lower incisor. I also attached sensors on the left and right mastoids as well as on the nasion or nose bridge. These last three sensors, the left/right mastoid, and the nasion/nose bridge were used to computationally correct for head movements in post-processing.

Once the sensors were attached, participants sat next to the NDI Wave field generator and read the target words in the carrier phrases from a computer monitor, located 50 cm outside of the EMA magnetic field. On each trial, the target word flashed on the screen for 500 ms, and then was shown in the carrier phrase. The target word embedded in the carrier phrase remained on the screen until the participant read the word and the experimenter pressed a button to accept the trial. The purpose of displaying the target word before eliciting it in the carrier phrase was to promote fluent pronunciation of the target word in its carrier phrase, and, in particular, to avoid a pause immediately before the target word. Speech acoustics were recorded concurrently at 22 kHz using a Sennheiser condenser microphone placed outside of the EMA magnetic field.

After completing the experimental trials, I recorded the occlusal plane of each participant and the location of the palate. The occlusal plane was recorded by attaching three NDI Wave sensors to a rigid object—a protractor—and having participants hold it between their teeth. The sensors on the protractor were attached in an equilateral triangle configuration and the protractor was oriented so that the mid-sagittal plane of the participant, as indicated by the sensors on the nasion and lips, bisected the triangle on the rigid object. Palate location was recorded using the NDI Wave palate probe. Participants traced the palate using the probe while the position of the probe was monitored using the real-time display of the NDI Wave system. The palate tracings provided a point of reference for visualizing the data but did not enter into any quantitative analysis of the data. Each participant completed at least 15 blocks and as many as 30 blocks, for a total of 1,090 tokens entering into the analysis.

As a post-processing procedure, the data was computationally corrected for head movements and rotated to the occlusal plane so that the bite of the teeth serves as the origin of the spatial coordinates. To eliminate high-frequency noise, all trajectories were then smoothed using Garcia's robust smoothing algorithm (Garcia, 2010). Finally, I calculated a lip aperture trajectory, as the Euclidean distance between the upper and lower lip sensors.

2.5.4. Analysis

The post-processed data was visualized in MVIEW, a Matlab-based program developed by Mark Tiede at Haskins Laboratories (Tiede, 2005). I used the lip aperture trajectory to identify labial gestures and the tongue blade (TB) trajectory to identify palatal gestures.

Gestural landmarks were parsed with reference to the velocity signal using the *findgest* function in MVIEW. Specifically, the gesture *Onset* and *Target* landmarks were labeled at 20% of the peak velocity in the movement toward constriction (See Figure 10). *Release* and *Offset* landmarks were labeled at a 20% threshold of peak velocity in the movement away from constriction.

I used these threshold values to index gestural landmarks instead of, e.g., velocity minima, because I am particularly interested in the temporal dimensions of the trajectories. Also, I have chosen the 20% threshold following the prevailing convention in articulatory studies (e.g., Hoole et al., 1994). Although the articulators rarely, if ever, stop moving during spontaneous speech, they are often slowed substantially when they are near phonologically relevant targets, giving the appearance of a “plateau” in the trajectory (see also the plateau at the constriction phase in the schematic diagrams in Figure 7 and Figure 8). During the plateau, a small variation in velocity, even on the order of magnitude of measurement error, < 1.0 mm (Berry, 2011), could have a substantial impact on the timing of the landmark. Defining landmarks as percentages of peak velocity, i.e., before velocity gets too low, helps to avoid this situation, essentially providing more reliable indices of gestural landmarks. For the palatal gestures, parsed using the TB sensor, I parsed gestures using the tangential velocity signal, based on movement in three dimensions. Since the lip aperture trajectory is a Euclidean distance (in 3D space), it is unidimensional.

Figure 10 shows one example of a labial gesture. The upper panel shows the positional signal, which in this case is lip aperture (in mm). The lower panel shows the corresponding velocity signal. The four articulatory landmarks are labeled on the positional signal with reference to the velocity peaks.

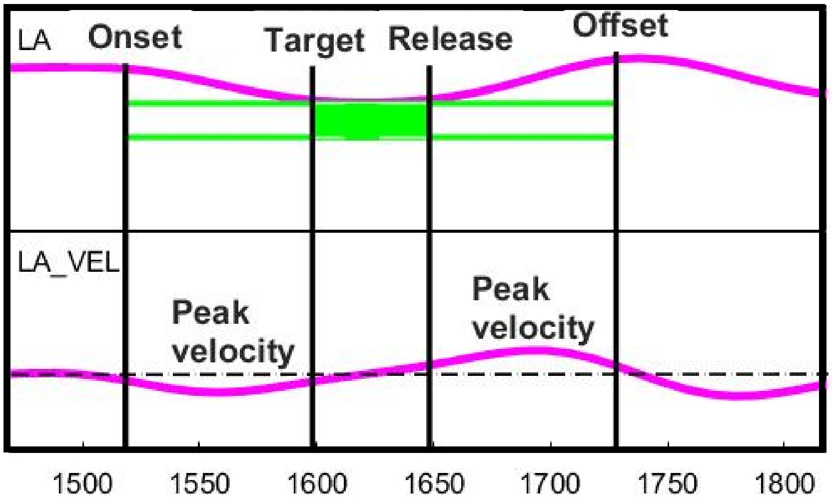


Figure 10: Example of gesture parse for a labial gesture. The gestural landmarks, Onset, Target, Release, Offset, are labeled at 20% thresholds of peak velocity.

Gestural landmarks, parsed as described above for the labial and palatal gestures of all target words, were used to calculate two intervals, which serve as the primary continuous measures in the analysis. These two intervals are schematized in Figure 11. *G_1 duration* was calculated by subtracting the timestamp of the *Onset* of the labial gesture from the *Offset* of the labial gesture. Accordingly, *G_1 duration*, a measure of intra-gestural timing, is always positive. The second interval, *onset-to-onset*, was calculated by subtracting the *Onset* of the labial gesture (*G_1*) from the *Onset* of the palatal gesture (*G_2*), providing a measure of the temporal lag between the two gestures. Note that when the two gestures start at the same time, the *onset-to-onset* interval is zero, i.e., no lag; likewise, when the palatal gesture starts before the labial gesture, the *onset-to-onset* interval will be negative; otherwise, *onset-to-onset* interval will be positive and a measure of temporal lag between the gestures. As positive values for the *onset-to-onset* interval are the most common scenario, I refer to the onset-to-onset measure as lag, i.e., *onset-to-onset* lag. Similarly, due to a tendency for temporal precedence of labial and palatal gestures, I refer to the target labial gesture in the materials as *G_1*, and the target palatal gesture as *G_2*.

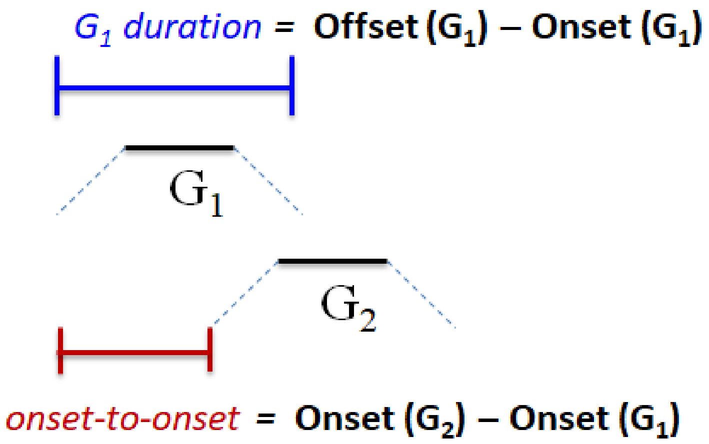


Figure 11: Schematic depiction of the two intervals, G1 duration and onset-to-onset lag, entering into the analysis. G1 refers to the labial gesture and G2 refers to the palatal gesture

Before proceeding with statistical analysis, I removed outliers that were greater than three standard deviations from the speaker-specific mean value of either *G_1 duration*, 8 tokens removed (0.7% of the data), or *onset-to-onset* lag, 6 tokens removed (0.5% of the data).

The main analysis of the data tests the hypothesis schematized in Figure 7, embodied in the stochastic models of Figure 8 and exemplified by simulations in Figure 9. As *G_1 duration* varies, I ask whether *onset-to-onset* lag will positively covary, as predicted by the segment sequence hypothesis, or whether these intervals will be statistically independent, as predicted by the complex segment hypothesis. I, therefore, treat *onset-to-onset* lag as a dependent variable, and evaluate whether *G_1 duration* is a significant predictor. Besides *G_1 duration*, there are other factors that could condition variation in onset-to-onset lag. Most notably, these include subject-specific factors, such as preferred speech rate, and item-specific factors, such as the lexical statistics and usage patterns of the specific items in my study. I factor these considerations into the analysis through the inclusion of random effects for speaker and item in a linear mixed-effects model, which

I fit to the data using the lme4 package (Bates et al., 2014) in R (Version 4.0.3). I calculated the residual deviation from the best-fitting model and eliminated outliers to the model that were greater than three standard deviations from the mean (following Baayen & Milin, 2010), resulting in the elimination of 23 additional outliers (2.1% of the data). The nested models were then re-fit to this data set, consisting of 1,053 tokens across speakers.

To a baseline model, consisting of random intercepts for subjects and items, I added fixed factors of interest incrementally. First, I added *G_1 duration*, then *language* (English vs. Russian, with Russian as the reference level), and finally the interaction between *G_1 duration* and *language*. This gives a set of four nested linear mixed-effects models. I evaluated the significance of each fixed factor through model comparison, considering whether the addition of the fixed factor provides a significant increase in the likelihood of the data and whether that increase is justified by the increased complexity of the model, for which I reference the Akaike Information Criterion (AIC). The AIC measures model fit while controlling for over-parameterization; a lower AIC value suggests a better model (Akaike, 1974; Burnham et al., 2011). The fixed factor of primary interest for the main hypothesis is the interaction term: *G_1 duration * language*. This is because G_1 duration is predicted to have a positive influence on onset-to-onset lag for English, since the target gestures behave phonologically as sequences (see Section 2.2.2 for arguments for English), but not for Russian, since the target gestures in Russian behave phonologically as complex segments (see Section 2.2.1 for arguments for Russian).

2.6. Results

In Section 2.4. I hypothesized that the gestures of complex segments are coordinated differently than gestures of segment sequences. The main analysis of the data tests the prediction of the stochastic models, exemplified by the simulations in Figure 9. As *G_1 duration* varies, I ask whether the *onset-to-onset* interval will covary, as predicted by the segment sequence hypothesis or whether these intervals will be statistically independent as predicted by the complex segment hypothesis. Since the data consists of English, where the target gestures form segment sequences, and Russian, where the target gestures form complex segments, I hypothesize that the influence of G_1 duration on onset-to-onset lag will differ across languages.

2.6.1. Kinematic trajectories and distribution

Before moving to the main results, which involve covariation between G_1 and onset-to-onset lag, I first examine the continuous trajectories of relevant articulators. Figure 12 provides a representative token, highlighting the target gestures, /b/ and /j/, as produced in ‘butte’ (English condition). The top panel shows the waveform. The second panel shows the lower lip, which is the primary determinant of the lip aperture trajectory for this subject. The bottom panel shows the tongue blade trajectory, which was used to parse the palatal gesture. For simplicity of display, only the vertical trajectories are shown. The onset and offset landmarks for the labial and the onset of the palatal gesture are also labeled. These labels show that the onset of the palatal gesture occurs, in this token, after the onset of the labial gesture but well before the offset of the labial gesture. Unsurprisingly, the palatal gesture starts during the labial closure. From a single token, however, it is not possible to test the hypothesis. That is, we currently do not have a method that would allow us to determine whether the control structure (dynamics) behind the kinematic data for a single token, such as this one, triggers the onset of the palatal gesture at the onset of the labial gesture (per the complex segment hypothesis) or whether the onset of the palatal gesture is instead triggered by the offset of labial gesture (per the segment sequence hypothesis). The token in Figure

12 is consistent with both hypotheses: complex segment timing with positive lag, as in Figure 1(c), or segment sequence timing with negative lag, as in Figure 7(d).

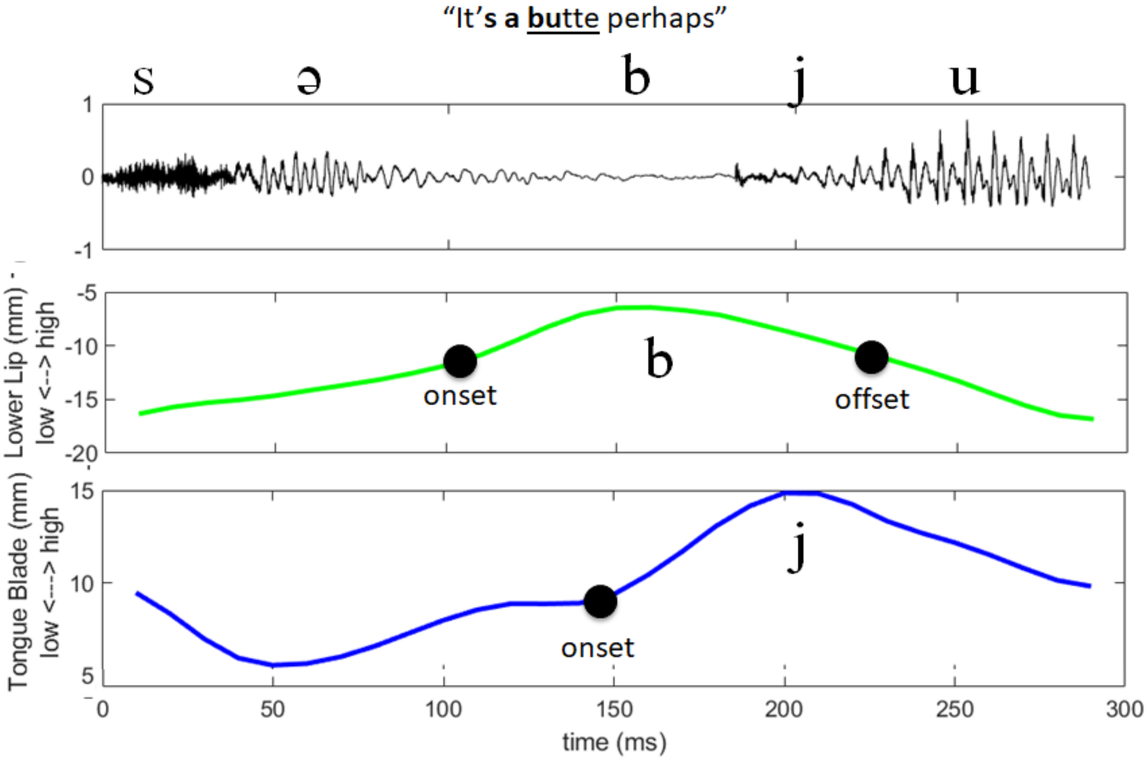


Figure 12: Example of a token of 'butte' from the English recordings. The top panel shows the waveform. The second panel shows the lower lip trajectory in the vertical dimension. The bottom panel shows the tongue blade, also in the vertical dimension. The three gestural landmarks relevant to calculating the intervals of interest (Figure 6) are labeled.

Figure 13 illustrates variability across kinematic trajectories for the token 'butte' as produced by the four English speakers in the study. The figure plots the Lip Aperture trajectory in the upper panels and the Tongue Blade (TB) trajectory in the lower panels. Each trajectory is a different color; the thick dotted line is the average trajectory. The figures plot trajectories from 100 ms before the onset landmark of the lip aperture gesture to 500 ms following this landmark, a temporal window of 600 ms. This window is long enough to observe the labial and palatal gestures for all tokens. The level of variability in both the timing and magnitude of the gestures varies by subject. For E2, most tokens occur tightly clustered around the mean; E1 shows more variability, and E3 and E4 show even more. Across speakers, the fall in the LA aperture trajectory, indicating the closing of the lips tends to (slightly) precede the rise of the TB for the palatal gesture. To facilitate comparison, vertical gray lines indicate when the LA trajectory starts to fall (based on the average) and when TB starts to rise (also based on the average).

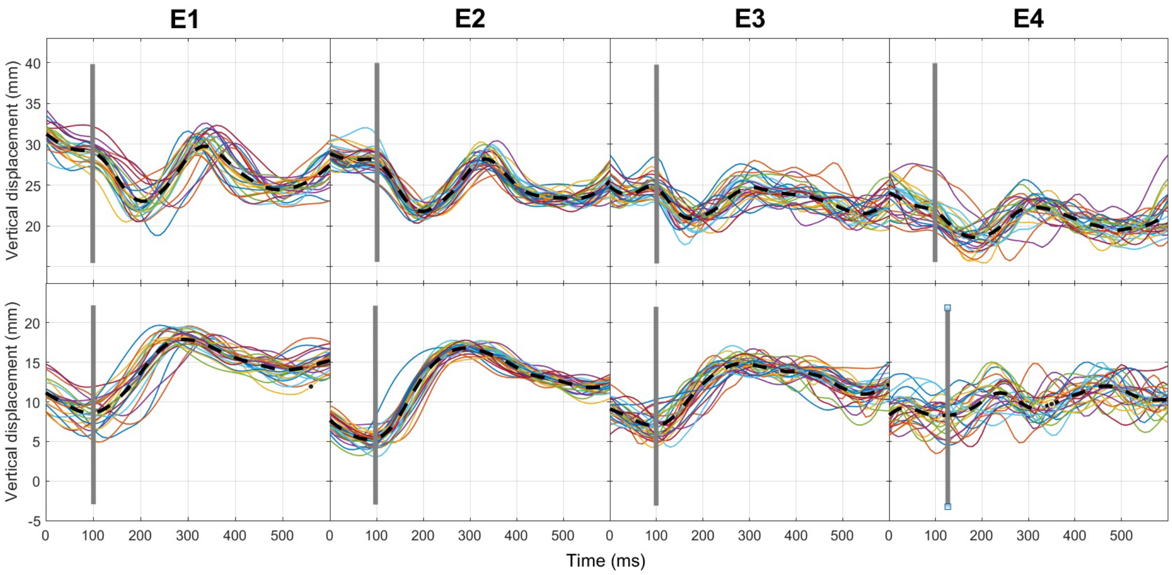


Figure 13: Tokens of /bjut/ from each English speaker

Figure 14 shows the same 600 ms window for the Russian token /bʲust/, as produced by four speakers. The level of variability in the magnitude of the gestures varies by subject as well. For R2, most tokens occur tightly clustered around the mean; R1 and R3 show more variability, and R4 shows even more. On the other hand, the relative timing of the gestures appears similar across speakers - the fall in the LA aperture trajectory tends to coincide with the rise of the TB for the palatal gesture.

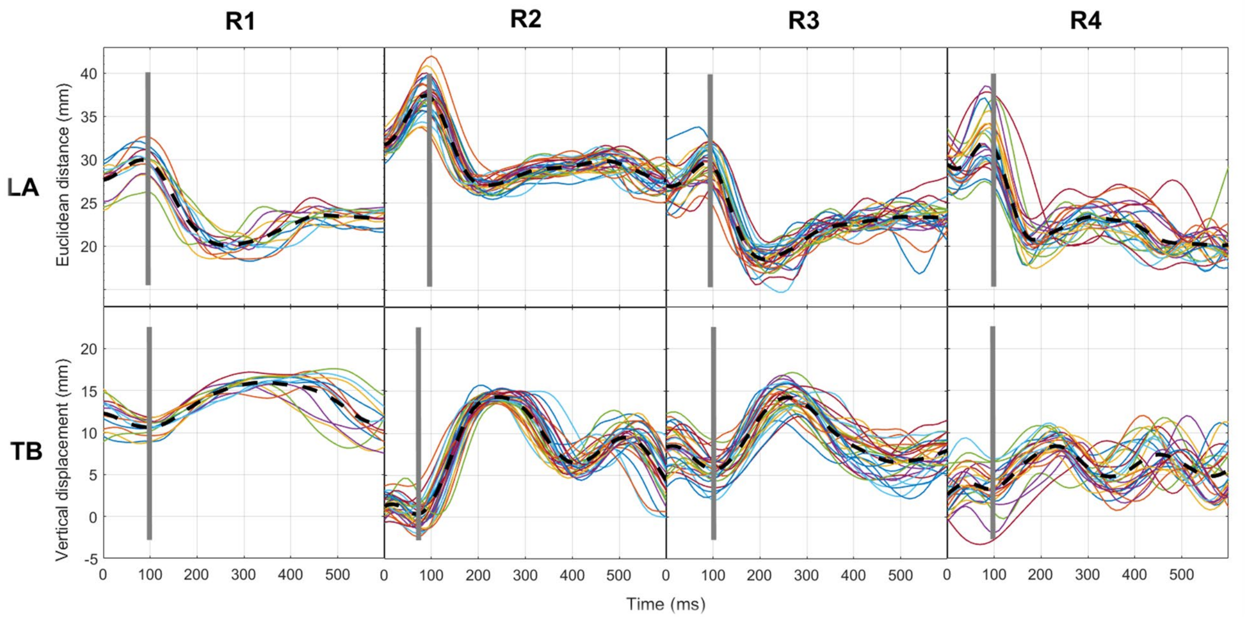


Figure 14: Tokens of /bʲust/ from each Russian speaker

Since the dependent measures in the analysis are temporal intervals and I am particularly interested in the correlation between intervals, I next present the distribution by language of the key continuous variables: G_1 duration (Figure 15) and onset-to-onset lag (Figure 17). The G_1 duration measures have a slight rightward skew, as is common for temporal measurements of speech associated with linguistic units. Notably, however, the distributions for English and Russian are heavily overlapped. The peak of the English distribution is at 201 milliseconds, with a standard deviation of 53 milliseconds; the peak of the Russian distribution, at 242 milliseconds, is within one standard deviation of the English peak. Thus, the average labial is similar in duration across English and Russian. For completeness, Figure 16 shows the distribution of G_2 (palatal gesture) duration by language. This measurement does not relate directly to any of the main hypotheses, but I include it for reference. The English data tends to have a longer palatal gesture than the Russian data. Finally, Figure 17 shows the distribution of onset-to-onset lag. Here too, both languages have similar mean values. However, the distributions differ in shape, with English having a long right tail.

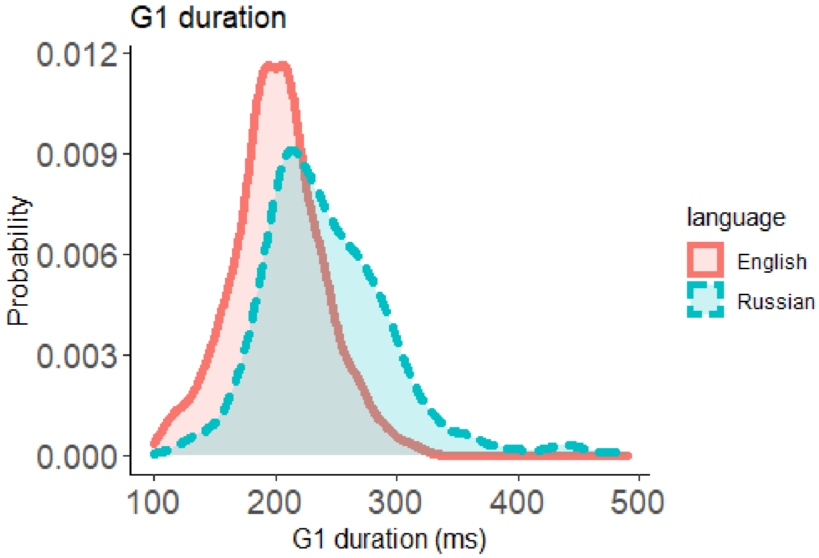


Figure 15: The distribution of G1 (labial consonant) duration by language

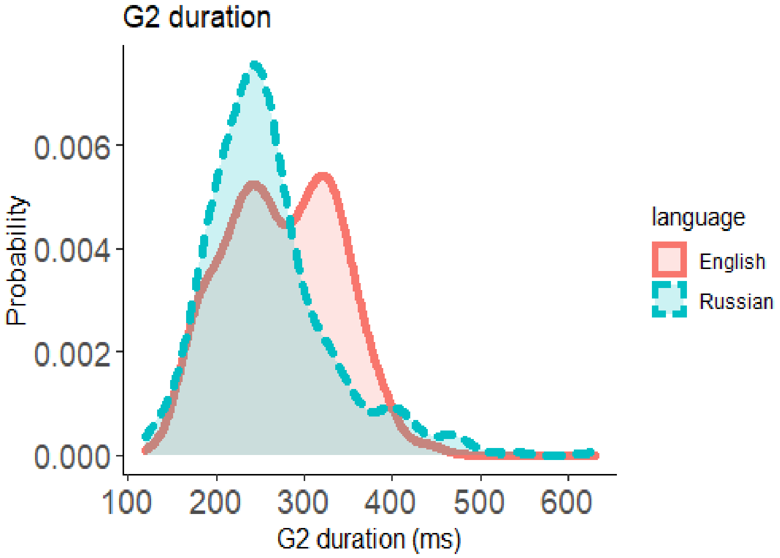


Figure 16: The distribution of G2 (palatal gesture) duration by language

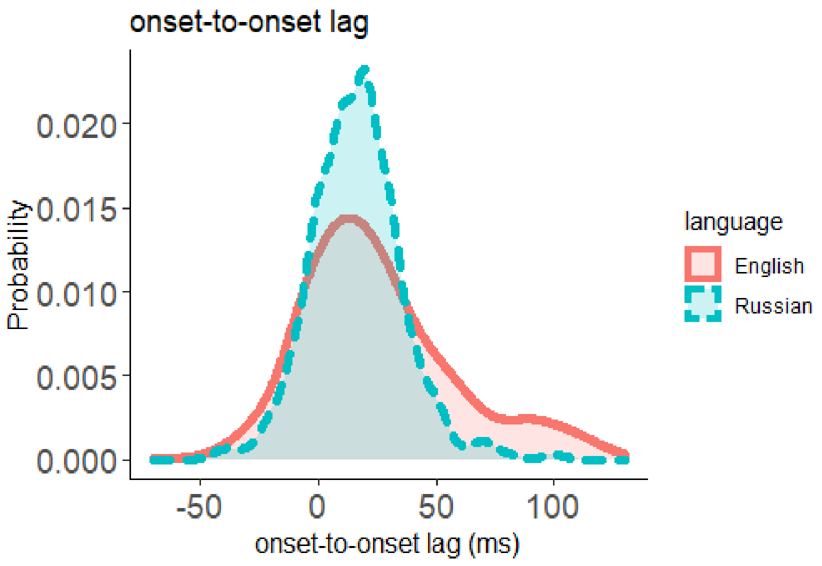


Figure 17: The distribution of onset-to-onset lag measurements by language

The Figures above indicate that, as expected, the palatal and labial gestures of English and Russian are quite similar. By considering how the variability summarized in Figure 15 relates to the variability in Figure 17, I can adjudicate between my competing hypotheses. The key insight is that the token-to-token kinematic variability is shaped uniquely by the dynamics. The dynamical control regime, formalized as a characteristic pattern of gestural coordination (Figure 7), that I have hypothesized for complex segments predicts that *G_1 duration* is independent of *onset-to-onset* lag (Figure 9(b), (d)). In contrast, the control structure for segment sequences predicts that these dimensions should be positively correlated (Figure 9(a), (c)). Crucially, it is natural variability in the kinematics that exposes patterns of gestural coordination characteristic of phonological structure: complex segments vs. segment sequences.

I have already shown that the distribution of G_1 duration, the duration of labial consonants, is similar in this data for both English and Russian, and that onset-to-onset lag distributions have a similar mean value. I now turn to the relation between these variables.

2.6.2. Temporal coordination

Figure 18 plots the relation between *G_1 duration* and *onset-to-onset* lag for each language. To illustrate the trend in the data, a least squares linear regression line is fit to each panel. The trends can be compared directly to the simulation results in Figure 9. For English, there is a positive correlation, as predicted by the segment sequence hypothesis. As G_1 duration increases, so too does onset-to-onset lag. For Russian, the regression line is nearly flat, showing only a slight upward trend, as predicted by the complex segment hypothesis. When compared to the simulation results in Figure 9, the English data most closely resemble Figure 9(c), segment sequences with negative lag, and the Russian data most closely resemble Figure 9(d), complex segments with positive lag.

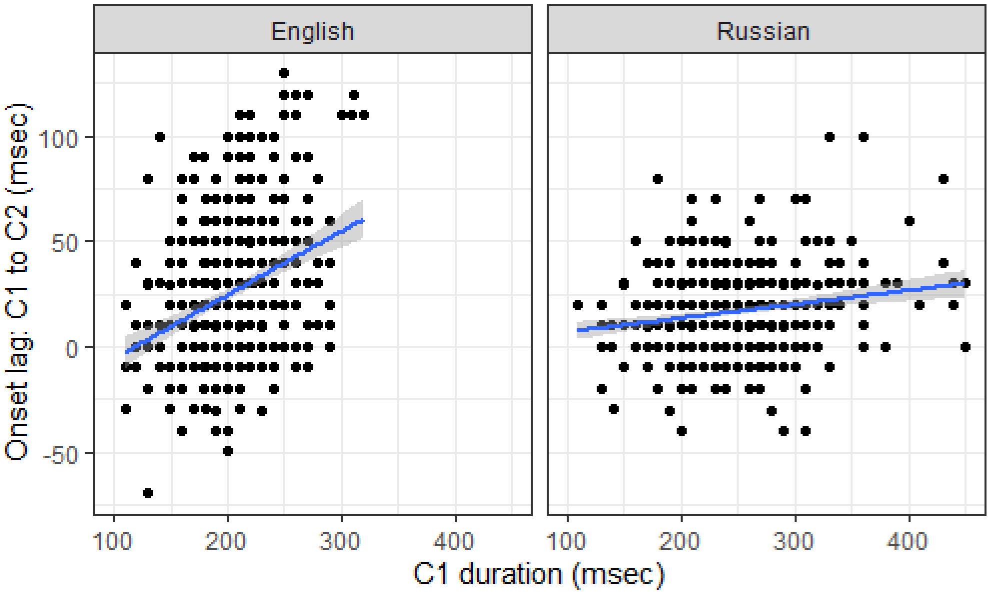


Figure 18: A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis) for each language – English (left) and Russian (right)

To assess the statistical significance of the trends in Figure 18, I fit a series of linear mixed-effects models to the data (for additional detail, see Section 2.5.5). As shown in Table 2, the addition of G_1 duration significantly improves the baseline model, which contains only random intercepts for subject and item. The addition of *language* as a fixed factor leads to additional modest improvement—the log-likelihood of the data given the model with *language* as a fixed effect (-4839.88) is greater than the log-likelihood of the simpler model, which includes only G_1 duration (-4842.37); moreover, the AIC decreases by about 3, from 9694.7 to 9691.8. In the final model, the addition of the interaction term leads to more substantial improvement ($\chi^2 = 47.3$, $p < 0.001$). The additional variance explained by the interaction term decreases AIC from 9691.8 to 9646.4 for the model with the *G_1 duration * language* interaction. Such a decrease in AIC of about 45 is sizable; to put this into context, Burnham and Anderson (1998) suggest that a difference in AIC of 9-10 is already big. The significant improvement contributed by the interaction term indicates that the influence of G_1 duration on onset-to-onset lag is different for the different language groups.

