Title
Vowel Intrusion in Turkish Onset Clusters

Author
Bellik, Jennifer Ann

Publication Date
2019

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VOWEL INTRUSION IN TURKISH ONSET CLUSTERS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Jennifer A. Bellik

June 2019

The dissertation of Jennifer A. Bellik is approved:

____
Professor Jaye Padgett, chair

____
Associate Professor Grant McGuire, co-chair

____
Assistant Professor Ryan Bennett

____
Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

VOWEL INTRUSION IN TURKISH ONSET CLUSTERS

by

Jennifer A. Bellik

Vowel insertion commonly occurs cross-linguistically to break up consonant clusters, particularly in loanwords. The most familiar form of this insertion is the categorical, phonological process of vowel epenthesis, which repairs syllable structure violations. However, gradient, phonetic vowel insertion also occurs in many languages. This phenomenon, known as vowel intrusion, occurs when speakers employ a gestural timing that results in an open transition between the consonants in the cluster. The resulting intrusive vowel is not a phonological segment, and therefore lacks gestural and durational targets.

This dissertation presents a case study of vowel intrusion in Turkish onset clusters, which occur in European loanwords. The non-lexical vowels in these clusters have previously been described as epenthetic vowels whose quality is categorically determined by the following lexical vowel, in a process of regressive vowel harmony. I present new evidence from experimental, corpus, and elicitation data, that the non-lexical vowel in

Turkish onset clusters lacks a gestural or durational target, and does not form a syllable nucleus—distinguishing characteristics of intrusive vowels. Gestural and acoustic data comes from an ultrasound production study, which shows that the non-lexical vowel <v> is gradiently present; does not achieve the same gestural and durational targets as lexical vowels; and is more affected by coarticulation.

A corpus study extends the investigation to other types of consonant clusters, and finds that the effect of the consonantal environment on the non-lexical vowel can be seen even in broad transcriptions. I also probe the syllabic and metrical status of <v> using two studies of text-setting of /CC/ words. These studies reveal that <v> is more variable and less likely to receive a beat than lexical vowels.

This evidence supports an analysis of Turkish /CC/ words as beginning with onset clusters. I present a coupled oscillator representation of vowel intrusion and onset clusters, using Optimality Theory and allowing the phonology to refers to both segments and gestures. I also argue that Turkish <v> has historically been reanalyzed as an underlying vowel, but that language-internal variation, as well as increasingly prevalent knowledge of source languages, today maintain its status as an intrusive vowel.

Acknowledgements

Graduate school is a lot of work, but if you're lucky, it can be a lot of fun too – it was for me. For that, I have many people to thank. First, my sincerest gratitude to the UC Santa Cruz Linguistics Department, which has been a generous, intellectually-stimulating, growth-promoting community to me during my graduate career. Particular thanks are due to my advisors and co-chairs, Jaye Padgett and Grant McGuire, whose invitation to work on the ultrasound investigation of Irish palatalization began my interest in ultrasound that became this dissertation, and who shepherded me and the project through to this point. Thank you both for your wisdom and encouragement throughout the process. Jaye, I especially thank you for making me write something down every step of the way; for your attentive reading of so many drafts; for telling me to text you so my emails wouldn't be lost in the sea of Vice Provost business; and for innumerable phonological insights. Grant, thank you for teaching me so much phonetics, plus how to work the ultrasound, for sharing Osa, and for telling me when something is good enough and it's time to move forward. Thank you both for making time for many meetings, and for always acknowledging that factors outside of academics do matter. Thanks also to Ryan Bennett for making time to hear about my research even while you were at Yale; for the scripts for analyzing ultrasound data; for the stimulating conversations, helpful comments, and paper suggestions. You've all been a great committee.

Thanks to Adrian Brasoveanu for deepening my understanding of R and linear regressions, and to Amanda Rysling for illuminating random effects structures and the source-filter model.

I'm so grateful to have been part of a department that supported and worked around my mothering-while-PhD-ing. Thank you to Clara Sherley-Appel for organizing a meal train when I had a baby in my first year, and thanks to everyone who brought us dinners. Thank you to everyone involved in scheduling that has let me be home to nurse and put babies to bed. An extra thank you to Ashley Hardisty and Maria Zimmer for helping me remember it's normal to have children and have to deal with all their issues. And special thanks to Donka Farkas for so often catching me in the hallway to encourage me in my journey as an academic with young children.

Thanks for helpful questions, comments, and suggestions about this research program are due to the members of Phlunch and the Winter 2016 edition of 219 (especially the inspiring Sandy Chung); to audiences at LASC 2016, BLS43 (2017), the LSA 2018 Annual Meeting, Stanford's P-Interest,

and Berkeley's Phorum; and to fellow UCSC Articulatory Phonologists Netta Ben-Meir and Maho Morimoto. Thanks to Lisa Davidson and to two anonymous reviewers for LabPhon, for helpful critique for what became Chapters 1 and 3. Thanks to Susan Lin for her time and her insightful comments as a committee member on the QE that grew into Ch. 1-3 here, and at conferences thereafter. Thanks to Junko Ito for teaching me phonology; for "professionalizing" my cohort in the Winter 2015 edition of 219; for encouraging me to publish; and last but not least, for telling me that word processors (not just LaTeX!) can support cross-references. That tip dramatically improved the underlying representation of this dissertation document (and probably its surface form, too).

To my fellow graduate students: it's been wonderful to be one of your number. Thank you to my cohort, Maho Morimoto, Deniz Rudin, Jason Ostrove, Kelsey Kraus, Scarlett Clothier-Goldschmidt, and Filippa Lindahl (in descending order of how long they were at UCSC with me)—it was a pleasure to take this journey with you. Thank you for your friendship. Maho, I'm so glad you stuck around for the PhD; thanks for always hanging on the P-side—and particularly for reading the task dynamics literature—with me. Kelsey, thank you for being a linguistics-auntie to my kids and thereby

helping me get linguistics work done. Nick Kalivoda, honorary cohort-member, you've been a great office-mate, syntax-prosody partner, and sanity-saving babysitter. Tom Roberts, thanks for keeping the prosmem project going with me, and for covering the CUNY presentation so I can get this thing written. Also, thank you to anyone I've TA-ed with who took the sections that made my childcare work out.

Thank you to the Turkish speakers in Santa Cruz for your help getting the data in this dissertation. Special thanks to Esra and Saadeddin Bozkurt, and Hatice and Ahmet Çiçek.

Thanks to my faith family at High Street Community Church for support, practical and spiritual. Thanks to all my neighbors in Family Student Housing, especially Dan and Lisa Spencer, Bethany and Ryan Hausen, and Jolene Gregory and Ranulfo, and all their children, for a wonderful community. You all make Santa Cruz home.

Finally, a series of huge thank-yous to my family for their love and support. Thanks to my parents, Carol and Collin Park, who nurtured my first interest in language, math, and programming, and took me to my first linguistics talk (in Santa Cruz!) while I was still in my early teens. Mom, thanks for driving over the mountains so many times, and watching the kids so many hours, over the last five years. I appreciate that you knew about phonetics and phonology long before I'd heard those words. Dad, thanks for teaching me about recursivity long before I took a syntax class, and getting me started on programming early. Thanks to my sister Sheri Park for a lifetime of close friendship, and being a listening ear on the phone and a helping hand around the house, particularly with the kids (yes, there is a theme in these acknowledgements…). Thanks to my extended family for believing I could do it, especially to my aunt Inga Park Okuna, whose stories of taking tests within days of giving birth gave me confidence to always keep moving forward in this PhD. Thank you to Feza and Semih Bellik, my dear parents-in-law and occasional Turkish language consultants, for your support and encouragement.

Thank you to my children, Isaac and Elizabeth Bellik, for all the storybooks and bubbles and sweet cards and walks in the forest and trips to the park—and for the onset cluster reductions and the island violations. You get me out of the office and you help me enjoy my time in the office. You make my days complete.

The biggest thanks are due to my husband Ozan Bellik: my constant supporter in matters emotional, practical, technical, and Turkish language.

Thanks for helping me get to work and for making sure I rest. Thanks for sharing the load with me, and thank you especially for prioritizing my work over your own so many times. You've helped me in so many ways through this PhD journey, and I'm thankful we get to keep to traveling together.

Chapter 1: Introduction

1.1 Intrusion and epenthesis

Cross-linguistically, underlying consonant clusters, particularly in loanwords, often surface with a vowel sound separating the two consonants. In its most familiar form, this vowel sound corresponds to an *epenthetic* vowel – an additional vocalic segment inserted by phonology to repair a consonant cluster that is prohibited by the language's phonology. Like underlying vowels, epenthetic vowels are phonological objects, so they typically do not differ acoustically from underlyingly present vowels, and they sometimes participate in other phonological processes. In particular, epenthetic vowels crucially participate in syllabification, since they repair illegal syllable structures (Hall 2011). The position and quality of epenthetic vowels are affected by the range of repairs and structures that are available elsewhere in the language, as well as perceptual factors (Fleischhacker 2005, Broselow 2015, Yun 2016).

To provide an example, epenthesis of a high vowel occurs in Turkish to repair illegal consonant clusters in codas (1). These illegal clusters (bolded) have rising or flat sonority, and occur in Arabic loanwords (Clements & Sezer 1982). The inserted high vowel (underlined) appears in the bare form of the word, or when the root is followed by a consonant initial suffix, but is absent when the consonant cluster is followed by a vowel-initial suffix, such as in the accusative case. Turkish orthography reflects these alternations. The coda-repairing vowel is obligatorily present in both speech and writing in the bare form of the word and before consonant-initial suffixes.

(1) Coda-repair in Turkish

	Root	Nominative	Accusative	Gloss
a.	/sɑ**br**/	[sɑ.ˈ**b<u>ɯ</u>r**]	[sɑ**b**.ˈ**r**ɯ]	‘patience’
b.	/dʒe**br**/	[dʒe.ˈ**b<u>i</u>r**]	[dʒe**b**.ˈ**r**i]	‘algebra’
c.	/bu**rn**/	[bu.ˈ**r<u>u</u>n**]	[bu**r**.ˈ**n**u]	‘nose’
d.	/ø**mr**/	[ø.ˈ**m<u>y</u>r**]	[ø**m**.ˈ**r**y]	‘life’

The inserted vowel in illegal coda clusters is a target for the phonological processes of syllabification, stress placement and vowel harmony. It forms the nucleus of a syllable, and allows the final consonant to be syllabified as a simple coda. Turkish stress placement is generally word-final, and like underlying vowels, inserted vowels in coda clusters receive stress when they occur in the final syllable.

Coda-repairing vowels are also subject to vowel harmony. The Turkish vowel inventory contains eight phonemes distinguished by [±high], [±back] and [±round]; all three features are relevant to harmony, which affects most suffix vowels. The backness of harmonizing vowels in Turkish is determined by rightward spreading from the nearest vowel in the root. The nearest root vowel also determines the roundness of high harmonizing vowels. Low vowels may trigger rounding harmony but are not targets for it. This harmony process can be seen in the variable realization of the accusative suffix in (1): [ɯ] following /a/, [i] following /e/, and [y] following /ø/. Like the accusative suffix, the inserted vowels in the nominative forms take their backness and rounding from the adjacent root vowel, which indicates that they are targets for vowel harmony. Since the Turkish coda-repairing vowel participates in syllabification, stress-assignment and vowel harmony, it has to be a phonological object. Thus, it must be an epenthetic vowel—a segment inserted during phonology, and mapped to a gesture during articulation.

However, an added vowel sound at the surface does not always correspond to an inserted phonological segment with its accompanying gesture. Studies in Articulatory Phonology (Browman & Goldstein 1993) have established that what sounds like insertion or deletion of a segment can be a side effect of gestural timing relations. When the gestures for adjacent consonants do not overlap, a vowel-like interconsonantal interval can result. The resulting *intrusive vowels* (term adopted from Hall 2003, 2006) can be schwa-like or can "copy" the quality of an adjacent vowel whose gesture overlaps the interval between consonants.

Intrusive vowels contrast with lexical and epenthetic vowels phonologically, gesturally and acoustically. Phonologically, intrusive vowels have no corresponding segment, so they cannot participate in phonological processes that target segments, such as vowel harmony and syllabification. Because intrusion does not result in a new vocalic segment, it does not alter syllable structure, and cannot be taken as repair of illegal syllable structures. When vowel intrusion occurs between two consonants, the two consonant gestures are still part of the same syllable. For instance, if vowel epenthesis occurs in complex codas, the maximal syllable for the language might be CVC. But if vowel *intrusion* occurs, then the two final consonants are still part of the same syllable, and CVCC syllables are permitted. Finally, since intrusive vowels do not form syllable nuclei, they also cannot be targets for stress assignment.

Hall (2003) uses external evidence to argue for the intrusive status of inserted “copy vowels” that repair sonorant-obstruent codas in Scottish Gaelic – they are invisible to syllable-counting in poetry, for example. Similar arguments suggest a gestural-timing origin for Dorsey’s Law vowels in Hocank (Winnebago) (Steriade 1990, Hall 2003), as well as inserted vowels in Finnish, Dutch, Q’eqchi’, and Mono (Hall 2003), and schwa-like vowels in Spanish codas (Bradley 2004, Schmeiser 2009) and Moroccan Arabic coda clusters (Gafos 2002).

Gesturally, intrusive vowels lack corresponding gestures and targets, so they differ articulatorily from phonologically present vowels. Figures 1.1 and 1.2 schematize some possible gestural sequences in an underlying /CrV/ word that is pronounced as [CVrV]. Figure 1.1 shows the gestural sequence for epenthesis. Epenthesis inserts a vowel segment in phonology, which is mapped to an additional gesture during articulation – represented by a bolded dashed line in Figure 1.1.

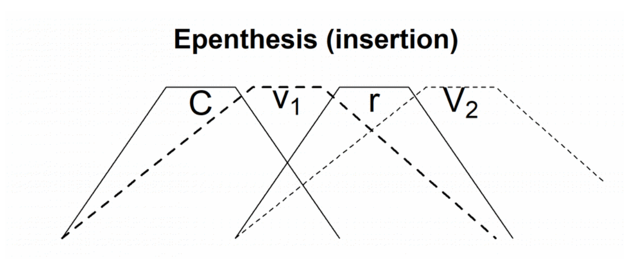


Figure 1.1: Gestural score for epenthesis

Figure 1.2 shows the different gestural sequence that produces intrusion. With intrusion, no segment is inserted in phonology, and no gesture is added in articulation, but the relative timing of the /C/ and /r/ gestures produces the percept of an intervening vowel <v>. This <v> can sound schwa-like when /V/ overlaps less with the interconsonantal interval (Figure 1.2a), or sound like a copy of the following /V/ when the /V/ gesture overlaps more (Figure 1.2b). Intermediate or more extreme alignments are also possible.

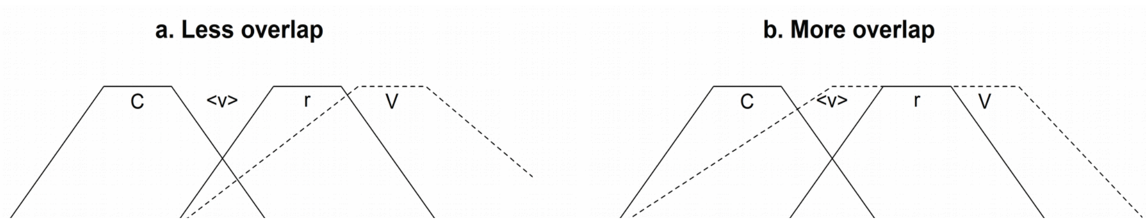


Figure 1.2: Gestural scores for vowel intrusion

Under a traditional division of phonology and phonetics, the phonological grammar cares about epenthesis, but is oblivious to intrusion. In the input to phonology, epenthetic vowels can be absent; but in the output of phonology, epenthetic and underlying vowels are indistinguishable. The segments in the output of phonology map onto gestural targets (C, V, and r), whether V is epenthetic or underlying. The gestures produce an acoustic result, which the listener perceives transparently as [CVr].

Intrusion, on the other hand, creates an opaque relationship between the phonological output and the acoustic output. This occurs when the output of phonology remains [Cr] (no epenthesis), which in turn maps onto a series of two consonantal gestures. Depending on the exact timing relations of how those gestures are produced, the listener may perceive an intrusive vowel <v> that was not present as a segment in the output of the speaker's phonology. Table 1.1 schematizes the stages between phonological input and acoustic output for underlying, epenthetic, and intrusive vowels.

Table 1.1: /Cr/ vs. [Cvr]

<table>
<tr><th></th><th>Input to phonology</th><th>Output of phonology[1]</th><th>Acoustic result</th></tr>
<tr><td>Underlying vowel</td><td>/CVr/</td><td rowspan="2">[CVr]</td><td rowspan="3">Cvr</td></tr>
<tr><td>Epenthetic vowel</td><td rowspan="2">/Cr/</td></tr>
<tr><td>Intrusive vowel</td><td>[Cr]</td></tr>
</table>

To distinguish intrusive and epenthetic inserted vowels, experiments have exploited the articulatory differences described above. For example, the intrusive schwas that break up illegal onset clusters like /zg/ in English are gesturally closer to /sk/ than to /sək/ (Davidson and Stone 2003), indicating that the acoustic schwa lacks its own gesture. In contrast, inserted schwas in Dutch, argued by Hall (2003) to be intrusive, have gestural consequences similar to lexical schwa, causing Warner et al. (2002) to interpret them as epenthetic instead.

Acoustically, vowel intrusion incompletely neutralizes the contrast between /CC/ and /CVC/. Since intrusive vowels have no durational target, they are typically shorter than lexical vowels. Since they have no gestural target, their formant values are more affected by coarticulation. Hall and Sue (2018) show that the "copy-vowels" in Hocank are indeed shorter than lexical vowels.

1 Although this table only includes segmental representations, this is not intended to limit the phonological representation to segments alone. Regular vowel intrusion is best understood as the result of phonological constraints on gestural alignment.

Davidson (2006) shows that intrusive schwas in English speakers' productions of non-native consonant sequences are likewise shorter than lexical schwas, as well as more affected by coarticulation with the following vowel.

Cross-linguistically, intrusive vowels typically occur across sonorants, share the quality of the vowel that is adjacent across the sonorant, do not contribute a syllable, and are sensitive to speech rate (Hall 2003, 2006; see also Fleischhacker 2005). These properties characterize complex onset repair in Turkish, in which an underlying consonant clusters optionally surfaces with an acoustic vowel breaking it up, as in (2). The hypothesized intrusive vowels are transcribed between <angle brackets>.

(2) Onset-repairing vowel insertion in Turkish (Clements & Sezer 1982)

a. /**pr**ens/	[p<i>rens]	' prince'
b. /**pr**ova/	[p<u>rova]	'test'
c. /**br**anda/	[b<ɯ>randa]	'canvas'
d. /**bl**uʒin/	[b<u>luʒin]~[b<y>lyʒin]	'blue jeans'

The onset-repairing vowel in these examples is invited by a stop+liquid cluster, and its quality is affected by the vowel that is adjacent over the liquid. Its presence is optional (Yıldız 2010) and affected by speech style (Clements & Sezer 1982). It is not present in orthography. These characteristics suggest that the onset-repairing vowel may be intrusive, an acoustic consequence of the open transition between consonant gestures.

1.2 Previous descriptions of Turkish onset cluster repair

Previous treatments of Turkish complex onset repair characterize it as the mirror image of complex coda repair. Both non-lexical vowels are described as epenthetic and harmonizing with the neighboring vowel (Yavaş 1980, Clements & Sezer 1982, Kaun 1999, Yıldız 2010). I review the specifics of each study below.

In Clements and Sezer's (1982; henceforth C&S) feature-spreading treatment of harmony and disharmony in Turkish, complex onsets are reported to surface faithfully in careful speech, but be broken up by epenthesis in casual speech. Likewise, Yıldız (2010) describes epenthesis in onset clusters as being in "free variation" with faithful productions. Both C&S and Yıldız (2010) characterize the onset-repairing vowel as a high vowel whose backness and rounding are determined by regressive harmony with the following vowel. C&S also report that the quality of the inserted vowel in onset cluster repair varies, and is affected by consonant place in ways that the coda-repairing epenthetic vowel is not. In particular, the onset-repairing vowel is reported to always be [+back] after dorsal consonants /k/ and /g/, and to optionally be [-back] following /s/, even in the absence of a [-back] lexical vowel to trigger harmony (C&S, Kabak 2011). This characterization of onset cluster repair as optional epenthesis (C&S, Yıldız 2010) predicts that some onset cluster tokens are repaired with phonologically present vowels, with durational and gestural targets like lexical vowels, while other onset cluster tokens contain no vowels, meaning the /CC/ cluster will have a categorically different durational target. In contrast, if onset cluster repair is actually gestural vowel intrusion, the variability is predicted to be gradient. Inter-speaker variation is also predicted: speakers who are familiar with coordinating the gestures for onset clusters in other languages (e.g., French or English) might be better able to closely coordinate the gestures in Turkish as well.

Experimental data on Turkish onset cluster repair comes from Yavaş (1980), Kaun (1999), and Bokhari et al. (2016). Yavaş (1980) reports experimental findings in which backness harmony holds without exception, although he finds that only high vowels trigger rounding harmony. His thirteen subjects, Turkish students at the University of Kansas, read a paragraph containing six nonsense words with word-initial consonant clusters: *treç, prat, dsop, tpük, bçöp, tzut*. Note that only the first two words, *treç* and *prat*, begin with clusters that clearly conform to sonority sequencing. On the basis of the conventionalized inserted vowel in real loanwords containing onset-clusters, Yavaş predicts backness harmony for all inserted vowels, and rounding harmony only when the trigger is a high vowel. Only one reading was elicited from each subject, so there is no information regarding intra-speaker consistency. However, interspeaker consistency was quite high. All thirteen subjects producing the same inserted vowel for the words *dsop, tpük, bçöp,* and *tzut*. For these words, the inserted vowel was always the predicted vowel: conforming to backness harmony in all cases, and conforming to rounding harmony when the trigger vowel was [+high] (in *tzut* [tuzut] and *tpük* [tüpük]).

Where the consonant cluster is less marked in terms of sonority sequencing: *treç* and *prat*, interspeaker agreement is imperfect. Twelve speakers produced the harmonic [tireç] and [pırat]. But one produced [tıreç], in which backness harmony has failed to apply. In addition, two speakers pronounced *prat* as [pirat] with a disharmonically [+front] inserted vowel. This may be the influence of statistical tendencies in the Turkish lexicon, which contains may borrowed words with i-a sequences, but relatively few words with ı as the first vowel. Unfortunately, with only two Cr- words to look at, and only one repetition from each speaker, no firm conclusions can be drawn from this variation. However, it is suggestive that the variability that did occur in Yavaş's results appears in clusters that would only be ruled out by a ban on complex onsets, rather than a general requirement that syllables conform to sonority sequencing. It seems possible that the vowel insertion in the stop+stop and stop+fricative clusters (*tpük, bçöp, dsop, tzut*) is true epenthesis, while the insertion in stop+r clusters (*treç, prat*) is intrusion; but there are not enough examples, and too much variation in the segments involved, to be confident about what factors are at work. Moreover, Yavaş does not report his criteria for determining the quality of the inserted vowel.