Table 2: Nested model comparison—each model is compared pairwise with a progressively more complex model, i.e., one additional degree of freedom. All additions lead to significant improvement and lowered AIC. The best-fitting model includes the interaction between G_1 duration and language

LME Model comparison (*onset-to-onset*~)	**Df**	**AIC**	**logLik**	**χ^2**	**Pr(>χ^2)**
1 + (1|subject)+(1|item)	4	9749.6	-4870.78	NA	NA
1 + G_1 duration + (1|subject)+(1|item)	5	9694.7	-4842.37	56.83	<0.001
1 + G_1 duration + language + (1|subject)+(1|item)	6	9691.8	-4839.88	4.97	0.026
1 + G_1 duration * language + (1|subject)+(1|item)	7	9646.4	-4816.22	47.33	<0.001

Table 3 summarizes the best fitting model. The intercept of ~6 ms approximates the average onset-to-onset lag, as observable in Figure 14, for Russian. The main effect of G1 duration is positive but very small (0.047 ms; $t = 2.03$, $p = 0.043$). This weak main effect for G1 duration is likely due to the highly significant interaction in the model between G1 duration and language. The combination of coefficients for *language* and the *G1*language* interaction, both highly significant, explain the differential effect across languages. The coefficient for *language* is -45.466 ms, which places the estimate for English much lower than the intercept value (Russian). The negative effect of *language* is offset by the positive *G1*language* interaction. For English only, the effect of G1 duration is large (0.265 ms) and highly significant ($t = 6.99$, $p < 0.0001$). For each millisecond increase in G1 duration, onset-to-onset lag in English relative to Russian increases by 0.265 milliseconds. This is the positive trend reflected in the English panel (left) of Figure 18.

Table 3: Summary of fixed factors in the best-fitting model (reference level for language = Russian)

	Estimate	**Std.Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	6.146	7.335	41	0.84	0.40692
G1_duration	0.047	0.023	700	2.03	0.043
language_English	-45.466	10.487	48	-4.34	0.00007
G1 duration*language	0.265	0.038	973	6.99	<0.00001

In sum, the statistical models confirm the trend observable in Figure 18. With respect to the predictions in Section 2.4, Russian palatalized consonants behave like complex segments while the English counterparts, although phonetically very similar to Russian in many respects, behave like segment sequences.

2.7. Discussion

2.7.1. Overview

In this chapter, I investigated temporal coordination in complex segments versus segment sequences in Russian and English to establish a quantification of palatalization in Russian. Both complex segments and segment sequences involve multiple gestures, in the sense of Articulatory Phonology (Browman & Goldstein, 1986 et seq.), where a gesture is both a unit of phonological contrast and a specification of articulatory dynamics. Moreover, the individual gestures involved in a contrast based on a simplex vs. complex segment distinction, e.g., /b/ vs. /b^j/, can be quite similar, even identical, to a contrast based on a single segment vs. segment sequence distinction, e.g., /b/ vs. /bj/. The distinct phonological behavior exhibited by complex segments (see Section 2.2) can be used to diagnose them as phonologically distinct from sequences. This chapter addressed whether there is also a revealing difference in how the component gestures of complex segments vs. segment sequences are coordinated in time. Such a difference could support a phonological distinction based not on the individual dynamics of the constituent gestures but their mode of coordination. A difference in gestural coordination conditions distinct kinematic patterns, providing a basis through which phonological structure can be diagnosed through a phonetic signal.

This chapter provided robust support for the temporal hypothesis. Results indicate that gestural coordination for complex segments (Russian) differs from segment sequences (English). Specifically, the Russian data (but not the English data) is consistent with the hypothesis that the constituent gestures of complex segments are coordinated according to their gesture onsets. The English data is instead consistent with the hypothesis that segment sequences are coordinated according to the offset of the first gesture and the onset of the second. These hypothesized coordination modes are roughly equivalent to synchronous (in-phase) and sequential (anti-phase) coordination, modes hypothesized to be intrinsically stable in speech (Nam et al., 2009), except that the possibility of a consistent +/-lag is incorporated into my models. In many ways, palatalized labials in Russian were phonetically similar to labial-glide sequences in English. This can be seen in, e.g., the measurements of gesture duration (Figure 15) and even in the kinematic trajectories (Figure 13: English, Figure 14: Russian). Moreover, the average degree of overlap between gestures, as indicated by the onset-to-onset lag measure, was also quite similar (Figure 17) and not significantly different. The key difference related to the hypothesis is that the languages differ in the relative timing of similar labial and palatal gestures. The predictions of this hypothesis were borne out in the data.

My approach to exposing differences in coordination makes use of the natural variation present in the data. Trial-by-trial variability in the duration of the labial consonant is correlated with onset-to-onset lag, as predicted, only for segment sequences (English) and not for complex segments (Russian). The positive correlation for segment sequences is predicted by the main hypothesis (Figure 9). Since, in the case of segment sequences, the second gesture is timed to the offset of the first, any increase in first gesture duration also delays the onset of the second gesture (relative to the onset of the first gesture). This is not the case for complex segments; by hypothesis, complex segments are coordinated with reference to gesture onsets. Therefore, variation in the first gesture duration is orthogonal to triggering the onset of the second gesture. The data presented here provide clear support, replicating patterns reported in Shaw et al. (2019), based on already collected data (see Section 2.3).

2.7.2. Why not just look within Russian?

In this chapter, I pursued a cross-language comparison between a case that is unambiguously a complex segment, the palatalized consonants of Russian, and a case that is unambiguously a segment sequence, consonant-glide sequences in English. However, since Russian exhibits a within-language contrast between C^j and C+j (e.g., /p^jok/ ‘bake (3ps past)’ vs. /p̲jot/ ‘drink (3ps pres)’), it might seem that my hypothesis could be tested within Russian. A problem with this is that the consonant in C+j is reported to be (at least variably) palatalized (Diehm, 1998; Kochetov, 2011; Suh & Hwang, 2016), resulting in a sequence of a complex segment and a glide, e.g., /pjot/ [p^jjot] ‘drink (3ps pres)’. Notably, since at least before labial consonants, there is not a three-way contrast between C^j, Cj, C^jj, a labial consonant before a palatal glide could freely vary between a plain and palatalized variant. Because of this possibility for variation, the within-language contrast between /C^j/ and /Cj/ would make for a less conclusive test of my main hypothesis. Indeed, given the claims that plain consonants are palatalized before a palatal glide (coarticulatory palatalization), I examine the phonetic realization of underlying and coarticulatory palatalization implementing the temporal diagnosis for complex segments and segment sequences (See Chapter 3). The cross-language approach to testing my main hypothesis allows us to avoid the complication of underlying vs. coarticulatory palatalization in Russian.

2.7.3. Why is there a slightly positive correlation between G1 duration and onset-to-onset lag for complex segments?

In Section 2.4, I hypothesize that the gestures of complex segments are coordinated differently than gestures of segment sequences. In particular, segment sequences are hypothesized to be coordinated with reference to the offset of the first gesture and the onset of the second, leading to a positive correlation between *G1 duration* and *onset-to-onset lag*. In contrast, the gestures of complex segments are hypothesized to be coordinated with reference only to gesture onsets, resulting in no correlation between them. Although there is a clear difference in the slope of the line between English and Russian, and the effect of *G1 duration* on *onset-to-onset lag* was significantly different across languages, the regression line for Russian was not entirely flat showing a slight upward trend (Figure 18). If palatalized consonants in Russian are predicted to show complex segment timing, why is there a slightly positive correlation between G1 duration and onset-to-onset lag for complex segments?

There are two possible explanations for this upward trend. First of all, it might be attributable to stochastic variation. The coupled oscillator model (Goldstein et al., 2006; Goldstein et al., 2009; Nam et al., 2009; Saltzman, Elliot et al., 2008) predicts that the correlation between *G1 duration* and *onset-to-onset lag* is unlikely to be zero due to stochastic variation in the intrinsic frequencies of the oscillators in the system of coupled oscillators. In fact, this slightly upward trend can also be observed in the simulations for complex segments from the stochastic modeling (See Figure 9). Another factor that may influence the correlation is speech rate. All else being equal, a positive correlation is expected between temporal intervals, because both will be influenced by a similar set of token-specific factors, such as, most notably, speech rate. This is true as well of *G1 duration* and *onset-to-onset lag*.

The following question is then where the cutoff is between the trend line that diagnoses a complex segment and the trend line that is representative of a segment sequence. This decision is subject to general procedures of statistical inference. It is still somewhat common to define thresholds of statistical significance. For example, we could say that the correlation is statistically significant if it crosses some threshold. Another way is to consider the value of the correlation predicted by a stochastic model, such as the one in this dissertation (See Figure 9), that is tuned to the data.

2.8. Summary

Evidence from articulatory kinematic data collected with Electromagnetic Articulography on Russian palatalized consonants and English consonant-glide sequences provided support for the hypothesis that complex segments differ from segment sequences in how the constituent gestures are coordinated. The gestures of complex segments, exemplified by palatalized consonants in Russian, are coordinated according to gesture onsets, such that the onset of one gesture provides the trigger to initiate the second gesture. The gestures of segment sequences in English, in contrast, are coordinated such that the offset of the first gesture triggers the onset of the second gesture. These distinct patterns of coordination can be masked in kinematic measures of temporal overlap, but are clearly revealed in patterns of covariation between temporal intervals. Token-by-token variability exposes distinct patterns of coordination unambiguously. This point was argued analytically, demonstrated through computational simulation, and verified in the experimental data. In this chapter, I examined temporal coordination in complex segments versus segment sequences in Russian and English, respectively, and established a way to quantify palatalization (or lack thereof) for consonants preceding a palatal glide. In the following chapter, the quantification of palatalization will be used to examine a case of putative phonological neutralization of palatalized consonants (underlying palatalization; e.g., /bʲ/) and plain consonants preceding a palatal glide (coarticulatory palatalization; e.g., /bj/) in Russian.

Chapter 3. Russian palatalization as incomplete neutralization

3.1. Introduction

Russian contrasts palatalized and plain (non-palatalized) consonants (so-called “soft” and “hard” consonants, respectively) (e.g., Avanesov, 1972; Kochetov, 2004; 2006; Padgett, 2001; 2003; Timberlake, 2004). The consonant inventory of Contemporary Standard Russian is illustrated in Figure 19. Palatalized and plain consonants are contrastive before back vowels both syllable-initially or word-initially (13a). The contrast is also maintained word-finally (13b). Before /i/, the contrast between palatalized and plain consonants is licensed by backing of /i/ to [ɨ] after plain consonants, as shown in (13c). The contrast is neutralized before /e/ with the exception of historical loanwords such as /kep/ [kep] ‘cap’ (e.g., Padgett, 2001; Padgett, 2003). In most cases, plain consonants are palatalized before /e/ as shown in (13d). In word-medial clusters, the contrast between palatalized and plain consonants is maintained in heterorganic medial clusters (13e), while the contrast is neutralized in homorganic medial clusters (13f) (Kochetov, 2006).

p	pʲ	t	tʲ			k
b	bʲ	d	dʲ			g
f	fʲ	s	sʲ	ʃ	ʃʲ	x
v	vʲ	z	zʲ	ʒ		
		ts		tʃʲ		
m	mʲ	n	nʲ			
		l	lʲ			
		r	rʲ			
				j		

Figure 19: Consonant inventory of Contemporary Standard Russian (adopted from Padgett, 2003, p.309)

(13) Contrast between palatalized and plain consonants (a-d from Padgett, 2001; e-f from Kochetov, 2006)

(a) Word-initial position (before back vowels)

/mʲat/ [**mʲ**at] ‘crumpled (past part.)’ vs. /mat/ [**m**at] ‘mat’

/vʲol/ [**vʲ**ol] ‘he led’ vs. /vol/ [**v**ol] ‘ox’

(b) Word-final position

/matʲ/ [ma**tʲ**] ‘mother’ vs. /mat/ [ma**t**] ‘mat’

/krofʲ/ [kro**fʲ**] ‘blood’ vs. /krof/ [kro**f**] ‘shelter’

(c) Before /i/

/bʲit/ [bʲ**i**t] ‘beaten’ vs. /bit/ [b**ɨ**t] ‘way of life

/vʲit/ [vʲ**i**t] ‘beaten’ vs. /vit/ [v**ɨ**t] ‘beaten’

(d) Before /e/

/sestʲ/ [**sʲ**estʲ] ‘to sit down’ *[**s**estʲ]
/petʲ/ [**pʲ**etʲ] ‘mother’ *[**p**etʲ]

(e) Heterorganic medial clusters

/katʲka/ [ka**tʲk**ɐ] ‘Katya (name; fam.)’vs. /katka/ [ka**tk**ɐ] ‘pail’

/rʲɛtʲka/ [rʲɛ**tʲk**ɐ] ‘radish’ vs. /rʲɛtka/ [rʲɛ**tk**ɐ] ‘rare’

(f) Homorganic medial clusters

/putʲ/ [pu**tʲ**] ‘way’ → /putʲ-nij/ [pu**tn**ɨj] ‘appropriate’

/pʲatʲ/ [pʲa**tʲ**] ‘five’ → /pʲitʲ-nattsatʲ/ [pʲɪ**tn**attsɐtʲ] ‘fifteen’

As discussed in Section 1.3, the contrast between palatalized and plain consonants in Russian is also neutralized when a plain consonant is followed by a glide (e.g., Kochetov, 2011). That is, both a palatalized consonant and a plain consonant preceding a palatal glide are realized as a palatalized consonant. For example, a plain consonant preceding a palatal glide in /p̲jot/ is realized as a palatalized consonant [pʲ], resulting in neutralization of the contrast between plain and palatalized consonants in Russian, as shown in (14). I refer to the palatalized consonants as underlying palatalization and the plain consonant preceding a palatal glide as coarticulatory palatalization.

(14)	Palatalized segments (Underlying palatalization)		Plain consonants preceding a palatal glide (Coarticulatory palatalization)
	/pʲatij/ [pʲatɨj] ‘fifth’	vs.	/pjanij/ [pʲjanɨj] ‘drunk’
	/pʲok/ [p̲ʲok] ‘bake (3ps past)’	vs.	/p̲jot/ [p̲ʲjot] ‘drink (3ps pres)’
	/bʲust/ [bʲust] ‘bust’	vs.	/bjut/ [bʲjut] ‘beat (3p pl)’
	/dʲatel/ [dʲatʲel] 'woodpecker'	vs.	/djakon/ [dʲjakon] 'deacon'
	/rʲadom/ [rʲadom] ‘near’	vs.	/rjanij/ [rʲjanɨj] ‘zealous’
	/lʲut/ [lʲut] ‘fierce’	vs.	/ljut/ [lʲjut] ‘pour (3p pl)’
	/sʲomga/ [sʲomga] ‘salmon’	vs.	/s-jomka/ [sjomka] ‘(film) shooting’

As noted in Footnote 5, some consonant-glide sequences are morphologically derived, (e.g., /pj-a-n-ij/ from /pʲi-tʲ/ ‘to drink’ via /i/-gliding), while others are underlying, (e.g., /djakon/ and /rjanij/, at least synchronically). In addition, consonant-glide sequences can occur morpheme-internally (as in the examples above) and across morphemes (prefix + stem and stem + suffix: e.g., /s-jom-k-a/, /brat-ja/) or words (preposition + stem; e.g., /s jamoj/ ‘with a pit’). Plain consonants before a palatal glide in tautomorphemic and stem + suffix sequences are realized as non-contrastively palatalized (e.g., /djakon/ [dʲjakon] ‘deacon’), with the exception of prefix-stem boundaries (e.g., /pod-jom/ [podjom] ‘rise, lift’). However, previous studies have described this coarticulatory palatalization as variable in the case of labial consonants, e.g., /pjanij/ [pʲjanɨj] ~ [pjanɨj] ‘drunk’ (Avanesov, 1972, pp. 348-377). The current dissertation focuses on the neutralization of palatalized and plain consonants in this context and examines phonetic realization of underlying and coarticulatory palatalization.

Previous studies have also reported that the “plain” stops possibly have a secondary articulation, involving retraction of the tongue dorsum (velarization/uvularization, see Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963). Skalozub (1963) is one of the early studies which systematically examined plain and palatalized consonants in Russian, using X-ray imaging, artificial palatography, odontography, and partial oscillography. Based on articulatory results from 4 Russian speakers, Skalozub (1963) argued that plain consonants, at least lateral /l/ and labial consonants, are velarized.

Recent ultrasound studies by Litvin (2014) and Roon and Whalen (2019) further confirmed that plain consonants in Russian are velarized (and/or uvularized). Litvin (2014) examined plain fricatives and /l/ across different vowel contexts [a] and [ɛ]. Ultrasound data from six Russian speakers revealed that regardless of vowel context /l/ and /f/ are uvularized and /s/ and /ʂ/ are either velarized or uvularized. Roon and Whalen (2019) have also shown that plain consonants in Russian are velarized (and/or uvularized), subject to intra-speaker variation. In particular, articulatory data from three Russian native speakers revealed that there are consistent and discernable dorsal gestures regardless of the manner and syllable position (initial vs. final), at least within labials [p, f, m], but the location of constriction varied by speaker (velar to uvular).

A question that arises from consideration of these findings is whether the neutralization between plain and palatalized segments in Russian is phonetically complete. If plain consonants have secondary velarization/uvularization, it is predicted that this secondary velarization/uvularization will have detectable effects on the coarticulatory palatalization occurring in consonant-glide sequences, distinguishing this palatalization from underlying palatalization. As discussed above, previous descriptions of the variable realizations of palatalization, at least for labial consonants, suggest that the contrast between palatalized and plain consonants in Russian may not be neutralized, or even if it is neutralized, the neutralization is incomplete.

To this end, this dissertation examines phonetic realization of underlying and coarticulatory palatalization, focusing on palatalized labial consonants and plain labial consonants situated in palatal-glide sequences. To resolve the issue of quantifying palatalization (see the discussion in Section 1.0), I explore the incompleteness of Russian palatalization using the temporal diagnosis of complex segments and segment sequences that I established in Chapter 2.

In Section 3.2, I review past acoustic and kinematic studies on Russian palatalization. I then lay out my hypotheses and predictions in Section 3.3. In particular, I hypothesize that the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) would lead to incomplete neutralization of the underlying palatalization and coarticulatory palatalization in Russian. Then, I transition to an empirical test of the hypotheses. In particular, I conducted an Electromagnetic Articulography (EMA) experiment examining temporal coordination and the spatial position of the tongue body for underlying and coarticulatory palatalization. The methods of the experiment are described in section 3.4, and the results are reported in section 3.5. The discussion and the summary are presented in sections 3.6 and 3.7, respectively.

3.2. Past results on the Russian palatalization

Independent of whether the contrast between underlying and coarticulatory palatalization is neutralized or not, the consonant-glide sequence itself is not necessarily identical to the palatalized consonant. In fact, previous studies reported that there is a perceivable difference between palatalized consonants (C^j) and consonant-glide sequences (Cj). For example, Ladefoged and Maddieson (1998, p. 364) reported that consonant-glide sequences (Cj) show short F2 steady-state duration, while the falling of F2 starts immediately after consonant release for palatalized consonants (C^j).

Similarly, Diehm (1998) examined acoustic characteristics of palatalized consonants (C^j) and consonant-glide sequences (Cj) in Russian produced by native speakers of Russian and learners of Russian. The results from eight Russian native speakers (4 male and 4 female) revealed that consonant-glide sequences (Cj) exhibited significantly higher F2 at the transition onset than palatalized consonants (C^j) (2704 Hz vs. 2362 Hz for female; 2233 Hz vs. 2012 Hz for male). In addition, she reported that consonant-glide sequences (Cj) showed significantly longer F2 steady-state duration than palatalized consonants (C^j) (117 ms vs. 33 ms for female mean value; 102 ms vs. 25ms for male mean).

In addition, Suh and Hwang (2016) also examined palatalized consonants (C^j) and consonant- glide sequences (Cj) in Russian and compared them with palatal glides in Korean. To measure glide duration, they first measured the vocalic duration comprising the j+V portion (from the onset of the vocoid to the offset of the vowel). Then, they calculated the durational ratio of the j+V portion to the pure vowel duration in CV. The results from five Russian native speakers revealed that the vocalic duration comprising the j+V portion of CjV syllables is significantly longer than the j+V portion of C^jV syllables.

These acoustic results confirm that there are salient acoustic cues to the difference between palatalized consonants (C^j) and consonant-glide sequences (Cj). These differences likely reflect the difference between the existence of a glide gesture as a secondary articulation and the glide gesture as a separate segment. Crucially, however, the acoustic differences between consonant-glide sequences (Cj) and palatalized consonants (C^j) do not provide any information as to whether the “plain” consonant in the consonant-glide sequences is palatalized or not. With regard to evaluating any incomplete neutralization, the requisite articulatory quantification of palatalization (or lack thereof) of the consonants preceding a palatal glide has yet to be determined.

Still, articulatory studies of Russian have shown differences that are consistent with the observations from acoustic data. For example, Kochetov (2006) examined the effect of syllable position on gestural organization, using kinematic data from EMMA (Electromagnetic Midsagittal Articulometer) to compare the articulatory patterns exhibited by a palatalized stop (/p^j/), a plain stop (/p/), and a palatal glide (/j/) in the productions of four native speakers of Russian. The results revealed that the palatal gesture is longer when it occurs as a segment in /p#j/ sequences than when it occurs as secondary palatalization in /p^j/. In addition, and of particular interest to the present study, Kochetov showed that the relative timing of the labial gesture and the palatal glide gesture in stop-glide sequences (/p#j/) differs from the relative timing of these gestures in palatalized stops like /p^j/. More specifically, the glide gesture is achieved later in the stop-glide sequence (/p#j/) than in the glide gesture for the palatalized stop /p^j/. This is illustrated in Figure 20. However, since the stop-glide sequence occurs across word boundaries, it is unclear whether the delayed glide gesture in the segment sequence is due to the characteristics of the segment sequence or from confounding effects that prosodic boundaries have on articulatory timing. Consequently, the difference in the delayed achievement lag for /p#j/ and /p^j/ is not a valid criterion for accessing incomplete neutralization of underlying and coarticulatory palatalization in Russian, nor is it a valid criterion for distinguishing complex segments and segment sequences more generally.

For these reasons, an ideal test to examine the incomplete neutralization of underlying and coarticulatory palatalization in Russian would better control for prosodic context. In the next section, I present my hypotheses regarding incomplete neutralization of underlying and coarticulatory palatalization in Russian, implementing the temporal diagnostics for complex segmenthood introduced in section 3.3.

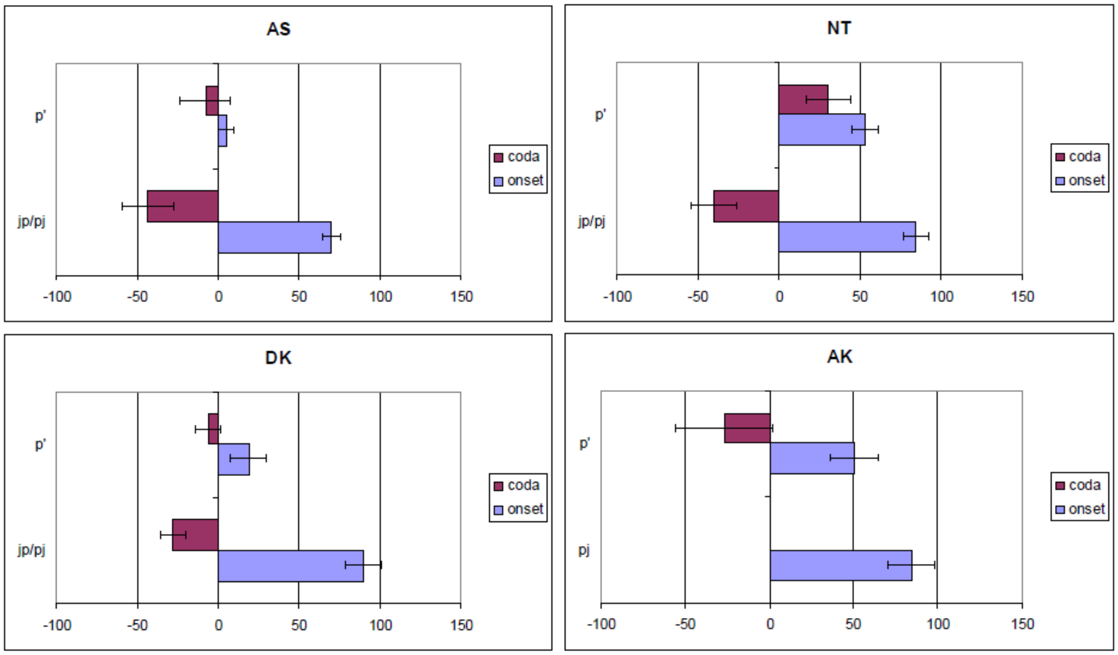


Figure 20: Mean values for achievement lag for /p^j/ in onset and coda, and for the sequences /p#j/ (compared with onset) and /j#p/ (compared with coda) in nonwords (adopted from Kochetov, 2006, p. 575)

3.3. Research questions and Predictions

The fundamental question of this section of this dissertation is whether two cases of Russian palatalization represent a case of incomplete neutralization. The research questions are as follows:

- Research question 1: Do underlying palatalization (e.g., /b^j/) and coarticulatory palatalization (e.g., /bj/) exhibit temporal coordination of complex segments?
- Research question 2: Do underlying palatalization (e.g., /b^j/) and coarticulatory palatalization (e.g., /bj/) exhibit spatial and/or temporal differences?

The first research question addresses whether two cases of Russian palatalization show neutralization. If plain consonants preceding a palatal glide (coarticulatory palatalization) are palatalized, this results in neutralization of underlying and coarticulatory palatalizations in Russian. I examine the neutralization using the temporal diagnostics of complex segments and segment sequences that are discussed in Chapter 2. That is, if Russian palatalization exhibits neutralization, both underlying and coarticulatory palatalizations will show the temporal coordination of complex segments. In contrast, if Russian palatalizations exhibit no neutralization, the underlying palatalization will show the temporal coordination of complex segments, while the coarticulatory palatalization will exhibit the temporal coordination of segment sequences.

The second research question addresses whether the neutralization is complete (if the neutralization exists). That is, if there are spatial and/or temporal differences between the underlying and coarticulatory palatalizations, it would be considered to be incomplete neutralization. Considering that plain consonants also have secondary velarization, I examine the completeness of the neutralization using the spatial position of the tongue body, as well as the temporal lag between the onset of the labial gesture and the onset of the palatal gesture.

Consequently, there are three possible outcomes depending on the temporal organization and spatial and/or temporal differences of underlying and coarticulatory palatalization: no neutralization, complete neutralization, and incomplete neutralization.

- No neutralization: Underlying palatalization shows temporal coordination of complex segments, while coarticulatory palatalization exhibits temporal coordination of segment sequences. Also, there are significant spatial and/or temporal differences of underlying and coarticulatory palatalization, and the differences are substantial.
- Complete neutralization: Both underlying and coarticulatory palatalizations show the temporal coordination of complex segments, and there are no significant spatial and/or

temporal differences of underlying and coarticulatory palatalization.

- Incomplete neutralization: Both underlying and coarticulatory palatalizations show the temporal coordination of complex segments, and yet there are small but significant spatial and/or temporal differences of underlying and coarticulatory palatalization.

Given that plain consonants have secondary velarization (Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963), I predict that the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) in coarticulatory palatalization will lead to incomplete neutralization of underlying and coarticulatory palatalization in Russian.

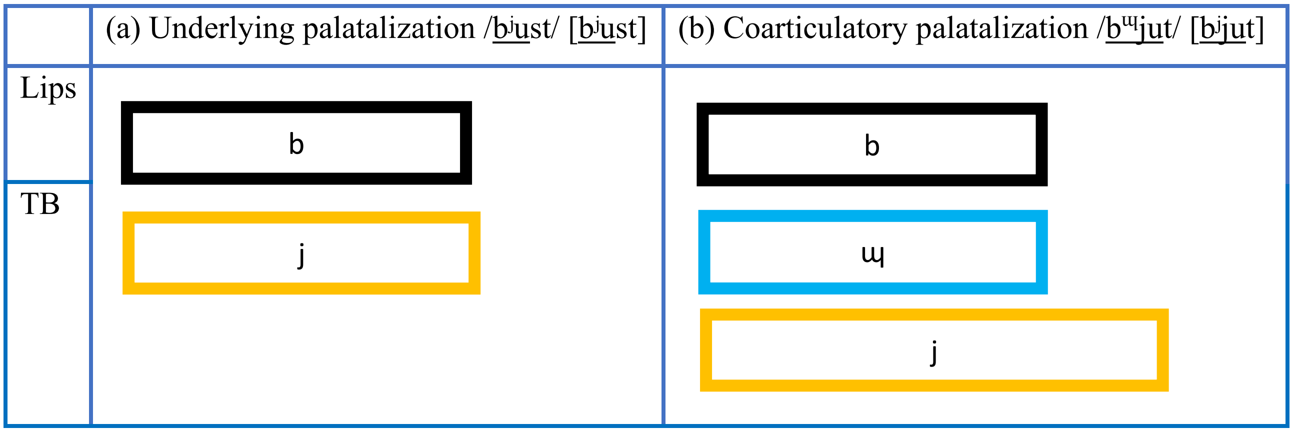


Figure 21: Predicted gestural scores for underlying (a) and coarticulatory palatalization (b) in Russian (incomplete neutralization)

As schematized in Figure 21 (a), there is a labial gesture and a palatal gesture for the underlying palatalization, while panel (b) shows that coarticulatory palatalization has a velar gesture on top of the labial and palatal gestures. The gestural overlap on the same tract variable (i.e., palatalization vs. velarization on the TB tract) would lead to gestural blending between these two gestures, resulting in a slightly more retracted tongue position for coarticulatory palatalization compared to underlying palatalization, which only has the palatal gesture on the TB tract.

Consequently, this difference would lead to incomplete neutralization between underlying and coarticulatory palatalization in Russian.

3.4. Methods

3.4.1. Participants

The same four native speakers of Russian who participated in the experiment described in Chapter 2 also participated in this experiment (3 male and 1 female). All speakers were in their 20s at the time of recording and living in the United States. The Russian speakers were born in Russia and moved to the United States as adults.

3.4.2. Materials

The materials included six closely matched pairs representing two conditions: palatalized consonants vs. plain consonants preceding a palatal glide (UNDERLYING vs. COARTICULATORY palatalization). In all cases, the primary word stress falls on the first syllable, and the vowel immediately following is either /u/ or /o/, as shown in Table 4. The carrier phase is shown in (15).

Table 4: Russian target words

Palatalized consonants (UNDERLYING palatalization)			**Consonant-glide sequences (COARTICULATORY palatalization)**		
word	**IPA**	**gloss**	**word**	**IPA**	**gloss**
пёк	/pʲok/	bake (3ps past)	пьёт	/pjot/	drink (3ps pres)
бюст	/bʲust/	bust (breast/sculpture)	бьют	/bjut/	beat (3pp pres)
мю	/mʲu/	Greek letter	Мью	/mju/	a Pokémon name
Фёдор	/fʲodor/	Fyodor (name)	фьорд	/fjord/	fjord
вёз	/vʲoz/	carry (3ps past)	вьёшь	/vjoʂ/	weave (2ps pres)
вёдра	/vʲodra/	bucket (pl)	вьёт	/vjotsa/	weave (3ps pres refl)

(15) Carrier phrases

Она ____ повторила [ʌˈna ____ pəftʌˈrʲilʌ]. ‘She ____ repeated.’

3.4.3. Procedure

Data collection was executed in the Phonetics Lab at Yale University. Following the same procedure as described in Section 2.5.3, the articulatory and acoustic data were simultaneously recorded by means of 5D Electromagnetic Articulography (EMA) and an audio-recording setup. To collect articulatory data, 9 sensors were attached to the participants. Sensors, attached to the upper and lower lips, jaw, tongue tip (TT), tongue blade (TB), and tongue dorsum (TD), were tracked using the NDI Wave Speech Production System. Reference sensors on the left/right mastoids and nasion were used to computationally correct for head movements. As a post-processing procedure, the data was computationally corrected for head movements and rotated to the occlusal plane so that the bite of the teeth serves as the origin of the spatial coordinates. I also calculated a lip aperture trajectory, as the Euclidean distance between the upper and lower lip sensors. See Section 2.5.3 for detailed descriptions of the procedure.