Following up on Yavaş's (1980) finding that high round vowels triggered rounding of the inserted vowel, but /o/ did not, Kaun (1999) presented nine subjects with a list of 109 loanwords beginning with consonant clusters, and asked them what vowel they would say them with. That is, this study involved simultaneous production and categorization. All inserted vowels were high vowels that matched the backness of the following lexical vowel. When the following lexical vowel was [+high, +round], inserted vowels were also consistently round. However, rounding varied between and among speakers when the trigger was low, which was interpreted as a height-agreement effect (Kaun 1999). This contrasts with the standard rounding harmony in Turkish, which is triggered by both low and high vowels, but only targets high vowels.

Bokhari et al. (2016) provide the only acoustic study of vowel insertion in Turkish to date. Their participants were four Turkish students at the University of Indiana, Bloomington, who read two repetitions each of 16 target words contained in a carrier sentence. Half the experimental target words contained a coda cluster to repair, and half contained an onset cluster to repair. Subjects also read control words containing an underlying /i/ in the same context. Four clusters were included: /br/, /dr/, /kr/, /kl/. The experimenters predicted that

[i] would be inserted in all target words. For onset clusters, this prediction was based on claims of harmony in the previous literature (reviewed above). For coda clusters, the quality of the inserted vowel in any given word is very consistent across speakers, and is generally determined by vowel harmony, although some roots require insertion of [i] after /a/, in violation of backness harmony. Bokhari et al. (2016) found that coda-repairing vowels did not differ significantly from underlying vowels, while onset-repairing vowels (coded as <i>) had a shorter duration, lower F2, and sometimes a higher F1 than underlying /i/.

Various confounds in the design, however, as well as the small sample size, make interpretation of the results somewhat tenuous. The coda-repairing vowels are orthographically present, and therefore visually present in the stimuli, unlike onset-repairing vowels. So differences between coda-repairing and onset-repairing vowels could be partially due to orthographic effects. Furthermore, the reasons for expecting [i] to be chosen as the repairing vowel varied from cluster to cluster. In the /br/ and /dr/ coda clusters, [i] is inserted despite the dictates of vowel harmony. (It is unclear why these phonologically exceptional words were chosen.) In other conditions, [i] insertion is expected because of the normal dictates of vowel harmony. More broadly speaking, differences in formants between coda-repairing and onset-repairing vowels could be due to differences in the vowel that precedes or follows the target vowel, since the backness and height of this vowel were not controlled. To take the /br/ condition as an example, the four words being compared are /bret/ (onset repair) ~ /birim/ (control for onset repair) ~ /kabir/ (coda repair) ~ /tabir/ (control for coda repair). The adjacent vowel varies between /e/, /i/, and /a/, meaning that the differences found between onset repair and other vowels could be due in part to coarticulation with the preceding or following vowel. The same goes for the consonants that were not part of the target syllable. In addition, the number of syllables in the target words varied, and the study size was small. Therefore, a more closely controlled follow-up study is desireable to confirm the suggestive findings from this pilot experiment.

1.2.1 Onset repair in a corpus

To supplement these studies, I conducted a pilot corpus study on the Turkish Electronic Living Lexicon (TELL; Inkelas et al. 2000). TELL consists of phonemic transcriptions of 17,500 Turkish lexemes produced by two native speakers of Istanbul Turkish. The data was collected by having these two speakers read through a dictionary and a list of place names, producing each lexeme in a variety of morphological contexts. Of the 415 tokens of word-initial onset clusters in TELL, 70% are transcribed with an inserted vowel. Looking specifically at stop+rhotic clusters, which will be the focus of the production experiment described in subsequent chapters, Speaker 1 has 189 input /Cr/

words, of which 135 are transcribed with vowel insertion (71.4% transcribed insertion rate among /Cr/-initial words). The Turkish rhotic, transcribed as /r/ in this paper, is typically realized as an alveolar tap (Göksel & Kerslake 2005:9, Lewis 1967:7).

Across all cluster types, by far the most commonly transcribed non-lexical vowel in TELL is [ɯ]. Non-lexical front vowels were transcribed in only 31% of the cases where a lexical front vowel trigger was available, contrary to the 100% application of backness harmony reported in Kaun (1999), Yavaş (1980), and Yıldız (2000). Non-lexical front vowels were only transcribed after /k/ or /g/ in a single token (*klişe* 'cliche' [kiliʃe]), supporting the claim that front vowels are not inserted after velar consonants (C&S, Kabak 2011). In addition, round non-lexical vowels were transcribed in only 36% of the tokens where a [+round] trigger was present. In line with the results from Kaun (1999) and Yavaş (1980), low round lexical vowels in TELL did not generally trigger rounding in the transcribed inserted vowels, although there were two exceptions where [u] was transcribed as being inserted before /o/. The low rate of transcribed harmony in TELL is surprising in light of some of the descriptions outlined above (possibly excepting C&S, who predict no insertion at all in careful speech), but fits well with the picture of onset cluster repair as vowel intrusion.

A more thorough study of vowel intrusion in TELL, following up on the results above, is reported in Ch. 5.

1.2.2 Discussion of previous work and corpus study

Although prior work largely describes onset cluster repair in Turkish as epenthesis accompanied by vowel harmony, it also reveals differences between onset cluster repair and other epenthesis and harmony in Turkish. Vowel insertion in onset clusters is variable (C&S, Yıldız 2010, TELL results here), and the inserted vowels may differ acoustically from underlying or epenthetic vowels (Bokhari et al. 2016). As observed by Kabak (2011), the harmony that affects onset-repairing vowels is not just the mirror image of the harmony that operates elsewhere in Turkish, because normal rounding harmony in Turkish is not affected by the height or backness of the harmony trigger. In Kaun's (1999) interpretation, the harmonic behavior of onset-repairing vowels reflects a different harmony process, driven by normally inactive constraints from Universal Grammar. The failure of low vowels to trigger harmony is ascribed to a requirement that the trigger and target agree in height. Nonetheless, these results are surprising from an epenthetic perspective, because low vowels are better harmony triggers than high vowels cross-linguistically (Kaun 1995). This is ascribed to the fact that the perceptual cues to the roundness of low vowels are weaker than the perceptual cues to the rounding of high vowels—grossly speaking, high vowels are rounder than low vowels. This articulatory fact suggests that coarticulation with a high round vowel is more likely to produce an impression of rounding on an intrusive vowel than coarticulation with a low round vowel. This is exactly the pattern of (apparent) harmony that emerges for the Turkish onset-repairing vowel (Kaun 1999, Yavaş 1980, corpus results above).

To sum up, where the coda-repairing vowel is obligatorily present in careful speech, casual speech, and in writing, the onset-repairing vowel is only optionally present in both speech and writing. Moreover, where the coda-repairing vowel participates obligatorily in vowel-harmony, the onset-repairing vowel reportedly participates in a variable, consonant-dependent fashion (Clements & Sezer 1982), and the regressive rounding harmony that affects it is sensitive to height factors that are irrelevant to the regular progressive rounding harmony (Yavaş 1980, Kaun 1999, Kabak 2011). Reviewing some of these differences, Kabak (2011) concludes that the apparent regressive harmony that affects onset-repairing vowels is not phonologized to the extent that progressive harmony in Turkish is, and should not be considered the same harmony process. These differences between inserted vowels in onsets and codas are explained if onset repair is vowel intrusion, while coda repair is epenthesis.

One possible problem with interpreting onset cluster repair in Turkish as vowel intrusion, however, is that vowel insertion is reported even in clusters containing two obstruents, particularly in Yavaş (1980). According to Hall (2003, 2006), vowel intrusion that copies the quality of an adjacent vowel occurs only across sonorant consonants. Hall (2003, 2006) attributes this to the gestures for sonorants being better able to overlap with vocalic gestures. In Turkish, the inserted vowel in onset clusters reportedly copies the backness and rounding of the following vowel, even across obstruents (Yavaş 1980, C&S, *inter alia*), contradicting Hall's (2003, 2006) generalization. Other work does report gesturally-driven vowel intrusion in obstruent + obstruent clusters (e.g., Gafos 2002, Davidson & Stone 2003, Davidson 2006), however. While it is possible that intrusion is only occurring across sonorants in Turkish onsets, and optional epenthesis is repairing obstruent + obstruent clusters, it also seems likely that intrusion across obstruents is possible, contra Hall's (2003) criteria, and it simply is less common, whether for the articulatory reasons Hall (2003) points to, or because a obstruent + sonorant cluster is less perceptually altered by vowel intrusion than an obstruent + obstruent cluster is (Fleischhacker 2005).

1.3 Terminology and notation

This dissertation will argues that vowels in onset clusters in Turkish are intrusive. However, to avoid presupposing that these vowels are intrusive, I refer to them as *non-lexical* vowels, since the consensus in the previous literature is that these vowels are not part of the underlying representation. To highlight the contrast between non-lexical and underlying vowels, I often refer to underlying vowels as *lexical* vowels, meaning they are part of the underlying representation of the word. Vowels hypothesized to be intrusive are written as <v> throughout the dissertation. Where an epenthetic vowel and an intrusive vowel in the same context are being contrasted, the intrusive vowel is notated as <v> while the epenthetic vowel is notated as [v]. Throughout, lexical vowels are written as /V/.

1.4 Roadmap of the dissertation

This dissertation probes the phonological status of onset repair in Turkish using a variety of methods. First, I present a production study with ultrasound and acoustic evidence. Chapter 2 presents ultrasound evidence that onset-repairing vowels differ gesturally from lexical vowels in ways that suggest the non-lexical vowels are targetless. These gestural differences have acoustic consequences, both in careful speech (Chapter 3) and casual speech (Chapter 4):

onset-repairing vowels are acoustically intermediate in their F1 and F2 values, compared to lexically [±back] and [±high] vowels (For example, before /i/, <v>'s F2 is higher than /ɯ/'s but lower than /i/'s, while its F1 is lower than /ɯ/'s but higher than /i/'s). Across speech styles and especially in deaccented casual speech, non-lexical vowels are more affected by their surrounding context. The greater gestural overlap of casual, deaccented speech causes the intrusive vowels to become more like the following lexical vowel, mimicking vowel harmony. Even in deaccented speech, however, the timing of C_1 and C_2 in a cluster yields an open transition; in fact, C-C timing is more stable in clusters (where the consonants are tautosyllabic) than in CVC words, where the two consonants are heterosyllabic onsets. These gestural and acoustic results support the hypothesis that Turkish onset cluster repair is the result of gestural timing relations, not insertion of a phonologial segment.

In addition to the production study, I also present a corpus study and two studies of text-setting. The corpus study more deeply investigates onset cluster repair in the Turkish Electronic Living Lexicon (Ch. 5), and confirms the findings of the acoustic study that onset cluster repair is pervasive, but not obligatory, and generally does not harmonize (contrary to previous reports). The TELL study included s+stop clusters, stop+l clusters, and fricative+r clusters, in addition to the /Cr/ clusters considered in the production study. I then examine the text-setting of non-lexical vowels in Turkish onset clusters in Chapter 6. I first analyze words beginning with onset clusters in forty Turkish songs (mostly pop music), then eliciting text-setting judgments for CC words in the Turkish version of “Happy Birthday To You.” I find that non-lexical vowels occupy less metrical space than lexical vowels for text-setting purposes, and their text-setting varies more across speakers than that of lexical vowels.

In Chapter 7, I synthesize these converging lines of evidence that onset cluster repair in Turkish is vowel intrusion, not epenthesis, and propose an Optimality Theoretic account of the phonological representation, gestural coordination, and the constraints on gestural timing that give rise to vowel intrusion in Turkish onset clusters. This analysis is couched in terms of a phonological representation that includes both segments and gestures, adapted from Gafos (2002) and Hall (2003), combined with a coupled oscillator model of syllable structure (Goldstein et al. 2007, 2008; Saltzman & Byrd 2000). This chapter also addresses the relevance of the intrusive vowel to harmony in Turkish, and the historical evolution of onset clusters in Turkish. Chapter 8 summarizes and concludes the dissertation.

This case-study of vowel intrusion in Turkish onset clusters contributes in four areas. First, it provides new, controlled Turkish data in the production study, by collecting repeated productions by multiple speakers of methodically chosen minimal pairs of words. Second, it probes the phonological status of the

Turkish onset-repairing vowel, thereby testing the validity of phonological arguments that have been made on the basis of its behavior. Turkish onset repair is significant for our understanding of both syllable structure and vowel harmony in Turkish. If onset repair is not phonological, then the traditional characterization of Turkish syllable structure as maximally CVC(C) needs to be revised, at least for loanwords. In addition, onset cluster repair provides the only counter-evidence to the traditional claim that harmony in Turkish is strictly left to right. If onset repair occurs outside of categorical phonology, then it is not relevant to harmony. Third, this study expands the knowledge-base for vowel intrusion by supplying phonetic detail about intrusive vowels in a variety of consonant and vowel contexts, and by exploring intrusion's interaction with vowel harmony. Finally, this dissertation's use of diverse data sources (ultrasound, acoustic, corpus, and text-setting) demonstrates how different methodologies, each incomplete on its own, together provide an enriched picture of a variable, gradient phenomenon.

Chapter 2: Ultrasound study

2.1 Gestural characteristics of targetless vowels

Given /CrV/ as the input to phonology, three classes of outputs are possible: insertion of a vowel segment between /C/ and /r/, which is epenthesis; an open transition between /C/ and /r/, resulting in vowel intrusion; and a close transition, resulting in a transparent complex onset. Vowel epenthesis and intrusion can be acoustically similar, in that they both include an interval of high amplitude periodicity between the two consonants, but they reflect different numbers of segments and gestures. In contrast, vowel intrusion and a close C-C transition are acoustically dissimilar, but both reflect the same series of gestures, albeit with different timing. Some possible gestural scores for these outputs are schematized in Figure 2.1. In these diagrams, each trapezoid represents a gesture, and each vertex of a trapezoid represents a gestural landmark: the onset of the gesture, achievement of its target, release of the constriction, and offset of the gesture (Gafos 2002). The degree of overlap between gestures, and whether there is a closed or open transition between them, is visually represented by the extent to which the trapezoids overlap in the diagram.

In vowel epenthesis, the first type of phonological output, the phonology yields [Cv_1rV_2], which is implemented as four distinct gestures, as shown in Figure 2.1a. The epenthetic vowel v_1 is shaped by its own gestural target, not just its context.

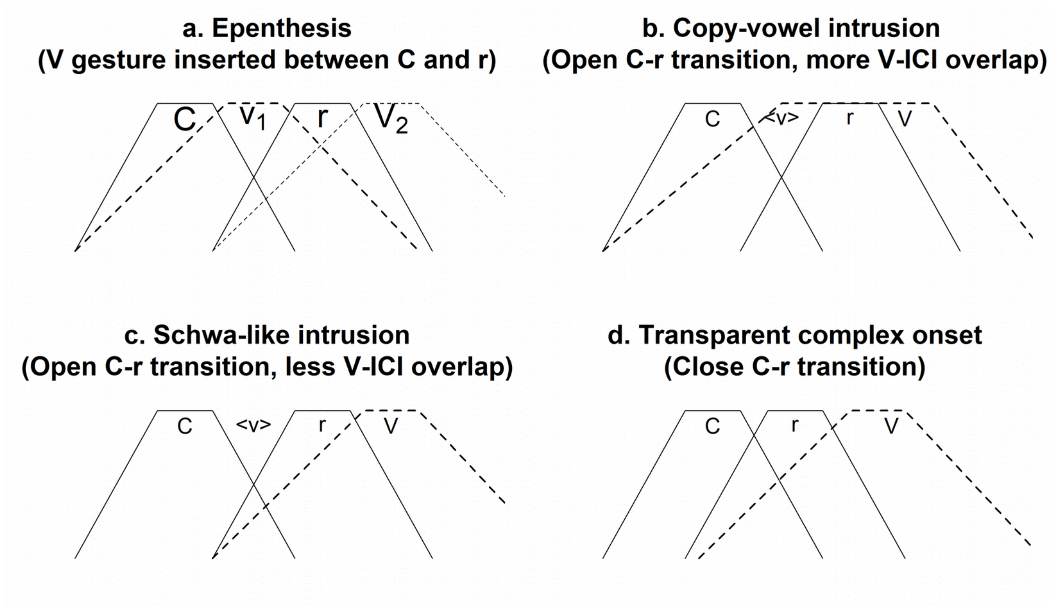


Figure 2.1: Four possible gestural scores for /CrV/

Alternatively, phonology may output [CrV], with no inserted vowel segment. This output is implemented with three gestures, not four: a gesture each for [C], [r], and [V]. Depending on the relative timing of the [C] and [r] gestures, the acoustic result may or may not contain an acoustic vocalic interval, the intrusive vowel. Under some gestural alignments, such as the one in (Figure 2.1b), /V/ achieves its target during the interconsonantal interval (ICI), and the ICI is long enough for the resulting open transition to sound like a copy of the following /V/. This is copy intrusion. /V/ is more likely to attain its target during the ICI if its gesture /V/ begins simultaneously with the onset of the syllable, as depicted in Figure 2.1b. Copy-vowel intrusion is most perceptually and gesturally similar to epenthesis (where the epenthetic vowel shares the features of the lexical vowel). Intrusive <v>s that sound like a copy of V_2 occur in Scots Gaelic (Hall 2003) and Hocank (Steriade 1990, Hall & Sue 2018), for example. Such copy vowels may be as long or even longer than lexical vowels, meaning the ICI can be just as long as a lexical vowel.

Another type of intrusion occurs when the ICI is shorter, and/or /V/ does not attain its target during the ICI (Figure 2.1c). For instance, if /V/ is aligned with the previous consonant (here, /r/), rather than with the first consonant in the syllable, then the voicing for /V/ will begin during the ICI, but /V/'s target may not be achieved during the ICI. This is particularly likely if the ICI is short, since /V/ will have less time to move toward its target. Schwa-like intrusive vowels, traditionally termed excresant, may be more affected by the preceding consonant's release than intrusive copy vowels are. In Turkish onset cluster repair, the inserted vowels are intermediate between schwa-like, excresant vowels and copy vowels, in that they reportedly share the backness of V_2 but not its height.

Finally, the same sequence of gestures but with greater overlap between [C] and [r], as in Figure 2.1d, results in a close transition with no intrusive vowel – the canonical complex onset, as in English.

In vowel intrusion, tongue position during the ICI will reflect the transition between the preceding consonant and the following vowel, since the intrusive vowel has no gestural target. Thus, targetless (intrusive) vowels are expected to differ gesturally from targeted (underlying or epenthetic) vowels in ways that reflect the coarticulatory pressures of the preceding and following consonants and vowel, whose gestures overlap (Alfonso & Baer 1982, Öhman 1966).

To look for these gestural differences between lexical vowels and onset-repairing vowels in Turkish, I conducted an ultrasound production study. This production experiment also addresses the lack of existing data on acoustic detail, intraspeaker variation, and the effect of the surrounding context on onset cluster repair in Turkish. The experiment presented here had three objectives: (1) to look for gestural and acoustic differences between lexical and non-lexical vowels, using ultrasound and audio recordings; (2) to establish whether apparent insertion in Turkish is a gradient or a categorical process, by examining the duration of the interval between C and /r/; (3) to determine the rate of acoustic insertion in onset clusters and the degree to which frontness or rounding spreads to the inserted vowel. This chapter addresses the gestural half of objective (1); Chapters 3 and 4 presented the acoustic results that address the other objectives as well.

2.2 Design and predictions

The ultrasound portion of the study was inspired by Davidson and Stone (2003), who use ultrasound to determine whether articulation of /zg/ clusters by English speakers as [zəg] represents phonological epenthesis, or phonetic intrusion. In addition to nonce words beginning with non-English zC clusters, like *zgama,* they elicited real English words containing a lexical schwa, such as *succumb*, and words with sC clusters like *scum*. This provided them with minimal triplets of gestural sequences, so that they could compare the gestural sequence in words with an inserted vowel (e.g. [zəgama]) to the lexical schwa and to the faithfully-produced cluster, which differed only in voicing from the unfaithfully produced cluster.

If the inserted vowel is epenthetic and has a gestural target, the gestures that produce the inserted vowel will resemble the gestures that produce the lexical schwa more closely than they resemble the sequence that produces the insertionless cluster. On the other hand, if the inserted vowel is intrusive, then the gestures that produce it will more closely resemble the gestures that produce the insertionless cluster. Davidson and Stone (2003) determine which gestures are more similar using L2 norms, a statistical measure that evaluates the height differences between two curves whose lengths have been normalized; a smaller L2 norm value means two curves are more similar to each other.

It is not possible to replicate Davidson and Stone's (2003) experimental design exactly for Turkish. There is no way to elicit Turkish minimal triplets, since all onset consonant clusters in Turkish are candidates for vowel insertion. However, the overall insight can still be used: if two vowels are both phonologically present, then their gestures will not be significantly different when they occur in the same environment, while if one vowel is phonologically present and the other is an artifact of gestural timing relations (an intrusive vowel), then even in the same environment, their gestures will be significantly different. Furthermore, the non-lexical vowels can be compared to harmonic and disharmonic lexical vowels to evaluate the degree to which harmony has applied to them. This idea is the basis for the experimental design, and particularly for the gestural analysis.

2.2.1 Design

Vowel insertion is reported to occur in all types of clusters in Turkish (including /s/+stop, obstruent+/l/). This was verified by a study of onset clusters in TELL, whose results are reported in Ch. 5. For the production experiment, /Cr/ clusters were chosen, because insertion is transcribed at a higher rate in /Cr/ clusters (71% in TELL) than in /sC/ clusters (42% in TELL).

In addition, surface harmonic effects resulting from vowel overlap are more likely to occur across a sonorant like Turkish /r/ (phonetically a tap) than across a stop (Hall 2003, 2006; see also Bradley 2004 for intrusion after a tap). /Cl/ clusters were avoided for ease of segmentation using spectrograms. The experiment had a 2 by 3 by 3 by 2 by 3 design (3).

(3) Experimental factors:
  [Lexical status: /Cr/ ~ /CVr/]
  x [C_1 = /b d g/]
  x [V_2 = /i a o/]
  x [Familiarity: Real ~ Nonce]
  x [Speech style: Careful, Casual mention 1, Casual mention 2]

The primary factor manipulated was the underlying syllable structure of the target word, and hence the lexical status of the vowel between /C/ and /r/: non-lexical vowels occurred in words beginning with a stop + /r/ onset cluster (/Cr/), and lexical vowels occurred in words beginning with a simple onset followed by an underlying vowel and /r/ (/Cvr/). The /Cvr/ words were included as controls so that non-lexical vowels in /Cr/ words could be compared to lexical vowels.

To ensure that the findings extend across all consonant and vowel places, and investigate claims of vowel harmony in the inserted vowel, three stop consonants (/b d g/ – voiced stops were chosen to avoid aspiration[2]) and three

2 In a pilot version of this experiment, voiceless stops /p t k/ were employed, but preliminary results showed that even underlying vowels were often completely devoiced following these

vowels (/i a o/[3]) were included. Both real, familiar words and completely unfamiliar nonce words were included, to check that insertion is a fully productive process. Finally, both careful and casual speech were included to test for the effect of speech style (C&S).

2.2.2 Predictions

Lexical vowels have phonologically specified gestural targets, while intrusive vowels are targetless. Hence, the position of the articulators during an intrusive vowel will be determined by the gestural demands of the surrounding vowels and consonants. In this experiment, then, we expect the surrounding consonants and following vowel to shape non-lexical vowels—hypothesized to be intrusive—more than they shape lexical vowels.

Previous work on gestural alignment has found that the vowel gesture for a syllable nucleus begins at the same time as the gesture for a simplex onset consonant does, or begins midway between the onsets of all the consonants in a

---

word-initial voiceless stops, due to aspiration. This would have made it impossible to use voicing during the interconsonantal interval as a criterion for the presence of a non-lexical vowel, so the design was revised to use /b d g/.

3 Originally /u/ was included as the third V_2 value, rather than /o/, but no sufficiently familiar words of the shape /bru-/ could be found, and so /o/ was selected instead. In some theories, /o/ is considered to be a better trigger of rounding harmony than /u/ (Kaun 1995), so /o/ also provides a better test of whether the quality of the inserted vowel is determined by phonological vowel harmony or by phonetic coarticulation. Additionally, /o/ as a non-high vowel provides more information about whether acoustic inserted vowels seem to share the height of the lexically present vowel, or only its backness and rounding.

complex onset (Byrd 1995, Browman & Goldstein 1988). If /Cr/ words begin with a complex onset, the V_2 gesture in this study will overlap the interconsonantal interval (ICI), pulling the tongue body toward V_2's following target during intrusive vowels. In contrast, in the control /CVr/ words, C is a simplex onset to V_1, while /r/ is a simplex onset to V_2. C and V_2 are in different syllables, then, and V_2 will not be phased to begin at the same time as C. Therefore, V_2's gesture will not influence the ICI (and V_1) as much in /CVr/ words as in /Cr/ words. I predict, then, that intrusive vowels in /Cr/ words will be shaped by V_2 more than lexical vowels in /CVr/ words.

The direction of expected differences between intrusive and underlying vowels will be determined by the particular vowel that drives anticipatory coarticulation in the ICI. In the discussion and analysis that follows, I compare predicted tongue body positions for intrusive vowels with those for /ɯ/ in the same context, based on the fact that <ɯ> is the vowel most commonly transcribed in underlying clusters in TELL (see Section 1.2.1 [pilot study] and Ch. 5 [full study]). I also compare intrusive vowels with the high vowels that regressive vowel harmony would demand, based on claims in the previous literature that onset cluster repair inserts a high vowel that conforms to backness and sometimes rounding harmony (Yavaş 1980, Clements & Sezer 1982, Kaun 2000, Yıldız 2010). Below, I spell out the predictions for the three following vowel conditions in this experiment: /i a o/.

2.2.2.1 Predictions for <v> before /i/

When the following vowel is /i/, then the tongue body is expected to be raised and fronted in anticipation of /i/'s [+high, -back] target. Consequently, before /i/, the tongue body should be fronter and potentially higher during intrusive <v> than during underlying [+high, +back] /ɯ/. (Although both /i/ and /ɯ/ are phonologically [+high], an MRI study by Kılıç and Öğüt [2004] found that /ɯ/ has a lower tongue body position than /i/.) Despite the fronting and raising pressure exerted by the following /i/, however, the tongue is predicted to be less front and high during a targetless <v> than during an underlying [+high, -back] /i/, because in an underlying cluster, a following /i/'s [+high, -back] target does not need to be attained during the targetless ICI, whereas an underlying /i/ that occurs in the same position must attain its target during that interval between C and /r/. Therefore, before /i/, we expect tongue body position in underlying clusters to be intermediate between that of underlying /ɯ/ and /i/.

2.2.2.2 Predictions for <v> before /a/

When the following vowel is /a/, anticipatory coarticulation will drive the tongue body to lower toward /a/'s [-high] target during the ICI. This predicts that a targetless vowel before /a/ will be lower than the [+high] /ɯ/. Since /ɯ/ matches /a/ in both backness and rounding, it is also harmonic with /

a/, so no other vowels are predicted before /a/, even under an account where harmony applies to the onset-repairing vowel.

2.2.2.3 Predictions for <v> before /o/

Finally, when the following vowel is /o/, the tongue body is predicted to be moving toward /o/'s [+back, -high] target during the ICI. As with intrusive vowels before /a/, intrusive vowels before /o/ may have a lower tongue position than high vowels like /ɯ/ and /u/ (predicted by vowel harmony). No differences in backness are predicted, since /o/, /ɯ/, and /u/ all have [+back] targets.

Differences in the degree of rounding are possible as well. If the lips are approaching /o/'s [+round] target, intrusive vowels before /o/ should be rounder than /ɯ/, although perhaps not as round as /u/. However, lip rounding is not observable from the ultrasound results, which image the tongue body only.

2.2.2.4 Summary of predictions

To summarize: if onset repairing vowels are targetless, we predict them to be fronter than /ɯ/ when they precede /i/, and lower than /ɯ/ when they precede /a/ or /o/. We also predict that targetless vowels will be less front than /i/, and less high than /u/. These vowels were chosen for comparison because they are predicted by epenthesis and harmony.

We might also expect the consonants that surround the vowel (C and /r/) to affect intrusive vowels more than they affect underlying vowels. Cross-linguistically, labial consonants are known to sometimes cause rounding on adjacent vowels; coronal consonants can cause fronting; and dorsal consonants can cause raising and/or backing (Padgett 2011). These consonant-to-vowel effects should be strongest in the case of targetless intrusive vowels. However, an intrusive vowel requires a sufficiently open vocal tract to sound vocalic at all, meaning that the tongue body cannot be too close to the closure for a preceding or following consonant. This limits the magnitude of the effect that the surrounding consonants can have. No such limit applies to the tongue's movement toward the following vowel target, however. We might therefore expect the differences between intrusive and underlying vowels to be driven more by the following vowel than by the surrounding consonants.

The degree to which V_2 can overlap the ICI, however, will still be shaped by C's demands on the tongue body. Vowel-based differences between intrusive and underlying vowels are predicted to be most pronounced when the preceding consonant is labial (/b/, in the experiment presented here), since the lips are able to move independently of the tongue body, meaning that V_2 can overlap with /b/ without interfering with its articulation. When the preceding consonant is coronal (here, /d/), its tongue tip target will limit movement of the tongue body toward V_2's target, since the tongue tip is coupled to the tongue body, with the results that V_2's impact on the ICI will be less pronounced than in labial conditions. Finally, V_2-driven differences between lexical and non-lexical vowels will be least pronounced when the preceding consonant is dorsal (/g/), since /g/'s tongue body target will most severely limit anticipatory movement toward the following vowel's tongue body target.

As a velar consonant, /g/ can be expected to contribute backing and/or raising to an adjacent vowel (Padgett 2011); this occurs, for example, in intrusive vowels in Maxakalí (Gudschinsky et al. 1970, Clements 1991, cited in Padgett 2011). At the same time, /g/'s articulation may be affected by the place of an immediately adjacent lexical vowel. As is common cross-linguistically, velar consonants in Turkish tend to palatalize in the context of an adjacent front vowel (Göksel & Kerslake 2005). In /gir/ words, then, /g/ will palatalize under the influence of the adjacent front v_1 /i/. But in /gr/ words, an intrusive vowel cannot pull /g/ forward; instead, /g/'s own tongue dorsum target will pull the targetless vowel backward.

2.3 Methods and procedure

2.3.1 Materials

A list of real and nonce words beginning with stop + /r/ clusters was constructed (Table 2.1, Experimental columns). Target words take the form /

$C_1(v_1)rV_2C_2.../$. Within each C_1-V_2 condition, C_2 was matched for major place of articulation. Stress was also controlled so that syllables that would be compared were all unstressed; stress falls on the final syllable (V_2 or later) in all words. Finally, the number of syllables was also controlled, such that all output forms in a C_1-V_2 condition are predicted to have the same number of syllables (not counting potential inserted vowels as syllabic).

Real /Cr/ words were chosen to be familiar, where possible. Familiarity was determined on the basis of a familiarity-rating survey conducted with three native speakers of Turkish (1 female, 2 male; ages 28 – 63), who did not participate in the experiment otherwise. Participants were asked to rate the familiarity of the words on a five-point scale, where 1 meant “I don't know this word at all” and 5 meant “I use this word regularly or learned it as a young child.” Instructions were presented in Turkish. A word was considered familiar if it received an average rating of at least 4 on the survey, with no participant giving it a rating of 1 or 2. Unfortunately, in the /dri-/ and /gro-/ conditions, no sufficiently familiar Turkish word was found, so the highest-rated available word was selected even though ratings were quite low (1 for *dripling* 'dribbling (as in basketball)'; 2.67 for *gros* 'gross (as opposed to net)').

Control words of the form /CVrV/ were created for every condition (Table 2.1, Control columns) so that non-lexical vowels in /Cr/ words could be compared to lexical vowels in the same context. The number of relevant /CVrV/ control words varies depending on the identity of V_2. The apparent insertion of [ɯ] is attested before all qualities of V_2, so control /CVrV/ words with $v_1 = /ɯ/$ were created in every V_2 condition (i.e., /Cɯri/, /Cɯra/ and /Cɯro/). In addition, [i] is reported to be inserted before /i/, and [u] is reported to be inserted before /o/, so I created control words where v_1 and V_2 were both [-back] or [+round] (/Ciri/ and /Curo/). Only [ɯ] is reported before /a/, since it is already harmonic for backness and rounding. As a result, there are fewer relevant control words in /a/ conditions.

While [ɯ-a] sequences are harmonic for both backness and rounding, [ɯ-i] sequences are disharmonic for backness, and [ɯ-o] sequences, for rounding. These disharmonic sequences are unattested as underlying sequences (except for *gardɯrop* 'wardrobe') in the corpora and dictionaries I consulted. This gap in the lexicon suggests that Turkish phonology prohibits these particular disharmonic vowel sequences. Onset-repairing vowel insertion creates them in surface forms, however. Therefore, the necessary /Cɯri/ and /Cɯro/ controls had to be nonce words, and it was not possible to maintain distinct real and nonce conditions in the control words. Instead, nonce control words were included in all conditions, and real words were also included when they existed, resulting in different numbers of control words depending on the condition. No familiarity ratings for control /CVr/ words were obtained, since so many nonce /CVr/ words were included, and since I expect familiarity levels to have no significant impact on the articulation of a lexically present V_1.

Stimuli are shown in Table 2.1. Unglossed items are nonce words. An asterisk following a word indicates that it is also being used as a /CVr/ match for a /Cr/ word in the real word condition, since no appropriately shaped real word could be found. Familiarity ratings for real /Cr/ words are shown in parentheses. There are 24 words each in the real and nonce word conditions, but 11 nonce /CVr/ words overlap between the two conditions, so the total number of distinct target words is 37.

In addition, 17 fillers (Table 2.2) were included, for a total of 54 target words. Because so many experimental items are nonce words, primarily real words were selected as fillers—mostly borrowings from English or French since all the familiar real words are borrowings.

Table 2.1. Stimuli for the production experiment.

C_1	V_2	Experimental		Control	
		Real /Cr/ *word (familiarity) 'gloss'*	Nonce /Cr/	v_1 = <ɯ>	v_1 ≠ <ɯ>
b	/i/	bri.fing (4) 'briefing'	bri.mi.ti	bɯ.ri.pis	bi.ri.m-in 'unit.your' bi.ri.bis
	/a/	bran.ʃ-ɯ (4.3) 'subject.ACC'	brat.ʧi.ten	bɯ.ran.ʤɯ *	–

C_1	V_2	Experimental		Control	
	/o/	bro.ʃyr (4.67) 'brochure'	bro.ʒør.le	bɯ.ro.ʒyn*	bu.ro.ʧyp*
d	/i/	drip.ling (1) 'dribbling'	drip.li.ke	dɯ.rib.le*	di.rim.-ler 'life.PL' di.rib.rit
	/a/	dra.ma (4) 'drama' *or* dra.m-a (3.7) 'drama.DAT'	dra.fa	dɯ.rap*	–
	/o/[4]	bor.dro-m (4) 'payroll.my'	lor.dro.pur	gar.dɯ.rop 'wardrobe'	nor.du.rof*
g	/i/	grip (5) 'influenza'	gri.vi	gɯ.rif*	gi.rim 'penetration' gi.riv
	/a/	gram (5) 'gram'	gra.bɯ	gɯ.rap*	–
	/o/	gro.s-u (2.7) 'gross.ACC'	gro.dol	gɯ.ron*	gu.rot*

4 A note about the /dro-/ cell: In all other C-V conditions, the consonant-cluster of interest is word-initial. But in the dro- condition, the cluster appears word-internally (/bordrom/ 'payroll.my'). This word was selected in order to maintain the same environment for the cluster as for the underlyingly present vowel, in order to be able to use the real word gardɯrop 'wardrobe' for the /Cɯro/ control word. Ultimately, however, this turned out to be a mistake, because the /rdr/ sequence that was intended as a coda /r/ followed by a complex onset was instead syllabified as a complex coda followed by a simplex onset. Consequently, the /dro/ condition was omitted from the analysis.

Table 2.2. Fillers for the production experiment

C1	V = /e/	V = /u, o/	V = /a/
Labial	merimit (nonce) meteorolozi 'meteorology'	provizjon 'commission'	marɯp (nonce) paralelkenar 'parallelogram'
Coronal	negatif 'negative' neptyn 'Neptune'	tuvalet 'toilet' turnike 'turnstile'	tablo 'painting' tansijon 'blood pressure'
Dorsal	kervan 'caravan' geometri 'geometry'	kuaför 'hair dresser'	kakao 'cocoa' karton 'carton'

In the first portion of the experiment, whose results are discussed in this chapter, participants were instructed to speak carefully, and both target and filler words were presented in the carrier sentence in (4), which includes slots for two target words (X and Y). The sentence was designed to elicit contrastive focus on the target words, to further enhance the carefulness of the elicited speech.

(4) Bana X deme, bana Y de.
me.dat X say.neg, me.dat Y say.
'Don't say X to me, say Y to me.'

Since the structure of the carrier sentence elicits an expectation of structural parallelism (that X and Y will be of the same grammatical category and case), X~Y pairs with the same case were selected. Also, within a given sentence, X and Y were either both nonce or both real.

In the second portion of the experiment (discussed in Chapter 4), participants were instructed to speak casually, as if talking with a family member or close friend, and a different carrier sentence was used (5).

(5) Fatma X_1 ve Y_1 dedi, Erhan da X_2 ve Y_2 dedi.
Fatma X_1 and Y_1 said, Erhan also X_2 and Y_2 said
'Fatma said X_1 and Y_1, Erhan also said X_2 and Y_2.'

The carrier sentence for the Casual condition was designed to elicit deaccentuation and concommitant hypoarticulation in the second mention of each target word (X_2 and Y_2). I set aside the Casual condition for the remainder of this chapter; a comparison of the two speech styles will be presented in Chapter 4.

To control for the possibility that prosodic factors would create a confounding difference in articulation between X and Y, half the repetitions employed an X-Y order, and the other half employed a Y-X order.

2.3.2 Participants

Six native speakers of Turkish (3 female: S4, S5, S7) were recruited from the University of California at Santa Cruz community. (A seventh (S1) participated in the pilot experiment, after which the design was significantly revised, so her data are not discussed.) S3 is bilingual in French and Turkish, so language effects may complicate the interpretation of his data. S6 lived in New Jersey, USA, for a year (age 4-5), but in Turkey otherwise. The remaining speakers all studied English in school during adolescence, but lived in Turkey, using Turkish as their primary language at home and work, until age 18 or later. Participants were paid $20 for their time.

2.3.3 Procedure

A consent form was provided in English. A language background questionnaire and experimental instructions were provided in Turkish. Participants were told that the purpose of the experiment was to study the way Turkish speakers pronounce words. Participants wore an Articulate Instruments Ultrasound Stabilization Headset (Wrench 2008) to stabilize the ultrasound probe. Recordings were made in a sound-attentuated booth using a shotgun microphone with a USB pre-amplifier connected to the ultrasound machine (Terason T3000 ultrasound system with a model 8MC3 probe, 45-60 frames per second). Subjects were asked to practice reading the instructions to get comfortable speaking with all the equipment, and were instructed to start the sentence over if they felt they had made a mistake. The experimenter also intervened when disfluencies or errors were noticed. In the Careful condition (the focus of this chapter), participants were requested to speak carefully and enunciate clearly, as if they were announcers on TRT (Turkish Radio and Television), whose broadcasters' careful articulation is famous in Turkey.

Stimuli were presented to subjects on a laptop screen, with the target words already embedded in the carrier sentences. One sentence was visible at a time. Participants read through a list of 27 sentences (each containing up to two target words) five times in Careful speech, and five times in Casual speech, for a total of fifteen repetitions of each target word (since the Casual carrier sentence elicited two instances of each target). Within the list, all sentences were randomized together, without any blocking of real vs. nonce words. After each reading of the sentence list, participants were offered the chance to take a break. At the end of the experiment, participants filled out a debriefing form with questions provided in Turkish as well as English. Responses indicated that participants had not identified the research question being investigated.

Acoustic annotation of the v_1 interval was conducted in Praat (Boersma & Weenink 2015) using TextGrids. The left edge of the interconsonantal interval (ICI) was marked from the beginning of the C_1 release burst, identified by a dramatic increase in amplitude. The right edge of the ICI was identified by the decrease in amplitude accompanying the onset of /r/. The ICI was further subdivided into the burst + VOT (annotated as "burst") and v_1, where v_1 was identified as the portion of the ICI that had high amplitude periodicity and formant structure. Sometimes no such formant structure occurred. Less commonly, high amplitude periodicity with formant structure sometimes occurred throughout the ICI. Representative spectrograms are shown in Figure

2.2 (underlying vowel), Figure 2.3 (non-lexical vowel), and Figure 2.4 (cluster with no vowel).

Although not analyzed here, the boundaries of the preceding consonant closure, following /r/ (phonetically a tap, sometimes fricated), and V_2 were also annotated. Consonant closure was identified by the dramatic drop in amplitude and total loss of formant structure. The left edge of /r/ was identified by a decrease in amplitude, usually accompanied by a loss of formant structure; its right edge and the beginning of V_2 were identified by the onset of high amplitude periodicity. Where /r/ was produced with frication, this frication was included in the /r/ interval, not the V_2 interval.

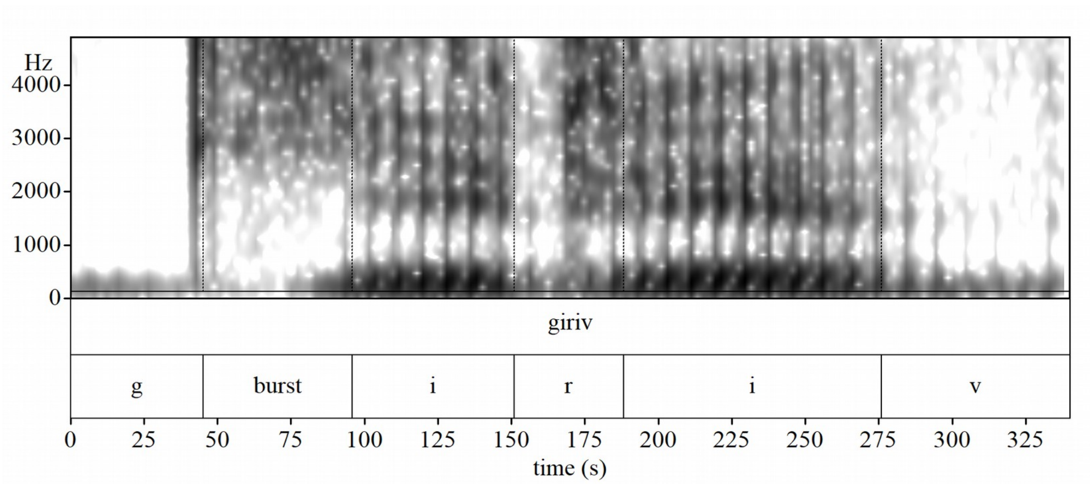


Figure 2.2: /Cvr/ token with an underlying vowel (from S3)

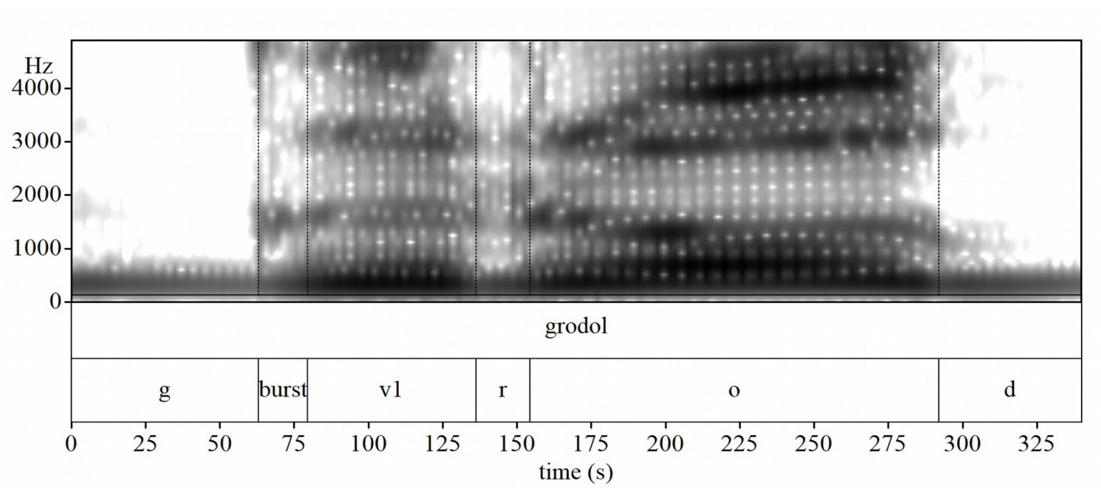


Figure 2.3: /Cr/ token containing an acoustic non-lexical vowel (from S4)

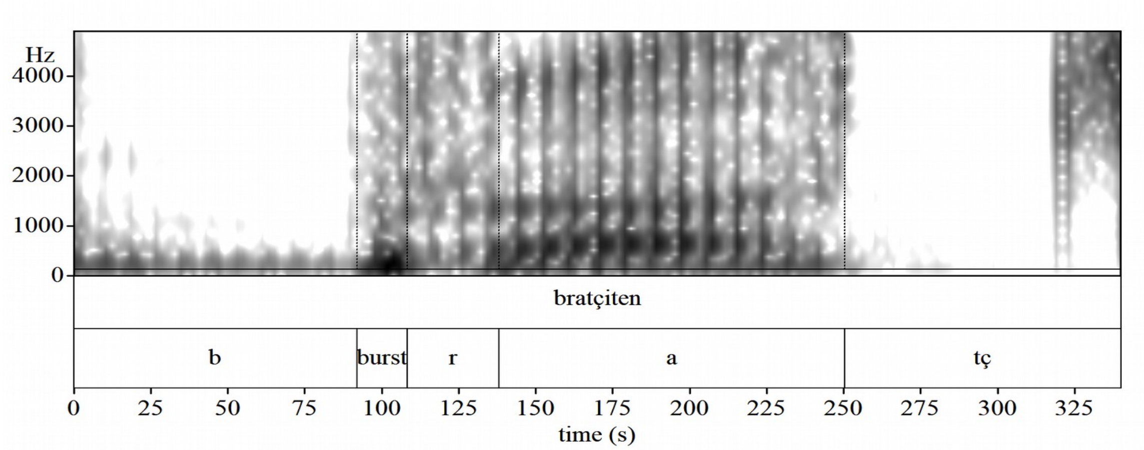


Figure 2.4: /Cr/ token with no acoustic insertion (from S3)

2.3.4 Data processing

For the gestural analysis, I used a Python script to select the three ultrasound frames best corresponding to beginning (onset), middle (midpoint) and end (offset) of the interconsonantal interval. This interval is quite short (generally 30-60 ms), so sometimes only one or two frames were captured. Tongue tracings for all available frames were made in Edgetrak (Li et al. 2005). Comparison of the frames from onset, midpoint and offset revealed few obvious differences between the different timepoints, except that in the third frame, the tongue tip is often raised relative to the preceding frames (reflecting the onset of the gesture for the following tap). Therefore, the analysis presented here focuses on the midpoint. Where no midpoint frame was captured, a frame that was closer in time to the onset or offset of the ICI was used instead (approximately

2% of the data); generally the onset frame provided a clearer image of the tongue (29 tokens) and was chosen in preference to the offset (6 tokens).