3.4.4. Analysis

As discussed in Section 2.5.4, the post-processed data was visualized in MVIEW (Tiede, 2005). Changes in Lip Aperture, computed as the Euclidean distance between the upper and lower lip sensors over time, were used to identify labial gestures. The TB sensor indexed the palatal gesture. Gestural landmarks were parsed with reference to the velocity signal using the *findgest* function in MVIEW. Specifically, the gesture *Onset* and *Target* landmarks were labeled at 20% of peak velocity in the movement toward constriction (See Figure 10). *Release* and *Offset* landmarks were labeled at a 20% threshold of peak velocity in the movement away from constriction. As illustrated in Figure 11, the two key temporal intervals computed from these articulatory landmarks were (1) *G_1 duration* defined as the interval from *Onset* to *Offset* of the labial gesture; and *onset-to-onset lag* is defined as the interval between the *Onset* of the labial gesture (*G_1*) and the *Onset* of the palatal gesture (*G_2*). See Section 2.5.4 for a detailed description. In addition to temporal coordination, the current study measured the spatial position of the TB sensors to assess any impact of underlying velarization/uvularization on the realization of coarticulatory palatalization. The spatial position of the TB sensors was normalized using z-scores for each speaker. Before proceeding with statistical analysis, I removed outliers that were greater than three standard deviations from the speaker-specific mean value of either *G_1 duration*, 7 tokens removed (0.6% of the data), or *onset-to-onset lag*, 18 tokens removed (1.6% of the data).

To examine the neutralization of Russian palatalization, the correlation between *onset-to-onset lag* and *G_1 duration* was analyzed. As *G_1 duration* varies, I ask whether *onset-to-onset lag* will positively covary, or whether these intervals will be statistically independent. As discussed in Section 3.3, if the contrast between a palatalized consonant (underlying palatalization) and a plain consonant preceding a palatal glide (coarticulatory palatalization) in Russian is preserved (no neutralization), underlying palatalization will show no correlation between consonant duration and *onset-to-onset lag*, while for coarticulatory palatalization, *onset-to-onset lag* will increase with *G_1 duration*, leading to a positive correlation between them. However, if the contrast is neutralized, both the underlying and coarticulatory palatalization will exhibit no correlation between *G_1 duration* and *onset-to-onset lag*.

I, therefore, treat *onset-to-onset lag* as a dependent variable and evaluate whether *G1 duration* and *Status* are significant predictors. *Speaker* and *Item* were added as random-effects in linear mixed-effects models, which I fit to the data using the *lme4* package in R (Bates et al., 2014). To a baseline model, consisting of random intercepts for subjects and items, I added fixed factors of interest incrementally. First, I added *G_1 duration*, then *Status* (UNDERLYING vs. COARTICULATORY, with UNDERLYING as the reference level), and finally the interaction between *G_1 duration* and *Status*. This gives a set of four nested linear mixed-effects models. I evaluated the significance of each fixed factor through model comparison. The fixed factor of primary interest is the interaction term: *G_1 duration * Status*. This is due to the effect of *G_1 duration* on *onset-to-onset lag*. This is because *G_1 duration* is predicted to have a positive influence on *onset-to-onset lag* for coarticulatory palatalization, but not for underlying palatalization if the contrast is not neutralized. On the other hand, if the contrast is neutralized, both palatalizations will exhibit the same pattern, showing no correlation between *G_1 duration* and *onset-to-onset lag*.

To assess the incompleteness of the neutralization, the current study also examined the effect of *Status* on the *TB position* at palatal gesture onset. If the neutralization is complete, there will be no difference in the *TB position* depending on the *Status*. However, if the neutralization is incomplete, the coarticulatory palatalization will exhibit a more retracted tongue position than the underlying palatalization. To test this, separate linear mixed-effects models were run with *TB position* as a dependent variable and *Status* as a significant predictor. *Speaker* and *Item* were added in the models as random-effects, and models were constructed in an incremental fashion by adding a fixed-effects factor. A similar approach was used for testing the effects of *Status* on *onset-to-onset lag*.

3.5. Results

3.5.1. Kinematic trajectories and distribution

I first examine the continuous kinematic trajectories of relevant articulators for underlying and coarticulatory palatalization in Russian. Figure 22 illustrates variability across kinematic trajectories for the token /bʲust/ (**UNDERLYING** palatalization), as produced by the four Russian speakers in the study. The figure plots the Lip Aperture trajectory in the upper panels and the Tongue Blade (TB) trajectory in the lower panels. Each trajectory is represented by a different color; the thick dotted line shows the average trajectory. As discussed in Section 2.6.1, a temporal window of the trajectories is 600 ms – from 100 ms before the onset landmark of the lip aperture gesture to 500 ms following this landmark. The level of variability in the magnitude of the gestures varies by subject. For R2, most tokens occur tightly clustered around the mean; R1 and R3 show more variability, and R4 shows even more. On the other hand, the relative timing of the gestures appears similar across speakers - the fall in the LA trajectory, indicating the closing of the lips tends to coincide with the rise of the TB for the palatal gesture. To facilitate comparison, vertical gray lines indicate when the LA trajectory starts to fall (based on the average) and when TB starts to rise (also based on the average).

Figure 23 shows kinematic trajectories for the token /bjut/ (**COARTICULATORY** palatalization). The level of variability in the magnitude of the gestures appears similar to the **UNDERLYING** palatalization. Regarding the relative timing of the gestures for the token /bjut/, the rise for the TB movement tends to follow shortly after the fall of the LA trajectory.

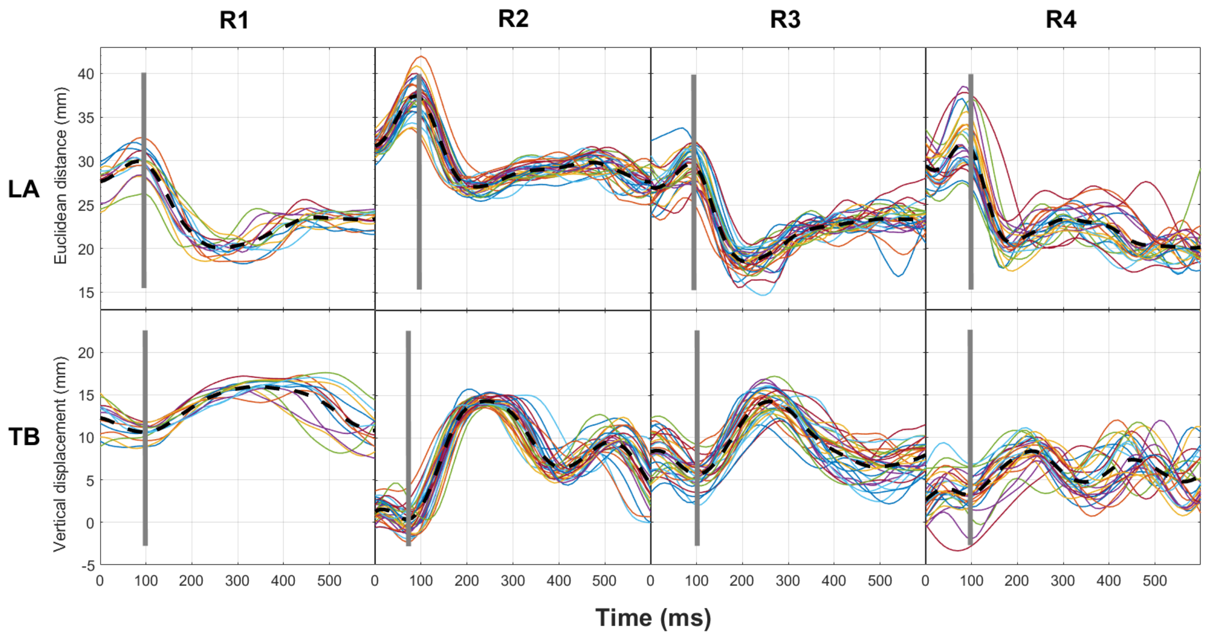


Figure 22: Tokens of /bʲust/ (UNDERLYING palatalization) from each Russian speaker

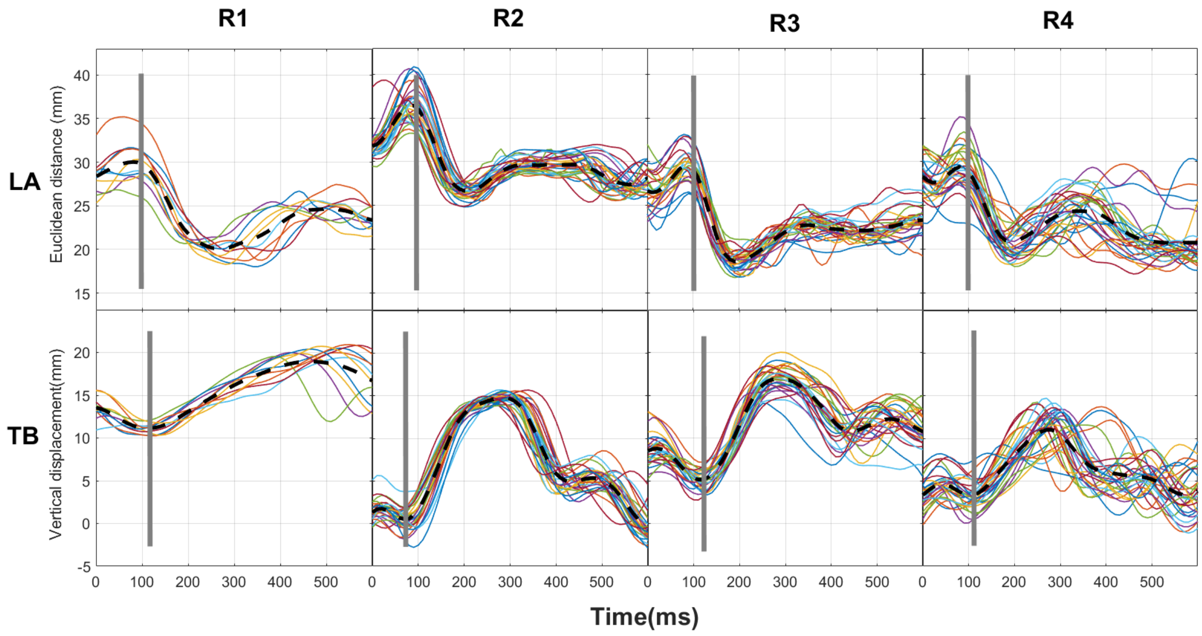


Figure 23: Tokens of /bjut/ (coarticulatory palatalization) from each Russian speaker

Next, I present the distribution of the continuous variables, the key intervals for the temporal coordination analysis: G_1 duration and onset-to-onset lag. In particular, I present the distribution by Status: **UNDERLYING** vs. **COARTICULATORY** palatalization. Also, for completeness, I plot the distribution of G_2 duration (palatal gesture) by Status. This measurement does not relate directly to any of the main hypotheses, but I have included it for reference.

As shown in Figure 24, the G_1 duration measures have a slightly right-skewed distribution with a long right tail, which is common for temporal measurements of speech associated with linguistic units. This is true for the distributions of palatal gesture duration as well as onset-to-onset lag. Notably, however, the distributions of G_1 duration for **UNDERLYING** and **COARTICULATORY** palatalization are heavily overlapped, with similar means and variance. On the other hand, as shown in Figure 25, **COARTICULATORY** palatalization tends to have a longer palatal gesture than **UNDERLYING** palatalization, consistent with the previous findings (Kochetov, 2006). Similarly, the distribution of onset-to-onset lag shows that **COARTICULATORY** palatalization tends to have a longer onset-to-onset lag than **UNDERLYING** palatalization and the distributions differ in shape, with **UNDERLYING** palatalization having a sharp peak with more values close to the mean.

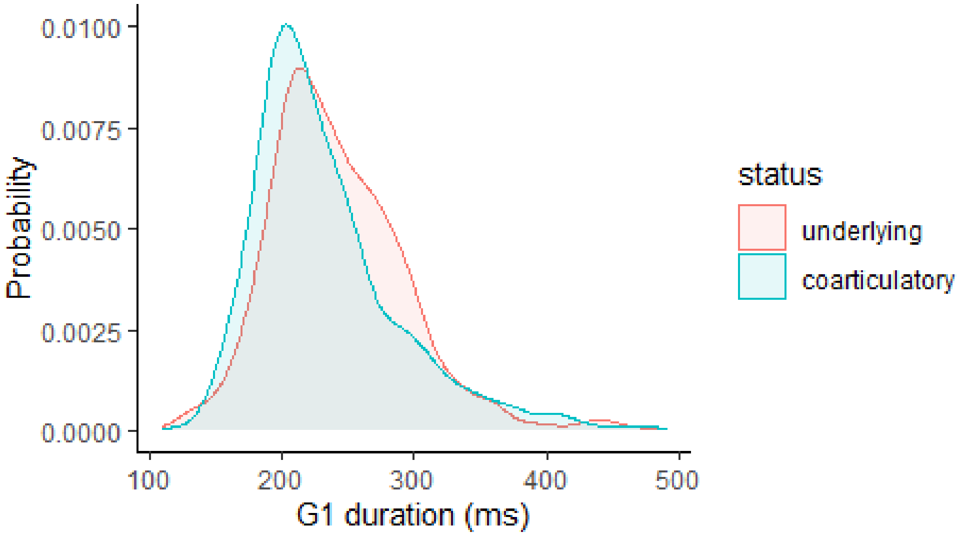


Figure 24: The distribution of G1 (labial consonant) duration by Status.

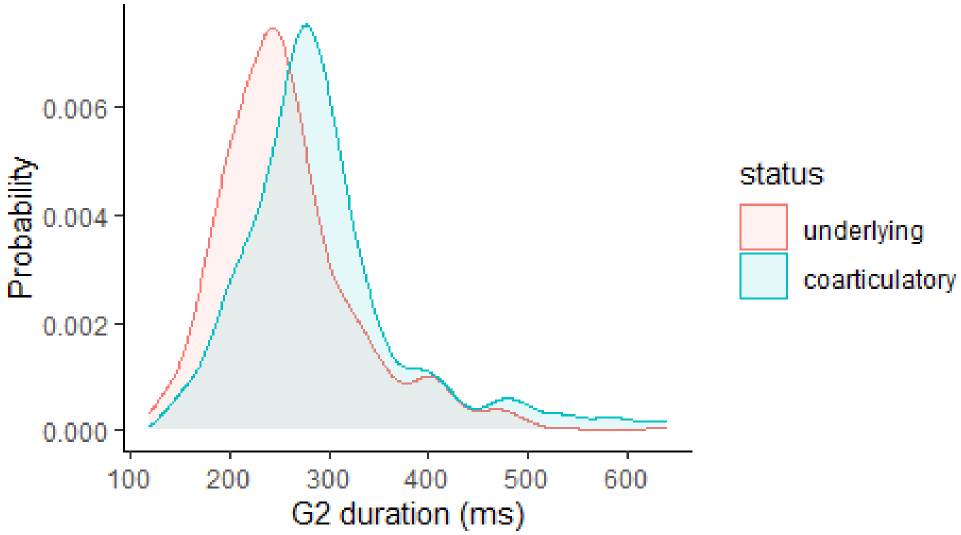


Figure 25: The distribution of G2 (palatal gesture) duration by Status.

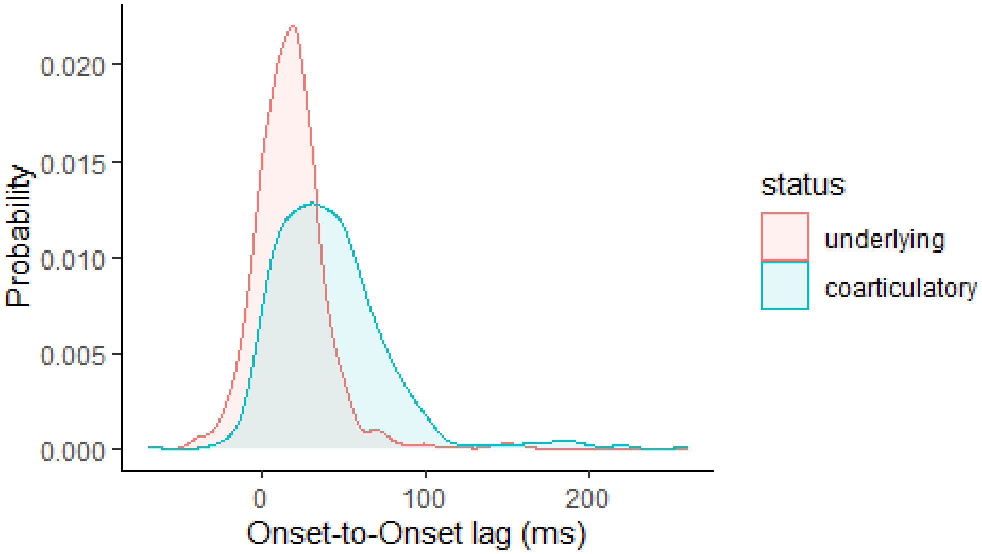


Figure 26: The distribution of onset-to-onset lag measurements by Status.

3.5.2. Temporal coordination

As discussed in Section 3.3, both the UNDERLYING and COARTICULATORY palatalization are expected to show no correlation between *G1 duration* and *onset-to-onset lag*, if both underlyingly and coarticulatory palatalized consonants are palatalized. If one (most likely the COARTICULATORY palatalization) turns out to behave like a segment sequence, then the *onset-to-onset lag* will increase with *G1 duration*, leading to a positive correlation between them.

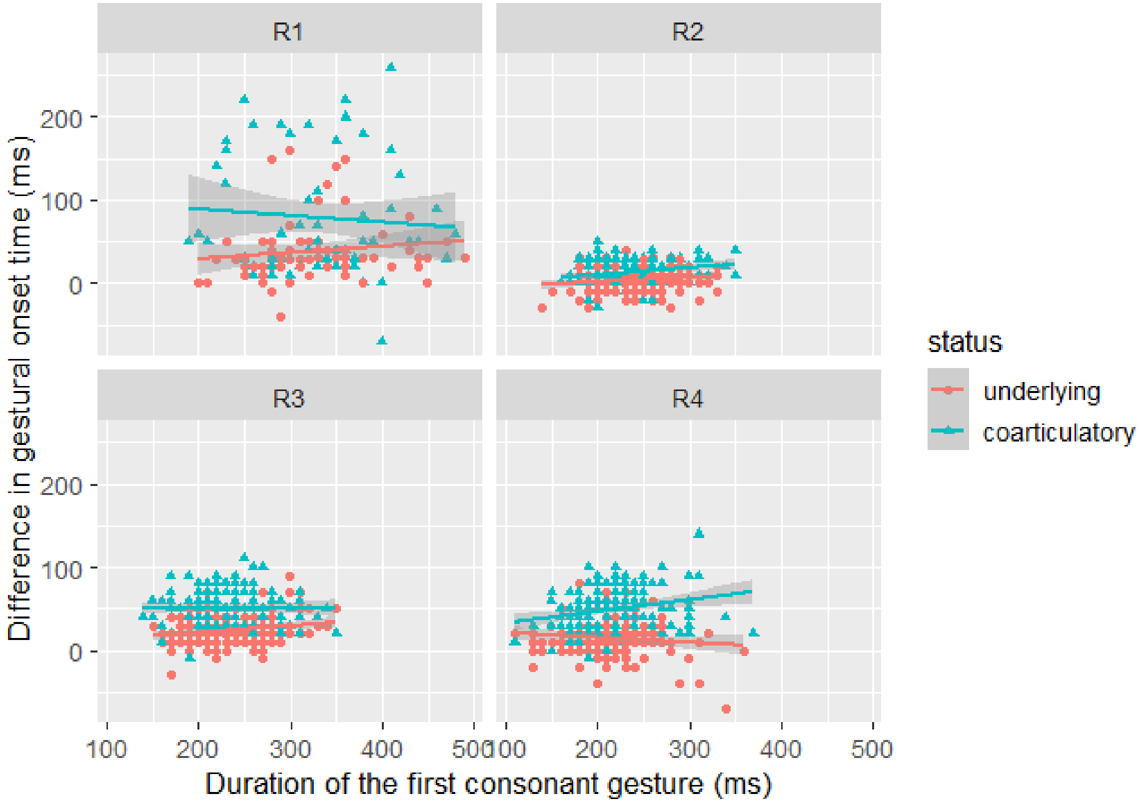


Figure 27: A scatter plot of the effect of *G1 duration* (x-axis) on *onset-to-onset lag* (y-axis) across *Status* for each speaker

Table 5: Summary of R^2 value of each regression line from Figure 27

Status	*R1*	*R2*	*R3*	*R4*
UNDERLYING	0.018	0.026	0.044	0.013
COARTICULATORY	0.0061	0.061	0.00001	0.052

Figure 27 plots the correlation between *G1 duration* (x-axis) and *onset-to-onset lag* (y-axis) across *Status* for each speaker. To illustrate the trend in the data, a least squares linear regression line is fit to each panel. The R^2 value of each regression line from Figure 27 is summarized in Table 5. The regression line is nearly flat, showing only a slight upward trend—precisely the pattern predicted for complex segments. Notably, this pattern was observed for both **UNDERLYING** and **COARTICULATORY** palatalization, indicating that plain consonants preceding glides (**COARTICULATORY** palatalization) are also palatalized. This suggests that the contrast between palatalized and plain consonants is neutralized in this context.

To assess the statistical significance of the trends in Figure 27, I fit a series of linear mixed-effects models to the data (for additional detail, see Section 3.4.4). As shown in Table 6, the addition of G_1 duration improves the baseline model, which contains only random effects of subject and item. The addition of *Status* as a fixed factor further improves the model ($\chi^2 = 22.5$, $p < 0.001$), indicating that the onset-to-onset lag significantly differs by *Status*, as observed in Figure 26. Crucially, in the final model, the addition of the interaction term does not improve the model ($\chi^2 = 1.06$, $p > 0.1$). The null improvement contributed by the interaction term indicates that the influence of G_1 duration on onset-to-onset lag is not different for **UNDERLYING** and **COARTICULATORY** palatalization.

Table 6: Temporal coordination – Nested model comparison

LME Model comparison (*onset-to-onset*~)	**Df**	**AIC**	**logLik**	χ^2	Pr(>χ^2)
1 + (1|subject)+(1|item)	4	10543	-5267.5	NA	NA
1 + G_1 duration + (1|subject)+(1|item)	5	10535	-5262.5	9.8397	<0.01
1 + G_1 duration + Status + (1|subject)+(1|item)	6	10514	-5251.3	22.527	<0.001
1 + G_1 duration * Status + (1|subject)+(1|item)	7	10516	-5250.7	1.0596	0.3033

Table 7 summarizes the best fitting model (onset-to-onset ~ G_1 duration + Status + (1|subject)+(1|item)). The main effect of G1 duration is significant ($t = 3.156$, $p < 0.01$), but very small (0.056 ms). This indicates that, for both types of palatalization in Russian, there is a small but positive correlation between G1 duration and onset-to-onset lag. In addition, the main effect of Status is significant ($t = 8.095$, $p < 0.001$). Specifically, **COARTICULATORY** palatalization is 25 ms longer than **UNDERLYING** palatalization in onset-to-onset lag.

Table 7: Temporal coordination – Summary of fixed factors in the best-fitting model (reference level for Status = Underlying)

	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	7.273	10.56	4.886	0.689	0.522
G1_duration	0.056	0.018	108.1	3.156	< 0.01
Status_Coarticulatory	25.32	3.128	10.12	8.095	< 0.001

In summary, the statistical models generally confirm the trend observable in Figure 27. There is a small but positive correlation between G1 duration and onset-to-onset lag, as predicted by the complex segment hypothesis (See section 2.4). Crucially, the null effect of the interaction between *Status* and *G1 duration* indicates that both UNDERLYING and COARTICULATORY palatalization have the temporal coordination of complex segments.

3.5.3. Articulatory evidence of incomplete neutralization

3.5.3.1. TB positions

Figure 28 shows the normalized horizontal position (front-back) of the TB sensors at the gestural onset across conditions. Positive and negative values on the y-axes illustrate the frontness and backness of the tongue body, respectively. The spatial position of the TB is more retracted for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization at the onset of the palatal gesture. As shown in Figure 29, this pattern generally holds across speakers.

To assess the statistical significance of the trends in Figure 28 and Figure 29, I fit a series of linear mixed-effects models to the data (for additional detail, see Section 3.4.4). As shown in Table 8, the addition of Status improves the baseline model, which contains only random effects of subject and item ($\chi^2 = 18.846$, $p < 0.001$), indicating that the TB position significantly differs by Status, as observed in Figure 28 and Figure 29. Specifically, the TB is 1.5 mm more retracted for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization at the onset of the palatal gesture, as shown in Table 9. This difference is consistent with the presence of a secondary tongue dorsum retraction gesture for plain stops. This suggests that some small residue of velarization/uvularization for plain stops persists in the **COARTICULATORY** condition, in line with the previous observations of an active tongue dorsum retraction gesture in the "plain" stops series.

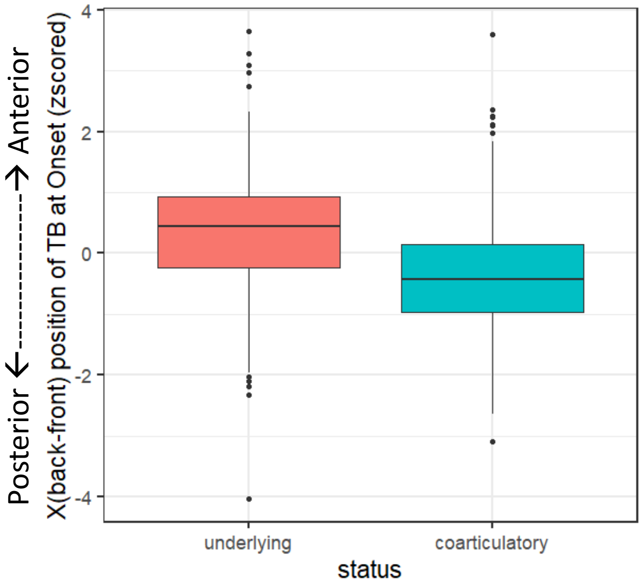


Figure 28: A boxplot of TB position (z-scored) at palatal gesture onset

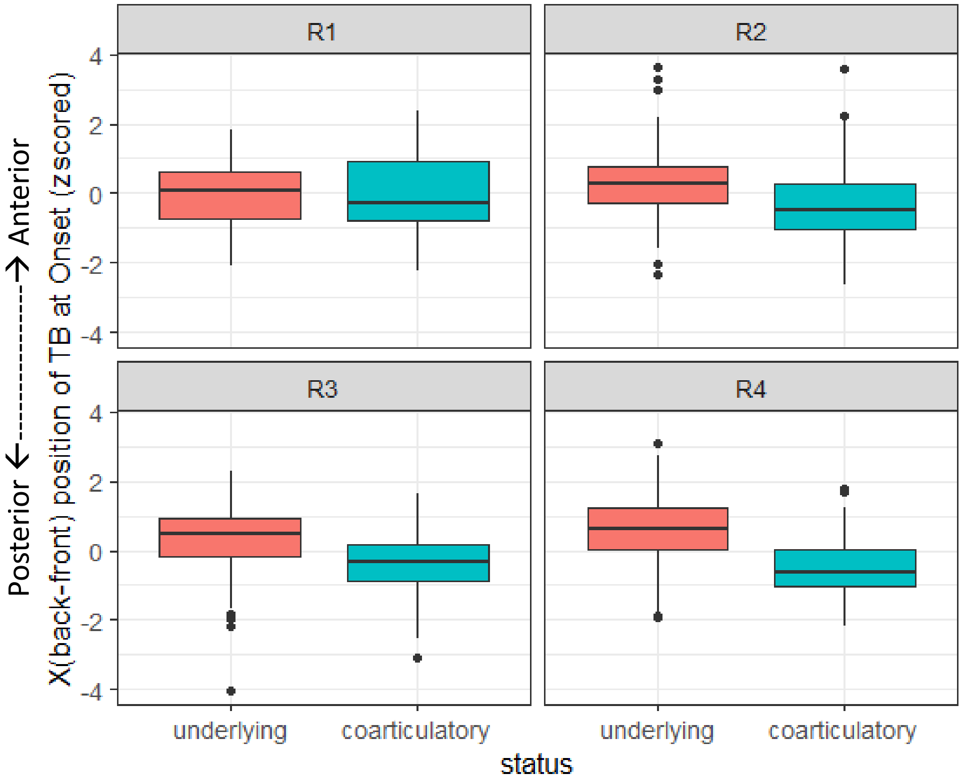


Figure 29: A boxplot of TB position (z-scored) at palatal gesture onset for each speaker

Table 8: TB position – Nested model comparison

TB	Df	AIC	logLik	χ^2	Pr(>χ^2)
1 +(1|speaker) + (1| sequence)	4	4626.5	-2309.3	NA	NA
1+status +(1|speaker) + (1| sequence)	5	4609.7	-2299.8	18.846	< 0.001

Table 9: TB position – Summary of fixed factors in the best-fitting model (reference level for Status = Underlying)

	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	-13.379	5.223	3.005	-2.562	< 0.1
Status_Coarticulatory	-1.515	0.225	10.034	-6.732	< 0.001

3.5.3.2. Onset lags

Figure 30 provides a box plot of onset-to-onset lag across *Status*. The lag between the gesture onsets is longer for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization. Figure 31 shows that this pattern is apparent across speakers.

Table 10 presents the comparison of linear mixed-effects (LME) models for onset-to-onset lag. The addition of *Status* improves the baseline model ($\chi^2 = 22.321$, $p < 0.001$), indicating that *Onset-to-onset lag* is significantly different across *Status*. Specifically, *Onset-to-onset lag* is approximately 25 ms longer for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization, as shown in Table 11.