After tongue tracings were completed, R (R Core Team 2016) was used to create smoothing spline ANOVAs (SSANOVA; Gu 2002, Davidson 2005) for each word, within subject. In order to compare tongue position during underlying clusters to tongue position during underlying vowels, SSANOVAs for words with underlying clusters and with underlying vowels within each C_1-V_2 combination were plotted together. Initial comparisons found minimal gestural distinctions between nonce and real words, so the nonce/real distinction was collapsed within each C_1-V_2 combination. Thus, each plot shows a contour for the underlying cluster, another for underlying /ɯ/ in the same context, and (where harmony would demand a different vowel) the underlying harmonic vowel in the same context (/i/ or /u/).

In the results below, tongue body position during underlying clusters is represented by an orange line; underlying /ɯ/ is represented by a light blue line; and where harmony demands a vowel other than /ɯ/, the harmonizing vowel (/i/ before /i/, /u/ before /o/) is shown in dark blue. Each speaker is shown separately, since there is sometimes significant interspeaker variation. The plots of these SSANOVAs also include 99% confidence intervals, shown by dashed lines. When the confidence intervals for two curves do not overlap, this indicates that the curves represent significantly different tongue positions. In most plots, the confidence intervals are so close to the main curve that they are hard to see. As is typical in ultrasound results, however, the position of the tongue tip and root is less certain, due to jaw shadow which limits imaging of those areas. This decreased certainty is reflected by the flared confidence intervals around the ends of the curves. The flaring at endpoints also reflects the method of calculating confidence intervals, which is sensitive to the square of distance from the midpoint on the x-axis.

2.4 Results

A comparison of SSANOVAs of tongue body position in underlying clusters vs. vowels largely bore out the predictions above. That is, tongue position in underlying clusters did differ significantly from tongue position in underlying vowels in the same context, in ways that show the greater influence of the following vowel on underlying clusters than on underlying vowels. Not every speaker conforms to the predictions in every condition, however. Also, as expected, a preceding /g/ obscures the effect of the following vowel on tongue body position.

Speakers can be grouped into three sub-patterns, according to the patterning of their acoustic (Ch. 3) and gestural results (this chapter). The early bilinguals S3 and S6 (first column of grouped SSANOVAs) tend to show the greatest differences between underlying clusters and underlying vowels. Late bilinguals S5 and S7 (second column) tend to show intermediate levels of gestural difference. Lastly, S4 (third column), the only monolingual in the study, patterns by herself. S2, an older male, is excluded here because his individual anatomy was not conducive to ultrasound imaging, resulting in poor data quality.

2.4.1 /i/ conditions

As discussed above, before /i/, tongue body position during a targetless vowel <v> is predicted to be intermediate in backness between that of /i/ and that of /ɯ/. This prediction was clearly borne out when the preceding consonant was labial (Figure 2.5). For all but one subject, the underlying cluster is intermediate in frontness between /ɯ/ and /i/. The exception is S4, whose /ɯ/ seems to be almost as front as her /i/, while her underlying cluster is higher and backer than either underlying vowel.

The same overall result holds when the preceding consonant is /d/, but the differences between /i/ and /ɯ/ are less dramatic, most likely due to the fronting effect of the coronal consonant (Figure 2.6). Nonetheless, for four out of five subjects, tongue body position during the cluster can be said to be intermediate in frontness between that of /i/ and /ɯ/. The exception is S7, whose underlying clusters largely overlap underlying /ɯ/. For all subjects, however, the cluster is lower than one (S3, S4, S7) or both (S5, S6) of the underlying vowels, either at its peak or throughout the contour. For S4, this height difference is much clearer than the difference in backness.

The pattern of clusters being intermediate between /i/ and /ɯ/ before /i/ is much less discernible when the preceding consonant is /g/ (Figure 2.7). Only S6's /gr/ clusters are intermediate between /i/ and /ɯ/. For S5, S7, and S4, tongue positions in /g_ri/ and /gɯri/ do not differ significantly. For S3[5], /g_ri/ is higher and backer than /ɯ/, perhaps indicating that /g/ is having a greater impact during the targetless interval than during underlying vowels. As a velar consonant, /g/ can be expected to contribute backing and/or raising to an adjacent vowel (Padgett 2011); this occurs, for example, in intrusive vowels in Maxakalí (Gudschinsky et al. 1970, Clements 1991, cited in Padgett 2011).

To summarize results for the conditions where the following vowel is /i/, we find that intrusive vowels are clearly intermediate between /ɯ/ and /i/ following /b/ and /d/, but much less clearly so following /g/, which exerts its own demands on the position of the tongue body.

5 The sharp dip near the tongue root in the SSANOVA for S3's /giri/ condition is an artifact of the averaging performed by the SSANOVA technique, combined with the fact that more of the tongue dorsum and root were traceable in some repetitions than in others.

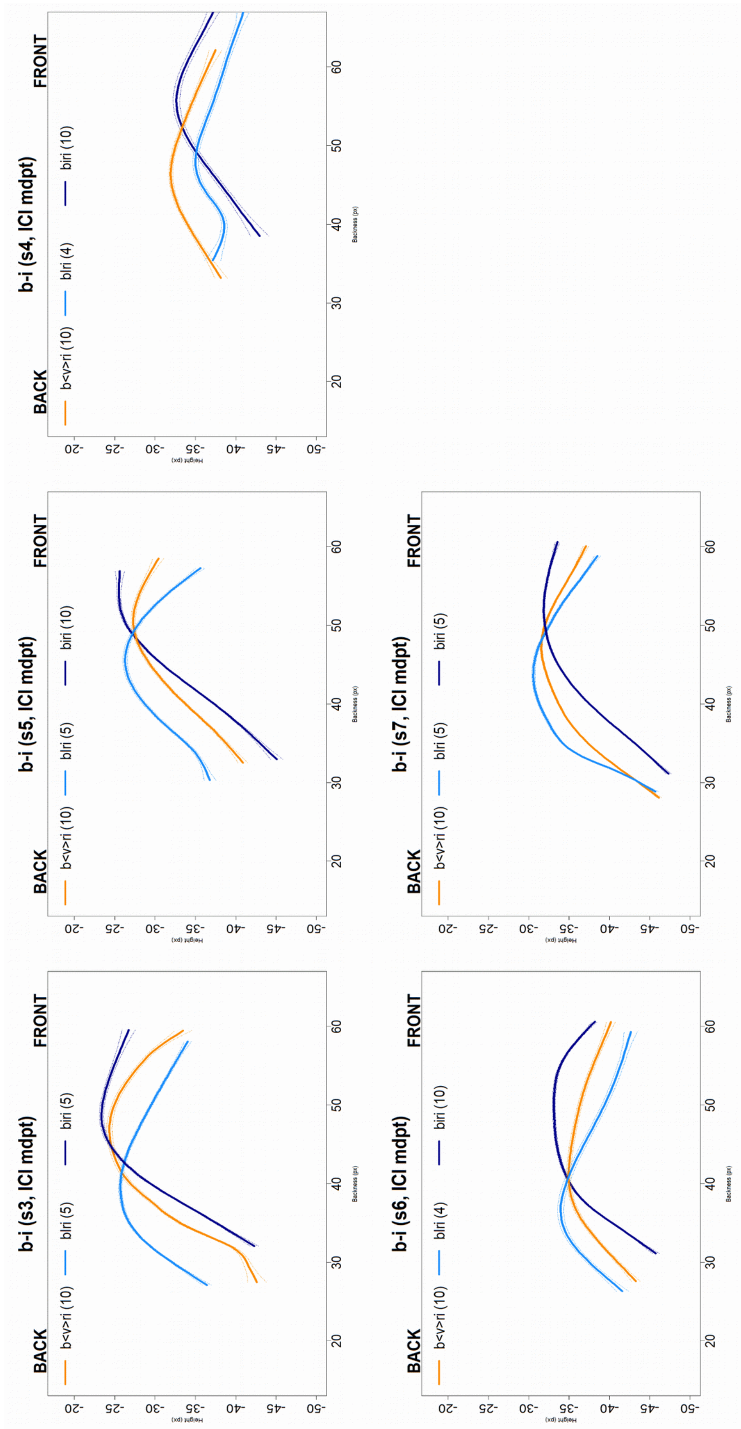


Figure 2.5: Tongue body position in the /b-i/ condition

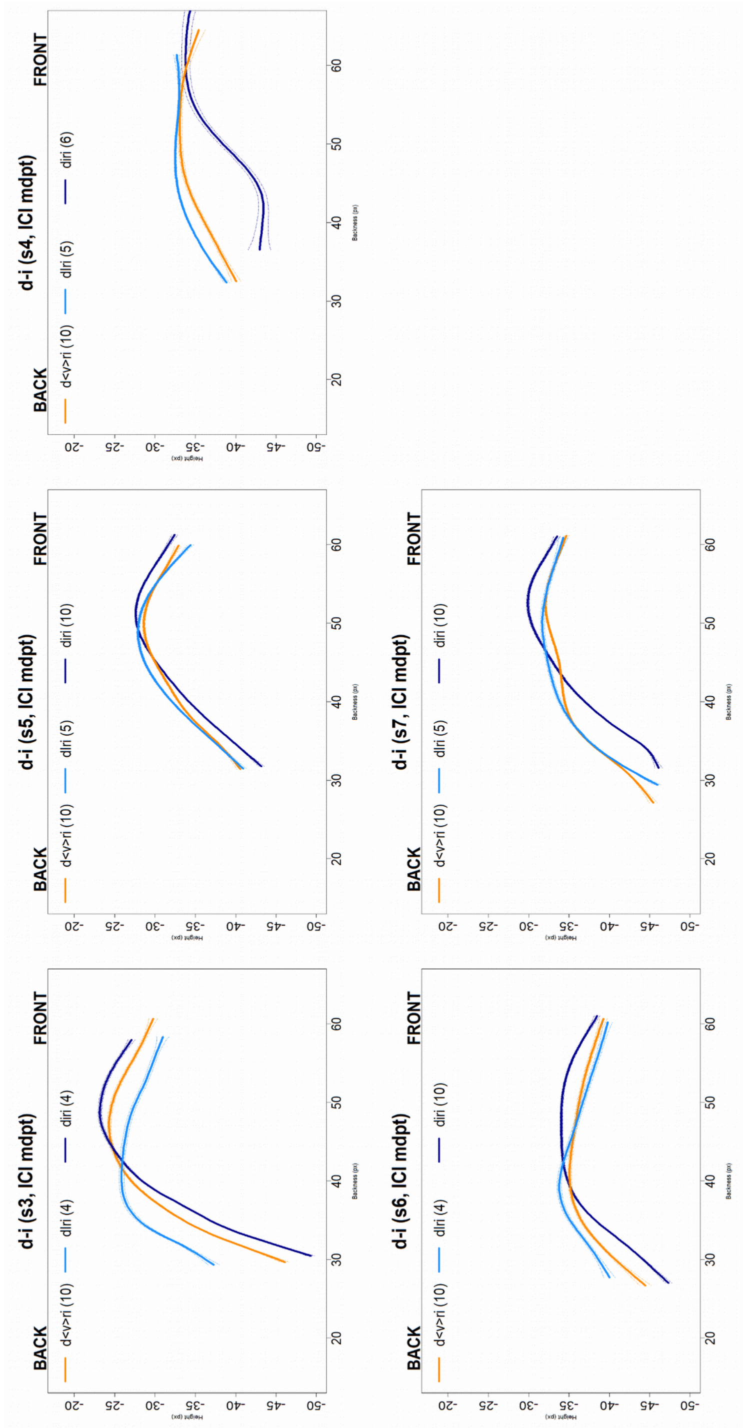


Figure 2.6: Tongue body position in the /d-i/ condition

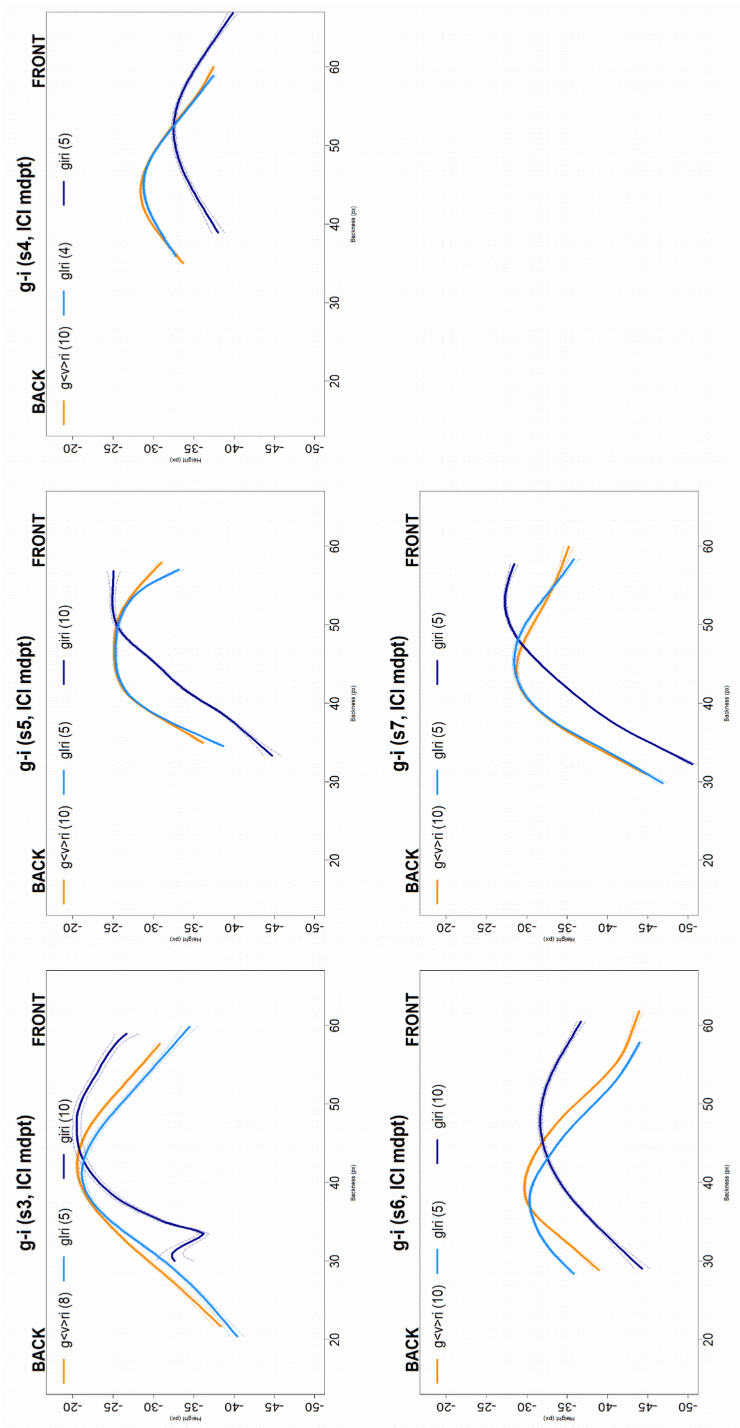


Figure 2.7: Tongue body position in g-i conditions

2.4.2 /a/ conditions

Turning now to conditions where /a/ follows, we predict that a following /a/ will result in lower tongue position during a targetless vowel than during the [+high] /ɯ/. As in the /i/ conditions, the prediction is most clearly borne out after /b/. Here, tongue position is significantly lower in underlying clusters for all subjects, indicated by the non-overlapping confidence intervals surrounding the cluster (orange) and /ɯ/ (blue) (Figure 2.8). The height difference is much greater for S3 and S6 than for the other three subjects, although it is significant for all five.

The same pattern holds after /d/, where S3 and S6's clusters are clearly lower, and S5 and S4's clusters are also significantly lower (Figure 2.9). For S7, however, underlying /dra/ and /dɯra/ do not differ significantly in tongue body position.

Finally, after /g/, we see significant differences for four out of five subjects, but mostly not in the direction predicted by anticipatory lowering (Figure 2.10[6]). S4 is the only subject who shows the predicted lowering in the cluster. The cluster is actually significantly higher than /ɯ/ for S3 and S7, suggesting that /g/'s velar closure has a greater effect on non-lexical than on

6 The sharp peak and dip in S5's /gɪra/ curve are artifacts of the SSANOVA smoothing method, caused by the small number of repetitions to average across, and the fact that the tongue tip was untraceable in one repetition, while the tongue root was untraceable in another.

lexical vowels. For S4 and S6, the tongue body is less back in clusters than in /ɯ/, perhaps reflecting movement toward /a/'s more central articulation.

Overall, tongue position in clusters with a following /a/ tends to be lower than in an /ɯ/; however, a preceding dorsal consonantal target interferes with this effect, and can even make /gra/ higher than /gɯra/. As in /i/ conditions, the following vowel has the greatest effect when the preceding consonant makes minimal demands on the position of the tongue body.

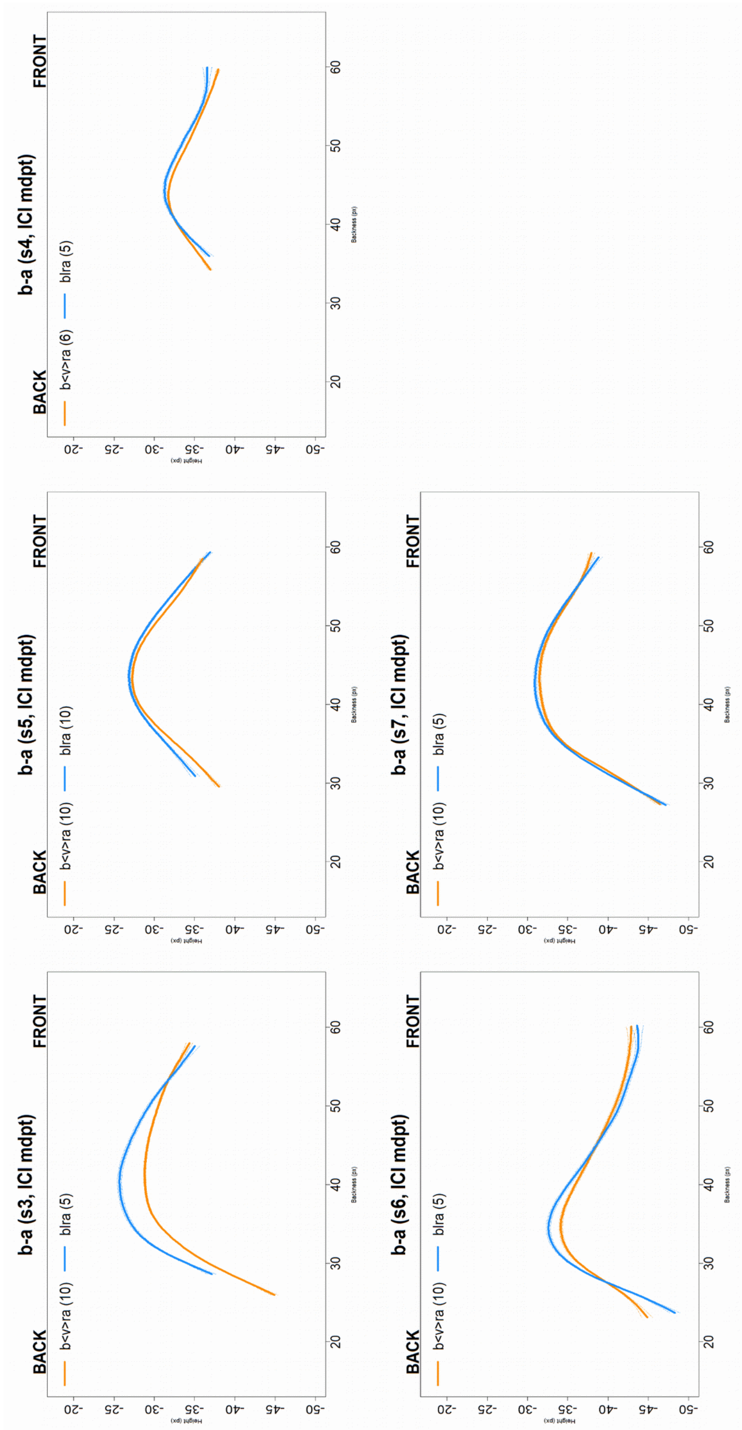


Figure 2.8: Tongue body position in the /b-a/ condition

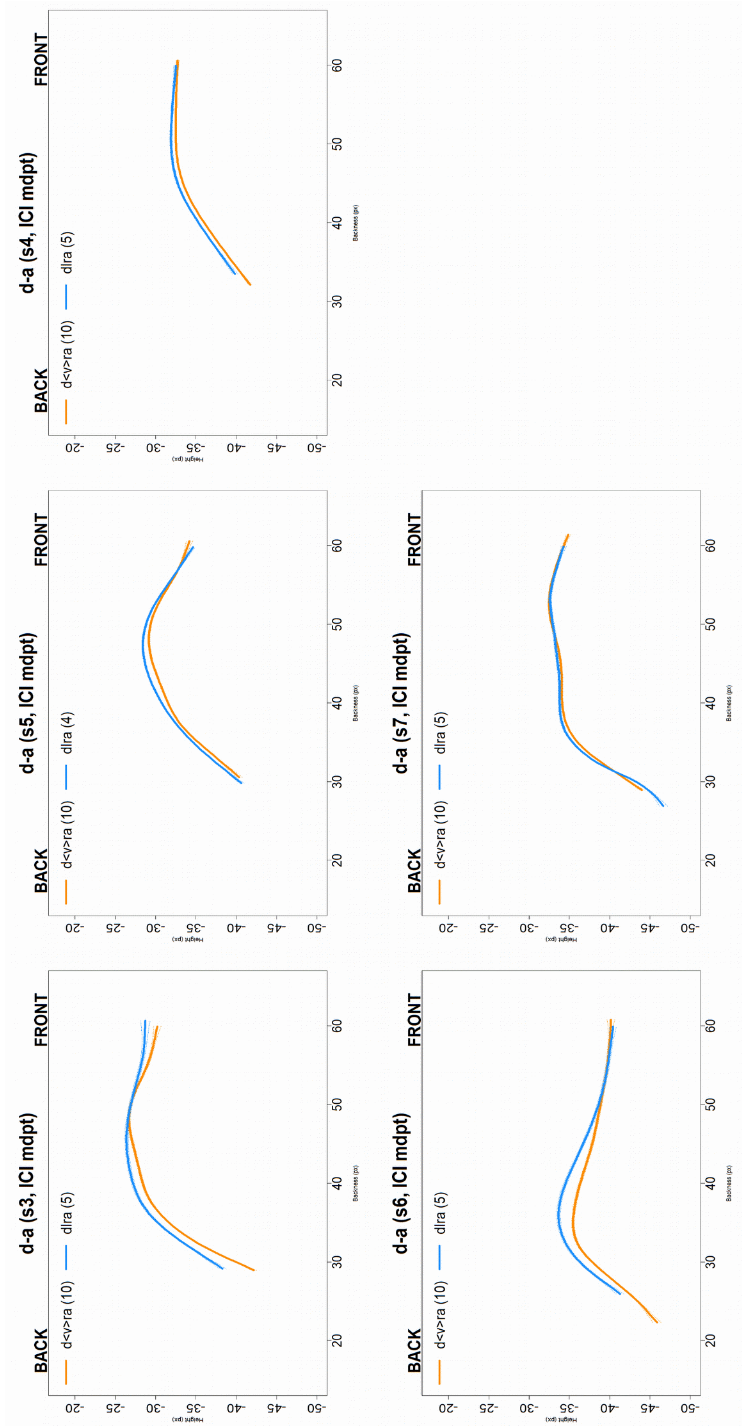


Figure 2.9: Tongue body position in the /d-a/ condition

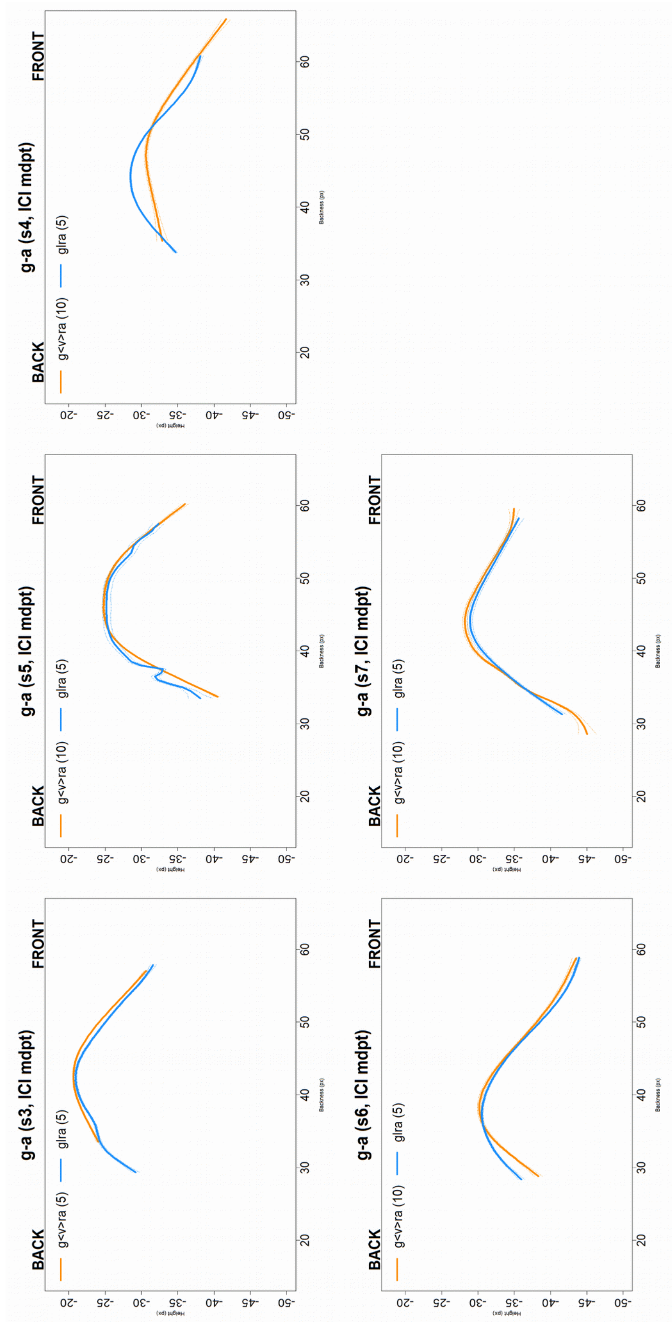


Figure 2.10: Tongue body position in the /g-a/ condition

2.4.3 /o/ conditions

When the following vowel is /o/, height differences similar to those in /a/ conditions are predicted. Results were mixed; the prediction is supported overall by results in the /b-o/ condition, but less clearly so in the /g-o/ condition. (The /d-o/ condition was omitted due to a confound in the experimental design; see Footnote 4, page 39.)

Examining the /b-o/ condition first, we find that tongue body position does tend to be lower in clusters than in underlying vowels (Figure 2.11). For S3 and S7, /b_ro/ is significantly lower than both /bɯro/ and /buro/. For S5 and S6, the cluster is lower than /ɯ/ only, and for S4, the cluster is lower than /u/ but higher than /ɯ/.

A complicating factor in this condition is that some subjects had a tendency to front the vowel preceding /o/, even though it was written as a [+back] vowel. This was most likely triggered by the fact that the target words chosen contained front rounded vowels in the syllable following /o/, in order to best match the most familiar real word that began with /bro/, namely, *broşür* /bro.ʃyr/ 'brochure.' The other words included in the condition were nonce words: *brojörle* /bro.ʒør.le/, *buroçüp* /bu.ro.ʧyp/, *bırojün* /bɯ.ro.ʒyn/. S4 in particular produced multiple tokens of *bırojün* that sounded more like [byroʒyn] than the intended [bɯroʒyn], which is probably the reason her /ɯ/ appears so far forward of her /u/ and underlying cluster. Similarly, S6's /u/ appears fronted, which probably reflects this speaker's tendency to pronounce intended [b**u**rotʃyp] as [b**y**rotʃyp]. The /o/ in the second syllable, however, was not affected by this fronting, which hints at the possibility that V_1 /u/ and /ɯ/ were affected because they are [+high], like the apparent trigger /y/ (a height agreement effect – cf. Kaun 1995).

Turning to the /g-o/ condition (Figure 2.12), /g_ro/ is indeed lower than /gɯro/ and /guro/ for S4 and possibly S7. For the other subjects, however, there are no significant differences in tongue body position between <v>, /ɯ/, and /u/. This illustrates again that /g/ influences tongue body position more than /b/ does.

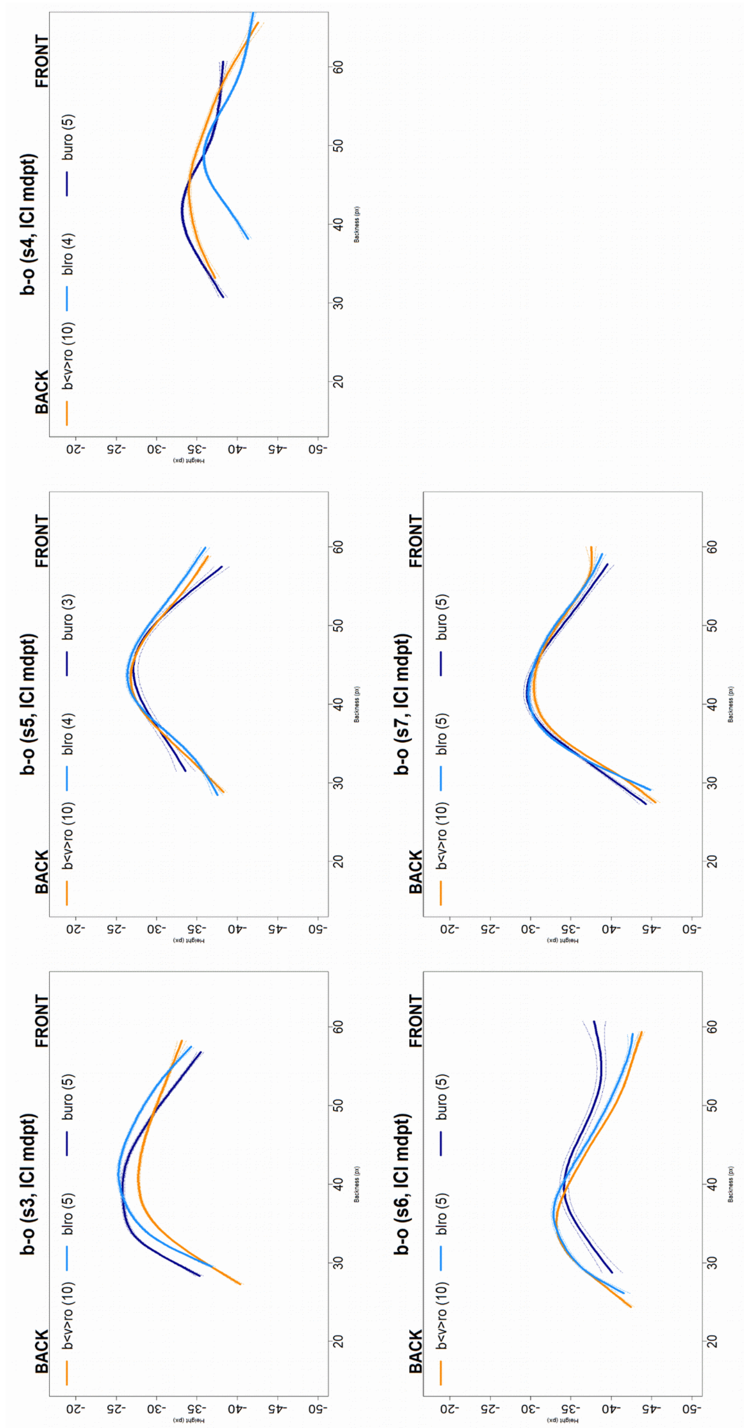


Figure 2.11: Tongue body position in the /b-o/ condition

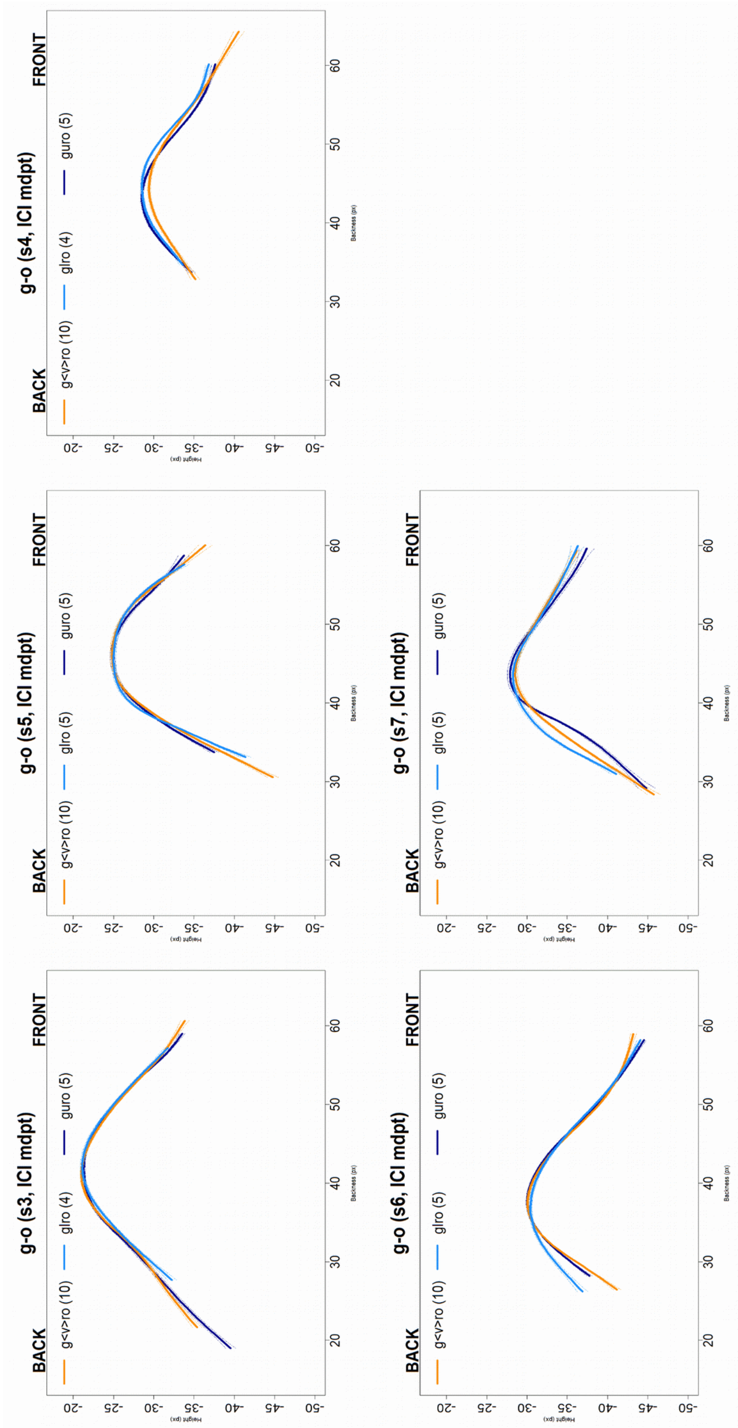


Figure 2.12: Tongue body position in the /g-o/ condition

2.5 Discussion

This study found that tongue body position during the ICI in underlying clusters is significantly different from tongue position in underlying vowels, even when those underlying clusters are acoustically broken up by a vocalic interval. In general, the anticipatory coarticulation with the following vowel influenced tongue position in the ICI of /Cr/ words more than it influenced tongue position in underlying vowels in /CVr/ words. The influence of V_2 was particularly clear when the preceding consonant was labial and therefore made no demands on tongue body position. When the preceding consonant was dorsal, its high, back target mostly blocked the lowering or fronting effect of the following vowel. Thus, in /b/ conditions, the differences in tongue body position were in the direction predicted by *a priori* phonetic expectations, whereas in /g/ conditions, differences were more subtle and variable. Gestural differences in the /d/ conditions were clearer than in /g/ conditions but less clear than in /b/ conditions. Overall, the results of the ultrasound study support the hypothesis that onset repairing vowels are targetless, and therefore their tongue position is determined by the coarticulatory effects of the preceding consonant and following vowel.

The tongue body positions found here indicate that, in underlying clusters, the tongue body is already moving toward its V_2 target during the ICI. However, it has not yet attained the following V_2's backness target, as shown by the fact that in /Cri/ words, the tongue is significantly backer during the ICI than it is during /Ciri/. Furthermore, the tongue has not attained its height target yet, since the tongue body is fairly high during the ICI even before a following mid or low vowel (/o/ or /a/). While a following non-high vowel does have a lowering effect, this effect is only clearly seen in the /b/ and to a lesser extent /d/ conditions. In the /g/ conditions, tongue body position during the ICI of clusters is not significantly lower than tongue body position in underlying high vowels for most subjects. This is particularly true for the /gro/ condition.

2.5.1 Gestural organization of Turkish onsets

These gradient distinctions between underlying vowels and underlying clusters, which vary according to the consonant context, imply that the gesture for V_2 is still in its onset phase during the ICI; it has not attained its target yet. In the language of Gafos (2002) or Hall (2003), this suggests a gestural alignment in which the release of the C_1 gesture is aligned with the onset of the V_2 gesture. Indeed, if C_1C_2 is syllabified as a complex onset, then we would expect the V_2 gesture to be coordinated with the C-center (Shaw et al. 2009, Browman & Goldstein 1988). Firmer conclusions about the relative timing of the gestures involved here, however, requires a study of tongue movement over the course of the C(v)rV sequence, going beyond the analysis of tongue position at a single point in time (the midpoint of the ICI, or C-center, examined here). Optical Flow Analysis (Barbosa & Vatikiotis-Bateson 2014, Hall et al. 2015), for example, could indicate a stable point corresponding to each gestural target in the sequence, and show when the fastest vertical or horizontal changes in tongue position are occurring. The precision of such a time sequence analysis would be limited by the frame rate of the ultrasound machine. In this experiment, the data was captured at a rate of 40-60 frames per second, meaning that each frame represents 17-40 ms of gestural information. Most ICIs in this experiment were 20-100 ms long, meaning that there is sometimes only one frame corresponding to the ICI. This issue is mitigated, however, by the fact that a time series analysis would examine the entire C(v)rV interval, which obviously has a much greater duration than the ICI alone.

The gestural coordination that produces intrusive vowels seems to be grammaticized in some languages (Gafos 2002, Hall 2003), and this is likely the case in Turkish as well. I return to this issue in Chapter 7, where I propose that the Turkish grammar of gestural timing prioritizes an anti-phase coordination between the two consonants in the cluster, which pushes them apart in time, over an in-phase coordination between C_1 and V, which seeks to synchronize their onsets. Interspeaker variation found in this study suggests that Turkish speakers vary in the gestural coordination they employ in onset clusters. This variation is manifested in the acoustic characteristics of the ICI as well (Ch. 3). I suggest in Ch. 7 that this could reflect individual differences in the coupling strengths assigned to the competing gestures in the onset, as well as individual differences in phonetic implementation.

2.5.2 Implications for harmony and syllable structure in Turkish

If onset cluster repairing vowels arise from gestural timing relations, rather than being epenthetic, then their behavior should not be used as the basis for arguments about the segmental phonology of Turkish, particularly vowel harmony. An intrusive vowel cannot be a target for phonological harmony since it is not a phonological object. This suggests that the reasoning behind studies like Kaun (1999)—where the harmonic behavior of the onset-repairing vowel is used to make claims about speakers' access to phonological constraints that are not active in the native lexicon—must be re-evaluated. Kaun (1999) may bear on the phonetic basis for phonological constraints, rather than the phonology of Turkish vowel harmony per se.

In addition, the non-harmonizing behavior of the inserted vowel cannot be used to bolster the traditional understanding of vowel-harmony in Turkish as a strictly left-to-right process (e.g., Lees, 1966; Underhill, 1986), since an intrusive vowel could never be a target for phonological harmony anyway;

neither can its occasional harmonic acoustics (actually due to coarticulation) be attributed to the emergence of a normally invisible right-to-left harmony process.

The phonological status of the vowels in Turkish onset clusters is also relevant to our understanding of Turkish syllable structure. If the onset-repairing vowel is not epenthetic, then there is no categorical prohibition of complex onsets in the foreign stratum of Turkish phonology. Rather, gestural timing relations create the percept of a vowel in a sequence that, phonologically speaking, remains a complex onset. In Chapter 6, I test this claim by probing Turkish speakers' mental representations of onset clusters with a syllable-counting task, and by examining text-setting of these non-lexical vowels in music.

2.5.3 Methodological contribution

Finally, this project also bears on the extensibility of Davidson & Stone (2003)'s methodology to other phonological problems. This study applies Davidson & Stone's experimental design but combines it with a more modern statistical technique, SSANOVAs, for a more nuanced analysis. This comparative ultrasound methodology was successful in probing the phonological status of Turkish onset-cluster repair. In addition, this study suggests that there can be a great deal of interspeaker variability in the articulation of sequences that are not contrastive in a language. Further research with more speakers could illuminate the factors that may structure this variability.

Chapter 3: Acoustic study (Careful speech)

Adapted from: Bellik, J. (2018). An acoustic study of vowel intrusion in Turkish onset clusters. Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 9(1), 16. DOI: http://doi.org/10.5334/labphon.112*

3.1 Acoustic characteristics of intrusive vowels

Given the gestural differences between intrusive and lexical vowels, we expect to see acoustic differences between them as well. Lacking a durational target, intrusive vowels are likely to be shorter than underlying vowels. Lacking gestural or acoustic targets, intrusive vowels' formant values will reflect the transition between surrounding consonants and vowels.

In addition, intrusive vowels should have a different distribution of durations than optional epenthetic vowels, since vowel intrusion is driven by gradient gestural alignment, in contrast to categorical epenthesis. As the result of a gradient process, the intrusive vowel's duration is predicted to have a unimodal distribution. In contrast, an optional categorical process is predicted to have a bimodal distribution, since the insertion would apply completely in some tokens and not at all in others.

In the case of Turkish onset cluster repair, we saw in the previous chapter (Ch. 2) that tongue body position during intrusive vowels is more affected by anticipatory coarticulation with the following vowel and preceding consonant than tongue body position during underlying vowels in the same context is. Although intrusive vowels sound, impressionistically, like the Turkish high back unrounded vowel /ɯ/ (see also the corpus study in Ch. 5), gesturally they are lower than /ɯ/ before /a/, and fronter than /ɯ/ before /i/, while still less front than lexical /i/. Acoustically, these gestural differences are predicted to trigger corresponding differences in F1 (a proxy for tongue body height) and F2 (a proxy for tongue backness). The acoustic study presented here shows that these predictions were borne out in the careful speech condition of the production study with six speakers.

3.2 Methods

To look for acoustic differences between lexical and non-lexical vowels in Turkish, I compared the duration, F1, and F2 of lexical and non-lexical vowels occurring between /b d g/ and /r/ in the first syllable of a word, and followed by V_2 = /i a o/. This data was collected in the production study whose design and methods are described in Chapter 2; the acoustic annotation procedure can be found in Section 2.3.3. This chapter focuses on findings from the Careful speech condition. A corresponding analysis of the Casual speech condition was also conducted, and can be found in Chapter 4.