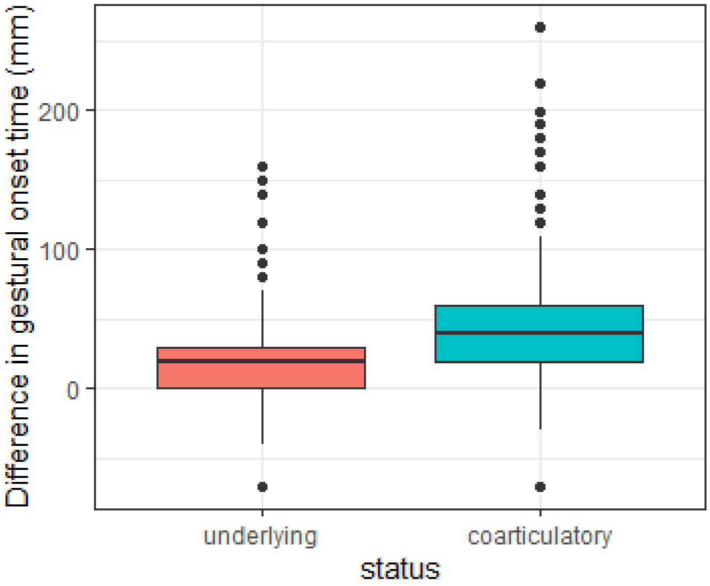


Figure 30: A boxplot of onset-to-onset lag across *Status*

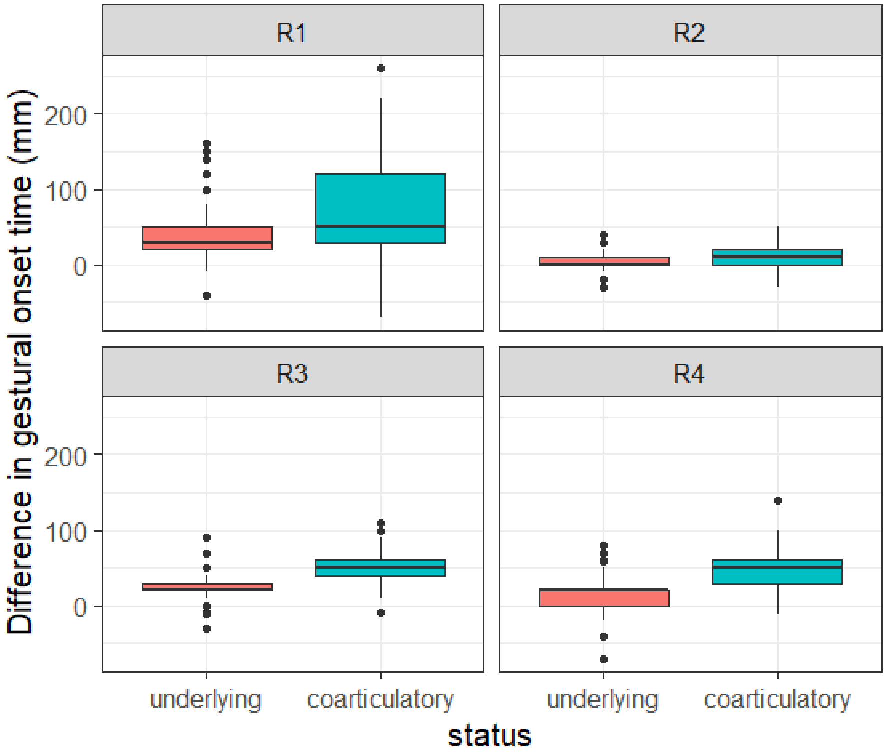


Figure 31: A boxplot of onset-to-onset lag across Status for each speaker

Table 10: Onset-to-onset lag – Nested model comparison

Onset lag	DF	AIC	LogLik	Chisq	Pr(>Chisq)
1 +(1|speaker) + (1| sequence)	4	10543	-5267.5	NA	NA
1+status +(1|speaker) + (1| sequence)	5	10523	-5256.3	22.321	< 0.001

Table 11: Onset-to-onset lag – Summary of fixed factors in the best-fitting model (reference level for Status = Underlying)

	Estimate	**Std.Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	21.543	10.460	3.232	2.060	0.125
Status_Coarticulatory	24.919	3.101	10.084	8.037	< 0.001

In sum, the findings from TB positions and onset-to-onset lag suggest that the delay of gestural onset in the **COARTICULATORY** condition might be attributable to the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization). That is, the spatial and temporal overlap between palatalization and velarization/uvularization leads to gestural blending, and, in turn, the blending of the two competing gestural forces delays the onset of the TB gesture.

3.5. Discussion

3.6.1. Overview

In this chapter, I explored a case of putative phonological neutralization of palatalized consonants (underlying palatalization; e.g., /b̲ʲ/) and plain consonants preceding a palatal glide (coarticulatory palatalization; e.g., /b̲j/) in Russian. The purpose of this chapter is to explore how complete the neutralization is between underlying palatalization and coarticulatory palatalization. To do so, I conducted an Electromagnetic Articulography (EMA) experiment and examined the temporal coordination and the spatial positions of articulators involving underlying and coarticulatory palatalizations in Russian.

I asked two research questions as follows: (1) Do underlying palatalization (e.g., /bʲ/) and coarticulatory palatalization (e.g., /bj/) exhibit temporal coordination of complex segments? (2) Do underlying palatalization (e.g., /bʲ/) and coarticulatory palatalization (e.g., /bj/) exhibit spatial and/or temporal differences? The first research question was regarding whether two cases of Russian palatalizations show neutralization. The second research question was regarding whether the neutralization is complete. There are three possible outcomes depending on the temporal organization and spatial and/or temporal differences of underlying and coarticulatory palatalization. If there is no neutralization, underlying palatalization will show the temporal coordination of complex segments, while coarticulatory palatalization exhibits the temporal coordination of segment sequences. On the other hand, if palatalized consonants and plain consonants preceding a palatal glide are both palatalized, both the underlying and coarticulatory palatalizations will exhibit no correlation between consonant duration and onset-to-onset lag. If the neutralization is phonetically complete, there will be no significant difference between the underlying and coarticulatory palatalizations in the spatial coordination of gestures. However, if the neutralization is phonetically incomplete, there will be a small but significant difference in the spatial coordination of gestures. Given that plain consonants have secondary velarization (Litvin, 2014; Roon & Whalen, 2019; Skalozub, 1963), I predicted that the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) in coarticulatory palatalization would lead to incomplete neutralization of underlying and coarticulatory palatalization in Russian. A key finding from my EMA study is as follows. There is a small but positive correlation between the G1 duration and the onset-to-onset lag, as predicted by the complex segment hypothesis (Figure 27), for both underlying and coarticulatory palatalization types. Crucially, the null effect of the interaction between *Status* and *G1 duration* indicates that both underlying and coarticulatory palatalizations have the temporal coordination of complex segments. This suggests that the contrast between a palatalized consonant and a plain consonant is neutralized to the palatal counterpart when a plain consonant is followed by a glide.

However, the underlying and coarticulatory palatalizations show small but significant phonetic differences in the temporal as well as spatial coordination of gestures. One of the findings is that the onset lag is longer for the coarticulatory palatalization than for the underlying palatalization. Furthermore, I also found residual evidence of an underlying tongue dorsum retraction for coarticulatory palatalization. In particular, the spatial position of the TB is more retracted for the coarticulatory palatalization than for the underlying palatalization at the onset of the palatal gesture. This is in line with previous findings of Russian plain consonants having secondary velarization. As predicted in Section 3.3, the gestural overlap on the same tract variable (i.e., palatalization vs. velarization) would lead to gestural blending between these two gestures. Accordingly, this results in a slightly more retracted tongue position for the coarticulatory palatalization compared to underlying palatalization, which only has the palatal gesture on the TB tract. Consequently, this difference leads to incomplete neutralization between the underlying and coarticulatory palatalizations in Russian.

3.6.2. Effect of back vowels

From the EMA data, I found residual evidence of an underlying tongue dorsum retraction for the coarticulatory palatalization. This finding is in line with the prediction that I made in Section 3.3. That is, the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) in the coarticulatory palatalization leads to incomplete neutralization of underlying and coarticulatory palatalization in Russian. However, an alternative explanation is that the retracted tongue position might be attributable to the blending of the palatal gesture and the following vowel gesture.

Consonant-to-vowel coarticulation is commonly found crosslinguistically, such as in English (e.g., Keating, 1993), Russian (e.g., Iskarous & Kavitskaya, 2010), French (e.g., Guitard-Ivent et al., 2021), Catalan (e.g., Recasens, 1985), and Algerian Arabic (e.g., Bouferroum & Boudraa, 2015). Given that the target vowels are all back vowels (/u/ and /o/) in the experiment, the retracted tongue position of the coarticulatory palatalization can result from the gestural blending of the palatal gesture and the gesture for a back vowel. However, further explanation is necessary for this account, since both types of palatalization have the same back vowels, but the retracted tongue position was observed only for the coarticulatory palatalization.

The coupled oscillator model hypothesizes that syllable structure is associated with a characteristic pattern of temporal coordination (Goldstein et al., 2006; Goldstein et al., 2009; Nam et al., 2009; Saltzman et al., 2008). For example, a gesture in a syllable onset is coordinated in-phase with the following vowel, with two gestures triggered at the same time. In contrast, a coda gesture is coordinated anti-phase with the preceding vowel, showing a sequential timing between the two gestures. Furthermore, the coupled oscillator model hypothesizes that multiple gestures in a syllable onset are coupled anti-phase with each other, along with both being in-phase with the vowel. Due to this competitive coupling, the vowel starts at the center of prevocalic consonants, the so-called "c-center effect" (Brunner et al., 2014; Byrd, 1995; Goldstein et al., 2007; Marin, 2013; Nam & Saltzman, 2003; Shaw et al., 2011).

Provided that the retracted TB position was observed at the onset of the gesture, the incomplete neutralization might be attributable to a different temporal coordination between underlying and coarticulatory palatalizations. If the gestures for /b/ and /j/ are coupled in-phase with each other for the underlying palatalization, and they are also coupled in-phase with the following vowel, the gestures for /j/ and /u/ will start at the same time for the underlying palatalization. In contrast, for the coarticulatory palatalization, if the gestures for /b/ and /j/ are coupled anti-phase with each other, and they are also coupled in-phase with the following vowel, the vowel gesture for /u/ starts before the palatal gesture for /j/ and continues concurrently for the coarticulatory palatalization, due to the c-center effect. In both cases, the temporal overlap between the palatal gesture and the following vowel gesture in the same tract variable (TB) will lead to gestural blending between them. Crucially, however, when the spatial position of the TB gestures is compared at the onset of the palatal gesture for both palatalizations, the coarticulatory palatalization, which has the tongue backing for /u/ starting earlier, will show a more retracted tongue position than the underlying palatalization, since the backing movement for /u/ will have already started before the palatal gesture starts for the coarticulatory palatalization. That is, due to the existence of the vowel gesture preceding the palatal gesture for the coarticulatory palatalization, the gestural blending of the tongue backing for /u/ and the fronting for /j/ may result in a more retracted tongue position at the onset of the TB gesture for the coarticulatory palatalization as compared to the tongue position for the underlying palatalization.

However, crucially, if the labial gesture and the palatal gesture are coordinated anti-phase for the coarticulatory palatalization, similar to segment sequences in English, they are expected to result in the temporal coordination of segment sequences, which is not the case in Russian. To demonstrate this, I present multiple articulatory models for underlying and coarticulatory palatalization using an articulatory-based synthesizer, Task Dynamic Application (TADA), and compare articulatory data to modeled stimuli in Chapter 4.

3.6.3. Does the blending produce the delayed onset?

As discussed above, the gestural blending between palatalization and velarization in the coarticulatory palatalization case may lead to a more retracted tongue position than would be expected for underlying palatalization. However, it is not clear what causes the delayed onset of the coarticulatory palatalization. Would it still be the blending that produces the delayed onset for the coarticulatory palatalization, or is there another factor?

There are three possible explanations for the delayed onset of the TB gesture for the coarticulatory palatalization. In the AP framework, the blending of the dynamical parameters of two gestures is predicted to produce an outcome that falls somewhere in-between the two gestures, depending on the strength of the two gestures in question (e.g., Browman & Goldstein, 1989; 1992). Consequently, the tongue body gesture starts at the same time for both palatalizations, but those blended gestures might not have the abrupt start that the underlying palatalization has. Provided that the *findgest* function in MVIEW parses gestural landmarks with reference to the velocity signal (Tiede, 2005), and the gesture *Onset* landmark was labeled at a 20% of peak velocity in the movement toward the constriction in my dissertation, if the blended gestures showed a more gradual start, it might not be enough for the *findgest* function to parse the gestural onset correctly. However, blending of the dynamical parameters of two gestures does not necessarily change the stiffness of the blended gestures, and thus this account also has an insufficient underlying theoretical basis.

Yet another possible explanation is to posit that the labial and palatal gestures are coordinated anti-phase, and the labial and velar gestures are coupled in-phase for the coarticulatory palatalization. In contrast, the labial and palatal gestures are coordinated in-phase for the underlying palatalization. Such a coordination will result in the delayed onset of the TB gesture for the coarticulatory palatalization compared to the underlying palatalization. However, as discussed in Section 3.6.2, the anti-phase coordination between the gestures for /b/ and /j/ is expected to produce the temporal coordination of segment sequences, which is not the case for the coarticulatory palatalization in Russian (see Chapter 4 for the test of this prediction).

Lastly, if the velar gesture starts before the palatal gesture for /j/ and continues concurrently for the coarticulatory palatalization, while the labial and palatal gestures are coordinated in-phase, these coordination relations may lead to delayed onset of the TB gesture for the coarticulatory palatalization and the temporal coordination of complex segments. That is, if the velar gesture is activated before the blending of the velar and palatal gestures starts, the *findgest* function in Mview may detect the brief activation of the velar gesture as a gesture, and then it may parse the rest as blended gestures. Consequently, the eccentric timing between the velar and palatal gestures may lead to a delayed onset of the TB gesture for the coarticulatory palatalization. If this is the case, the *findgest* function in Mview will parse the gesture the same way with the modeled (synthesized) articulatory data. The details about the computational models and the results are presented in Chapter 4.

3.7. Summary

Russian contrasts palatalized and plain (non-palatalized) consonants, but this contrast is reported to be neutralized when a plain consonant is followed by a glide (e.g., Kochetov, 2011). In this chapter, I explored the neutralization of palatalized consonants (underlying palatalization; e.g., /b̲ʲ/) and plain consonants preceding a palatal glide (coarticulatory palatalization; e.g., /b̲j/) in Russian using Electromagnetic Articulography (EMA).

A key finding from the EMA experiment is that both the underlying and coarticulatory palatalizations exhibited temporal coordination of complex segments, showing no correlation between consonant duration and onset-to-onset lag. This suggests that the contrast between a palatalized consonant and a plain consonant is neutralized to the palatal counterpart when a plain consonant is followed by a glide. Crucially, however, the neutralization of the palatal-plain contrast is phonetically incomplete. In particular, I found that the tongue body was significantly more retracted for the coarticulatory palatalization than for the underlying palatalization at the onset of the palatal gesture, but the difference was small (1.5 mm). In addition, *Onset-to-onset lag* is significantly longer for the coarticulatory palatalization than for the underlying palatalization. These small but significant differences suggest that the neutralization of the palatal-plain contrast is phonetically incomplete. Furthermore, the residual evidence of an underlying tongue dorsum retraction for the coarticulatory palatalization is in line with previous findings of Russian plain consonants having secondary velarization.

Chapter 4. Articulatory modeling of Russian palatalization as incomplete neutralization

In Chapter 3, I hypothesized that the gestural blending of two secondary articulation gestures (palatalization and velarization/uvularization) in the coarticulatory palatalization condition leads to incomplete neutralization of underlying and coarticulatory palatalization in Russian. Experimental evidence from EMA supports my hypothesis in that underlying and coarticulatory palatalization exhibit inter-gestural coordination characteristic of complex segments, but there is residual evidence of an underlying tongue dorsum retraction (velarization/uvularization) gesture for the coarticulatory palatalization. In this chapter, I explore multiple gestural models to represent underlying and coarticulatory palatalization in Russian, using TADA, and evaluate each model by comparing the simulations from each model against the results from the EMA recordings.

4.1. Introduction: Task Dynamic Application

Task Dynamic Application (henceforth, TADA) is a MATLAB-based software for simulating the gestural representations of utterances and generating acoustic output (Nam et al., 2004; 2006; 2012). Based on the Task Dynamic model of inter-articulator coordination in speech (e.g., Saltzman & Munhall, 1989), TADA implements the following models that feed one another as shown in Figure 32.

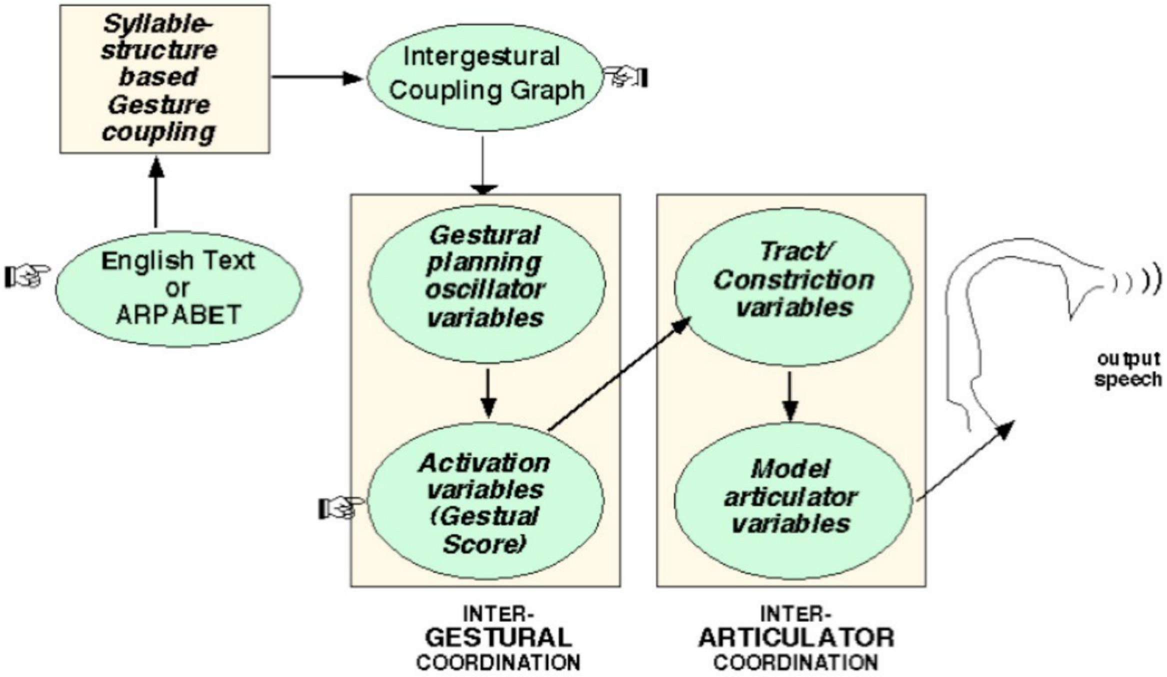
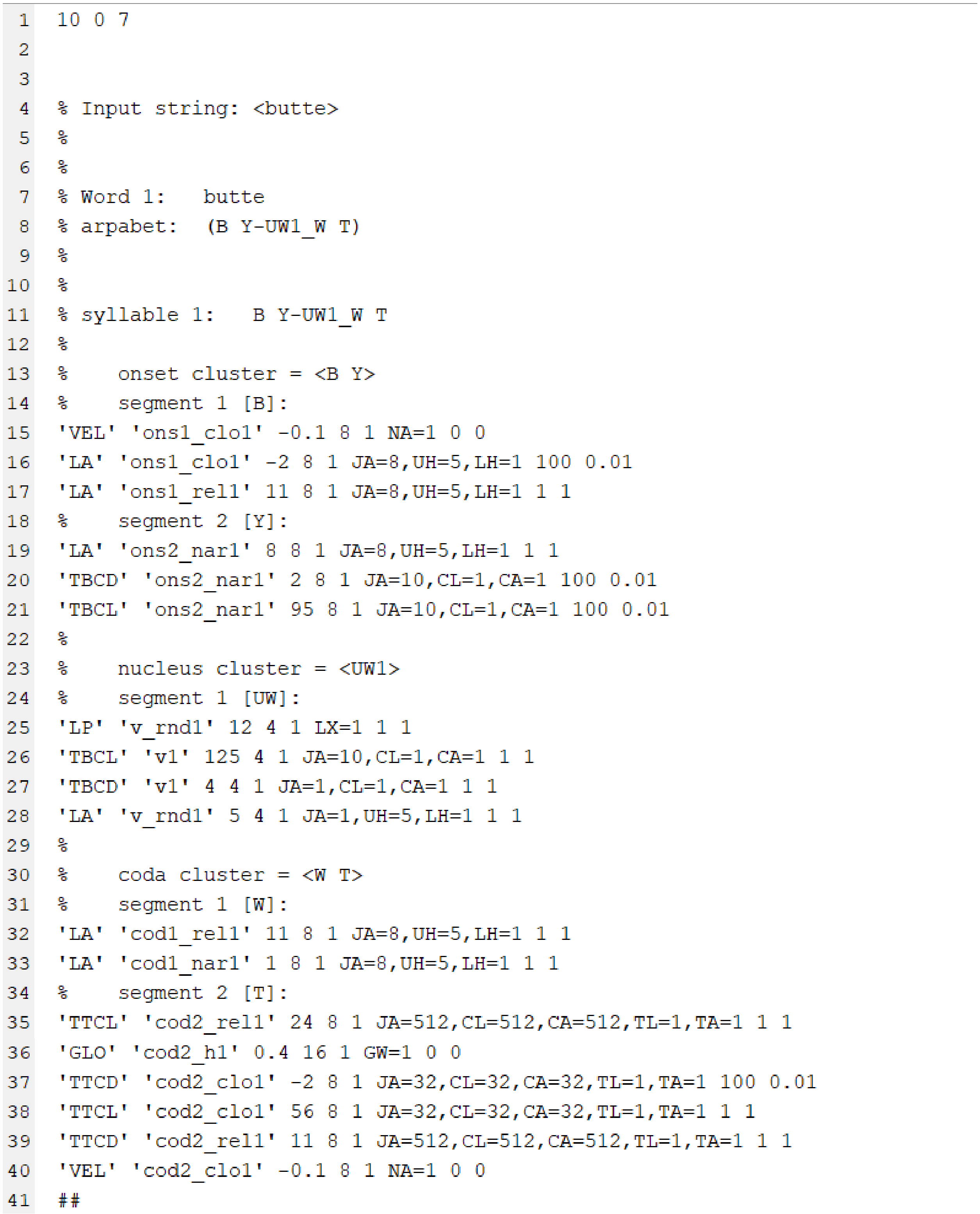


Figure 32: Information flow through TADA models (adopted from Nam et al., 2006, p.2)

First, with a text string as input, the Syllable structure-based gesture coupling model generates an intergestural coupling graph, which feeds into the coupled oscillator model of inter-gestural coordination. In turn, the coupled oscillator model generates a gestural score, which specifies inter-gestural coordination and becomes an input of the task dynamic model of inter-articulator coordination. Then, the task dynamic model generates the vocal tract constriction variables and the articulatory degrees of freedom. The outcome feeds into Configurable Articulatory Synthesis (CASY) to compute a time-varying vocal tract area function and the resonance frequencies and bandwidths corresponding to those area functions. Finally, taking these as an input, Sensimetrics' HLsyn synthesizer generates the acoustic output. Notably, these models can run separately and independently when an input is provided for each model.

In TADA, coupling graphs can be created through the GEST menu by using either English text or ARPABET as an input, resulting in TV<id>.O and PH<id>.O files. TV<id>.O file contains gestural specifications. It consists of the specification of control parameters of each gesture and a label specifying the oscillator that controls the activation of that gesture (corresponds to an oscillator label in the PH.O file). PH<id>.O file, on the other hand, contains timing oscillator and coupling specifications. It consists of dynamical parameter values, the phases for activation and deactivation of gestures, and the coupling parameters for oscillator pairs.

As an example, I created TVbutte.O and PHbutte.O files through the GEST menu in TADA for the English word ‘butte.’ As shown in Figure 33, TVbutte.O consists of a list of gestures and their positions within syllables, and each line contains dynamical parameters, articulator weights, and blending parameters. As shown in lines 18-21 in Figure 33, the second onset /j/ in the word *butte* contains three tract variables. Two are for the TB gesture, one specifying the constriction degree (TBCD) the other the constriction location (TBCL). There is also a gesture for lip aperture (LA).



```
10 0 7


% Input string: <butte>
%
%
% Word 1:   butte
% arpabet:  (B Y-UW1_W T)
%
%
% syllable 1:   B Y-UW1_W T
%
%    onset cluster = <B Y>
%    segment 1 [B]:
'VEL' 'ons1_clo1' -0.1 8 1 NA=1 0 0
'LA' 'ons1_clo1' -2 8 1 JA=8,UH=5,LH=1 100 0.01
'LA' 'ons1_rel1' 11 8 1 JA=8,UH=5,LH=1 1 1
%    segment 2 [Y]:
'LA' 'ons2_nar1' 8 8 1 JA=8,UH=5,LH=1 1 1
'TBCD' 'ons2_nar1' 2 8 1 JA=10,CL=1,CA=1 100 0.01
'TBCL' 'ons2_nar1' 95 8 1 JA=10,CL=1,CA=1 100 0.01
%
%    nucleus cluster = <UW1>
%    segment 1 [UW]:
'LP' 'v_rnd1' 12 4 1 LX=1 1 1
'TBCL' 'v1' 125 4 1 JA=10,CL=1,CA=1 1 1
'TBCD' 'v1' 4 4 1 JA=1,CL=1,CA=1 1 1
'LA' 'v_rnd1' 5 4 1 JA=1,UH=5,LH=1 1 1
%
%    coda cluster = <W T>
%    segment 1 [W]:
'LA' 'cod1_rel1' 11 8 1 JA=8,UH=5,LH=1 1 1
'LA' 'cod1_nar1' 1 8 1 JA=8,UH=5,LH=1 1 1
%    segment 2 [T]:
'TTCL' 'cod2_rel1' 24 8 1 JA=512,CL=512,CA=512,TL=1,TA=1 1 1
'GLO' 'cod2_h1' 0.4 16 1 GW=1 0 0
'TTCD' 'cod2_clo1' -2 8 1 JA=32,CL=32,CA=32,TL=1,TA=1 100 0.01
'TTCL' 'cod2_clo1' 56 8 1 JA=32,CL=32,CA=32,TL=1,TA=1 1 1
'TTCD' 'cod2_rel1' 11 8 1 JA=512,CL=512,CA=512,TL=1,TA=1 1 1
'VEL' 'cod2_clo1' -0.1 8 1 NA=1 0 0
##
```

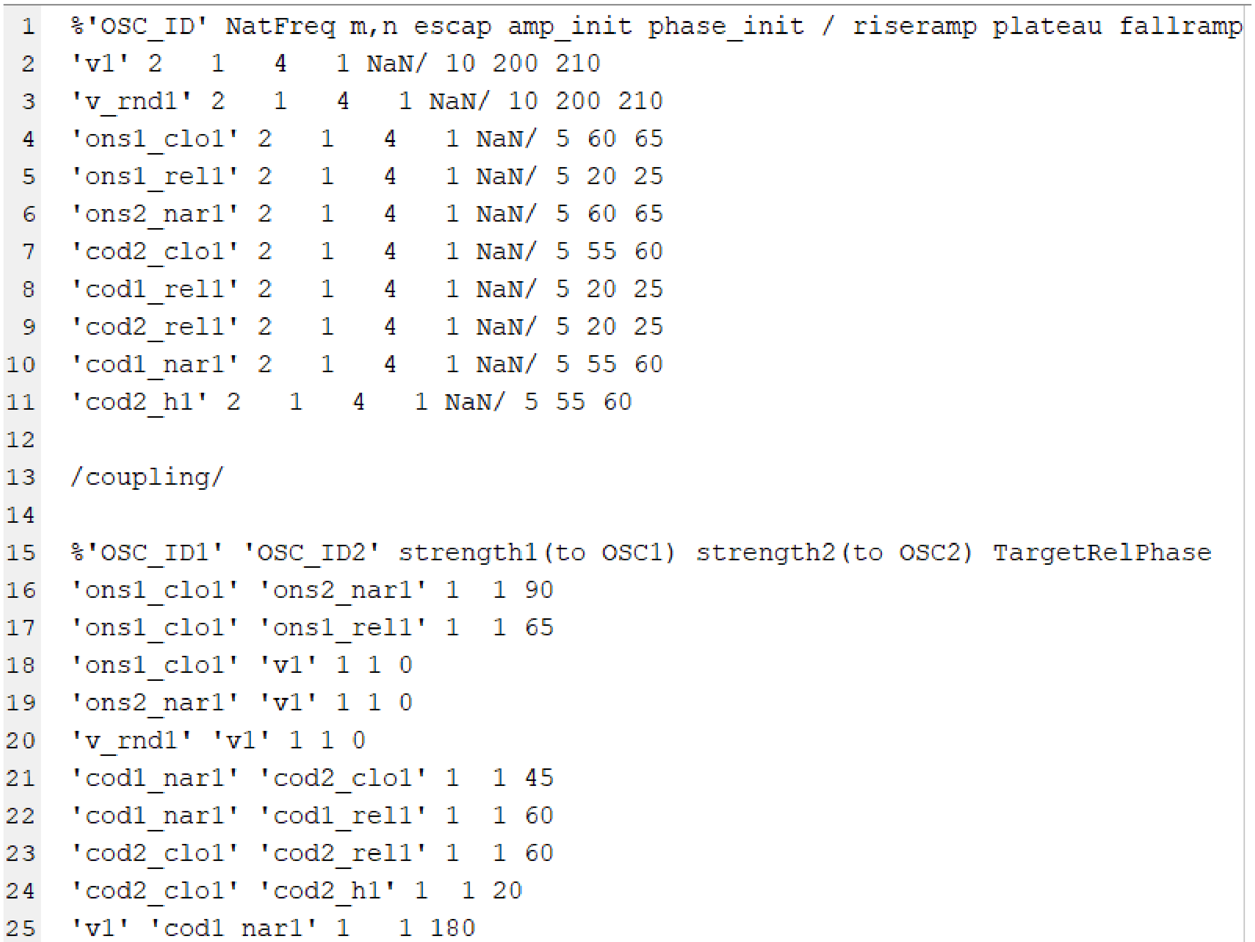
Figure 33: TVbutte.O

Each line in Figure 33 consists of the following information: *TV_name Osc_ID target freq damp art_wts alpha beta*. In line 21, for example, *TV_name* is '*TBCL,*' and *Osc_ID* is *'ons2_nar1'* which corresponds to an oscillator label in the PHbutte.O file (See line 6 in Figure 34). *target* is 95 degrees since the target specification for palatal is 95 degrees (c.f., target specifications for CD is in mm; e.g., see line 20). *freq* specifies the stiffness of a gesture in which the default is 8 Hz for gestures associated with consonants (c.f., the default for vowels is set to 4 Hz; e.g., see lines 25-28). This stiffness parameter is used as one of the ways to elicit temporal variation in Section 4.2.2. *damp* refers to the Damping ratio which is set to 1 by default. *art_wts* specifies an articulator weight where the higher value indicates that the articulator is "heavy" and less likely to move in the production of a constriction. *alpha* specifies the blending strength of the gesture. A higher value indicates a stronger blending strength. In line 21, *alpha* is set to 100 *beta* is set to the reciprocal of alpha, 0.01 (c.f., when *alpha* is zero, *beta* is also set to zero; e.g., see line 15).

Figure 34 illustrates PHbutte.O file. This file is divided into two sub-sections: the parameters of timing oscillators (line 1 – 11) and coupling specifications (line 13 – 25). Each line of oscillator parameters consists of the following information: *'OSC_ID' NatFreq m:n escap amp_init phase_init / riseramp plateau fallramp.* In line 6, for example, *Osc_ID* is *'ons2_nar1'* which is identical to an oscillator label in the TVbutte.O file. *NatFreq* refers to the natural frequency of a limit cycle oscillator, which is set to 2 Hz for all oscillators in the example file. This natural frequency parameter is used to elicit temporal variation (See Section 4.2.2). *m:n* refers to the ratio of the natural frequency of oscillator pairs which determines the generalized relative phase. In the example file, *m:n* is set to 1 since the natural frequency of oscillators corresponding to vowel and consonant gestures is equal to 2 Hz. *escap* refers to the oscillator escapement, which is set to

4. *amp_init* is the amplitude at time t_0, which is set to 1. *phase-init* refers to the oscillator phase at time t_0. In the example file, *phase-init* is set to NaN which means that a random phase is chosen. *riseramp, plateau,* and *fallramp* refer to activation and de-activation phases. For example, in line 4, 5 degrees for *riseramp* mean that the activation of the gesture is started from a value of 0 at 0 degrees to a maximum value of 1 at 5 degrees. 60 degrees for *plateau* indicate that the gesture stays at the maximum level until the phase reaches 60 degrees. 65 degrees for *fallramp* means that the activation of gesture goes down to reach a value of 0 again at 65 degrees.

Each line of coupling specifications consists of the following information: *'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase*. ‘OSC_ID1’ and ‘OSC_ID2’ refer to a pair of oscillator labels. *strength1(to OSC1)* specifies the relative coupling strength from osc2 onto osc1, and *strength2(to OSC2)* specifies the coupling strength from osc1 onto osc2. *TargetRelPhase* refers to a target relative phase for the two oscillators. For example, line 18 shows the coupling specifications between *'ons1_clo1'* (Onset C_1) and *'v1'* (vowel), and their relative coupling strength is equal to 1. Their target relative phase is 0 degrees, which indicates the in-phase relation between the onset and the vowel (c.f., line 25 shows anti-phase relation between the vowel and the coda, specified as 180 degrees; See e.g., Goldstein et al., 2006; Nam et al., 2009).



```
%'OSC_ID' NatFreq m,n escap amp_init phase_init / riseramp plateau fallramp
'v1' 2   1   4   1 NaN/ 10 200 210
'v_rnd1' 2   1   4   1 NaN/ 10 200 210
'ons1_clo1' 2   1   4   1 NaN/ 5 60 65
'ons1_rel1' 2   1   4   1 NaN/ 5 20 25
'ons2_nar1' 2   1   4   1 NaN/ 5 60 65
'cod2_clo1' 2   1   4   1 NaN/ 5 55 60
'cod1_rel1' 2   1   4   1 NaN/ 5 20 25
'cod2_rel1' 2   1   4   1 NaN/ 5 20 25
'cod1_nar1' 2   1   4   1 NaN/ 5 55 60
'cod2_h1' 2   1   4   1 NaN/ 5 55 60

/coupling/

%'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
'ons1_clo1' 'ons2_nar1' 1  1 90
'ons1_clo1' 'ons1_rel1' 1  1 65
'ons1_clo1' 'v1' 1 1 0
'ons2_nar1' 'v1' 1 1 0
'v_rnd1' 'v1' 1 1 0
'cod1_nar1' 'cod2_clo1' 1  1 45
'cod1_nar1' 'cod1_rel1' 1  1 60
'cod2_clo1' 'cod2_rel1' 1  1 60
'cod2_clo1' 'cod2_h1' 1  1 20
'v1' 'cod1 nar1' 1   1 180
```

Figure 34: PHbutte.O

Based on the TV<id>.O and PH<id>.O files, a gestural score can be computed through TV computation (by clicking the [TV] button in TADA), as shown in Figure 35. The gestural score can be saved as a TV<id>.G file, which contains a gestural score with timing information for each gesture. The gestural score and output time functions can also be saved in .mat file format. This .mat file can be visualized in MVIEW and gestural landmarks can also be parsed with reference to the velocity signal using the *findgest* function in MVIEW (Tiede, 2005). This makes the comparison possible between the simulations from computational modeling and the results from the EMA recordings.

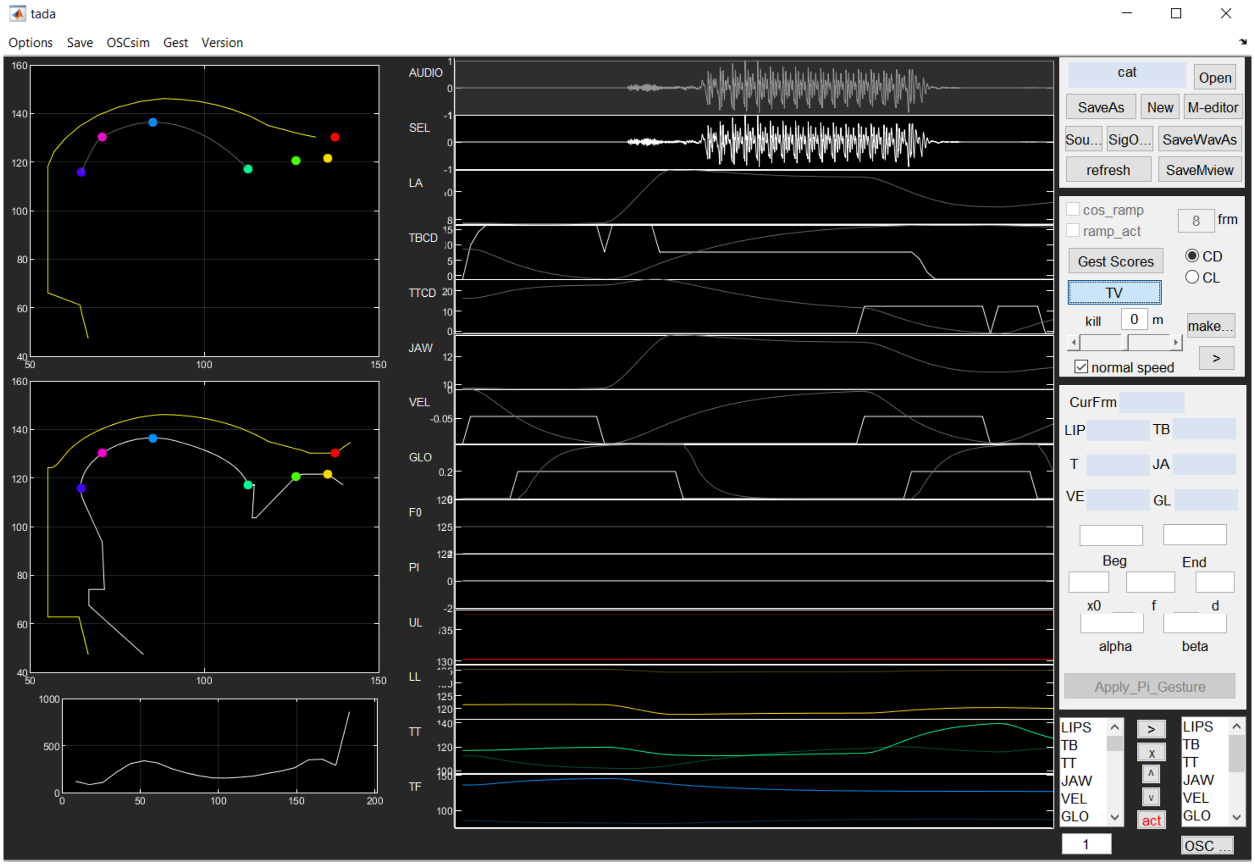


Figure 35: Gestural score in TADA

The structure of this chapter is as follows. In Section 4.2, I review ways to introduce temporal variation in simulations. I then lay out four gestural models, which may produce the same results that were observed in the EMA study, and I outline gestural and coupling specifications for each model in Section 4.3. The results from TADA simulations are reported in section 4.4. The discussion and the summary are presented in sections 4.5 and 4.6, respectively.

4.2. Temporal variation in simulations

The goal of this chapter is to find a gestural model to represent underlying and coarticulatory palatalization in Russian. To do so, I evaluate multiple gestural models by comparing the simulations from each model against the results from the EMA recordings. However, since my approach to exposing differences in coordination makes use of the natural variation present in the data (See Section 2.7 for more details), it is necessary to elicit temporal variation in the simulations as well. In this section, I evaluated two methods to elicit temporal variation in TADA.

4.2.1. Modulating stiffness

To elicit temporal variation in G1 duration, I manipulate the stiffness of /b/ by modulating *freq* in TV<id>.O file. As discussed in Section 4.1, *freq* specifies the stiffness of a gesture, in which the stiffness is set to 8 Hz for gestures associated with consonants, and 4 Hz for gestures associated with vowels. To test the effectiveness of this manipulation, I modulate the stiffness of /b/ in the English word /bjut/ ‘butte’ as a test word and examine the temporal coordination of the word. Table 12 shows the variation in G_1 duration for /b/ in ‘butte’ which is derived by the changes in stiffness from 3 to 20 Hz.

Given that /bj/ in ‘butte’ is an obvious case of a segment sequence, this simulation should result in a positive correlation between G_1 duration and onset-to-onset lag (See also Figure 18 in Section 2.6). Crucially, however, and as shown in Figure 36, *G_1 duration* is independent of *onset-to-onset* lag, when I modulate the stiffness of /b/ to elicit temporal variation. In addition, considering that the natural variation for G_1 duration present in the EMA data is between 100 ms and 500 ms, the manipulation of stiffness does not produce enough variation to compare the simulations from each model against the results from the EMA recordings. In sum, manipulating the stiffness of /b/ not only fails to produce enough variation but also fails to predict the temporal coordination of segment sequence for English. In the following section, I examine another way to elicit temporal variation in TADA simulations.

Table 12: Stiffness and G1 duration

stiffness	G1 duration	stiffness	G1 duration
3	155	12	140
4	155	13	135
5	155	14	135
6	155	15	130
7	150	16	130
8	145	17	130
9	145	18	125
10	140	19	125
11	140	20	125

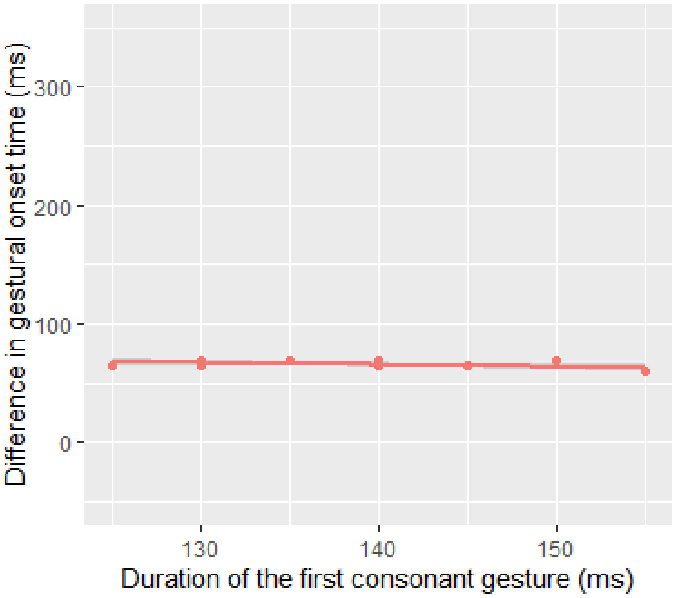


Figure 36: A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis) for Stiffness modulation

4.2.2. Modulating natural frequency

Another method to elicit temporal variation in TADA simulations is to manipulate *NatFreq* in the PH<id>.O file. As discussed in Section 4.1, *NatFreq* refers to the natural frequency of a limit cycle oscillator, which is set to 2 Hz for all oscillators as a default. To test the effectiveness of this method, I modulate the natural frequency of all oscillators in the English word /bjut/ 'butte' as a test word. Table 13 shows the variation in G_1 duration for /b/ in 'butte' which is derived by the changes in natural frequency from 0.5 to 3 Hz.

Table 13: Natural frequency and G1 duration

NatFreq	G1 duration	NatFreq	G1 duration	NatFreq	G1 duration
0.5	430	1.4	190	2.3	135
0.6	370	1.5	190	2.4	135
0.7	330	1.6	175	2.5	135
0.8	285	1.7	170	2.6	130
0.9	260	1.8	165	2.7	125
1	240	1.9	150	2.8	125
1.1	225	2	145	2.9	125
1.2	210	2.1	145	3	120
1.3	195	2.2	140		

Unlike the result from Section 4.2.1, the manipulation of natural frequency produces enough variation (from 120 ms to 430 ms) to compare the simulations from each model against the results from the EMA recordings. More importantly, as shown in Figure 37, variability in G_1 duration via natural frequency modulation turns out to be correlated with onset-to-onset lag, as predicted by the segment sequence hypothesis. Thus, as a method to elicit temporal variation in the simulations, I modulate the natural frequency for all oscillators from 0.5 to 3 Hz for each gestural model and compare the simulations from each model against the results from the EMA recordings. In the following section, I lay out four gestural models.

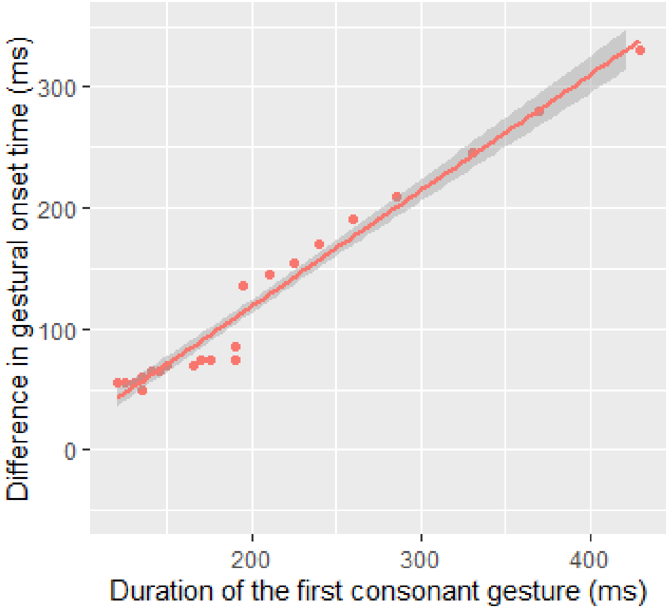


Figure 37: A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis) – Natural frequency modulation

4.3. Gestural models for underlying and coarticulatory palatalization in Russian

To find the best gestural models to represent the underlying and coarticulatory palatalization in Russian, I chose a pair of target words that were used in the EMA study, /b^{j}ust/ and /bjut/, and propose the following four models, which may produce the same results that were observed in the EMA study.

4.3.1. Model 1

As discussed in 3.6.2, the incomplete neutralization of palatalization types in Russian might be attributable to the blending of the palatal gesture and the following vowel gesture. Model 1 is proposed to examine whether gestural blending between /j/ and /u/ alone can lead to a retracted tongue position at the onset of TB gesture and delayed onset-to-onset lag for the coarticulatory palatalization, as well as no effect of variation in consonant duration on onset-to-onset lag. However, since both types (i.e., both sources of) palatalization have the same back vowels, but the retracted tongue position was observed only for the coarticulatory palatalization, there should be other underlying differences that could lead to incomplete neutralization.

In Model 1, I posit a different temporal coordination for both types of palatalization. Gestural models for /bʲust/ and /bjut/ are visualized in Figure 38. The main difference between /bʲust/ and /bjut/ is in the phasing relation between onset consonant gestures. In /bʲust/, /b/ and /j/ are coupled in-phase with each other, while they are coupled 90 degree phase in /bjut/. For /bjut/, given that the vowel gesture is coupled in-phase to both onset consonant gestures, the vowel gesture for /u/ precedes the palatal gesture for /j/ due to the c-center effect (See e.g., Shaw et al., 2011). Consequently, the gestural blending between the tongue backing for /u/ and the fronting for /j/ may result in a retracted tongue position at the onset of the TB gesture, and delayed onset-to-onset lag for the coarticulatory palatalization.

I created the gestural score for /bʲust/ and /bjut/ in the following way. First, I created the coupling graph through the GEST menu in TADA by using ARPABET: (B Y-UW_S T) for /bʲust/ and (B Y-UW_T) for /bjut/. ARPABET and the corresponding IPA symbols are listed in Table 14. The same gestural specifications are used for /bʲust/ and /bjut/ as summarized in Table 15. Then, for /bʲust/, I modified the phasing relation between /b/ and /j/ to be in-phase by changing *TargetRelPhase* to 0 in the PHbʲust.O file (See line 16 in Figure 39). The phasing relation for /bjut/ is not modified. The coupling specifications for /bʲust/ and /bjut/ are shown in Figure 39 and Figure

40, respectively. Based on the coupling graphs, I created the gesture scores for /bʲust/ and /bjut/ through TV computation in TADA and saved them in *.mat* format.

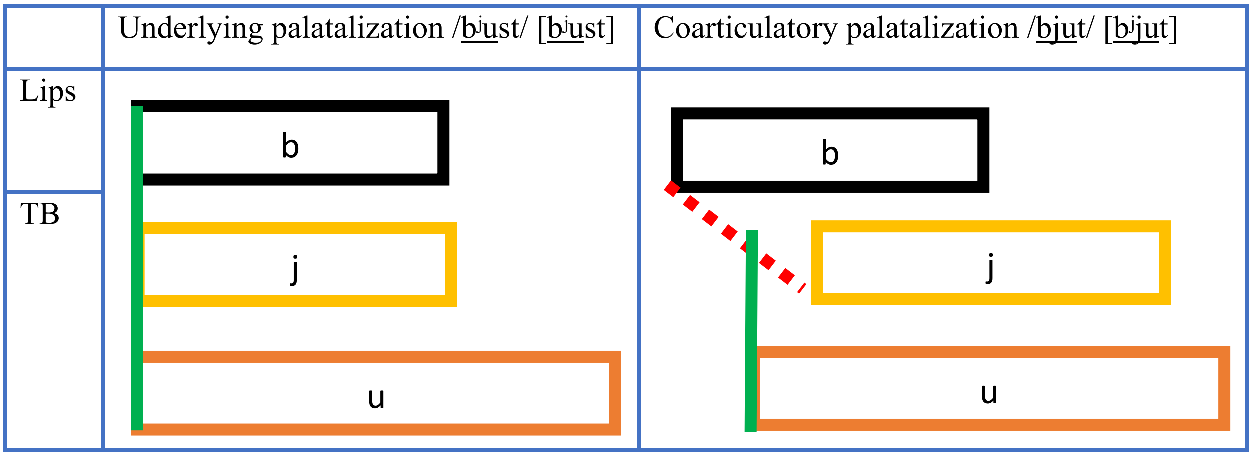


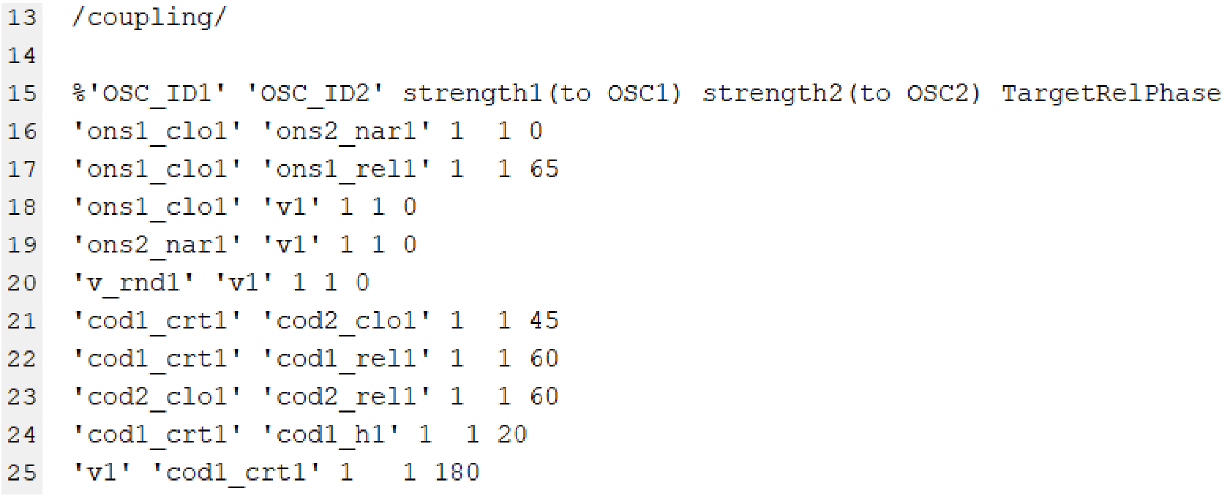
Figure 38: Gestural models for /bʲust/ and /bjut/. The green solid lines indicate in-phase coupling, and the red dotted line indicates 90-degree phase coupling.

Table 14: ARPABET

IPA	ARPABET	IPA	ARPABET	IPA	ARPABET	IPA	ARPABET	IPA	ARPABET
/p/	P	/f/	F	/m/	M	/i/	IY	/u/	UW
/b/	B	/v/	V	/n/	N	/ɪ/	IH	/ʊ/	UH
/t/	T	/θ/	TH	/ŋ/	NX	/e/	EY	/o/	OW
/d/	D	/ð/	DH	/l/	L	/ɛ/	EH	/ɔ/	AO
/k/	K	/s/	S	/ɹ/	R	/æ/	AE	/a/	AA
/g/	G	/z/	Z	/j/	Y	/ʌ/	AH	/aɪ/	AI
/tʃ/	CH	/ʃ/	SH	/w/	W	/ə/	AX	/aʊ/	AW
/dʒ/	JH	/ʒ/	ZH	/h/	HH	/ɚ/	ER	/ɔɪ/	OY

Table 15: Gestural specifications for /b/, /j/, and /u/

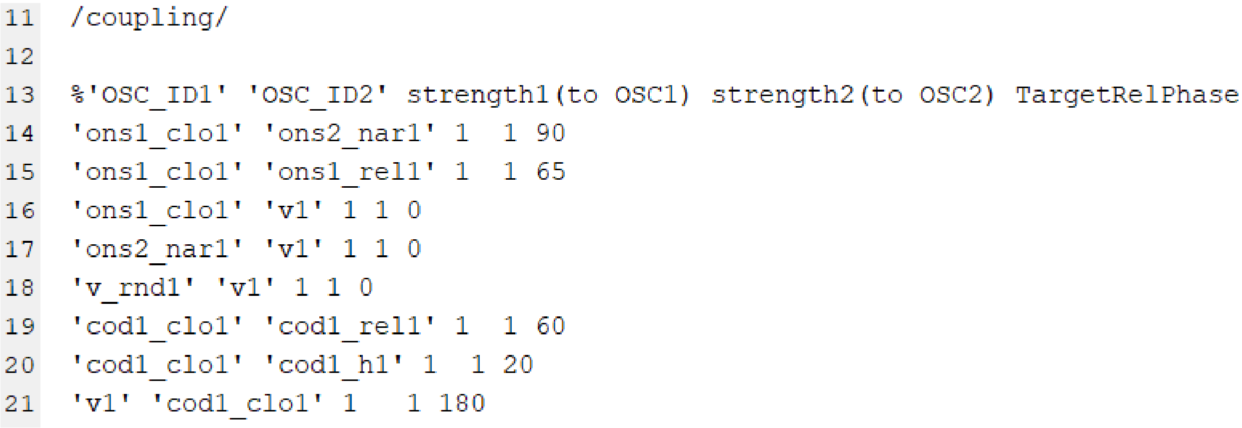
IPA	Organ	OSC_ID	TV	Constrict	Target	Stiff	Blending
/b/	Lips	ons1_clo1	LA	CLO	-2	8	100
	Lips	ons1_rel1	LA	REL	11	8	1
	Velum	ons1_clo1	VEL	CLO	-0.1	8	0
/j/	TB	ons2_nar1	TBCL	PAL	95	8	100
	TB	ons2_nar1	TBCD	NAR	2	8	100
	Lips	ons2_nar1	LA	V	8	8	1
/u/	TB	v1	TBCL	UVU/VEL	125	4	1
	TB	v1	TBCD	V	2	4	1
	Lips	v_rnd1	LP	PRO	12	4	1
	Lips	v_rnd1	LA	NAR	5	4	1



```
/coupling/

%'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
'ons1_clo1' 'ons2_nar1' 1  1 0
'ons1_clo1' 'ons1_rel1' 1  1 65
'ons1_clo1' 'v1' 1 1 0
'ons2_nar1' 'v1' 1 1 0
'v_rnd1' 'v1' 1 1 0
'cod1_crt1' 'cod2_clo1' 1  1 45
'cod1_crt1' 'cod1_rel1' 1  1 60
'cod2_clo1' 'cod2_rel1' 1  1 60
'cod1_crt1' 'cod1_h1' 1  1 20
'v1' 'cod1_crt1' 1   1 180
```

Figure 39: Coupling specifications for /bʲust/



```
/coupling/

%'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
'ons1_clo1' 'ons2_nar1' 1  1 90
'ons1_clo1' 'ons1_rel1' 1  1 65
'ons1_clo1' 'v1' 1 1 0
'ons2_nar1' 'v1' 1 1 0
'v_rnd1' 'v1' 1 1 0
'cod1_clo1' 'cod1_rel1' 1  1 60
'cod1_clo1' 'cod1_h1' 1  1 20
'v1' 'cod1_clo1' 1   1 180
```

Figure 40: Coupling specifications for /bjut/

4.3.2. Model 2

As hypothesized in Section 3.3, Model 2 is proposed to test the prediction that the incomplete neutralization observed in Russian palatalization patterns is attributable to the gestural blending between the velar gesture and the palatal gesture. In this model, I kept the 90-degree phase timing between /b/ and /j/ from Model 1, and then added the velar gesture, which is coordinated in-phase with the labial gesture. That is, this model is different from Model 1 in that there is a gesture for the secondary velarization in the coarticulatory palatalization. Gestural models for /b^{j}ust/ and /b^{ɰ}jut/_90-degree-phase are schematized in Figure 41. The gestural blending of the tongue backing for /ɰ/ and /u/ and the fronting for /j/ may result in a retracted tongue position at the onset of TB gesture, and delayed onset-to-onset lag for the coarticulatory palatalization.

The gestural score for /b^{ɰ}jut/_90-degree-phase is created in the following way. First, I created the coupling graph through the GEST menu in TADA by using ARPABET (B W Y-UW_T). To create /ɰ/ from the gestural specifications for /w/, I delete gestural specifications for labial gestures in the TVb^{ɰ}jut_anti.O file. The gestural specifications for /ɰ/ are summarized in Table 16. See Table 15 for the gestural specifications of /b/, /j/, and /u/. Then, I also modified the phasing relation between /b/ and /ɰ/ to be in-phase by changing *TargetRelPhase* to 0 in the PHbᶭjut_anti.O file (See line 15 in Figure 42). See Section 4.2.1.1 for the gestural and coupling specifications for /bʲust/.

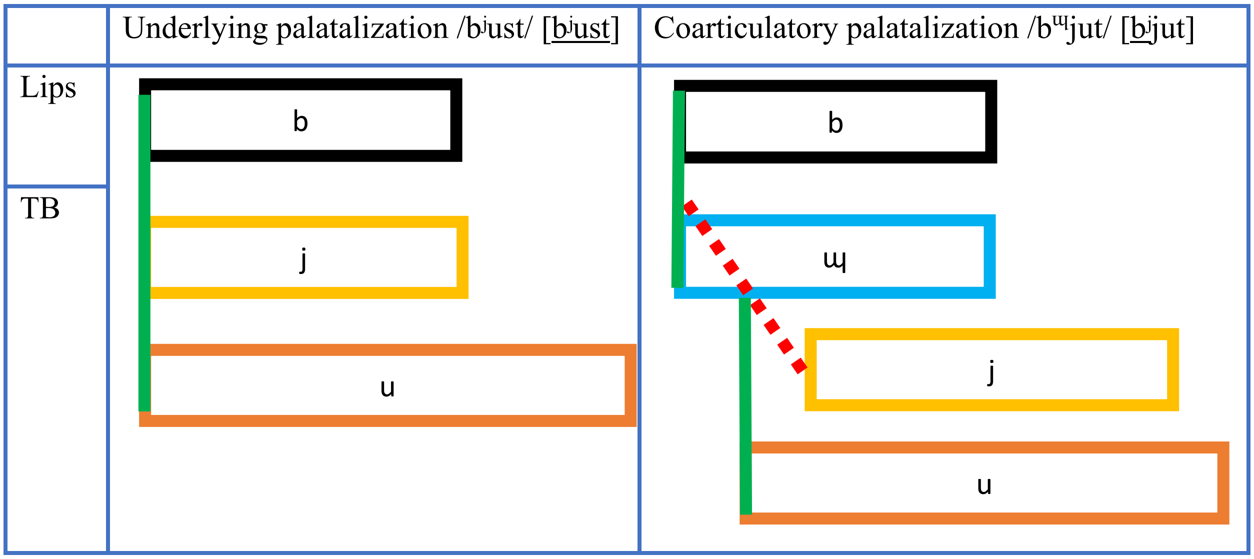
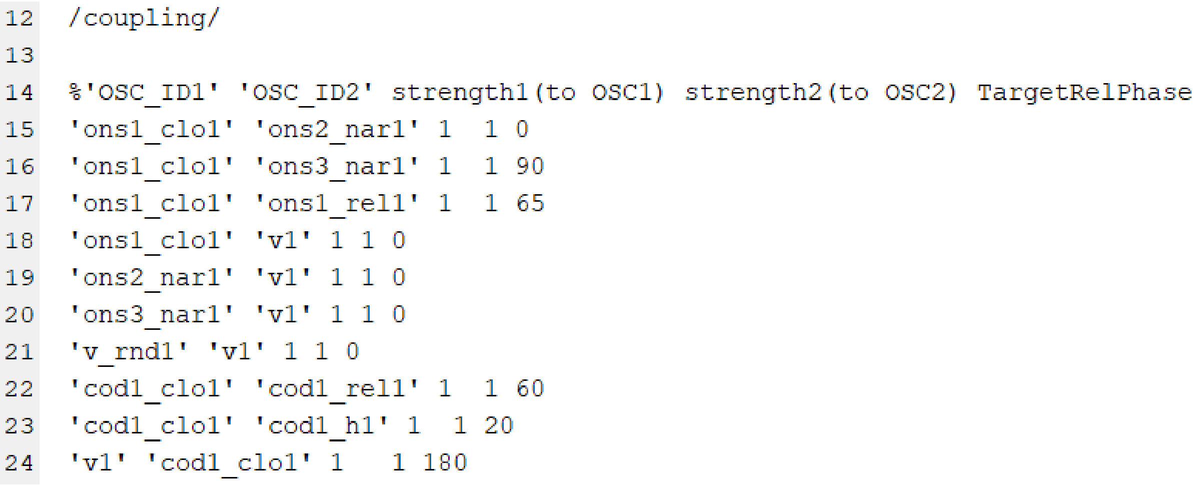


Figure 41: Gestural models for /bʲust/ and /bᶭjut/_90-degree-phase. The green solid lines indicate in-phase coupling, and the red dotted line indicates 90-degree-phase coupling.



```
12  /coupling/
13
14  %'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
15  'ons1_clo1' 'ons2_nar1' 1  1 0
16  'ons1_clo1' 'ons3_nar1' 1  1 90
17  'ons1_clo1' 'ons1_rel1' 1  1 65
18  'ons1_clo1' 'v1' 1 1 0
19  'ons2_nar1' 'v1' 1 1 0
20  'ons3_nar1' 'v1' 1 1 0
21  'v_rnd1' 'v1' 1 1 0
22  'cod1_clo1' 'cod1_rel1' 1  1 60
23  'cod1_clo1' 'cod1_h1' 1  1 20
24  'v1' 'cod1_clo1' 1   1 180
```

Figure 42: Coupling specifications for /bᶭjut/_90-degree-phase

Table 16: Gestural specifications for /ɰ/

IPA	Organ	OSC_ID	TV	Constrict	Target	Stiff	Blending
/ɰ/	TB	ons2_nar1	TBCL	UVU/VEL	125	8	10
	TB	ons2_nar1	TBCD	NAR	2	8	100

4.3.3. Model 3

Model 3 is similar to Model 2 except that gestures for /b/, /ɰ/, and /j/ are all coupled in-phase with each other for the coarticulatory palatalization. Gestural models for /bʲust/ and /bᶭjut/_in-phase are visualized in Figure 43. The gestural score for /bᶭjut/_in-phase is created in the following way. First, I take the coupling graph for /bᶭjut/_90-degree-phase and modify the phasing relation between /b/ and /j/ to be in-phase by changing *TargetRelPhase* to 0 in the PHbᶭjut_in.O file (See line 16 in Figure 44). The gestural specifications for /bᶭjut/_in-phase are the same as /bᶭjut/_90-degree-phase (See Table 15 and 16). Also, see Section 4.2.1.1 for the gestural and coupling specifications for /bʲust/.