Measurements of F1 and F2 were taken at the midpoint of v_1 using a Praat script, and converted from Hertz to Bark using the formula from Traunmüller (1990). I excluded nine tokens which stood out as outliers when the vowels were plotted; their F1 in Bark was less than 1.9 (*burozyn, burotʃyp* and *gurot* from S3; *branʃɯ* from S7) or greater than 5.2 (*bratʃiten, brozørle* and *driplike* from S3; *dirimler*, S4; *burozyn*, S7). Five tokens in which V_2 was mispronounced were also excluded (three repetitions of *bratʃiten* (S4) and one each of *driplike* (S4) and *gɯron* (S3), all nonce words). Finally, the *dro-* condition was excluded because the /dr/ cluster occurred word-medially and was not syllabified as a complex onset in many cases (See Footnote 2 in Ch. 2). All other clusters were word-initial. The resulting dataset contained 936 tokens from six speakers.

3.3 Results

No differences were found between real/familiar words and nonce words, so real and nonce words are treated together throughout the analysis. I find that onset cluster repair is acoustically variable and gradient (3.3.1). Non-lexical vowels tend to be acoustically [ɯ]-like, rather than conforming to vowel harmony (3.3.2). However, non-lexical vowels display significant differences in duration, F1 and F2 from both harmonic lexical vowels (3.3.3 - 3.3.4) and from lexical /ɯ/ (3.3.5).

3.3.1 Gradience in onset repair

Using R (R Core Team 2016), the distributions of two duration-measures for underlying clusters were plotted (Figure 3.1, solid line): the ICI, and the portion of the ICI with high amplitude periodicity with formant structure (which I refer to as the vowel). The distribution of lexical vowels is also plotted for comparison (dashed line). If onset cluster repair is gradient vowel intrusion, the interconsonantal interval (ICI) is predicted to have a unimodal distribution of durations. But if onset cluster repair is optional but categorical epenthesis, ICI durations are predicted to have a bimodal distribution (one mode for insertion and one mode for no insertion). This is the case, for example, with the lingual gesture in devoiced vowels in Japanese: EMA data indicates that it is present in some tokens, absent in others (Shaw & Kawahara 2018).

In Turkish, both duration distributions appear unimodal, suggesting that acoustic insertion is a gradient process, not an optional categorical one. (The secondary mode in the smoothed density curve for vowel durations is an artifact of coding the underlying clusters that were produced with no vowel as having a vowel duration of 0ms; of course no negative durations are possible.) The ICI is shorter in underlying clusters than in underlying vowels, though the underlying vowels themselves are quite short as well (mean total duration = 74.1ms, mean duration of high amplitude periodicity = 53.6ms).

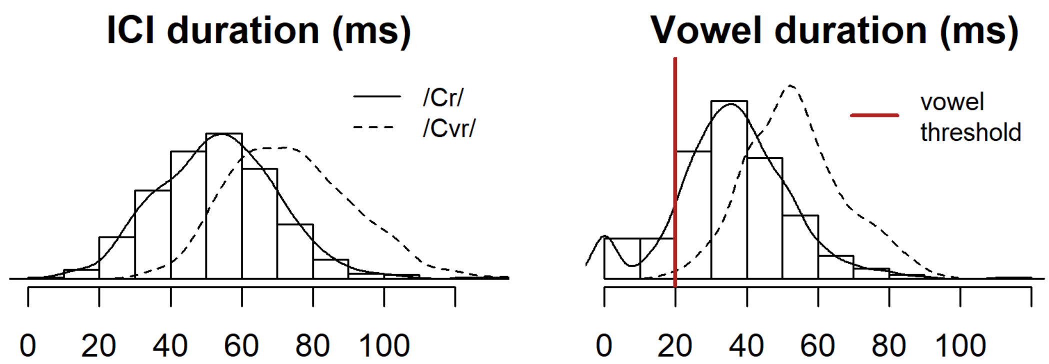


Figure 3.1. Duration of ICI and of v_1

For purposes of comparing acoustic non-lexical vowels to lexical vowels, a vowel-duration threshold (shown in red in Figure 3.1) was established at 20ms[1], since the histogram of non-lexical vowels reveals a sharp change between the 20-30ms and 10-20ms bins, and all but two underlying vowels are longer than 20ms. Clusters produced with at least 20ms of high amplitude periodicity with formant structure were coded as containing an acoustic non-lexical vowel. With this criterion, acoustic vowel insertion occurs in 88.3% of the underlying clusters, with the insertion rate varying between subjects (Figure 3.2).

1 If we take a more conservative approach and place the threshold midway between the mean of the lexical vowel duration distribution (57.4ms) and the mean of a hypothesized no-insertion distribution centered on 0 (i.e., at 28.6ms), all results are essentially the same.

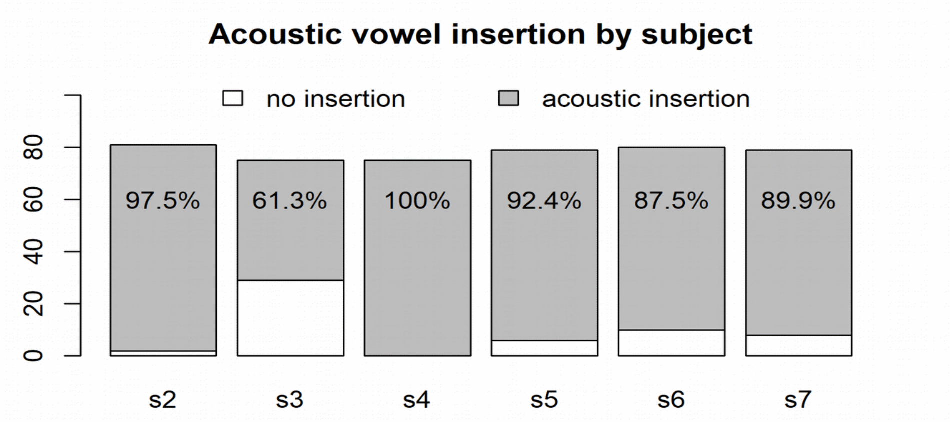


Figure 3.2: Acoustic insertion in onset clusters by subject

3.3.2 Vowel plots and F1~F2 of v_1

Using R (R Core Team, 2016) and the 'ellipse' package (Murdoch & Chow, 2018), both lexical and non-lexical v_1 tokens were plotted in F1~F2 space (Figure 3.3). Non-lexical vowels are plotted as open circles.

Consistent with Kiliç and Öğüt's (2004) report that /ɯ/ is more mid/central than other Turkish back vowels, /ɯ/'s F1 and F2 values are intermediate between those for /i/ and for /u/. Most non-lexical vowels (open circles) lie within the distribution of lexical /ɯ/ (purple circles and ellipse), and few of them lie within the distribution of lexical /i/ (red squares and ellipse). Acoustically speaking, then, harmony does not seem to have applied to the onset-repairing vowels in a categorical, consistent way, contra previous descriptions. These differences between lexical and non-lexical vowels are investigated below.

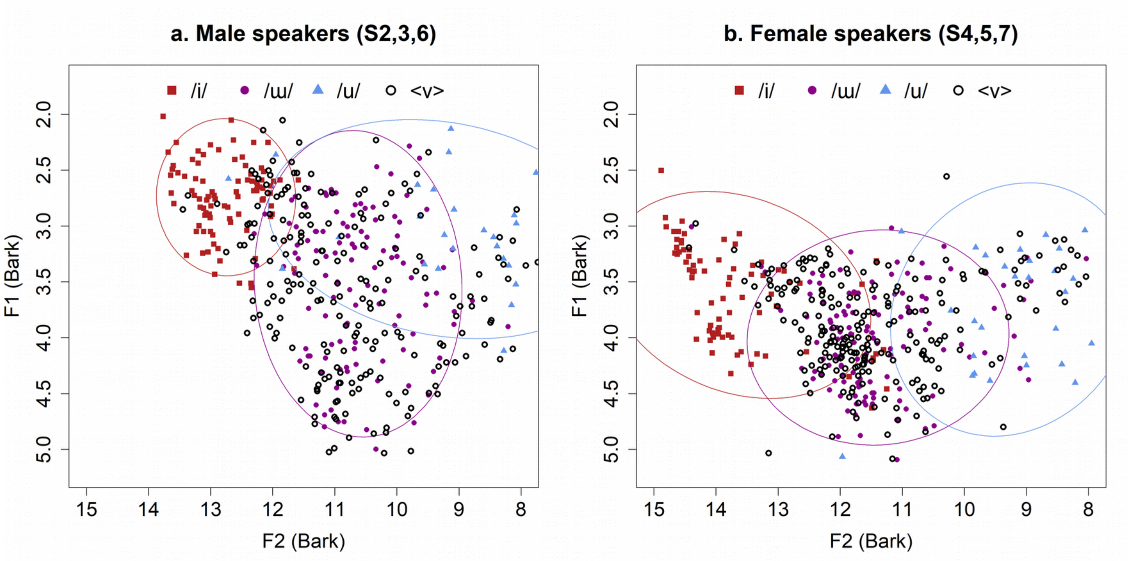


Figure 3.3. Vowel plots.

3.3.3 Acoustic differences between lexical and non-lexical vowels

If the non-lexical vowels are intrusive, as hypothesized here, they will lack the durational and gestural targets associated with true vowels, and are predicted to be shorter and more subject to coarticulation than lexical vowels. According to the standard epenthetic theory of onset-cluster repair in Turkish, non-lexical vowels are subject to backness and rounding harmony. For purposes of testing that hypothesis, I treat the non-lexical vowels accordingly: as <i> before /i/, <ɯ> before /a/, and <u> before /o/ (cf. C&S, Yildiz 2001).

Linear mixed effects models of duration, F1 and F2 were computed using R (R Core Team 2016) and lmerTest (Kuznetsova et al. 2012). All models included fixed effects for the preceding consonant, the category of the following vowel (V_2), and the hypothesized category of v_1, as well as random intercepts for subject and item. Models representing the intrusive hypothesis additionally included the lexical status of v_1 (whether the word underlyingly begins with /Cr/ or /CVr/), along with one or more interactions. Since the epenthetic hypothesis predicts no acoustic differences between lexical and non-lexical vowels, the models representing this hypothesis did not include v_1's lexical status as a factor. Models are shown in the Appendix at the end of this dissertation.

3.3.3.1 Duration

Three separate measures of duration were analyzed: the duration of the whole interconsonantal interval, the vocalic portion of the ICI, and the burst combined with any additional positive VOT. For all measures, models that included v_1's lexical status performed better than models that did not in maximum likelihood ratio tests (all *ps* < 0.001).

Duration of the interconsonantal interval (ICI): In the best model of ICI duration (`duration.ICI.model`), lexical status had a significant main effect, with the ICI being significantly shorter in non-lexical vowels than lexical vowels ($\beta = -9.01$, SE = 2.74, $p < 0.005$). There was also a significant interaction between lexical status and v_1 category—<i> and <u> are shorter than their lexical counterparts by an additional 12.19ms (SE = 3.5, $p < 0.005$) and 10.97 ms (SE = 4.39, $p < 0.05$) respectively. The interaction is visualized in Figure

3.4a, where each line is a different v_1 category. In essence, lexical /i u ɯ/ are more distinct from each other than the three categories of non-lexical vowel.

In addition, the ICI is longer before /i/ than before /a/ (β = 11.20, SE = 3.01, $p < 0.005$), perhaps reflecting a trade-off effect where a longer V_2 like /a/ results in a shorter v_1. ICI duration is also significantly shorter after /b/ than /d/ (β = -4.97 ms, SE = 1.9, $p < 0.05$) and longer after /g/ than /d/ (β = 10.27 ms, SE = 1.85, $p < 0.001$), as predicted by previous work on place effects on VOT (Cho & Ladefoged 1999) and gestural coordination (Yip 2013).

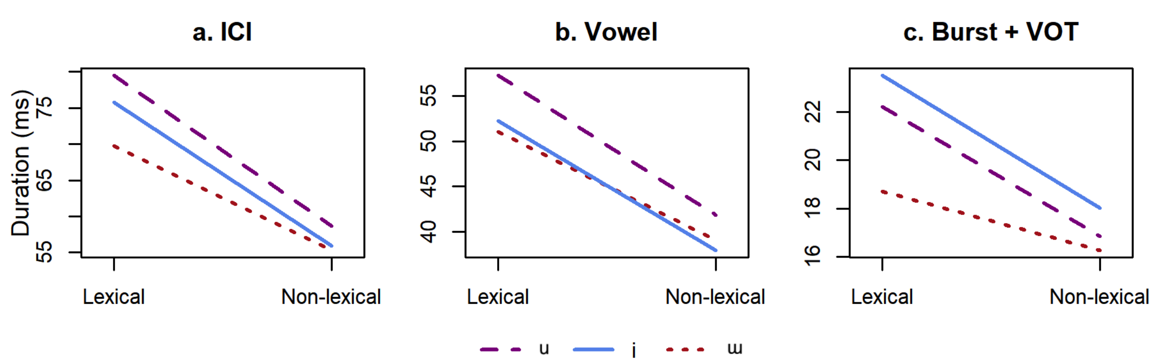


Figure 3.4: Interaction of lexical status and v_1 category for three different measures of duration.

Duration of v_1: Analysis of v_1, the portion of the ICI that has high amplitude periodicity with formant structure, again found significant main and interaction effects of lexical status (`duration.vowel.model`). In non-lexical vowels, the vocalic interlude is shorter (β = -7.98, SE = 2.09, p = 0.001). As illustrated in Figure 3.4b, non-lexical <i> and <u> are particularly short again (β = -8.38, SE = 2.70, p < 0.01; β = -7.09, SE = 3.37, p < 0.05). Also, like the ICI, the vowel is longer before /i/ than /a/ (β = 9.20, SE = 2.31, p < 0.001), as well as slightly longer after /g/ (β = 4.54 ms, SE = 1.41, p < 0.005) than after /d/.

Duration of burst+VOT: In the analysis of the duration of the consonant burst plus any additional positive VOT (`duration.burst.model`), there was a significant main effect of lexical status (β = -3.89, SE = 1.02, p < 0.001). Unsurprisingly, burst durations were shorter for /b/ (β = -4.80, SE = 1.25, p < 0.001) and longer for /g/ (β = 5.52, SE = 1.24, p < 0.005) than /d/. In addition, burst+VOT was longer in /i/ than /ɯ/ (β = 3.60, SE = 1.42, p < 0.05), perhaps due to the tongue's higher position in /i/. VOT is known to be larger before high vowels than mid or low vowels (Klatt 1975). The interaction of lexical status and v_1 category did not reach significance, but is shown in Figure 3.4c for comparison to the other duration measures.

3.3.3.2 F1

The hypothesis that non-lexical vowels are intrusive also predicts differences in their formant values. Formant values were measured at the midpoint of the high amplitude portion of the ICI with periodicity and formant structure. The best model of F1 (`harmony.F1.model1`) included an interaction between lexical status and hypothesized v_1 category, and outperformed the epenthetic model that excluded lexical status ($\chi^2(3) = 25.6$, $p < 0.001$). The interaction between lexical status and hypothesized v_1 category was significant ($\beta = 20.05$, SE = 9.55, $p < 0.05$). The higher F1 of non-lexical <i> suggests that it lacks lexical /i/'s [+high] target.

The model shows the expected main effects of surrounding context: F1 is significantly lower for /i/ (β = -56.47 Hz, SE = 6.72, $p < 0.001$) than for /ɯ/, which is known to be lower than /i/ in Turkish (Kiliç & Öğüt 2004); and lower after /g/ than /d/ (β = -18.43, SE = 5.01, $p < 0.005$). The effect of the following vowel was also significant: F1 is lower when /i/ follows (β = -18.35, SE = 8.17, $p < 0.05$) and when /o/ follows (β = -38.43, SE = 9.63, $p < 0.001$), compared to /a/.

3.3.3.3 F2

The best model of F2 (`harmony.F2.model1`) also included an interaction between lexical status and hypothesized v_1 category, and it outperformed the epenthetic model that excludes lexical status in maximum likelihood ratio test ($\chi^2(3) = 32.58$, $p < 0.001$). There was a significant interaction between lexical status and v_1 category, showing that F2 is lower for non-lexical <i> than for underlying /i/ (β = -320.52 Hz, SE = 75.33, $p < 0.001$). This suggests that non-lexical <i> is backer than lexical /i/, perhaps lacking /i/'s [+front] target.

As expected, though less relevant to the hypotheses of this dissertation, the model also shows that F2 is higher in /i/ than /ɯ/ (β = 393.70 Hz, SE = 42.44, $p < 0.0001$) and lower in /u/ than /ɯ/ (β = -202.93 Hz, SE = 76.70, $p < 0.05$), and that a following /i/ raises F2 (β = 162.94, SE = 62.35, $p < 0.05$).

3.3.4 Assuming frontness harmony but no rounding harmony

The models above found significant differences between non-lexical vowels and harmonic lexical vowels in the same context. However, as discussed above, some previous experiments suggest that rounding harmony in onset cluster repair may only be triggered by high vowels (Yavas 1980, Kaun 1999). To take this possibility into account, non-lexical vowels were recoded as <i> before /i/ and <ɯ> before /a/ and /o/. Modeling of F1 and F2 under these assumptions recapitulated the effects described above, with non-lexical <i> having a higher F1 (`harmony.F1.model2`: β = 21.03, SE = 7.56, $p < 0.001$) and lower F2 than /i/ (`harmony.F2.model2`: β = -248.63, SE = 70.49, $p < 0.005$), as well as the expected effects of preceding consonant and following vowel.

3.3.5 Assuming no harmony

Logically, the differences between lexical and non-lexical vowels reported above could also result from epenthesis applying but harmony not applying, in which case all epenthetic vowels would be [ɯ]. To address this possibility, non-lexical vowels were again recoded, this time treating all of them as <ɯ>, and the analyses above were repeated. Data was subsetted to exclude /i/ and /u/, since these were no longer relevant.

3.3.5.1 F1

The best model of F1 assuming no harmony (`noharmony.F1.model`) includes an interaction between lexical status and V_2. This model was significantly better than the epenthetic model that did not include lexical status as a factor ($\chi^2(3) = 16.12$, $p < 0.005$). F1 is lower in non-lexical vowels followed by /i/ ($\beta = -34.04$, SE = 9.42, $p < 0.005$). This interaction effect suggests greater anticipatory coarticulation in the non-lexical vowel, since a following /i/ lowers F1 in non-lexical vowels more than in lexical /ɯ/.

Less relevantly, the model also shows main effects of a preceding /g/ ($\beta = -20.49$, SE = 5.30, $p < 0.005$), following /i/ ($\beta = -17.95$, SE = 7.46, $p < 0.05$), and following /o/ ($\beta = -38.79$, SE = 8.84, $p < 0.001$).

3.3.5.2 F2

The best harmony-free model of F2 (`noharmony.F2.model`) performed better than the epenthetic model in a maximum likelihood test ($\chi^2(3) = 20.16$, $p < 0.001$), and includes an interaction between lexical status and place of the following vowel. Before /o/, non-lexical vowels have a lower F2 than lexical vowels (β = -158.13, SE = 49.34, p < 0.01) – further evidence that non-lexical vowels are more affected by anticipatory coarticulation. The model also shows the expected main effects from vowel and consonant context.

3.3.6 Summary of acoustic differences

Model comparison found that the lexical status of v_1 significantly improved model performance for duration, F1 and F2. Non-lexical vowels are shorter than their underlying counterparts, a result predicted if non-lexical vowels are not true vowels, only the acoustic consequence of an open transition between consonant gestures, which has no durational or acoustic target. In addition, non-lexical vowels are acoustically intermediate between the harmonizing vowels /i/ and /u/ and the non-harmonizing /ɯ/ in their F1 (Figure 3.5) and F2 (Figure 3.6). This means that acoustic differences between lexical and non-lexical vowels are found regardless of whether harmony is assumed to have applied.

The differences between lexical and non-lexical vowels are particularly clear before /i/, where non-lexical vowels had higher F1 and lower F2 than lexical /i/, but lower F1 and higher F2 than lexical /ɯ/ (Figure 3.5a, Figure 3.6a). This suggests non-lexical vowels are more centralized than /i/ but also more affected by anticipatory raising and fronting for the following /i/ than lexical /ɯ/ is. Likewise, non-lexical vowels before /o/ had a significantly lower

F2 than lexical /ɯ/ (Figure 3.6c). This suggests that non-lexical vowels are more affected by anticipatory rounding for /o/. A lowered F2 is also compatible with increased coarticulatory backing before /o/, although rounding is the more salient difference between /o/ and /ɯ/, since both are phonologically [+back]. These observations are compatible with the hypothesis that the non-lexical vowels are targetless.

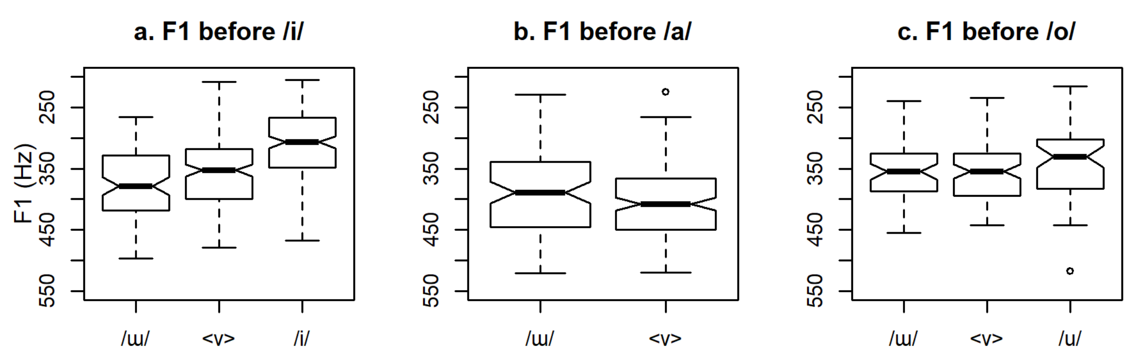


Figure 3.5. Effect of lexical status on F1.