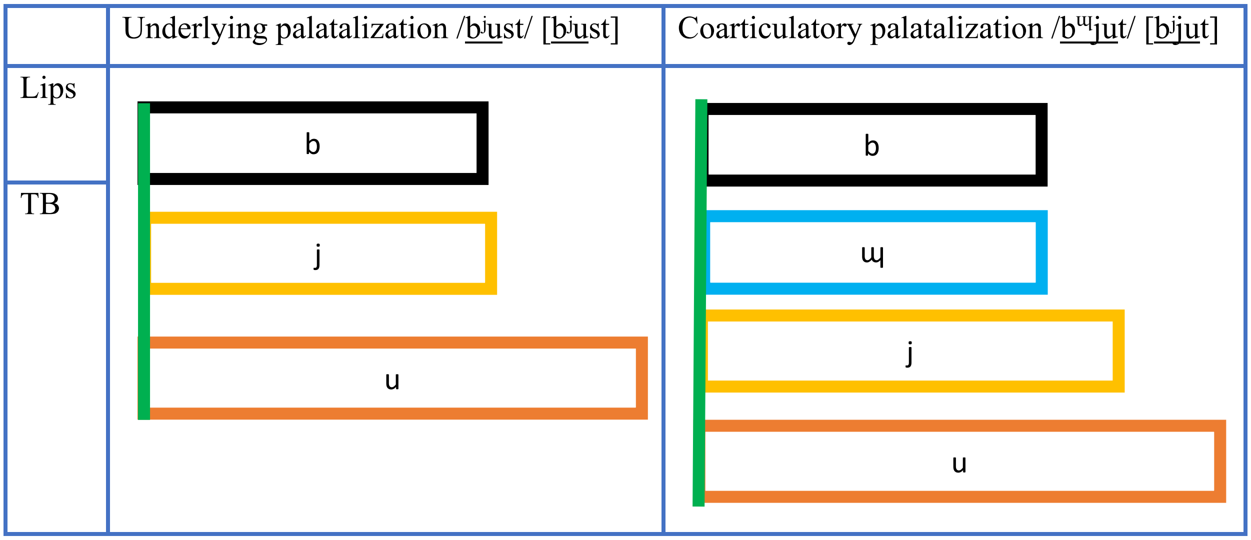
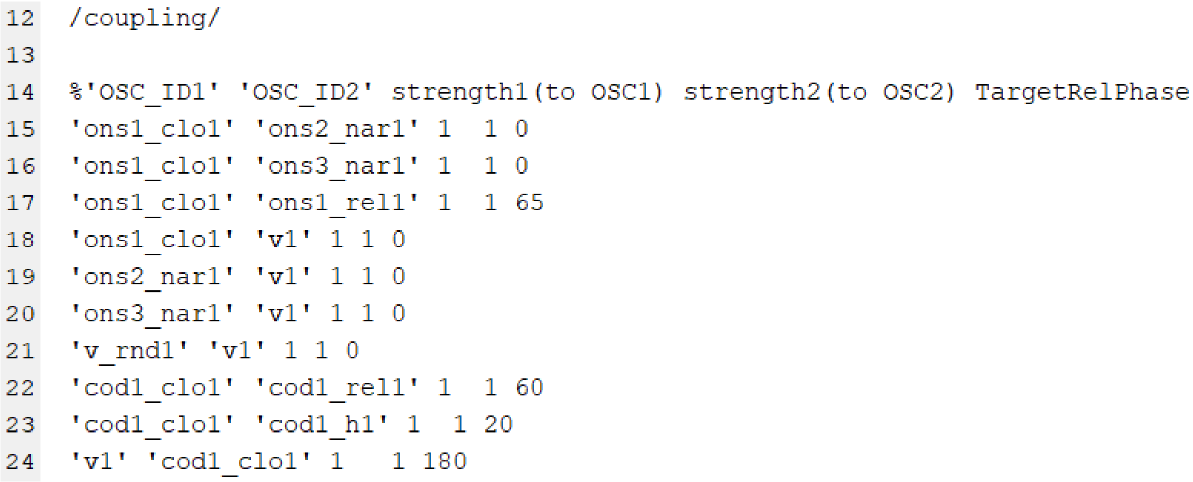


Figure 43: Gestural models for /bʲust/ and /bᶭjut/_in-phase. The green solid lines indicate in-phase coupling.



```
12  /coupling/
13
14  %'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
15  'ons1_clo1' 'ons2_nar1' 1  1 0
16  'ons1_clo1' 'ons3_nar1' 1  1 0
17  'ons1_clo1' 'ons1_rel1' 1  1 65
18  'ons1_clo1' 'v1' 1 1 0
19  'ons2_nar1' 'v1' 1 1 0
20  'ons3_nar1' 'v1' 1 1 0
21  'v_rnd1' 'v1' 1 1 0
22  'cod1_clo1' 'cod1_rel1' 1  1 60
23  'cod1_clo1' 'cod1_h1' 1  1 20
24  'v1' 'cod1_clo1' 1   1 180
```

Figure 44: Coupling specifications for /b^{ɰ}jut/_in-phase

4.3.4. Model 4

As discussed in Section 3.6.3, the velar gesture may start before the palatal gesture for /j/ and continues concurrently for the coarticulatory palatalization. To test this, I posit an eccentric timing between /ɰ/ and /j/ for coarticulatory palatalization in Model 4, as schematized in Figure 45. To create the gestural score for /b^{ɰ}jut/_eccentric, I take the coupling graph for /b^{ɰ}jut/_in-phase and modify the phasing relation between /ɰ/ and /j/ to have eccentric timing by changing *TargetRelPhase* from 0 degrees to 45 degrees in the PHb^{ɰ}jut_eccentric.O file (See line 15 in Figure 46). The gestural specifications for /b^{ɰ}jut/_eccentric are the same as /b^{ɰ}jut/_in-phase (See Table 15 and 16). Also, see Section 4.2.1.1 for the gestural and coupling specifications for /b^{j}ust/.

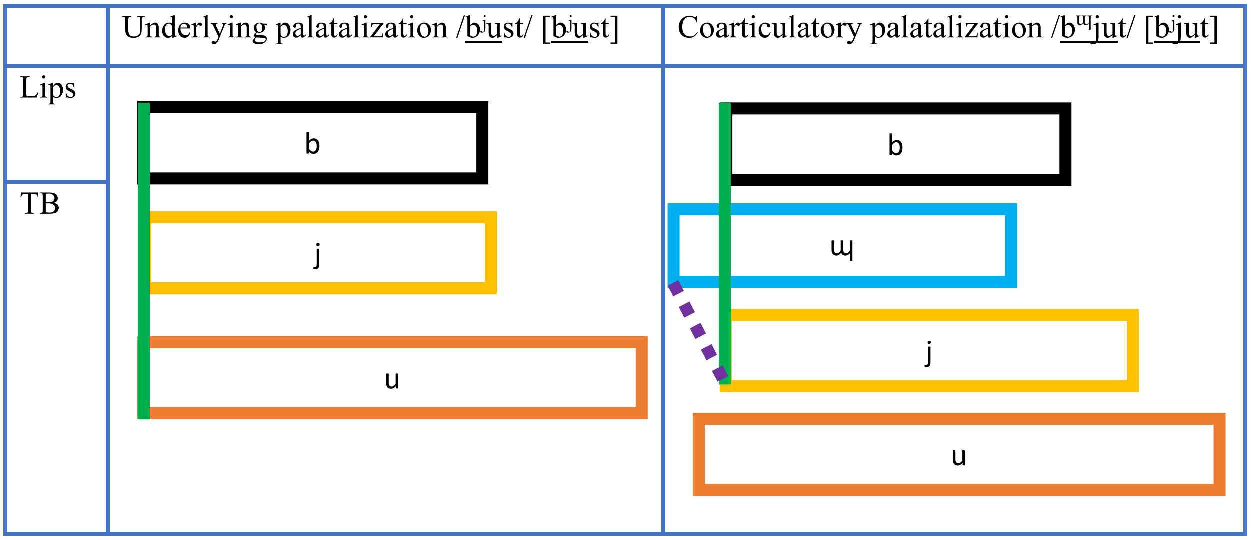
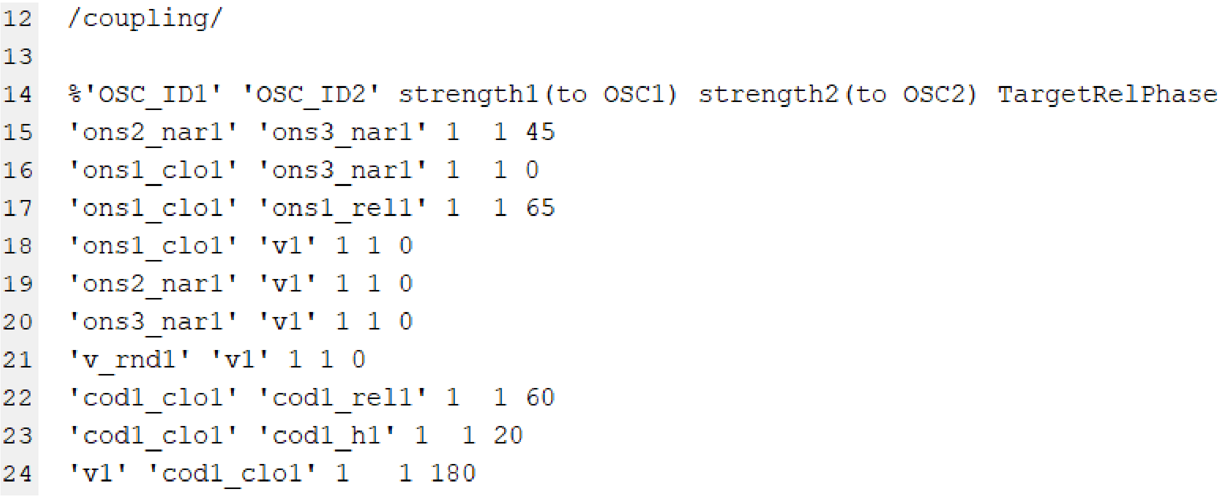


Figure 45: Gestural models for /bʲust/ and /bᶭjut/_eccentric. The green solid lines indicate in-phase coupling, and the purple dotted line indicates eccentric timing.



```
/coupling/

%'OSC_ID1' 'OSC_ID2' strength1(to OSC1) strength2(to OSC2) TargetRelPhase
'ons2_nar1' 'ons3_nar1' 1  1 45
'ons1_clo1' 'ons3_nar1' 1  1 0
'ons1_clo1' 'ons1_rel1' 1  1 65
'ons1_clo1' 'v1' 1 1 0
'ons2_nar1' 'v1' 1 1 0
'ons3_nar1' 'v1' 1 1 0
'v_rnd1' 'v1' 1 1 0
'cod1_clo1' 'cod1_rel1' 1  1 60
'cod1_clo1' 'cod1_h1' 1  1 20
'v1' 'cod1_clo1' 1   1 180
```

Figure 46: Coupling specifications for /bɯjut/_eccentric

4.4. Results from TADA simulations

A key finding from Chapter 3 is that the palatal-plain contrast in this context is neutralized, but more importantly, this neutralization is phonetically incomplete. In particular, both palatalizations exhibit the temporal coordination of complex segments (a nearly flat regression line with a slight upward trend), suggesting that plain consonants in the coarticulatory palatalization context are also palatalized. However, I also found residual evidence of an underlying tongue dorsum retraction for the coarticulatory palatalization. The spatial position of the TB is more retracted for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization at the onset of the palatal gesture. In addition, the lag between the gesture onsets is longer for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization. In this section, I evaluate each model that I proposed in Section 4.3 by comparing the simulations from each model against the results from the EMA recordings.

4.4.1. Model 1

As discussed in Section 4.3.1, Model 1 is proposed to examine whether gestural blending between /j/ and /u/ alone can lead to incomplete neutralization in Russian palatalization. As visualized in Figure 38, Model 1 posits that /b/ and /j/ are coupled in-phase with each other in /b^{j}ust/ (**UNDERLYING** palatalization), while they are coupled 90 degree phase in /bjut/ (**COARTICULATORY** palatalization).

Figure 47 plots the correlation between *G1 duration* (x-axis) and *onset-to-onset lag* (y-axis) across *Status*. To illustrate the trend in the data, a least-squares linear regression line is fit to each panel. The regression line for /b^{j}ust/ (**UNDERLYING** palatalization) is nearly flat, precisely the pattern predicted for complex segments. Crucially, however, simulations for **COARTICULATORY** palatalization from Model 1 turn out to behave like a segment sequence. The *onset-to-onset lag* increases with *G1 duration*, leading to a positive correlation between them. That is, gestural blending between /j/ and /u/ leads to no neutralization between **UNDERLYING** and **COARTICULATORY** palatalization, unlike the results from the EMA recording.

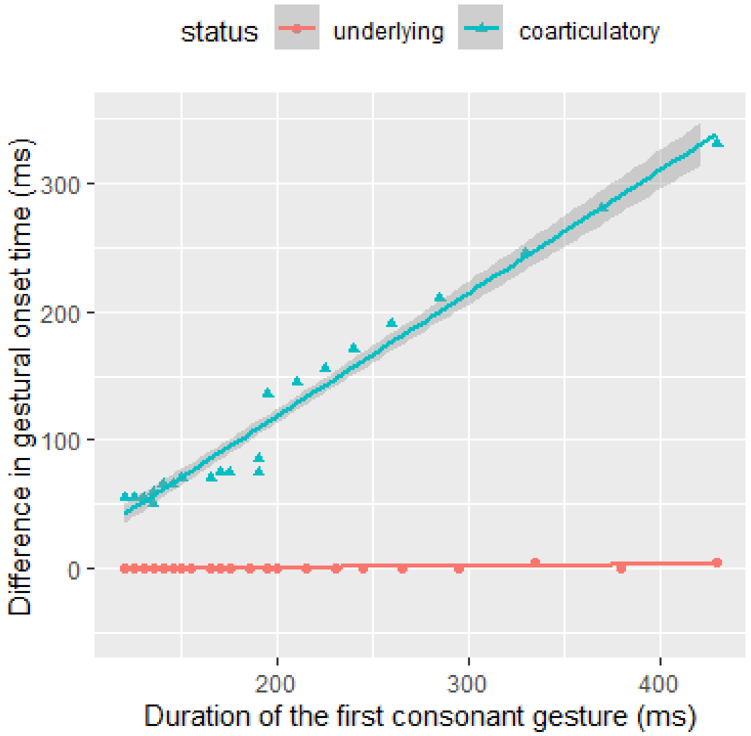


Figure 47: Model 1 – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

Figure 48 shows the horizontal position (front-back) of the TB sensors at the gestural onset. Positive and negative values on the y-axes illustrate the frontness and backness of the tongue body, respectively. The spatial position of the TB tends to be slightly more retracted for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization at the onset of the palatal gesture. However, considering that the difference in TB position is approximately 1.5 mm in the EMA results (See Section 3.5.3.1), this difference is too small. Figure 49 provides a box plot of onset-to-onset lag across *Status*. The lag between the gesture onsets is longer for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization.

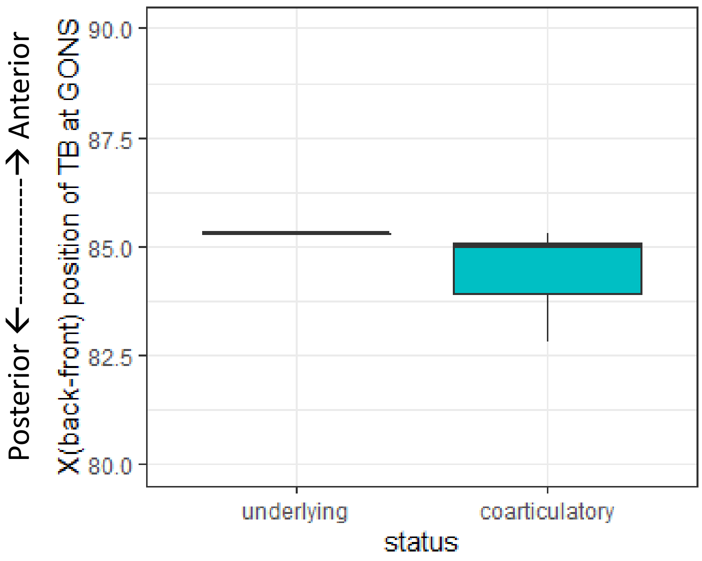


Figure 48: Model 1 – A boxplot of TB position (mm) at palatal gesture onset

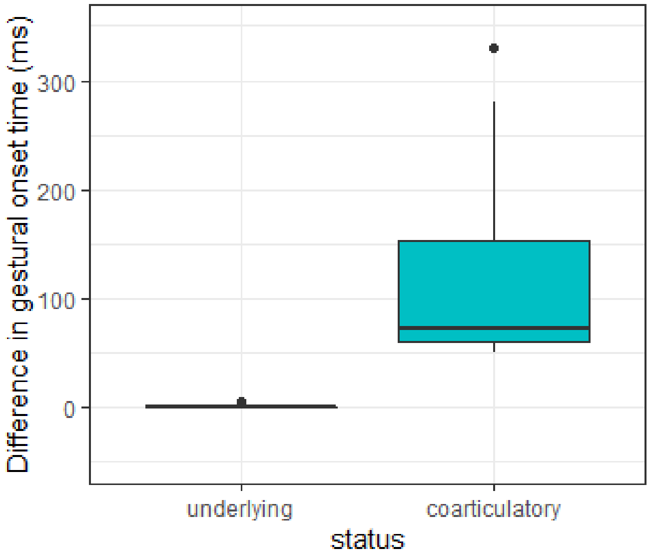


Figure 49: Model 1 – A boxplot of onset-to-onset lag across Status

In sum, simulations from Model 1 produce delayed onset-to-onset lag for the coarticulatory palatalization, but crucially, fail to produce the temporal coordination of **UNDERLYING** and **COARTICULATORY** palatalization in Russian, as well as a retracted tongue position at the onset of

TB gesture. In the next section, I review the results of simulations from Model 2.

4.4.2. Model 2

For Model 2, I added a gesture for the secondary velarization in **COARTICULATORY** palatalization (See Section 4.3.2 and Figure 41). In particular, Model 2 posits that gestures for /b/ and /j/ are coupled in-phase with each other in /bʲust/ (**UNDERLYING** palatalization), while gestures for /b/ and /j/ are coupled 90 degree phase with each other and gestures for /b/ and /ᵚ/ are coupled in-phase in /bᵚjut/_90-degree-phase (**COARTICULATORY** palatalization).

As shown in Figure 50, the regression line for /bʲust/ (**UNDERLYING** palatalization) is nearly flat, precisely the pattern predicted for complex segments, while/bᵚjut/_90-degree-phase shows a positive correlation between the onset-to-onset lag and G1 duration. That is, simulations for coarticulatory palatalization from Model 2 turn out to behave like a segment sequence, similar to Model 1. However, unlike Model 1, the spatial position of the TB is more retracted for the coarticulatory palatalization than for the underlying palatalization at the onset of the palatal gesture, as shown in Figure 51. The lag between the gesture onsets is longer for the coarticulatory palatalization than for the underlying palatalization (See Figure 52).

In sum, simulations from Model 2 produce a more retracted tongue position and delayed onset-to-onset lag for the coarticulatory palatalization, but crucially, fail to produce the temporal coordination of **COARTICULATORY** palatalization in Russian observed in the EMA data. In the next section, I review the results of simulations from Model 3.

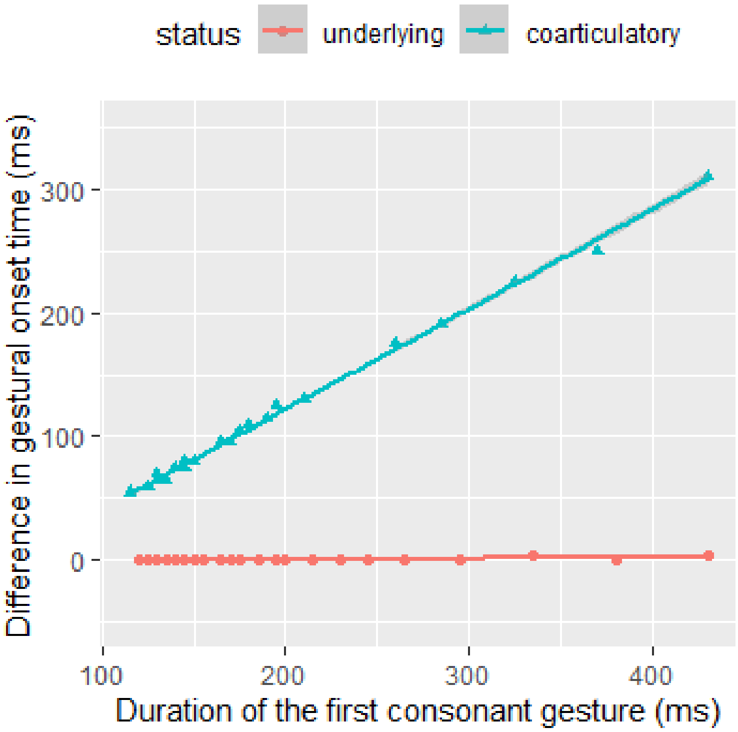


Figure 50: Model 2 – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

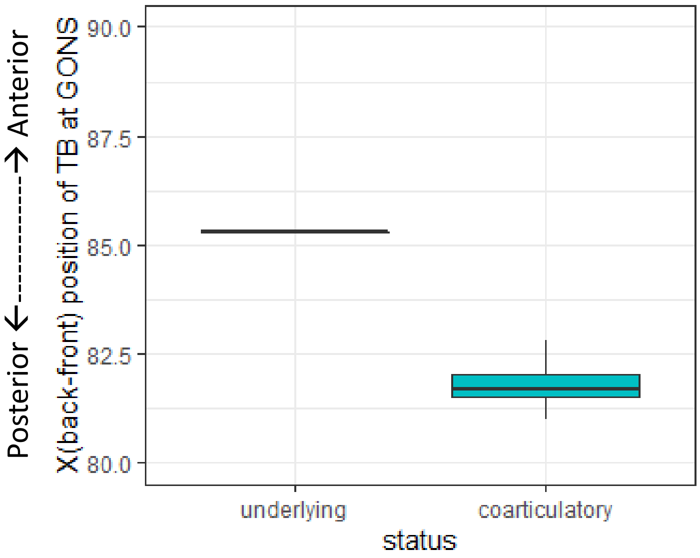


Figure 51: Model 2 – A boxplot of TB position (mm) at palatal gesture onset

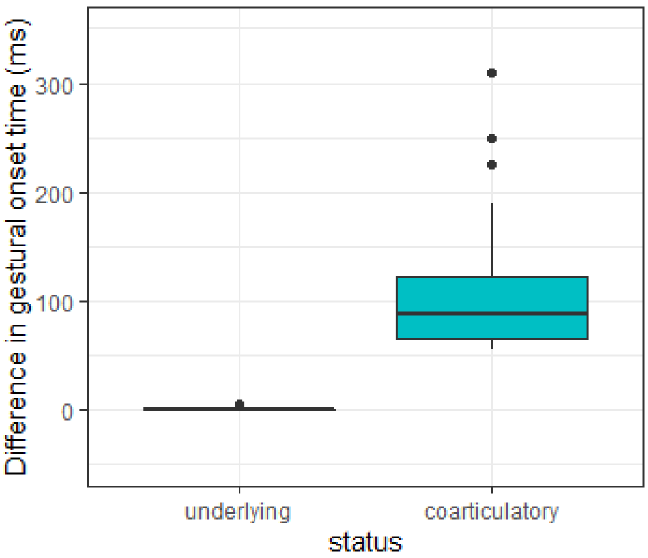


Figure 52: Model 2 – A boxplot of onset-to-onset lag across Status

4.4.3. Model 3

As discussed in Section 4.3.3, Model 3 is proposed to examine whether in-phase coupling among /b/, /ɰ/, and /j/, and gestural blending between /ɰ/ and /j/ can lead to incomplete neutralization in Russian palatalization. Gestural models for /bʲust/ and /bᶭjut/_in-phase are visualized in Figure 43.

As shown in Figure 53, the regression lines for both /bʲust/ (UNDERLYING palatalization) /bᶭjut/_in-phase (COARTICULATORY palatalization) are nearly flat, suggesting that simulations for both types of palatalization from Model 3 turn out to behave like a complex segment, similar to the results from the EMA recordings. Crucially, however, simulations from Model 3 show a difference between UNDERLYING and COARTICULATORY palatalization in neither the spatial position of the TB, nor in the lag between the gesture onsets (See Figure 54 and 55, respectively).

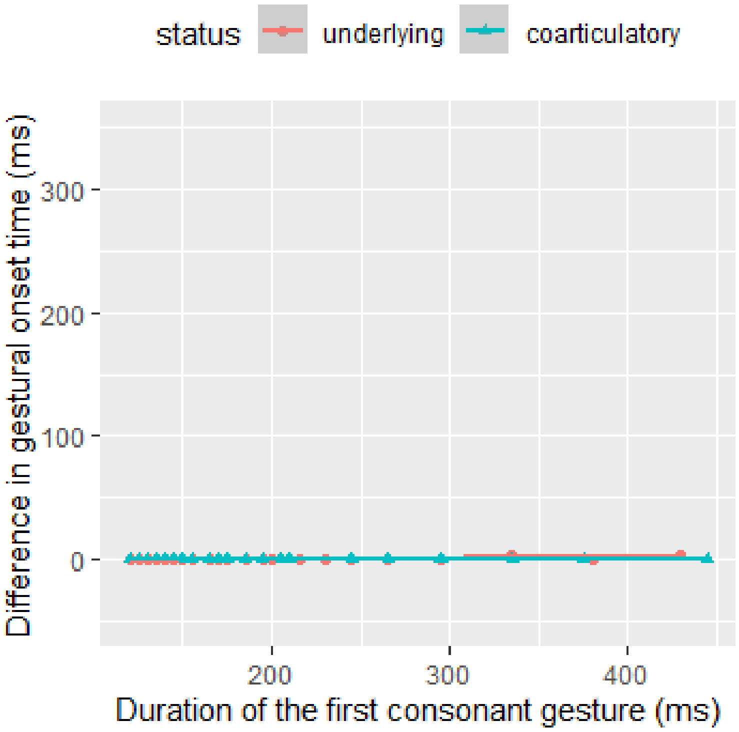


Figure 53: Model 3 – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

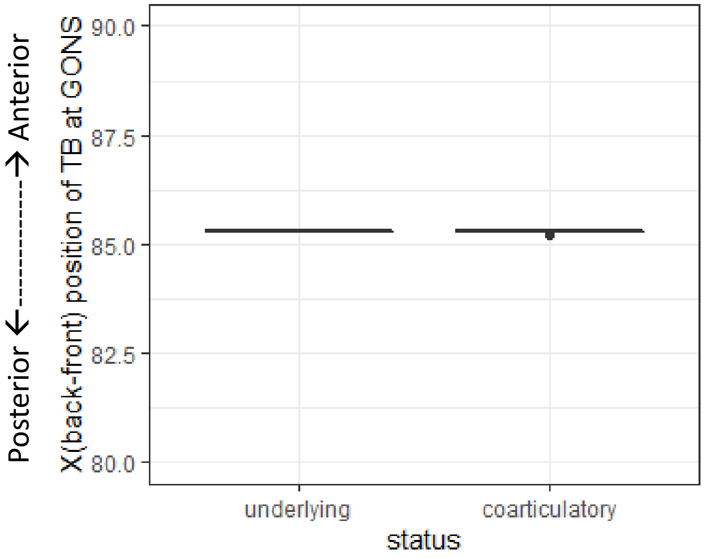


Figure 54: Model 3 – A boxplot of TB position (mm) at palatal gesture onset

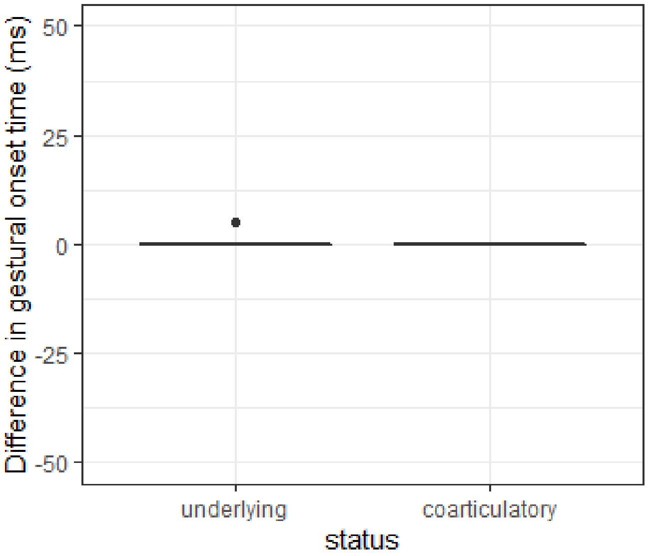


Figure 55: Model 3 – A boxplot of onset-to-onset lag across Status

In sum, simulations from Model 3 produce the temporal coordination of coarticulatory palatalization in Russian observed in the EMA data. However, they fail to produce a more retracted tongue position and delayed onset-to-onset lag for the coarticulatory palatalization. The results of simulations from Model 4 are presented in the next section.

4.4.4. Model 4

For Model 4, I posit an eccentric timing between /ɰ/ and /j/ for coarticulatory palatalization as schematized in Figure 45. In particular, gestures for /ɰ / and /j/ are coupled eccentric-phase (45 degrees) with each other, and gestures for /b/ and /j/ are coupled in-phase in /b^{ɰ}jut/_eccentric (**COARTICULATORY** palatalization).

As shown in Figure 56, for both **UNDERLYING** and **COARTICULATORY** palatalization, the regression line is nearly flat (although **COARTICULATORY** palatalization shows only a slight upward trend). That is, simulations for both **UNDERLYING** and **COARTICULATORY** palatalization from Model 4 show temporal coordination of a complex segment similar to simulations from Model 3. Furthermore, the spatial position of the TB is more retracted for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization at the onset of the palatal gesture, as shown in Figure 57. The lag between the gesture onsets is also longer for the **COARTICULATORY** palatalization than for the **UNDERLYING** palatalization (See Figure 58).