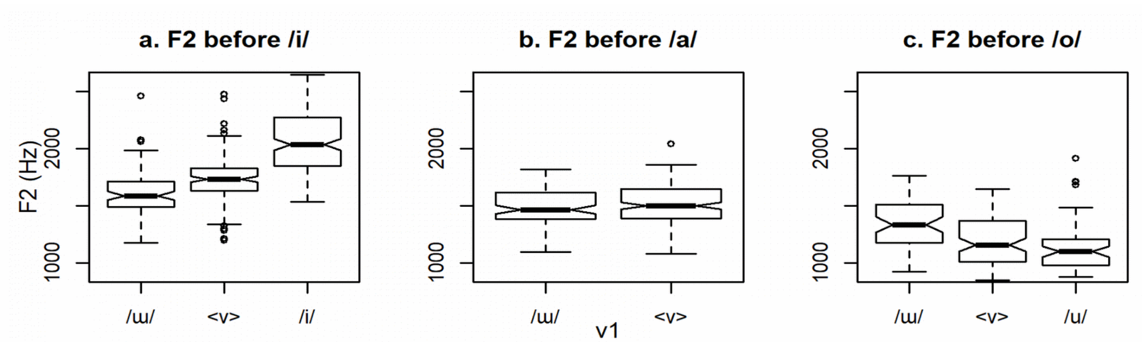


Figure 3.6. Effect of lexical status on F2.

3.4 Discussion

This chapter has presented acoustic evidence that the non-lexical vowels in underlying onset clusters in Turkish result from gradient, gestural intrusion, and consequently lack durational and gestural targets. In the production experiment, 88% of underlying onset clusters in Turkish are produced with an acoustic inserted vowel. Though this high rate of acoustic insertion is contrary to C&S (1982)'s report that vowel insertion does not occur in careful or formal speech, it is consistent with the overall landscape of data on Turkish onset cluster repair, since all other work on this topic reports plenty of insertion in laboratory speech, and does not mention an effect of speech style (Yavaş 1980, Kaun 1999, Yıldız 2010, Bokhari et al. 2016). The ICI in words with underlying clusters has a unimodal duration, as predicted by a gestural account, where the duration of the ICI is determined by the degree of gestural overlap, not by an optional categorical insertion with an accompanying durational target. This indicates that apparent insertion is a gradient process that can “apply” to a range of degrees, not a categorical but optional process as previously described.

The quality of non-lexical vowels is also gradiently determined by the surrounding gestural context. Non-lexical vowels are acoustically intermediate between harmonizing and non-harmonizing lexical vowels, with F1 and F2 differences being most significant before /i/. Generally, the non-lexical vowels are /ɯ/-like in careful speech. These acoustic differences show that the vowels appearing in onset clusters are definitely not participating in backness harmony. This implies they are not participating in rounding harmony, either, although the acoustic differences between lexical /u/ and the non-lexical vowels did not reach significance. The next chapter (Ch. 4) will show that non-lexical vowels take on somewhat more of the backness of the following vowel in casual speech, where gestural overlap increases, although even in casual speech, their F1 and F2 values still differ significantly from lexical vowels'.

To summarize, vowel intrusion does not completely neutralize the distinction between /CC/ and /CVC/ in Turkish. Rather, the non-lexical vowels in Turkish onset clusters are shorter than lexical vowels; are more affected by the surrounding context; and do not participate in vowel harmony. Moreover, the Turkish lexicon lacks the structures that would be created if the intrusive vowel were taken to be true inserted [ɯ] (i.e., forms containing underlying disharmonic sequences [ɯ i] and [ɯ o] – see Section 2.3.1), suggesting that the Turkish grammar actually rules out such sequences. These observations argue that non-lexical vowels in Turkish onset clusters are intrusive vowels. Lacking their own gestural targets, the acoustics of these intrusive vowels are determined by their context. There is no insertion of a vowel gesture even in clusters that are produced with an acoustic vowel between the two consonants; instead, this intrusive vowel represents a period when the closure of the first consonant has been released but the closure of the second consonant has not been completed. Meanwhile, the tongue body is already moving toward the following vowel's target, such that the formant values during the ICI are shaped by that target.

This interpretation is readily represented in an Articulatory Phonology (Browman and Goldstein 1993) framework, where gestures within a syllable are coordinated together in time. This kind of gestural coordination can be represented in the grammar, as in Gafos (2002) and Hall (2003). For example, the gestural coordination that produces vowel intrusion in Turkish onset clusters could be modeled with a constraint aligning the onset of C_2 with the release of C_1, or with an anti-phase relation between C_1 and C_2 with little to no competition from a C_1—V in-phase relation. Chapter 7 elaborates on the possible phasing relations of the gestures in Turkish /CC/ words.

3.4.1 Interspeaker variation and cross-linguistic implications

Given that all onset clusters in Turkish come from loanwords, an anonymous reviewer of Bellik (2018) asked whether it might be the case that the Turkish phonological grammar originally prohibited onset clusters and repaired them with epenthesis, and has changed (or is changing) to permit onset clusters, even if they are realized with an open transition between the consonants. We can also consider the possibility that the transition went in the opposite direction, from initially attempting to produce borrowed onset clusters with a foreign-like gestural coordination, to later producing intrusive vowels, and finally toward reanalyzing the intrusive vowels as epenthetic and integrating the loanwords into the native phonological grammar, which prohibits onset clusters.

I would like to propose that both scenarios played out in different segments of the population. It seems likely that there has always been variation in Turkish speakers' realization and representation of onset clusters in loanwords, based on individuals' degree of exposure to the source languages. Post-hoc examination of the inter-speaker variation in this experiment provides tentative support for this: synchronically, the degree of exposure to languages with onset clusters seemed to predict the degree to which clusters contrasted with /CVC/ sequences[2]. Speakers roughly fall into three groups: categorical differentiators, gradient differentiators, and neutralizers, echoing the pattern in Hall (2013).

First, speakers who are experienced with languages like French or English are likely to be aware that <CC> spellings represent underlying clusters, and to succeed in producing the borrowed words with a French- or English-like gestural timing. In this study, S3 and S6 had early exposure to languages with onset clusters, and insert acoustic vowels less frequently than the other speakers (S2,

2 Although see Zsiga (2011) for a case where L2 English proficiency does not predict the degree of transfer from L1, with Korean speakers applying word-final nasalization to stops in English.

4, 5, 7). These early bilinguals also have bimodal distributions of ICI durations. Diachronically, when these /CC/ loanwords originally entered Turkish from prestige languages like French, they were probably used by bilinguals who were highly conscious of their status as loanwords, and (variably) able to achieve a foreign-like gestural coordination. These bilinguals may even have been code-switching, and would have been aware that the borrowings began with /CC/, not /CVC/.

Second, hearing /CC/ loanwords in the speech of bilinguals, other speakers with less foreign language experience may recognize the underlying clusters, but fail to achieve a foreign-like gestural timing. This situation is comparable to English speakers producing illegal onset clusters in Shaw & Davidson (2011). This would produce the gradient differentiation of non-lexical and lexical vowels found in the experiment as a whole, and exemplified in the data of S5 and S7, as well as S2[3]. These speakers could also be adapting an existing gestural coordination relation and its accompanying motor plan, perhaps one that governs the timing of onsets of adjacent syllables. Loanword phonology seeks to adapt a loanword to the existing phonological structures of

3 For S2, unlike S5 and S7, lexical status is not a significant predictor of vowel duration. For S2, lexical vowels had a mean duration of 39.55ms, and non-lexical vowels had a mean duration of 37.11ms. The difference was not significant ($t(462.03) = 1.34$, $p = 0.18$). But S2 does exhibit F1 and F2 differences between lexical and non-lexical vowels. Intrusive models outperform epenthetic models, whether harmony is assumed (F1: $\chi^2(3) = 40.62$, $p < 0.001$. F2: $\chi^2(3) = 42.04$, $p < 0.001$) or not (F1: $\chi^2(3) = 12.2$, $p < 0.01$. F2: $\chi^2(3) = 7.99$, $p < 0.05$).

the borrowing language, recycling native phonological processes to do so. We might expect a similar strategy of recycling native patterns at the level of gestural coordination as well. Cross-linguistically, this predicts that a language transitioning from a simpler syllable structure to a more complex syllable structure would exhibit vowel intrusion in the course of the transition, as speakers articulate complex new syllables by repurposing a limited set of existing gestural coordination plans that result in low overlap between consonant gestures.

Third, the presence of acoustic intrusive vowels in some tokens of complex onsets could result in some speakers reanalyzing the borrowed words as /CVC/. This occurred in the transcription task in Davidson (2007), where listeners sometimes transcribed [CəC] (containing a transitional schwa) as CVC. In Turkish, such reanalysis may be the source of orthographic alternations like *stil* ~ *sitil* 'style' and *klup* ~ *kulup* 'club'. Walter (2018) also chronicles some loans which entered Ottoman Turkish and today are so integrated that they are spelled CVC instead of CC (e.g., *pırasa* 'leek' from Greek *prason*). Today, even Turkish monolinguals commonly use /CC/ loanwords, and might be expected to reanalyze intrusive vowels as underlying vowels. Anecdotally, Turkish children who are learning to write tend to write the intrusive vowel in onset clusters, and must be taught not to; this suggests that they are reanalyzing the words as starting with /CVC/ rather than /CC/. The prescriptive spelling of onset clusters in loanwords without an orthographic vowel probably works to maintain their representation as complex onsets – a hypothesis that should be investigated in future research.

Listeners who interpret the acoustic vowels they have heard as underlying vowels would not differentiate lexical and non-lexical vowels in their speech, either. This appears to describe the one monolingual speaker in this study, S4, and, to a lesser extent, her husband S2. Both S2 and S4 are from a smaller town in the province of Antalya, and exhibit a higher rate of acoustic insertion than speakers of the “standard”, urban/Istanbul dialect (S3, 5, 6, 7) (Figure 3.2), as well as non-lexical vowels that are not significantly shorter than lexical vowels. While S2 exhibits F1 and F2 differences between lexical and non-lexical vowels (see Footnote 3), S4's non-lexical vowels do not differ significantly from lexical /ɯ/ in duration, F1, or F2.[4] This suggests that S4 may have reanalyzed the vowels in onset clusters as underlying /ɯ/. However, it cannot be that all S4’s vowels in onset clusters result from reanalyzing those vowels as part of the underlying representations of familiar words, because even S4 also produced acoustic vowels in novel nonce forms. That is, the insertion process generalizes

4 For S4, the mean duration of lexical vowels was 48.34ms, and the mean duration of non-lexical vowels was 47.47ms. This difference was not significant in a t-test ($t(472.82) = 0.34$, $p > 0.5$). Intrusive models of the duration and F1 of S4’s vowels were also not significantly better than epenthetic models ($p > 0.05$), whether or not harmony was assumed. For F2, if harmony is assumed, lexical status is a significant predictor ($\chi^2(3) = 9.88$, $p = 0.02$), but if no harmony is assumed, it is not ($\chi2(3) = 2.84$, $p = 0.42$).

beyond the conventionalized forms, even for speakers who do not acoustically differentiate lexical and non-lexical vowels. If onset cluster repair is not intrusion for S4, it must involve epenthesis, not only reanalysis.

To summarize, there is considerable interspeaker variation even in this small sample, which I suggest reflects variation among different speakers' grammars of gestural alignment. Some speakers apparently allow complex onsets and often achieve a close /CC/ coordination that does not produce an intrusive vowel. Other speakers also seem to allow complex onsets but employ a different gestural timing with less /CC/ overlap, resulting in a gradient distinction between the lexical and non-lexical vowels. Finally, some speakers do not differentiate lexical and non-lexical vowels; their grammars employ a /CC/ coordination with even less overlap, possibly because they still prohibit complex onsets and require epenthesis.

These three production strategies – categorical differentiation, gradient differentiation, and complete neutralization – correspond to the three strategies employed by different speakers in producing epenthetic vowels in Lebanese Arabic. Some speakers differentiate categorically from lexical vowels, some gradiently, and some not at all (Gouskova & Hall 2009, Hall 2013). However, the Lebanese speakers in Hall (2013) who differentiate epenthetic and lexical vowels did so consistently across items and repetitions, which is not the case for speakers in this study. Also, in Lebanese Arabic, the interspeaker variation is unlikely to be tied to proficiency in a foreign language, since the epenthesis is not occurring in loanwords. One plausible reason for these differences between Lebanese Arabic and Turkish is that the insertion processes are phonologized to different degrees in the different languages. Vowel insertion in Lebanese is true phonological epenthesis, with interspeaker variation either in the degree of neutralization (Gouskova & Hall 2009), or in the degree of cross-dialect influence (Hall 2013). Each speaker of Lebanese Arabic realizes their epenthetic vowels in a predictable way.

In contrast, vowel insertion in Turkish is intrusion, produced by a gestural alignment that may be phonologized to different degrees for different speakers. It is not the case that each Turkish speaker realizes consonant clusters in a consistent way. The acoustic variation within speakers could reflect gradience and ambiguity in speakers' mental representations, as in Gradient Symbolic Computation (Smolensky et al. 2013, Smolensky & Goldrick 2016). Mental representations could be ambiguous between /CC/ and /CVC/, or could be solidly /CC/ but ambiguous as to the specific gestural coordination between the consonant gestures. Alternately, within-speaker variation could reflect failure to consistently achieve a targeted coordination, or other phonetic factors like speech rate. These conclusions are necessarily tentative, however, since the number of speakers here is so small. A future investigation of the factors that shape this intra- and inter-speaker variability could shed additional light on the mental representation of onset clusters for Turkish speakers, with possible implications for our understanding of loanword adaptation and its diachronic stability.

Furthermore, a perceptual study of Turkish onset cluster repair could clarify whether Turkish speakers, particularly monolinguals, are able to distinguish lexical and non-lexical vowels. If Turkish speakers use the acoustic differences to identify intrusive vowels in complex onsets, that would be a point in favor of an analysis where the gestural coordination that produces vowel intrusion is in fact grammaticized in Turkish, and maintained through perceptual cues. A perceptual study would also shed light on the ways in which factors like language-specific phonetic knowledge and the acoustic similarity of the stimuli, which have been shown to affect English speakers' perception of vowel intrusion (Davidson 2007, Davidson & Shaw 2012), also predict cross-linguistic perception of illegal consonant sequences.

Appendix to Ch. 3: Models of Careful speech

This appendix presents the fixed effects coefficients tables for the models referred to in Ch. 3.

A. Duration

In duration models, non-lexical vowels are coded as <ɯ> before /a/, <i> before /i/, and <u> before /o/.

duration.ICI.model = Duration of ICI, assuming harmony

Formula: vdur_ici ~ /Cr/ * v1 + c + v2 + (1 | subj) + (1 | target_word)

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	62.275	3.171	19.660	19.638	2.20e-14	***
/Cr/=TRUE	-9.102	2.740	21.060	-3.322	0.00323	**
v1=[i]	4.804	2.460	38.420	1.953	0.05815	.
v1=[u]	6.847	3.830	25.370	1.788	0.08580	.
C1=/b/	-4.971	1.866	22.050	-2.664	0.01415	*
C1=/g/	10.273	1.853	21.750	5.545	1.48e-05	***
V2=/i/	11.201	3.013	23.680	3.717	0.00109	**
V2=/o/	3.164	3.562	21.380	0.888	0.38442	
/Cr/=TRUE:v1=[i]	-12.193	3.530	22.030	-3.454	0.00226	**
/Cr/=TRUE:v1=[u]	-10.977	4.391	23.300	-2.500	0.01990	*

duration.vowel.model = Duration of vowel, assuming harmony

Formula: vdur ~ Cr * v1 + c + v2 + (1 | subj) + (1 | target_word)

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	45.286	2.918	12.490	15.520	1.56e-09	***
/Cr/=TRUE	-7.983	2.088	20.840	-3.823	0.001001	**
v1=[i]	1.883	1.935	33.010	0.973	0.337622	
v1=[u]	3.164	2.952	24.510	1.072	0.294264	
C1=/b/	-0.236	1.426	21.980	-0.165	0.870130	
C1=/g/	4.541	1.415	21.550	3.209	0.004117	**
V2=/i/	9.195	2.312	22.750	3.977	0.000606	***
V2=/o/	1.080	2.716	20.870	0.398	0.694940	
/Cr/=TRUE:v1=[i]	-8.379	2.698	21.710	-3.106	0.005212	**
/Cr/=TRUE:v1=[u]	-7.093	3.370	23.410	-2.105	0.046264	*

duration.burst.model = Duration of burst + positive VOT, assuming harmony

```
Formula: vdur_burst ~ Cr + c + v1 + v2 + (1 | subj) + (1 | target_word)
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	19.0594	2.2717	10.5200	8.390	5.57e-06	***
/Cr/=TRUE	-3.8869	1.0230	24.1400	-3.800	0.000866	***
C1=/b/	-4.8037	1.2469	22.6000	-3.852	0.000831	***
C1=/g/	5.5225	1.2394	22.6000	4.456	0.000187	***
v1=[i]	3.5997	1.4184	59.9300	2.538	0.013769	*
v1=[u]	2.7034	2.1611	27.6000	1.251	0.221453	
V2=/i/	-0.8843	1.6381	34.7700	-0.540	0.592774	
V2=/o/	0.3989	2.1423	24.6800	0.186	0.853793	

B. F1 and F2, assuming backness and rounding harmony

In the first set of F1 and F2 models, non-lexical vowels are coded as <ɯ> before /a/, <i>before /i/, and <u> before /o/.

harmony.F1.model1 = F1 of vowel, assuming backness and rounding harmony

```
Formula: vf1 ~ Cr * v1 + v2 + c + (1 | subj) + (1 | target_word)
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	398.771	21.301	6.100	18.721	1.29e-06	***
/Cr/=TRUE	10.210	7.408	20.230	1.378	0.183200	
v1=[i]	-56.473	6.717	35.360	-8.407	5.96e-10	***
v1=[u]	-11.131	10.391	24.180	-1.071	0.294632	
V2=/i/	-18.346	8.165	22.560	-2.247	0.034732	*
V2=/o/	-38.430	9.634	20.460	-3.989	0.000696	***
C1=/b/	-4.417	5.049	21.230	-0.875	0.391477	
C1=/g/	-18.425	5.012	20.900	-3.676	0.001415	**
/Cr/=TRUE:v1=[i]	20.050	9.553	21.140	2.099	0.048016	*
/Cr/=TRUE:v1=[u]	-0.844	11.896	22.480	-0.071	0.944068	

harmony.F2.model1 = F2 of vowel, assuming backness and rounding harmony

```
Formula: vf2 ~ Cr * v1 + v2 + c + (1 | subj) + (1 | target_word)
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	1477.53	71.56	18.96	20.646	1.87e-14	***
/Cr/=TRUE	11.75	59.35	21.83	0.198	0.844837	
v1=[i]	393.70	42.44	99.48	9.277	4.00e-15	***
v1=[u]	-202.93	76.70	34.42	-2.646	0.012202	*
V2=/i/	162.94	62.35	26.55	2.613	0.014575	*
V2=/o/	-141.07	76.11	23.79	-1.854	0.076240	.
C1=/b/	-70.54	39.96	22.64	-1.765	0.091022	.
C1=/g/	81.68	39.72	22.79	2.057	0.051333	.
/Cr/=TRUE:v1=[i]	-320.52	75.33	23.09	-4.255	0.000296	***
/Cr/=TRUE:v1=[u]	14.57	92.78	23.81	0.157	0.876536	

C. F1 and F2, assuming backness harmony

In the second set of F1 and F2 models, non-lexical vowels are coded as <ɯ> before /a/ and /o/, and as <i> before /i/.

harmony.F1.model2 = F1 of vowel, assuming backness harmony but not rounding harmony

```
Formula: vf1 ~ Cr * v1 + v2 + c + (1 | subj) + (1 | target_word)
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	399.489	21.146	5.660	18.892	2.49e-06	***
/Cr/=TRUE	8.934	5.274	21.862	1.694	0.104473	
v1=[i]	-55.827	6.130	31.303	-9.108	2.60e-10	***
V2=/i/	-19.450	6.835	22.797	-2.846	0.009207	**
V2=/o/	-39.261	5.523	25.429	-7.109	1.71e-07	***
C1=/b/	-2.754	4.571	21.130	-0.603	0.553246	
C1=/g/	-20.085	4.530	20.647	-4.434	0.000238	***
/Cr/=TRUE:v1=[i]	21.028	7.556	21.925	2.783	0.010865	*