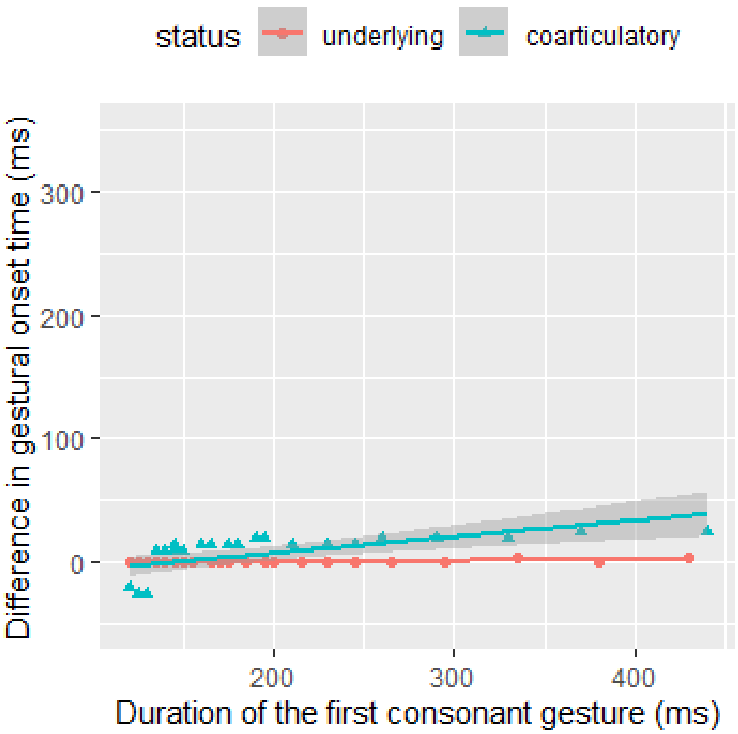


Figure 56: Model 4 – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

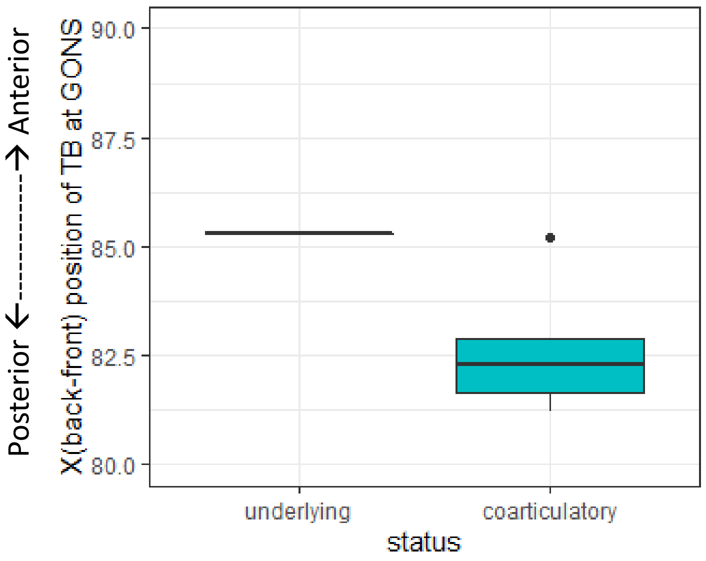


Figure 57: Model 4 – A boxplot of TB position (mm) at palatal gesture onset

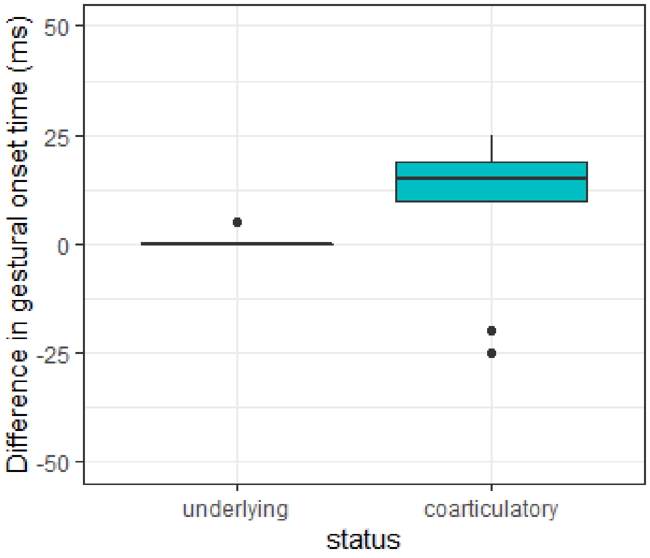


Figure 58: Model 4 – A boxplot of onset-to-onset lag across Status

In this section, I compare the simulations from each gestural model against the results from the EMA recordings. All the models I evaluated in this section except for Model 4 either fail to produce the temporal coordination of **COARTICULATORY** palatalization observed in the EMA data (Model 1 and Model 2) or fail to produce a more retracted tongue position and delayed onset-to-onset lag for **COARTICULATORY** palatalization (Model 3). However, simulations from Model 4 produce the same results that were observed in the EMA data: a retracted tongue position at the onset of TB gesture and delayed onset-to-onset lag for **COARTICULATORY** palatalization, as well as no effect of variation in consonant duration on onset-to-onset lag.

4.5. Discussion

4.5.1. Overview

The goal of this chapter was to find the best gestural models to represent the underlying and coarticulatory palatalization in Russian. To do so, I explored four gestural models of underlying and coarticulatory palatalization in Russian using an articulatory-based synthesizer, Task Dynamic Application (TADA), and compared the simulations from each model against the results from the EMA recordings.

Model 1 was proposed to examine whether gestural blending between /j/ and /u/ alone can lead to incomplete neutralization of Russian palatalization types. The rest of the models contained both velar and palatal gestures for the coarticulatory palatalization with different phasing relations with the other gestures. Model 2 and Model 3 were proposed to test the prediction that the incomplete neutralization is attributable to gestural blending between a velar gesture and a palatal gesture. The only difference between Model 2 and Model 3 was that prevocalic gestures in Model 3 were coordinated in phase with each other. Lastly, Model 4 was proposed to test whether the velar gesture starts before the palatal gesture and whether this eccentric timing leads to incomplete neutralization.

All the models I evaluated in this chapter except for Model 4 either failed to produce the temporal coordination of coarticulatory palatalization observed in the EMA data (Model 1 and

Model 2) or failed to produce a more retracted tongue position and delayed onset-to-onset lag for coarticulatory palatalization (Model 3). However, simulations from Model 4 produced the same results that were observed in the EMA data: a retracted tongue position at the onset of TB gesture and delayed onset-to-onset lag for coarticulatory palatalization, as well as no effect of variation in consonant duration on onset-to-onset lag.

4.5.2. Temporal coordination of two secondary articulation gestures

The coupled oscillator model hypothesizes that multiple gestures in a syllable onset are coupled anti-phase with each other, along with both being in-phase with the following vowel (Goldstein et al., 2006; Goldstein et al., 2009; Nam et al., 2009; Saltzman et al., 2008). Regarding this complex onset timing, the vowel starts at the center of prevocalic consonants (the so-called c-center effect) due to the competitive coupling. The complex onset timing has been found in English (e.g., Browman & Goldstein, 1988), French (e.g., Kühnert et al., 2006), Georgian (e.g., Goldstein et al., 2007), Serbian (e.g., Tilsen et al., 2012), etc. In contrast, recent studies revealed that some languages show simple onset timing for onset consonant clusters (Goldstein et al., 2007; Hermes et al., 2017; Shaw et al., 2011; Tilsen et al., 2012). Regarding this simple onset timing, onset consonants are coupled anti-phase with each other, while only the right-most prevocalic consonant is coupled in-phase with the vowel. The simple onset timing has been found in Moroccan Arabic (Shaw et al., 2011), Tashlhiyt Berber (Hermes et al., 2017), Hebrew (Tilsen et al., 2012), Montreal French (Tilsen et al., 2012), etc.

In this dissertation, Model 4 which produces the same results that were observed in the EMA data, resembles the simple onset timing. That is, like the simple onset timing in which only the rightmost prevocalic consonant is coordinated in-phase with the vowel, Model 4 shows that among the secondary articulations only the palatal gesture is coupled in-phase with the labial gesture. Considering there are two secondary articulations involved in cases of coarticulatory palatalization, it is possible that these two secondary articulation gestures are coupled in eccentric-phase with each other, just like onset clusters are anti-phase with each other.

Although simulations with gestural blending between the velar and palatal gestures and eccentric timing showed the same results that were observed in the EMA data, this does not necessarily mean that this is the only model that can produce these results. Another possible scenario is that the eccentric timing between the velar and palatal gestures might be attributable to a competitive coupling between the two gestures, such that they are both timed in-phase to the labial gesture but anti-phase (180°) to each other. However, the competitive coupling also failed to produce a more retracted tongue position and delayed onset-to-onset lag for coarticulatory palatalization observed in the EMA data (See Appendix 1). I also explored other competitive coupling graphs with four different phasing relations (90°, 65°, 45°, and 20°) for the velar and palatal gestures, but they also failed to produce a more retracted tongue position and delayed onset-to-onset lag for coarticulatory palatalization. I reported the simulations from the competitive coupling graph with a relative phase of 20° of a gesture's oscillator as an example in Appendix 1.

4.6. Summary

In this chapter, I explored four gestural models of underlying and coarticulatory palatalization in Russian using an articulatory-based synthesizer, Task Dynamic Application (TADA), and compared the simulations from each model against the results from the EMA recordings. In the gestural model, which exhibited a similar outcome with the EMA recordings, the palatal and velar gestures are set to be coupled eccentric-phase (45°) with each other, and the palatal gesture is set to be coupled in-phase with the labial gesture for coarticulatory palatalization. In contrast, for underlying palatalization, there is no velarization gesture, and the labial and palatal gestures are set to be coupled in phase with each other. Simulations from this model showed a temporal coordination of complex segments for both underlying and coarticulatory palatalizations, as well as a retracted tongue position at the onset of TB gesture and delayed onset-to-onset lag for coarticulatory palatalization. The simulations from the computational modeling supported that gestural blending between the velar and palatal gestures and their eccentric timing may lead to incomplete neutralization of underlying and coarticulatory palatalizations in Russian. These results are consistent with previous findings regarding secondary velarization/uvularization in plain consonants, as well as the EMA results presented in Chapter 3.

Chapter 5. Summary and general discussion

5.1. Summary

Incomplete neutralization refers to small but significant phonetic traces of underlying contrasts in phonologically neutralizing contexts. It has been found for final devoicing in many languages (e.g., Port & O’Dell, 1985), flapping in American English (Herd et al., 2010), vowel epenthesis in Levantine Arabic (Gouskova & Hall, 2009), among other patterns. Due to the difficulty of incorporating it into the grammar, however, incomplete neutralization presents serious challenges to most phonological models. To address this issue, I have examined incomplete neutralization in the Articulatory Phonology framework, investigating underlying palatalization and coarticulatory palatalization in Russian as a test case.

In Chapter 2, I first established a quantification of palatalization in Russian, by examining temporal coordination in complex segments versus segment sequences, with the Russian palatalized consonants as a representative case of complex segments and the English consonant-palatal glide sequences as a representative case of segment sequences. I hypothesized that complex segments differ from segment sequences in terms of how constituent articulatory gestures are coordinated in time. Following Shaw and colleagues (2019), I derived the following predictions: if the gesture onsets are timed to each other (complex segment), a longer G1 duration will not delay the G2 onset, leading to no correlation between G1 duration and temporal lag between gesture onsets. In contrast, if G2 is timed to some gestural landmark later in the unfolding of G1 (segment sequence), increases in G1 duration will delay the onset of G2, increasing the onset lag. Notably, these predictions relate variation in one phonetic dimension, G1 duration, to variability in another, the lag between gesture onsets. Results from an EMA experiment confirmed the predictions, showing no correlation between G1 duration and onset lag for complex segments, and showing a positive correlation for segment sequences. Thus, in Chapter 2, I showed that it is the pattern of covariation between phonetic dimensions that uniquely distinguishes phonological structures on the basis of coordination.

Implementing this temporal diagnostic in Chapter 3, I explored the phonetic realization of two palatalization patterns in Russian. In Russian, the contrast between a palatalized consonant (e.g., /lj/) and a “plain” consonant (e.g., /l/) is reported to be neutralized to the palatal counterpart when a plain consonant is followed by a glide (e.g., Kochetov, 2011). That is, the coarticulatory palatalization of the plain stop in the environment preceding palatal glides results in apparent neutralization of the palatalized vs. plain contrast in, e.g., /lʲut/ [lʲut] ‘fierce’ (underlying palatalization) vs. /ljut/ [lʲjut] ‘pour (3p pl)’ (coarticulatory palatalization). However, previous studies have reported that plain consonants may actually feature secondary velarization/uvularization (e.g., Roon & Whalen, 2019). A question that arises from consideration of these patterns is whether the apparent neutralization is phonetically complete. The hypothesis that I pursued is that it is not, and that the distinction between underlying and coarticulatory palatalization is in fact maintained by subtle spatio-temporal differences that result from the gestural blending of palatalization and velarization/uvularization gestures. If the contrast is effectively neutralized but the neutralization is phonetically incomplete, both types of palatalization would exhibit the temporal coordination associated with complex segments, but showing phonetic traces of the underlying contrast.

A key finding from the EMA experiment in Chapter 3 is as follows: both underlying and coarticulatory palatalizations exhibit inter-gestural coordination characteristic of complex segments. However, the spatial position is significantly more retracted for the coarticulatory palatalization than for the underlying palatalization at the onset of the palatal gesture, showing residual evidence of an underlying tongue dorsum retraction gesture. This suggests that the gestural blending of palatalization and velarization/uvularization leads to incomplete neutralization of underlying palatalization and coarticulatory palatalization in Russian.

In Chapter 4, to determine the gestural configuration that gives rise to the observed patterns, I explored four gestural models of underlying and coarticulatory palatalization in Russian using an articulatory-based synthesizer, Task Dynamic Application (TADA). Then, I compared the simulations from each model against the results from Electromagnetic Articulography (EMA) recordings. In Model 1, I modeled the coarticulatory palatalization as a ‘blending’ of two gestures for palatalization and the following back vowel, with a c-center timing for the vowel. This is in contrast to underlying palatalization, which I modeled using an in-phase timing between the palatal and vowel gestures. The rest of the models contain both velar and palatal gestures for the coarticulatory palatalization with different phasing relations with the other gestures for coarticulatory palatalization. In contrast, underlying palatalization was modeled without the gestural specifications for velarization (the gestural specifications for underlying palatalization are the same across models). In particular, I modeled the coarticulatory palatalization as a ‘blending’ of two gestures for velarization and palatalization with a 90-degree-phase timing between them (Model 2), with an in-phase timing between them (Model 3), and with an eccentric phase between them (Model 4). Among other models, simulations from Model 4 only produced the same results that were observed in the EMA data: a retracted tongue position at the onset of TB gesture and delayed onset-to-onset lag for coarticulatory palatalization, as well as no effect of variation in consonant duration on onset-to-onset lag.

Thus, the results from both EMA experimentation and computational modelling clearly show that the underlying and coarticulatory palatalization contrast in Russian represents a case of incomplete neutralization. Moreover, and crucial to the goals of this dissertation, it was shown that this case of incomplete neutralization can be modeled successfully as gestural blending in the Articulatory Phonology framework. Notably, and to the best of my knowledge, the present study is the first to systematically examine underlying and coarticulatory palatalization using kinematic data. In the following sections, I will discuss some important implications and possible applications of my findings for temporal coordination of complex segments and incomplete neutralization within the AP framework. More importantly, however, my dissertation provides new insights for interpreting incomplete neutralization in the AP framework.

5.2. Scope of the complex segmenthood hypothesis

The temporal diagnostics for complex segments and segment sequences presented in Chapter 2 have substantial potential to be applied to other cases of complex segments and segment sequences. My definition of a complex segment (from Section 2.4) is any segment that involves multiple articulatory gestures. This definition encompasses cases of secondary articulations, such as the palatalized consonants that are the empirical focus of Chapter 2, as well as cases sometimes termed "doubly articulated stops", such as /k͡p/, "contour segments" including affricates, e.g., /p͜s/, and others that are not so obvious.

For example, most gestural analyses of laterals, e.g., /l/, involve multiple gestures, whether a tongue tip and tongue dorsum gesture, as in Browman & Goldstein (1995), or more direct control of lateral channel formation, as in Ying et al. (2021). Since there are multiple gestures in /l/, I could ask if those gestures are coordinated according to the temporal diagnosis for complex segments.

One apparent problem for applying the complex segment diagnostic to /l/ is that the synchronicity of tongue tip and tongue dorsum kinematic movements, as tracked in the mid-sagittal plane, is sensitive to syllable position, showing greater synchronicity in syllable onset position than in syllable coda position (Sproat & Fujimura, 1993). This would be a problem if the temporal diagnostics predict that /l/ is a complex segment in syllable onset position and a sequence in coda position, while phonological behavior remains consistent across positions. However, as I have illustrated in Figure 7 and Figure 9, gestural overlap can be disassociated from coordination. Moreover, Ying et al. (2021) show that the timing of lateral channel formation in Australian English is temporally stable across syllable positions, even as the relative timing between tongue tip and tongue dorsum movements varies (as it does in American English and other varieties). This finding supports an analysis of /l/ as composed of a tongue tip gesture and a tongue blade lateralization gesture, which may indeed be coordinated as a complex segment across positions. In other words, the tongue dorsum retraction might not be under active control, but a side effect of other gestures, a proposal first raised by Sproat and Fujimura (1993).

The loss of /l/ in New Zealand English (i.e., /l/ vocalization) fits nicely into this discussion. There appears to be a stage in which active control of lateral channel formation gives way to a different gestural control structure involving tongue tip advancement and tongue dorsum retraction (Strycharczuk et al., 2020). This stage of development is similar to Browman and Goldstein's (1995b) proposal for American English. Interestingly, this gestural control structure might not be stable, as it precipitates the loss of the tongue tip gesture. Viewed from the standpoint of my hypothesis for complex segments, I could see the New Zealand development as a transition from /l/ as a complex segment (with tongue tip and tongue blade lateralization gestures) reinterpreted as a segment sequence (with a tongue dorsum retraction gesture followed by a tongue tip gesture) and then as a single (simplex) segment (just tongue dorsum retraction gesture).

More broadly, if I fail to identify the phonetic dimension under gestural control, I might not be able to diagnose coordination. The criteria for identifying gestures are twofold: a gesture (i) supports phonological contrast and (ii) specifies the dynamics of some phonetic dimension. To evaluate coordination, it is crucial to first establish the constituent gestures. This point is relevant as I seek to test the hypothesis on new cases of potential complex segments.

The phonetic dimensions of gestural control in early work in Articulatory Phonology were limited to a relatively small number of articulatory parameters, but have expanded over the years as demanded by empirical evidence. For example, the tongue blade lateralization gesture in Ying et al. (2021) was not one of the original eight dimensions of gestural control (known as "tract variables" in the Articulatory Phonology framework). Aerodynamic gestures (McGowan & Saltzman, 1995) and acoustic gestures have also been proposed to explain a wider range of phonological contrasts and experimental data. For example, f0, an acoustic parameter, is now widely assumed to be a dimension of gestural control in lexical tone (Gao, 2008; Geissler et al., 2021; Hu, 2016; Karlin, 2018b; Zhang et al., 2019) and pitch accent (Karlin, 2018a; Zsiga, Elizabeth & Zec, 2013) languages. Moreover, f0 has been shown in many cases to interact in coordination in the same way as other gestures. Identifying the dimensions of contrast and of phonetic control, i.e., gestures, is a prerequisite to evaluating inter-gestural coordination.

In sum, I think there is substantial potential for the hypothesis presented in Section 2.4 to be generalized across a wide range of segments, and even to serve as a diagnostic for complex segmenthood in cases for which revealing phonological evidence may otherwise be lacking. As a first pass, I chose a test case that is uncontroversial in its phonological status and for which I have good a priori knowledge of the dimensions of phonetic control.

5.3. Application of a gestural overlap account to final devoicing

In the case of final devoicing, voicing contrast is preserved in the word-initial and word-medial positions. However, in the word-final position, both underlying voiced and underlying voiceless obstruents surface as voiceless. In German, for example, the voicing contrast of alveolar stops is neutralized in word-final positions, while the contrast is preserved in word-medial positions as shown in (16).

(16) Examples of final devoicing in German

Rat [ʁa:**t**] *'council'*	Räte [ʁæ:**t**ə] *'councils'*
Rad [ʁa:**t**] *'wheel'*	Räder [ʁæ:**d**ɐ] *'wheels'*

Previous studies have provided much evidence that such phonological neutralization is phonetically incomplete in many languages (See Section 1.2.1 for more discussion). In German, for example, Port and O'Dell (1985) found that the underlying voiced obstruents have shorter final stop closure durations, a shorter release burst, a longer preceding vowel, and/or more extensive voicing into closure than the underlying voiceless obstruents. The amount of difference, albeit statistically significant, was very small in magnitude, on the order of 10–20 milliseconds at most.

Still, incomplete neutralization presents serious challenges to most phonological models such as serial models of phonology as well as non-serial models of phonology. For example, under the standard view of phonology (serial models of phonology), phonological rules have to be applied before low-level phonetic implementation rules. However, regarding incomplete neutralization, if the phonological rules are applied first, it is impossible to apply the phonetic implementation rules, since the contrast has already been neutralized. Even in a constraint-based approach such as optimality theory, where there are no serial phonological rules, it is difficult to incorporate incomplete neutralization into the model. That is, OT also predicts that there are no phonetic differences between an underlyingly voiceless and voiced consonants, unless one posits different representations for underlyingly voiceless and voiced consonants, [t] and [d̥], respectively (e.g., turbidity theory; Van Oostendorp, 2008). However, positing different representations in OT may partially solve the issue with final devoicing, but this may not work with other cases of incomplete neutralization, such as palatalization in Russian.

This dissertation provides a promising solution to a long-standing problem by showing that at least some cases of incomplete neutralization can be modeled as gestural overlap in the AP framework. The gestural overlap account can also be applied to the most representative case of incomplete neutralization, final devoicing. In particular, the incomplete neutralization between underlyingly voiceless and voiced consonants can be modeled as gestural overlap and sound change.

In the case of final devoicing, I posit that an underlyingly voiceless consonant has an underlying glottal opening gesture, while devoicing of an underlyingly voiced consonant comes from an adjacent glottal opening gesture to mark a prosodic boundary. For example, Figure 59 and Figure 60 show gestural scores for /ʁa**t**/ and /ʁa**d**/, respectively. For both cases, there is an adjacent glottal opening gesture to mark a prosodic boundary overlapped with gestures /t/ or /d/. However, since this underlyingly voiceless consonant has an underlying glottal opening gesture, the gestural overlap between gestures for /t/ and the pause gesture does not affect the articulation of /t/.

In contrast, the gestural overlap between gestures for /d/ and the pause gesture leads to devoicing of an underlyingly voiced consonant. However, this latter overlap does not yield the identical output to the underlyingly voiceless counterpart. Instead, the gestural overlap between gestures for /d/ and the pause gesture is expected to yield a longer preceding vowel and more extensive voicing into closure than the underlying voiceless obstruents since the pause gesture has less overlap with the preceding vowel than the overlap between the underlying glottal opening gesture for /t/ and the preceding vowel. Moreover, difference in gestural specifications between /d/ and /t/ will lead to shorter final stop closure durations and a shorter release burst for /d/ in comparison with /t/ (i.e., difference in activation and deactivation phases). Therefore, this difference may lead to incomplete neutralization of underlyingly voiced and voiceless consonants in word-final positions. On the other hand, when these obstruents occur before a vowel, there is no pause gesture associated with them, and consequently the contrast is preserved in this case. This is well illustrated in Figure 60 which shows partial gestural scores for /ʁa**ta**/ (left) and /ʁa**da**/ (right).

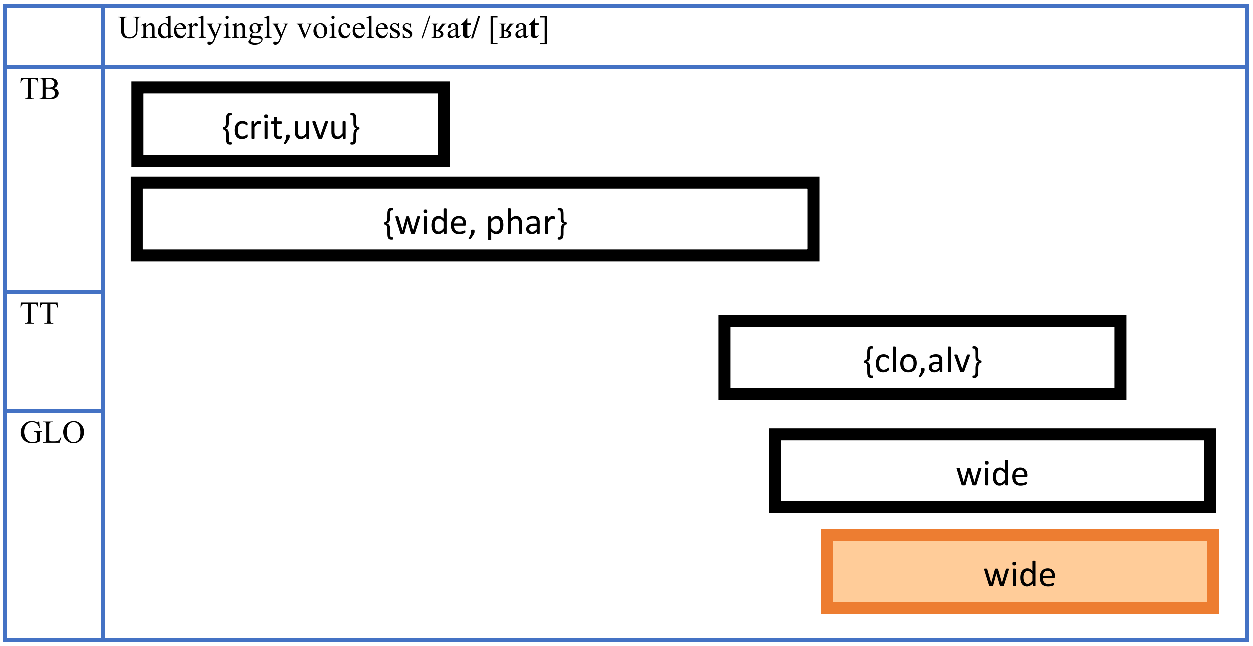


Figure 59: The gestural scores for /ʁat/. The gestures for /ʁat/ are shown in unshaded black boxes, and the glottal opening gesture to mark a prosodic boundary is shown in shaded orange boxes

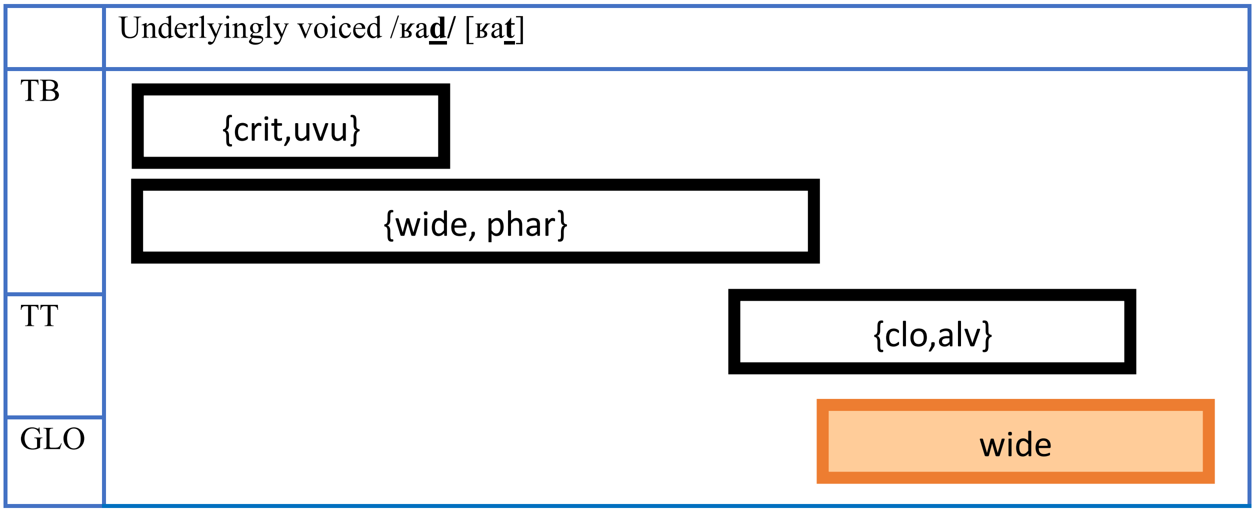


Figure 60: The gestural scores for /ʁad/. The gestures for /ʁad/ are shown in unshaded black boxes, and the glottal opening gesture to mark a prosodic boundary is shown in shaded orange boxes

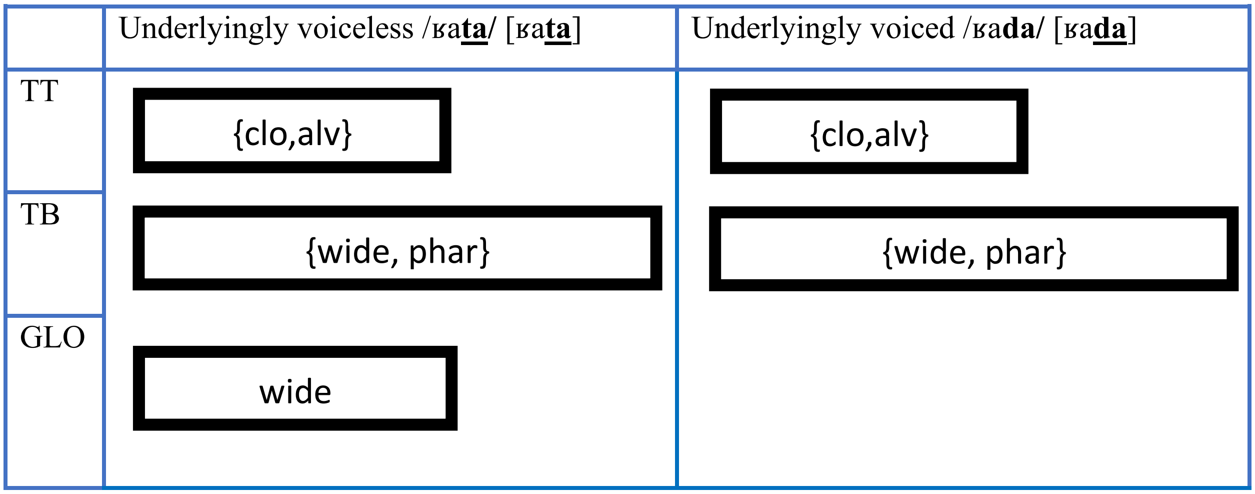


Figure 61: The partial gestural scores for /ʁata/ (left) and /ʁada/ (right)

However, recent studies on pause posture suggest that the pause posture occurs only at strong prosodic boundaries (Katsika, 2012; Katsika et al., 2014). This raises a concern on whether it is reasonable to posit a glottal opening gesture to mark a prosodic boundary at every word-final position to explain word-final devoicing. Moreover, some languages show final devoicing even in syllable-final word-medial positions. As shown in (17), for example, the voicing contrast of labio-velar fricatives is preserved syllable initially, while the contrast is neutralized in syllable-final word-medial positions in German. These examples may serve as counterexamples to the proposal since it does not make sense to posit a glottal opening gesture in syllable-final word-medial positions.

(17) Examples of syllable-final devoicing in German (adopted from Beckman et al., 2009, p. 236)

sur**f**en [səː.**f**n̩] *'surf INF'*	sur**f**te [səː**f**.te] *'surf 1/3SG PAST'*
kur**v**en [kʊr.**v**n̩] *'curve INF'*	kur**v**te [kʊr**f**.te] *'curve 1/3SG PAST'*

This is where a sound change account is necessary. Blevins (2004; 2006) argued that final devoicing results from sound change, reflecting "an emergent property of sound systems." She noted that final devoicing has occurred across unrelated languages such as Indo-European languages (e.g., German, Catalan, Russian), Turkic (e.g., Turkish), Semitic (e.g., Chadic Arabic), and Cushitic (e.g., Afar). Moreover, final devoicing in Afar, Chadic Arabic, Russian, Ingush, Turkish, Old Chinese, and Malay did not inherit final devoicing from the proto-languages since the proto-languages feature the voicing contrast in word-final position.

Blevins (2004; 2006) saw that final devoicing is a natural phonological development based on physiological and perceptual factors which favor voiceless obstruents and disfavor voiced ones. In addition, she presented a few other phonetic sources of final devoicing such as laryngeal spreading and closing gestures at phrase boundaries. Then, she predicted that in the early stages final devoicing will occur only before a pause or phase-finally and the direction of final devoicing will be utterance > phrase > word > syllable. That is, ample exposure to phrase-final devoicing will lead to overgeneralization of the pattern by learners from phrase-final to word-final then to syllable-final positions. She argued that some languages such as Nigerian Arabic and Gulf Arabic are in the early stages of final devoicing, since final devoicing tends to occur before pauses in Nigerian Arabic and in utterance-final position in Gulf Arabic.

However, her discussion was limited to interpreting final devoicing as sound change. Here, I posit that incomplete neutralization between underlyingly voiceless and voiced consonants can also be modeled as gestural overlap and sound change. The scenario for the development of incomplete neutralization is as follows. In Stage 1, this incomplete neutralization occurs only in the phrase-final position due to gestural overlap between glottal opening gestures to mark a prosodic boundary and gestures involving an obstruent. (See Figure 59 and Figure 60 for more detail about interpreting incomplete neutralization as gestural overlap). Possibly, Nigerian Arabic and Gulf Arabic are in the early stages of incomplete neutralization. Then, learners overgeneralize the pattern from phrase-final to word-final (Stage 2). That is, people reinterpret the glottal opening gesture to be a part of underlyingly voiced consonants in the word-final position. Crucially, however, listeners may retain relative timing between the glottal opening gesture and the gestures for underlying voiced consonants when they overgeneralize this pattern to word final positions. Consequently, this leads to incomplete neutralization of underlyingly voiceless and voiced consonants in the word-final position. Dhaasanac, Chardic Arabic, and Maltese, which show final devoicing only word-finally, are possibly in these intermediate stages of incomplete neutralization as well. In Stage 3, over time, learners overgeneralize the pattern even further to syllable-final positions. This results in the reinterpretation of the glottal opening gesture to be a part of the underlyingly voiced consonants in the syllable-final position, leading to incomplete neutralization in the syllable-final word-medial position.

In contrast, when a vowel or a sonorant occurs at the phrase-final position, it is also expected to have an overlapping glottal opening gesture to mark the prosodic boundary, leading to devoicing of vowels and sonorants at phrase boundaries. In fact, this phenomenon has been observed in a final devoicing language (e.g., final-vowel devoicing in Bulgarian: Andreeva & Koreman, 2003), as well as in a non-final devoicing language (e.g., final-sonorant devoicing in Icelandic: Dehé, 2014; final-vowel devoicing in French: Smith, 2003). However, devoicing of vowels and sonorants at word-final positions or at syllable-final positions is not nearly as prevalent as devoicing of obstruents at word-final positions, but the reasons for this asymmetry is unclear.

One possibility is that there may be less overlap between the glottal opening gesture and the vowel/sonorant gestures in a given language in comparison with final obstruents, resulting in less extensive devoicing of vowels and sonorants at prosodic boundaries. If that is the case, listeners may be less likely to reinterpret the glottal opening gesture as an underlying gesture for vowels/sonorants, and final vowel/sonorant devoicing remains as a phonetic process driven by prosody.

However, if a given language exhibits an amount of phonetic devoicing of vowels/sonorants that is comparable to the amount of devoicing of final obstruents, listeners should be equally likely to reinterpret phonetically devoiced final vowels/sonorants as they are to reinterpret phonetically devoiced obstruents. However, phonetically devoiced final vowels/sonorants can also be a source of deletion process as well, as cues to perceive vowels and sonorants become weaker when they are devoiced. That is, listeners may reinterpret final vowels/sonorants devoicing as deletion of vowels/sonorants. Therefore, there are two diachronic paths that phonetically devoiced final vowels/sonorants may take: being reanalysed as phonologically devoiced vowels/sonorants, or as phonologically deleted vowels/sonorants. These different scenarios may characterize the divergent development of Woleaian and Trukese from

Micronesian: the former shows devoicing of vowels at word final positions, while the latter exhibits deletion of vowels (Blevins, 2018).

Returning to the discussion on final-obstruent devoicing, recent studies reported that an exposure to a non-devoicing second language increases the incompleteness of final devoicing in the first language (e.g., L1 Bulgarian & L2 English: Bishop et al., 2019; L1 Russian & L2 English: Dmitrieva et al., 2010). The gestural overlap account can also provide an insight into these patterns. Considering that English does not exhibit final devoicing, it is assumed that there is less overlap between gestures for obstruents and the pause gesture. As Bulgarian or Russian learners of English are exposed to this temporal coordination, they may have learned this coordination and applied it to their native languages to a certain degree, resulting in increased incompleteness of the neutralization in their L1. However, a future study would be necessary to investigate how an exposure to L2 affects temporal coordination of L1. In sum, my dissertation has provided a promising solution to incomplete neutralization, which presents serious challenges to most phonological models. I showed that final devoicing, the representative case of incomplete neutralization, can also be modeled as gestural overlap and sound change. This gestural overlap account for final devoicing can be possibly tested using a combination of EMA and nasopharyngeal endoscopy. This gestural overlap account is incompatible with most phonological models except for AP, since AP is the only model that bridges abstract phonological representations and continuous physical movement.

5.4. Conclusion

Incomplete neutralization has presented serious challenges to most all phonological models. This dissertation explored the incomplete neutralization of Russian palatalization within the

Articulatory Phonology framework. One hypothesis I pursued is that some cases of incomplete neutralization are the result of gestural blending of two competing gestural forces. To quantify palatalization in Russian, I first examined temporal coordination in complex segments versus segment sequences. Evidence from articulatory kinematic data collected with Electromagnetic Articulography on Russian palatalized consonants and English consonant-glide sequences provided support for the hypothesis that complex segments differ from segment sequences in how the constituent gestures are coordinated. Implementing this temporal diagnostic, I explored the phonetic realization of Russian palatalization.

Through simulations from computational modeling and comparisons with physiological data from EMA, I tested this hypothesis on patterns of palatalization in Russian. A key finding of my dissertation was that gestural blending of two secondary articulation gestures, palatalization and velarization/uvularization, does in fact result in incomplete neutralization of underlying and coarticulatory palatalization in Russian. The current dissertation offers an explanation for incomplete neutralization patterns by showing that at least some cases of incomplete neutralization can be modeled as gestural overlap in the AP framework. There is substantial potential for the gestural overlap account to generalize across a wide range of incomplete neutralization.

Appendix

As discussed in Section 4.5.2, the eccentric timing between the velar and palatal gestures might be attributable to a competitive coupling between the two gestures. That is, the velar and palatal gestures are both timed in-phase to the labial gesture, but anti-phase to each other. As shown in Figure 62, the regression lines for both /bʲust/ (**UNDERLYING** palatalization) /bᵚjut/_in-phase (**COARTICULATORY** palatalization) are nearly flat (although **COARTICULATORY** palatalization shows only a slight upward trend). That is, simulations for both **UNDERLYING** and **COARTICULATORY** palatalization from the competitive coupling show temporal coordination of a complex segment, similar to the results from the EMA recordings. However, as shown in Figure 63 (left), simulations from the competitive coupling showed no difference between **UNDERLYING** and **COARTICULATORY** palatalization in the spatial position of the TB. Crucially, the lag between the gesture onsets was longer for the **UNDERLYING** palatalization than for the **COARTICULATORY** palatalization (Figure 63 right), which is the opposite result from the EMA recording. The same results were found in the simulations from the competitive coupling graph with relative phase 20° of a gesture's oscillator (See Figure 64 and Figure 65).

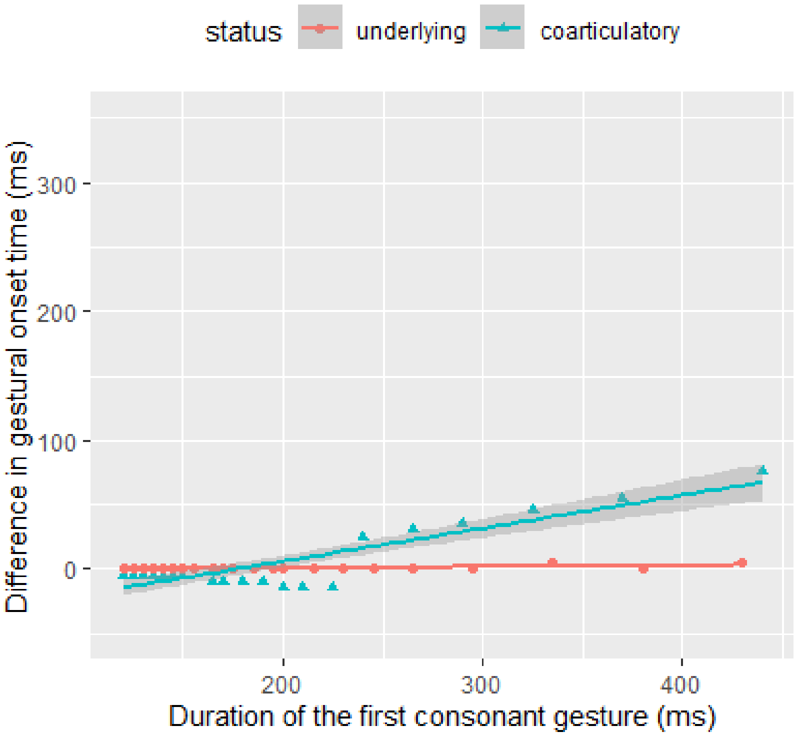


Figure 62: Competitive coupling (180°) – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

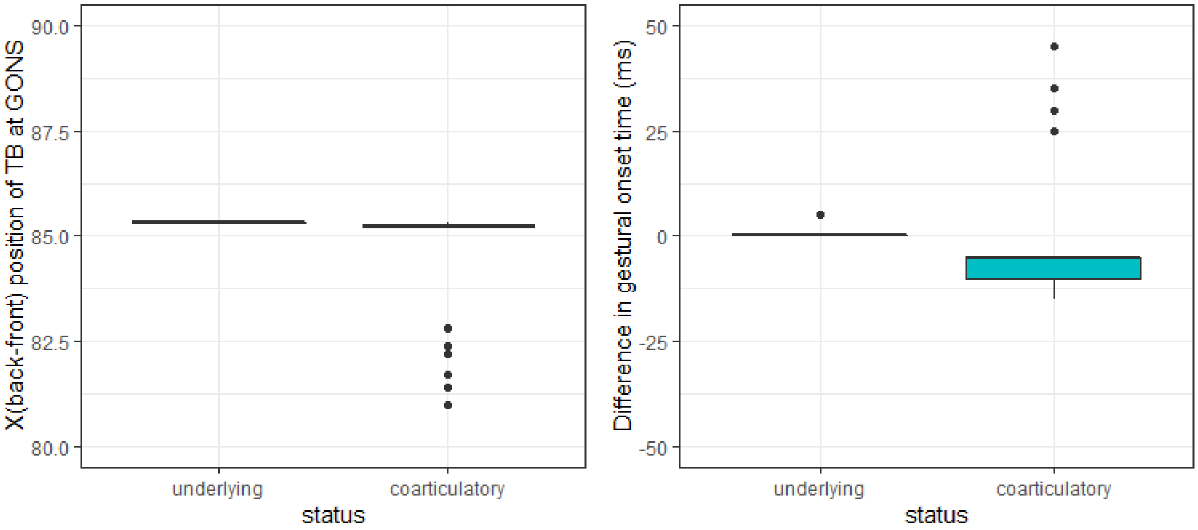


Figure 63: Competitive coupling (180°) – TB position (mm) at palatal gesture onset (left), and onset-to-onset lag (right)

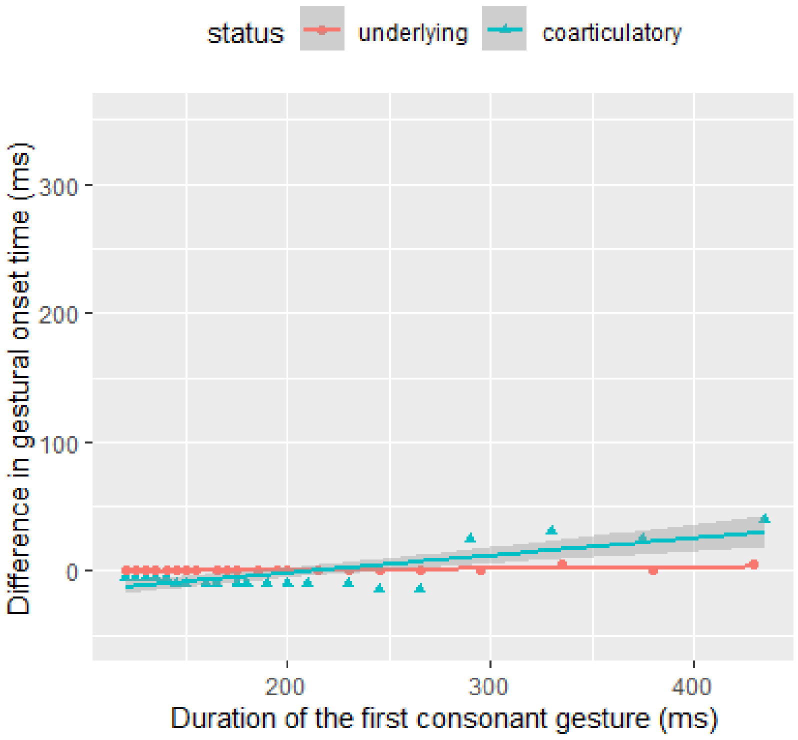


Figure 64: Competitive coupling (20°) – A scatter plot of the effect of G1 duration (x-axis) on onset-to-onset lag (y-axis)

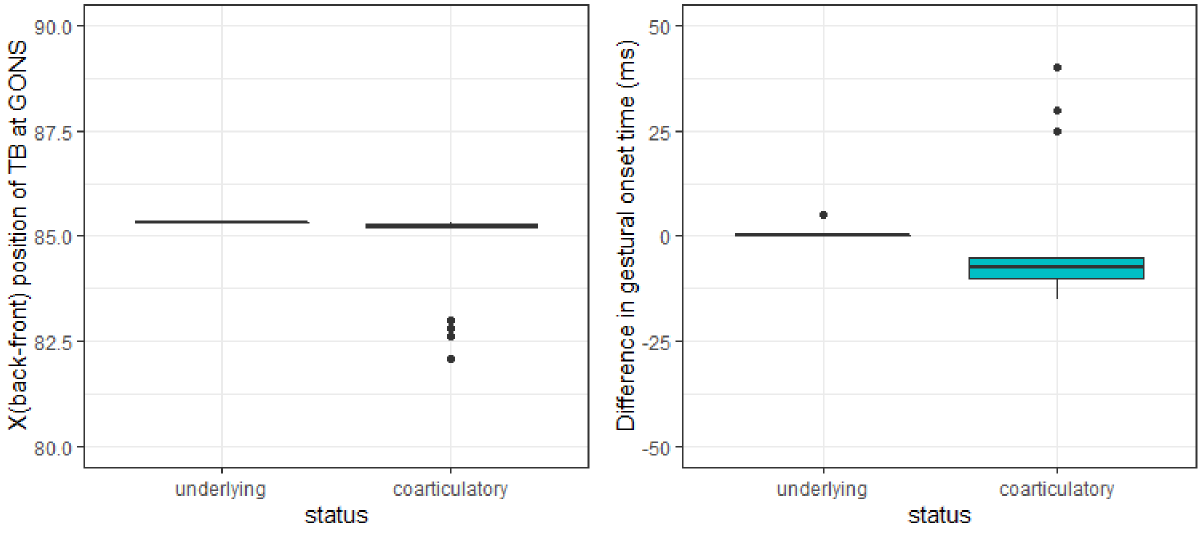


Figure 65: Competitive coupling (20°) – TB position (mm) at palatal gesture onset (left), and onset-to-onset lag (right)

References

Abdul-Karim, K. W. (1980). *Aspects of the Phonology of Lebanese Arabic* http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqil:res_ver=0.2&res_id=xri:ilcs-us&rft_id=xri:ilcs:rec:abell:R02383210

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*(1), 41-61.

Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. *Language,* , 326-344.

Andreeva, B., & Koreman, J. (2003). The Status of Vowel Devoicing in Bulgarian: Phonetic or Phonological? Paper presented at the *Formal Description of Slavic Languages: The Fifth Conference (FDSL5), Leipzig,* 81-91.

Avanesov, R. I. (1972). Russian Literary Pronunciation [Russkoe literaturnoe proiznoshenie]. *Education, M,*

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37*(1), 94-117.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

Barlow, J. A. (2001). Individual differences in the production of initial consonant sequences in Pig Latin. *Lingua, 111*(9), 667-696.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823,* Beckman, J., Jessen, M., & Ringen, C. (2009). German fricatives: Coda devoicing or positional faithfulness? *Phonology, 26*(2), 231-268.

Bernštejn, N. A. (1967). *The co-ordination and regulation of movements*. Pergamon Press.

Berry, J. J. (2011). Accuracy of the NDI wave speech research system.

Bishop, J. B. (2007). Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration. Paper presented at the *Proceedings of the International Congress of Phonetic Sciences, , 16* 1765-1768.

Bishop, J., Oh, S., & Zhou, C. (2019). Final devoicing in Bulgarian: Incomplete neutralization and L2 experience. *In Proceedings of the 19th International Congress of Phonetic Sciences,* , 652-656.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology.

Blevins, J. (2018). Evolutionary phonology and the life cycle of voiceless sonorants. *Typological Hierarchies in Synchrony and Diachrony,* , 29-60.

Bombien, L., Mooshammer, C., Hoole, P., & Kühnert, B. (2010). Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics, 38*(3), 388-403.

Borroff, M. L. (2007). *A landmark underspecification account of the patterning of glottal stop*. State University of New York at Stony Brook.

Bouferroum, O., & Boudraa, M. (2015). CV coarticulation, locus and locus equation perspective on the invariance issue involving Algerian Arabic consonants. *Journal of Phonetics, 50*, 120-135.

Braver, A. (2013). *Degrees of Incompleteness in Neutralization: Paradigm Uniformity in a Phonetics with Weighted Constraints*

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology, 36*(1), 1-36.

Braver, A., & Kawahara, S. (2016). Incomplete Neutralization in Japanese Monomoraic Lengthening. *Proceedings of the Annual Meetings on Phonology, 2*10.3765/amp.v2i0.3757

Brockhaus, W. (1995). *Final devoicing in the phonology of German* http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=006861236&sequence=000002&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology, 3*, 219-252.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*(2-4), 140-155.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(2), 201-251. 10.1017/S0952675700001019

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18*(3), 299-320.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica, 49*(3-4), 155-180. 10.1159/000261913

Browman, C. P., & Goldstein, L. (1995a). Dynamics and articulatory phonology. *Mind as Motion, 175*, 193.

Browman, C. P., & Goldstein, L. (1995b). Gestural syllable position effects in American English. *Producing Speech: Contemporary Issues,* , 19-33.

Browman, C. P., & 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.

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology, 5*(4), 403-454.

Burnham, K. P., & Anderson, D. R. (1998). Practical use of the information-theoretic approach. *Model selection and inference* (pp. 75-117). Springer.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), 23-35.

Bybee, J. L. (1994). A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics, 5*(4), 285-306. 10.1515/cogl.1994.5.4.285

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425-455. 10.1080/01690969508407111

Bybee, J. L. (2001). Phonology and language use Cambridge: Cambridge UP.

Byrd, D. (1995). C-Centers Revisited. *Phonetica, 52*(4), 285-306. 10.1159/000262183

Charles-Luce, J., & Dinnsen, D. A. (1987). A reanalysis of Catalan devoicing. *Journal of Phonetics, 15*(2), 187-190.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120-141.

Cohen-Priva, U. (2017). Informativity and the actuation of lenition. *Language, 93*(3), 569-597.

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances, 5*(9), eaaw2594.

Davidson, L., & Roon, K. (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association, 38*(2), 137-165.

Davis, S., & Hammond, M. (1995). On the status of onglides in American English. *Phonology, 12*(2), 159-182.

Dehé, N. (2014). Final devoicing of/l/in Reykjavík Icelandic. Paper presented at the *SpeechProsody7,* 757-761.

Diehm, E. E. (1998). *Gestures and linguistic function in learning Russian: Production and perception studies of Russian palatalized consonants*. The Ohio State University.

Dinnsen, D. A., & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics, 12*(1), 49-60.

Dmitrieva, O. (2005). Final devoicing in Russian: Acoustic evidence of incomplete neutralization. *The Journal of the Acoustical Society of America, 117*(4), 2570.

Dmitrieva, O., Jongman, A., & Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics, 38*(3), 483-492.

Durvasula, K., Ruthan, M., Heidenreich, S., & Lin, Y. -. (2021). Probing Syllable Structure Through Acoustic Measurements: Case-studies on American English and Jazani Arabic. *Phonology,* , 1-30.

Ernestus, M., & Baayen, H. (2007). Intraparadigmatic effects on the perception of voice. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4, 286*, 153.

Ernestus, M., & Baayen, H. (2009). The functionality of incomplete neutralization in Dutch: The case of the past-tense formation. *Laboratory phonology 8* (pp. 27-50). De Gruyter Mouton.

Feinsilver, L. M. (1961). On Yiddish Shm. *American Speech, 36*(4), 302-303.

Feinstein, M. H. (1979). Prenasalization and Syllable Structure. *Linguistic Inquiry, 10*(2), 245-278. https://www.jstor.org/stable/4178108

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America, 101*(6), 3728-3740.

Fougeron, C., & Steriade, D. (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? Paper presented at the *Fifth European Conference on Speech Communication and Technology,*

Fourakis, M., & Iverson, G. K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica, 41*(3), 140-149.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory, 20*(2), 269-337.

Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University, CT,*

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis, 54*(4), 1167-1178.

Geissler, C., Shaw, J. A., Fang, H., & Tiede, M. (2021). Eccentric CV timing across speakers of diaspora Tibetan with and without lexical tone contrasts. Paper presented at the *Proceedings of the 12th International Seminar on Speech Production, Yale University, 4pgs,*

Gerfen, C. (2002). Andalusian codas. *Probus, 14*(2), 247-277. 10.1515/prbs.2002.010

Gerfen, C., & Hall, K. C. (2001). Coda aspiration and incomplete neutralization in Eastern Andalusian Spanish. University of North Carolina at Chapel Hill.

Goldstein, L. (2011). Back to the past tense in English. *Representing Language: Essays in Honor of Judith Aissen,* , 69-88.

Goldstein, L., Byrd, D., & 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,* , 215-249.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. Paper presented at the *Proceedings of the XVIth International Congress of Phonetic Sciences,* 241-244.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). *Coupled oscillator planning model of speech timing and syllable structure*. HAL CCSD.

Gouskova, M., & Hall, N. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. *Phonological Argumentation: Essays on Evidence and Motivation,* , 203-225.

Gouskova, M., & Stanton, J. (2021). Learning complex segments. *Language, 97*(1), 151-193.

Guitard-Ivent, F., Turco, G., & Fougeron, C. (2021). Domain-initial effects on C-to-V and V-to-V coarticulation in French: A corpus-based study. *Journal of Phonetics, 87*, 101057.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*(5), 347-356.

Herbert, R. K. (1986). *Language universals, markedness theory, and natural phonetic processes*. Walter de Gruyter.

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of/t/and/d/flaps in American English. *Journal of Phonetics, 38*(4), 504-516.

Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics, 64*, 127-144.

Hoole, P., Mooshammer, C., & Tillmann, H. G. (1994). Kinematic analysis of vowel production in German. Paper presented at the *Third International Conference on Spoken Language Processing,*

Hu, F. (2016). Tones are not abstract autosegmentals. Paper presented at the *Speech Prosody, , 2016* 302-306.

Idsardi, W. J., & Raimy, E. (2005). Remarks on language play. *Ms., U.of Delaware,*

Iskarous, K., & Kavitskaya, D. (2010). The interaction between contrast, prosody, and coarticulation in structuring phonetic variability. *Journal of Phonetics, 38*(4), 625-639.

Iverson, G. K., & Salmons, J. C. (2011). Final devoicing and final laryngeal neutralization. *The Blackwell Companion to Phonology,* , 1-22.

Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics, 17*(4), 317-325.

Karlin, R. (2018a). Effects of syllable onset on the timing of pitch accent in Belgrade Serbian. *Proceedings of the Linguistic Society of America, 3*(1), 36-1–15.

Karlin, R. (2018b). *Towards an articulatory model of tone: a cross-linguistic investigation*. Cornell University.

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

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics, 44*, 62-82.

Keating, P. (1993). Phonetic representation of palatalization versus fronting. *UCLA Working Papers in Phonetics, 85*, 6-21.

Kharlamov, V. (2012). *Incomplete neutralization and task effects in experimentally-elicited speech: Evidence from the production and perception of word-final devoicing in Russian*. University of Ottawa (Canada).

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics, 43*, 47-56.

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics, 38*(2), 185-196.

Kochetov, A. (2004). Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries. *Language and Speech, 47*(4), 351-382.

Kochetov, A. (2006). Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica, 63*(2-3), 113-148.

Kochetov, A. (2009). Syllable position effects and gestural organization: Articulatory evidence from Russia. *Laboratory Phonology 8* (pp. 565-588). De Gruyter Mouton.

Kochetov, A. (2011). Palatalization. *The Blackwell Companion to Phonology,* , 1-25.

Kochetov, A. (2013). *Production, Perception, and Phonotactic Patterns: A Case of Contrastive Palatalization*. Routledge.

Kröger, B. J., Schröder, G., & 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.

Kugler, P. N., Kelso, J. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. *The Development of Movement Control and Coordination, 5*, 1-78.

Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. Paper presented at the *7th International Seminar on Speech Production (ISSP),* pp. 327-334.

Ladefoged, P., & Maddieson, I. (1998). The sounds of the world's languages. *Language, 74*(2), 374-376.

Litvin, N. (2014). No title. *An Ultrasound Investigation of Secondary Velarization in Russian,*

Lombardi, L. (1990). The nonlinear organization of the affricate. *Natural Language & Linguistic Theory, 8*(3), 375-425.

Maddieson, I. (1989). Prenasalized stops and speech timing. *Journal of the International Phonetic Association, 19*(2), 57-66.

Maddieson, I., & Ladefoged, P. (1993). Phonetics of partially nasal consonants. *Nasals, nasalization, and the velum* (pp. 251-301). Elsevier.

Marin, S. (2012). Romanian 'blended'vowels: A production model of incomplete neutralization. *Journal of Portuguese Linguistics, 11*(2)

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics, 41*(3-4), 211-227.

Mascaró, J. (1987). Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics, 15*(2), 183-186.

McCarthy, J., & Prince, A. (1986). Prosodic Morphology (RuCCS Technical Report Series TR-3). *Rutgers University,*

McGowan, R. S., & Saltzman, E. L. (1995). Incorporating aerodynamic and laryngeal components into task dynamics. *Journal of Phonetics, 23*(1-2), 255-269.

Nam, H., Goldstein, L., Browman, C., Rubin, P., Proctor, M., & Saltzman, E. (2006). *TADA (task dynamics application) manual*. New Haven, CT: Haskins Laboratories.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to phonological complexity* (pp. 297-328). De Gruyter Mouton.

Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America, 115*(5), 2430.

Nam, H., Mitra, V., Tiede, M., Hasegawa-Johnson, M., Espy-Wilson, C., Saltzman, E., & Goldstein, L. (2012). A procedure for estimating gestural scores from speech acoustics. *The Journal of the Acoustical Society of America, 132*(6), 3980-3989.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. Paper presented at the *Proceedings of the 15th International Congress of Phonetic Sciences, , 1*(2253-2256)

Nevins, A., & Vaux, B. (2003). Metalinguistic, shmetalinguistic: The phonology of shmreduplication. Paper presented at the *Proceedings from the Annual Meeting of the Chicago Linguistic Society, , 39*(1) 702-721.

O'Dell, M., & Port, R. (1983). Discrimination of word-final voicing in German. *The Journal of the Acoustical Society of America, 73*(S1), S31.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. *The Role of Speech Perception in Phonology, ,* 187-218.

Padgett, J. (2003). The emergence of contrastive palatalization in Russian. *Optimality Theory and language change* (pp. 307-335). Springer.

Parrell, B., & Lammert, A. C. (2019). Bridging Dynamical Systems and Optimal Trajectory Approaches to Speech Motor Control With Dynamic Movement Primitives. *Frontiers in Psychology, 10*, 2251.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica, 58*(4), 254-275.

Piroth, H. G., & Janker, P. M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics, 32*(1), 81-109.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics, 13*(4), 455-471.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*(4), 257-282.

Pouplier, M. (2020). Articulatory Phonology. *Oxford Research Encyclopedia of Linguistics* ()

Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness. *Journal of Phonetics, 64*, 108-126.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech, 28*(2), 97-114.

Riehl, A. (2008). No title. *The Phonology and Phonetics of Nasal-Obstruent Sequences,*

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics, 43*, 11-25.

Roon, K. D., & Whalen, D. H. (2019). Velarization of Russian labial consonants. Paper presented at the *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia,* Ruthan, M. Q., Durvasula, K., & Lin, Y. (2019). Temporal Coordination and Sonority of Jazani Arabic Word-Initial Clusters. Paper presented at the *Proceedings of the Annual Meetings on Phonology, , 7*

Sagey, E. C. (1986). No title. *The Representation of Features and Relations in Non-Linear Phonology,*

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382.

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. Paper presented at the *Proceedings of the 4th International Conference on Speech Prosody (Speech Prosody 2008), Campinas, Brazil,* 175-184.

Sharf, D. J. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech, 5*(1), 26-30.

Shaw, J. A., & Chen, W. (2019). Spatially conditioned speech timing: Evidence and implications. *Frontiers in Psychology, 10*, 2726.

Shaw, J. A., Durvasula, K., & Kochetov, A. (2019). The temporal basis of complex segments. Paper presented at the *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia,* 676-680.

Shaw, J. A., & Gafos, A. I. (2015). Stochastic time models of syllable structure. *PloS One, 10*(5), e0124714.

Shaw, J. A., Gafos, A. I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology,* , 455-490.

Shaw, J. A., & Kawahara, S. (2019). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech, 62*(1), 80-114.

Silverman, D. D. (1997). *Phasing and recoverability*. Taylor & Francis.

Skalozub, L. G. (1963). *Palatogrammy i rentgenogrammy soglasnych fonem russkogo literaturnogo jazyka [Palatograms and X-ray images of Russian consonants].* Izdat. Kievskogo Univ.

Slowiaczek, L. M., & Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics, 13*(3), 325-341.

Smith, C. L. (2003). Vowel devoicing in contemporary French. *Journal of French Language Studies, 13*(2), 177-194.

Sorensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology, 28*(4), 188-215.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English/l/and its implications for phonetic implementation. *Journal of Phonetics, 21*(3), 291-311.

Strycharczuk, P., Derrick, D., & Shaw, J. (2020). Locating de-lateralization in the pathway of sound changes affecting coda/l.

Suh, Y., & Hwang, J. (2016). The Korean Prevocalic Palatal Glide: A Comparison with the Russian Glide and Palatalization. *Phonetica, 73*(2), 85-100.

Tiede, M. K. (2005). *MVIEW: software for visualization and analysis of concurrently recorded movement data.* New Haven, CT: Haskins Laboratories.

Tiede, M. K., Perkell, J., Zandipour, M., & Matthies, M. (2001). Gestural timing effects in the ‘‘perfect memory’’ sequence observed under three rates by electromagnetometry. *The Journal of the Acoustical Society of America, 110*(5), 2657.

Tilsen, S., Zec, D., Bjorndahl, C., Butler, B., L'Esperance, M., Fisher, A., Heimisdottir, L., Renwick, M., & Sanker, C. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology, 2012*, 51-81.

Timberlake, A. (2004). *A reference grammar of Russian*. Cambridge University Press.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*(8), 938.

Umeda, N. (1977). Consonant duration in American English. *The Journal of the Acoustical Society of America, 61*(3), 846-858.

Van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua, 118*(9), 1362-1374.

Vaux, B. (2011). Language games. *The Handbook of Phonological Theory,* , 722-750.

Vinogradov, G. S., Ivanova, E. A., & Gindin, S. I. (2005). Detskie tajnye jazyki: Kratkij ocherk [Children's secret languages: An overview]. , 16, . *Russkij Jazyk, 16*, 14-25.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004a). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*(2), 251-276.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004b). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*(2), 251-276.

Westbury, J. R., Turner, G., & Dembowski, J. (1994). X-ray microbeam speech production database user's handbook. *University of Wisconsin,* Ying, J., Shaw, J. A., Carignan, C., Proctor, M., Derrick, D., & Best, C. T. (2021). Evidence for active control of tongue lateralization in Australian English/l. *Journal of Phonetics, 86*, 101039.

Zhang, M., Geissler, C., & Shaw, J. A. (2019). Gestural representations of tone in Mandarin: evidence from timing alternations. Paper presented at the *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019,* 1803-1807.

Zsiga, E. C. (1995). An Acoustic and Electropalatographic Study of Lexical and Postlexical Palatalization in American English. *Phonology and Phonetic Evidence* (pp. 282). Cambridge UP.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of Phonetics, 28*(1), 69-102. 10.1006/jpho.2000.0109

Zsiga, E., & Zec, D. (2013). Contextual evidence for the representation of pitch accents in Standard Serbian. *Language and Speech, 56*(1), 69-104.