harmony.F2.model2 = F2 of vowel, assuming backness harmony but not rounding harmony

```
Formula: vf2 ~ Cr * v1 + v2 + c + (1 | subj) + (1 | target_word)
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	1522.80	73.96	15.67	20.589	9.16e-13	***
/Cr/=TRUE	-55.87	50.08	22.13	-1.116	0.27656	
v1=[i]	378.82	42.38	132.05	8.938	3.11e-15	***
V2=/i/	126.67	60.75	28.05	2.085	0.04629	*
V2=/o/	-249.10	50.59	23.03	-4.924	5.61e-05	***
C1=/b/	-74.78	43.45	22.66	-1.721	0.09888	.
C1=/g/	85.18	43.13	22.92	1.975	0.06045	.
/Cr/=TRUE:v1=[i]	-248.63	70.49	23.65	-3.527	0.00175	**

D. F1 and F2, assuming no harmony

In these models, all non-lexical vowels are coded as <ɯ>, regardless of V2.

noharmony.F1.model = F1 of vowel, assuming all non-lexical vowels are [ɯ] (no harmony)

Formula: vf1 ~ Cr * v2 + c + (1 | subj) + (1 | target_word)

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	397.074	22.277	5.841	17.825	2.58e-06	***
/Cr/=TRUE	10.784	6.772	12.049	1.593	0.137148	
V2=/i/	-17.954	7.459	12.988	-2.407	0.031689	*
V2=/o/	-38.790	8.840	11.714	-4.388	0.000935	***
C1=/b/	3.222	5.446	14.724	0.592	0.563091	
C1=/g/	-20.487	5.303	14.091	-3.863	0.001703	**
/Cr/=TRUE:V2=/i/	-34.037	9.415	13.528	-3.615	0.002959	**
/Cr/=TRUE:V2=/o/	-4.518	11.187	13.978	-0.404	0.692435	

noharmony.F2.model = F2 of vowel, assuming all non-lexical vowels are [ɯ] (no harmony)

Formula: vf2 ~ Cr * v2 + c + (1 | subj) + (1 | target_word)

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	1525.806	53.215	8.989	28.672	3.80e-10	***
/Cr/=TRUE	16.697	30.244	13.469	0.552	0.58994	
V2=/i/	122.929	32.997	15.074	3.725	0.00202	**
V2=/o/	-110.282	39.590	13.157	-2.786	0.01530	*
C1=/b/	-176.019	23.878	15.991	-7.372	1.58e-06	***
C1=/g/	41.953	23.317	15.766	1.799	0.09115	.
/Cr/=TRUE:V2=/i/	71.535	41.550	15.211	1.722	0.10540	
/Cr/=TRUE:V2=/o/	-158.133	49.339	14.926	-3.205	0.00593	**

Chapter 4: Acoustic effects of speech style

4.1 Casual speech and gestural overlap

Gestural overlap is expected to vary based on speech style. In casual, hypoarticulated speech, gestures overlap more, producing greater coarticulatory effects compared to careful, hyperarticulated speech, where gestures tend to pull apart (Browman & Goldstein 1990, Lindblom 1983, Gay 1981). If onset cluster repair is driven by gestural timing relations, then we would expect it to behave differently in casual speech (more gestural overlap) than in careful speech (less gestural overlap). One possibility is that the gestures for the two consonants will overlap more, changing the open C-C transition to a more closed one. Sufficient C-C overlap could eliminate the intrusive vowel entirely. This seems unlikely for Turkish onset cluster repair, however, since casual speech reportedly increases the incidence of vowel insertion (Clements & Sezer 1982). Such a high degree of C-C overlap could be blocked by phonological constraints on gestural alignment or phasing. In the terms of Gafos (2002) or Hall (2003), the language might have a highly ranked constraint that aligns the release of C_1 with the onset of C_2, preventing C_2 from attaining closure before C_1's closure is released. To say the same thing in the terms of the coupled oscillator model of syllable structure (Goldstein et al. 2007, Goldstein et al. 2008), C_1 and C_2 might have an anti-phase coupling relationship (as expected for consonants in a complex onset), without the balancing pull of both C_1 and C_2 being coupled in-phase to V (which produces the C-center effect in English-type complex onsets). That is, C_1 might function as an extrasyllabic consonant. Alternately, C_1 might be syllabified as part of a complex onset, but with a lower coupling strength between C_1 and V than is seen in, e.g., English. Either of these coupling graphs would result in C_1 being pushed away from C_2, even in fast speech, thus maintaining an open transition between consonants, and leaving room for an intrusive vowel. It is also possible that the strength of the C_1-V coupling could fluctuate with speech style, such that casual speech reflects a more native-like coordination pattern, while careful speech includes coordination patterns that are restricted to the foreign stratum of the lexicon (Ito & Mester 1995), analogous to the speech-rate driven shifts in Dutch rhythmic structure proposed by Schreuder & Gilbers (2004).

If casual speech causes a less extreme increase in C-C overlap, this could result in an intrusive vowel whose acoustics are more strongly influenced by C_1, since its offset would overlap a greater proportion of the ICI. In general, the surrounding consonants and vowels shape the acoustic and gestural characteristics of intrusive vowels to a greater extent than they shape targeted, lexical vowels. This predicts that the preceding consonant would have a stronger effect on non-lexical vowels than on lexical vowels. This prediction was not borne out in the acoustic study of careful speech (Ch. 3); however, in casual speech, greater C-C overlap could allow an interaction between preceding consonant and the vowel's lexical status to become statistically significant where it was not previously.

Another possible consequence of greater gestural overlap is that the gesture for the following lexical vowel will overlap the ICI more, causing the intrusive vowel to take on more of the following vowel's backness and rounding than it does in careful speech. Unlike increased C_1-C_2 overlap, greater overlap of V with C_1 and/or C_2 in a C_1C_2V sequence is unlikely to be ruled out by constraints on gestural alignment/phasing, because such constraints prefer a closer alignment of onset consonants with the vowel nucleus. In the coupled oscillator model, onset consonant gestures are coupled in-phase with the gesture for the nucleus, meaning that the coupling graph prefers for onset consonant gestures and nucleus gestures to begin simultaneously. This is, in fact, proposed to be the key difference between onset and coda consonants. Coda consonants are said to be locally coupled anti-phase with nucleus or with a preceding coda consonant, meaning a nucleus-adjacent coda will begin as the vowel gesture ends, and subsequent coda consonants will begin as their immediate predecessors end (Goldstein et al. 2008, O'Dell & Nieminen 2008).

In the studies of careful speech presented in Ch. 2-3, the hypothesized targetless intrusive vowels before /i/ were found to be intermediate between a targeted front vowel /i/ and a targeted back vowel /ɯ/, both in terms of tongue body position (Ch. 2) and F2 values (Ch. 3). In casual speech, the gesture for the following /i/ may overlap with the ICI more, with the result that <i> would be more /i/-like in casual speech than careful speech. Similarly, non-lexical vowels before /o/ may have a lowered F2 in casual speech, if the lip rounding gesture for /o/ overlaps the ICI more, and non-lexical vowels before /a/ may have an increased F1, if the gesture for the low vowel /a/ begins earlier in casual speech than in careful speech.

If casual speech causes greater overlap of V_2 with the ICI, it might explain previous claims about the quality of the inserted vowel. According to Clements & Sezer (1982), onset cluster repair is restricted to casual speech, and inserts a harmonizing vowel (which they consider to be epenthetic). Other work on Turkish onset cluster repair also characterizes the inserted vowel as subject to regressive vowel harmony with the following lexical vowel, but does not mention an effect of speech style (Yavaş 1980, Kaun 2000, Yıldız 2010). In contrast to previous descriptions, I did not find evidence of regular regressive vowel harmony in onset cluster repair in studies of TELL (Ch. 1, Ch. 5) or in careful speech in the production study (Ch. 2, 3). In my careful speech data, onset cluster repair did generally apply, but the resulting vowel sound was closer to the schwa-like /ɯ/ than to /i/ or /u/, even when regressive harmony with V_2 would have required the inserted vowel to be front or round. (Whether regressive harmony is ever active in Turkish is an open question, whose answer I suspect to be "no"; see Kabak [2011] for more on this point.) The reduced, centralized quality of the intrusive vowel in the careful speech data reported in this dissertation may reflect the lessened gestural overlap between the following lexical vowel V_2 and the ICI. It is possible, therefore, that casual speech data may be more in line with previous descriptions.

4.2 Methods

To determine whether these predicted effects of speech style do obtain for onset cluster repair in Turkish, I collected casual speech in a production experiment. The design, materials, and participants were identical to those for the Careful speech condition. The procedure was as described in Ch. 2; Casual speech data was collected after the Careful speech. Participants were instructed to speak casually, as if talking to a friend or family member, and the frame sentence was changed to elicit deaccented productions of the target words.

(1) Fatma X_1 ve Y_1 dedi, Erhan da X_2 ve Y_2 dedi.
'Fatma said X_1 and Y_1, Erhan also said X_2 and Y_2.'

Thus, each production contained two repetitions of each target word. The first instance (X_1 and Y_1) is expected to be spoken casually but with focus, while the second instance (X_2 and Y_2) were expected to be deaccented as well as spoken casually. Hence, the greatest gestural overlap and hypoarticulation are expected to occur in the second mention. The pitch contour and TextGrid for a sample utterance are shown in Figure 4.1. In practice, the Casual condition speech was still fairly formal, due to the nature of the task (reading) and the environment (laboratory with ultrasound helmet). Nonetheless, second mentions were found to differ acoustically from Careful speech in ways that indicate they were deaccented as intended.

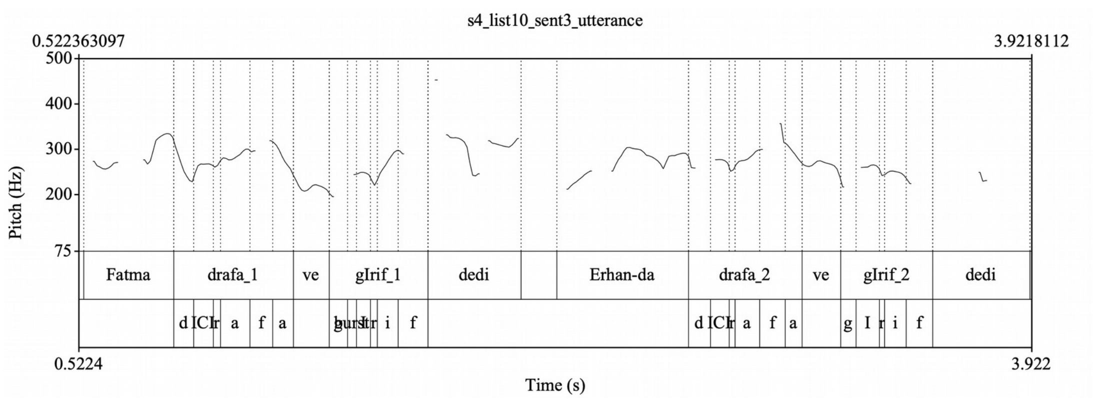


Figure 4.1: Sample Casual condition utterance from S4

4.2.1 Data

Eight tokens (one each of *biranci, drama, grabi,* and *drafa*; two each of *bratçiten* and *driplike*) produced with the wrong V_2 were excluded, six from S4 and two from S5. In addition, fifteen tokens that were outliers in terms of their F1 were excluded (two with F1 $<$ 1.0 Bark, 13 with F1 $>$ 6.0 Bark), leaving

1890 tokens from six subjects. Bark was selected as the unit for setting the exclusion threshold because of its more transparent relationship to perceived vowel quality, compared to Hertz.

4.3 Results

As in the analysis of careful speech, the distributions of two duration measures were plotted to determine whether onset cluster repair applies gradiently (predicting unimodal distributions), or categorically but optionally (predicting bimodal distributions). The first duration measure used was the duration of the entire ICI, which includes the consonant burst and any positive VOT, as well as any high amplitude periodicity that occurred in the ICI. The distribution of the durations of the ICI in underlying clusters (/Cr/) is shown in the lefthand panel of Figure 4.2; for comparison, I also include the duration of the ICI in words containing underlying vowels between /C/ and /r/ (/Cvr/). The second duration measure employed here is the duration of high amplitude periodicity within the ICI (i.e., the vowel). The durations of non-lexical and lexical vowels are plotted in Figure 4.2. Both duration measures are unimodal, supporting the interpretation that vowel insertion is a gradient process. Non-lexical vowels and ICIs are shorter than lexical vowels and ICIs. These results are similar to those for the Careful speech condition (Figure 4.3).

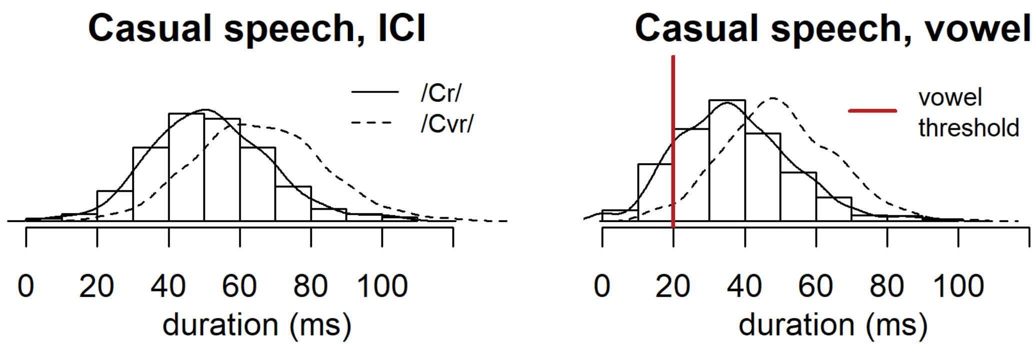


Figure 4.2: Duration distributions, Casual speech

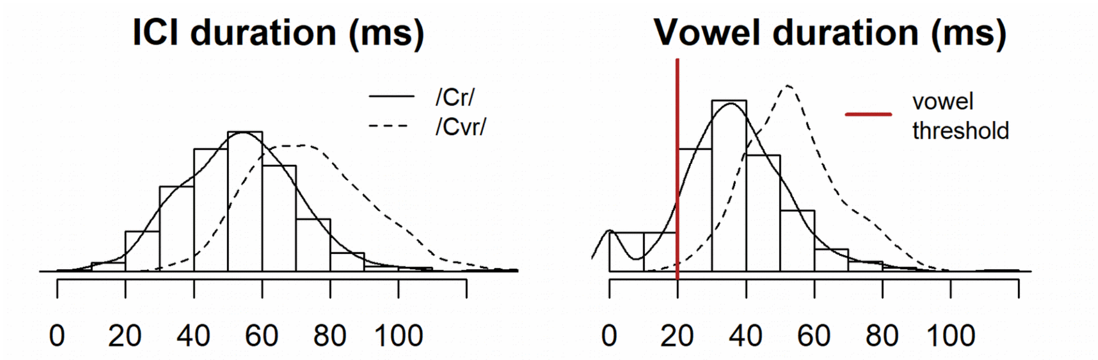


Figure 4.3: Careful speech duration distributions, for comparison.

For purposes of comparing insertion rates across speech style conditions, I employed the same threshold as in the Careful condition for determining whether a token contained an acoustic inserted vowel or not (20 ms of high amplitude formant structure during the ICI, shown by the red line in the righthand panel of Figure 4.2). As in the Careful speech condition, few lexical vowels but more non-lexical vowels fall below this threshold.

Vowel insertion rates varied by subject from as low as 60% (S3) to as high as 99% (S4), with an interspeaker average of 87%—practically the same as the 88% average insertion rate in Careful speech. The differences in insertion rate across speech style conditions are neither large, nor systematic: some speakers show more insertion in Casual speech, others show less.

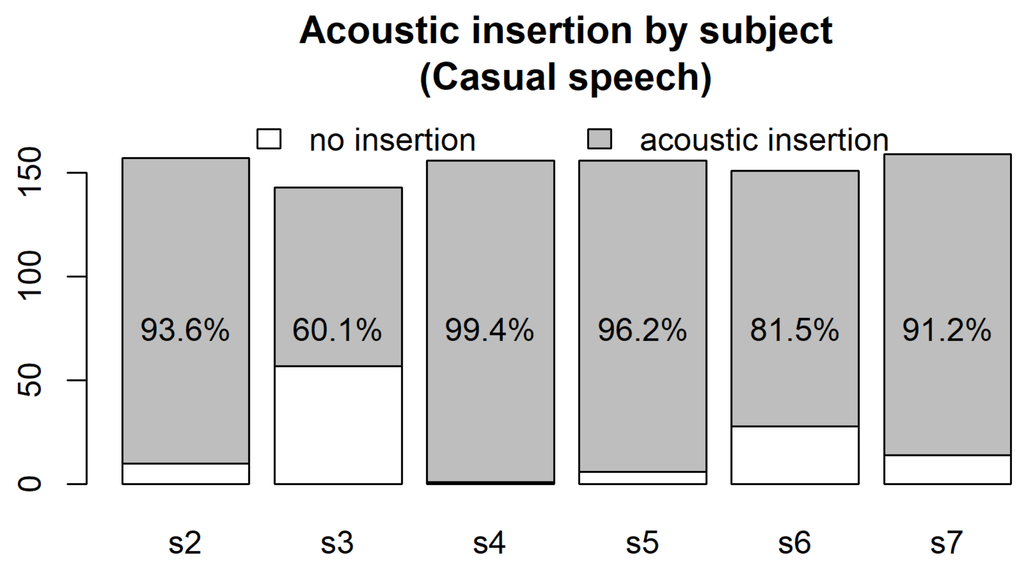


Figure 4.4: Insertion rates in Casual speech

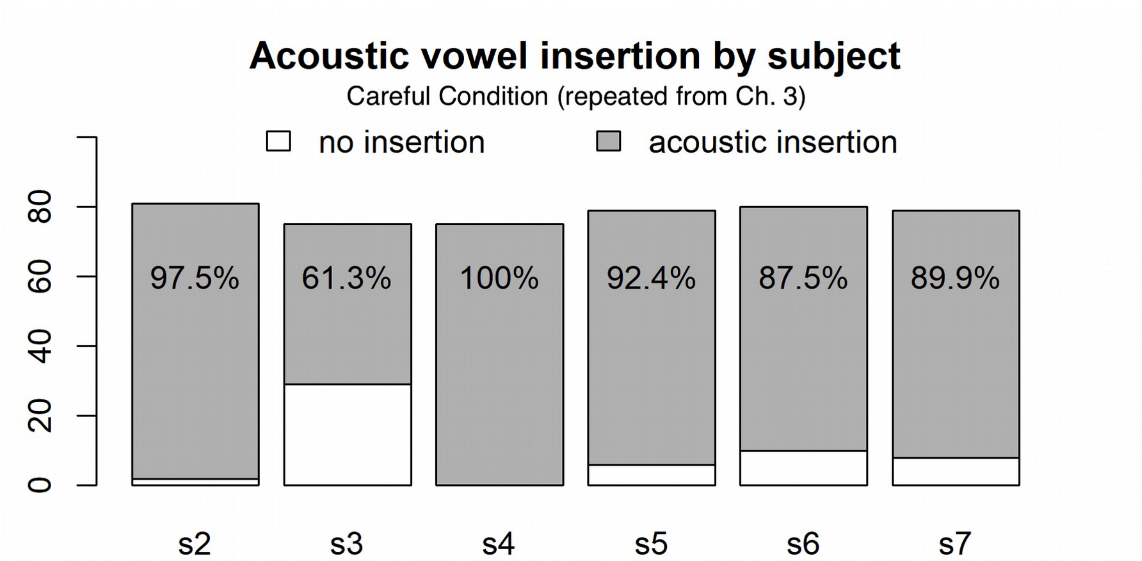


Figure 4.5: Insertion rates in Careful condition, for comparison

Figure 4.6 shows the duration of underlying and non-lexical vowels across the three speech conditions (Careful, Casual first mention, and Casual second mention). For underlying vowels, duration in Casual speech, first mention, is not significantly different from duration in Careful speech. But in Casual second mentions, $/v_1/$ duration is significantly shorter than either Careful speech or first mentions in Casual speech. This shortening of lexical vowels shows that Casual second mentions are deaccented, as intended.

In contrast, non-lexical vowels in first mentions in Casual speech are longer than non-lexical vowels in either Careful speech or Casual second mentions. Deaccented non-lexical vowels are not significantly shorter than non-lexical vowels in Careful speech.

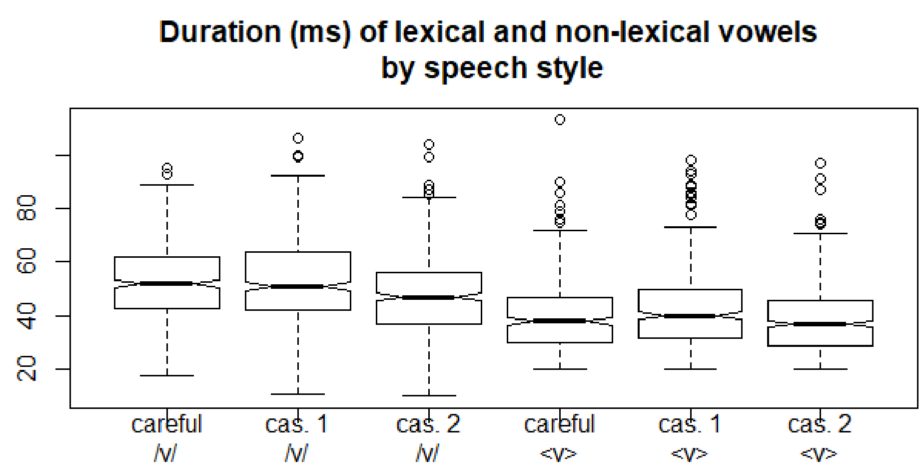


Figure 4.6: Duration and speech style

These results suggest that the distinction between Careful and Casual speech is less relevant than the distinction between first and second mentions within Casual speech in this study. This may indicate that participants were not employing a truly casual speaking style even in the Casual condition. It is notoriously difficult to elicit truly natural, casual speech in laboratory conditions, and the difficulty of speaking casually was no doubt heightened by the fact that participants were wearing the ultrasound probe stabilization helmet – hardly conducive to imagining oneself speaking casually with a friend. Within the "Casual" condition, however, first mentions of both lexical and non-lexical vowels are longer than second mentions, showing that second mentions were deaccented as intended.

4.3.1 Analysis of durational differences

To further explore the impact of speech style on duration, I conducted a mixed effects linear regression analysis of vowel duration, using R, lmer, and lmerTest (R Core Team 2016, Bates et al. 2015, Kuznetsova et al. 2016). Fixed effects were: speech style, preceding consonant /b d g/, following vowel /i a o/, and the vowel's lexical status and hypothesized category. Since non-lexical vowels have no underlying category, they were assigned to the hypothesized categories [i] before V_2 = /i/, [u] before V_2 = /o/, and [ɯ] before V_2 = /a/, according to the predictions of Clements & Sezer (1982) and Yıldız (2010). The models included random slopes and intercepts by subject for lexical status, V_2, and speech style. Random slopes for the other fixed effects were omitted so that the model would converge, on the assumption that the effect of C and v_1 would be more stable across subjects than insertion status and mention would.

The best-performing model included three interactions: C with speech style, lexical status with V_2, and lexical status with speech style. It outperforms a model without the vowel's lexical status in a maximum likelihood ratio test ($\chi^2(11) = 759.56$, $p < 0.0001$).

Table 4.1: Fixed effects table, vowel duration model

```
Formula: vdur ~ c * mention + v1 + /Cr/ * v2 + /Cr/ * mention
   + (/Cr/ + v2 + mention | subj)
                    B      SE       df t value Pr(>|t|)
(Intercept)     43.94    2.06     8.60  21.339 9.67e-09 ***
c=/b/           -0.23    0.99  2589.40  -0.238  0.81226
c=/g/            4.89    0.97  2589.30   5.048 4.76e-07 ***
Cas1             0.20    1.73    14.70   0.116  0.90952
Cas2            -6.21    1.60    18.40  -3.873  0.00108 **
v1=i             6.15    0.82  2591.90   7.484 9.86e-14 ***
v1=u             5.28    1.21  2588.90   4.365 1.32e-05 ***
/Cr/            -7.76    3.09     7.30  -2.513  0.03878 *
v2=/i/           8.21    2.17     7.90   3.793  0.00538 **
v2=/o/           0.70    1.47    15.50   0.479  0.63876
c=/b/:Cas1      -3.29    1.38  2589.50  -2.387  0.01707 *
c=/g/:Cas1       0.03    1.35  2588.90   0.024  0.98091
c=/b/:Cas2      -0.35    1.39  2589.40  -0.248  0.80415
c=/g/:Cas2       0.90    1.36  2589.20   0.660  0.50945
/Cr/:v2=/i/   -10.13    1.12  2593.70  -9.060  < 2e-16 ***
/Cr/:v2=/o/    -9.03    1.40  2570.00  -6.434 1.47e-10 ***
Cas1:/Cr/        3.26    1.06  2578.00   3.076  0.00212 **
Cas2:/Cr/        4.60    1.08  2586.60   4.265 2.07e-05 ***
```

Speech style had a significant main effect, with vowels in the deaccented second mention being 6.21 ms shorter (SE = 1.60, $p = 0.001$) than in Careful speech. Lexical status also had a significant main effect, with non-lexical vowels being 7.76 ms shorter than lexical vowels (SE = 3.06, $p < 0.05$). But Casual non-lexical vowels were longer than would be expected given those main effects alone, in both first ($\beta = 3.26$, SE $= 1.06$, $p < 0.005$) and second mentions ($\beta = 4.60$, SE $= 1.08$, $p < 0.001$). That is, the duration of non-lexical vowels is more stable under changes in speech style ($\approx$ degree of gestural overlap) than that of lexical vowels. The model, then, supports the pattern shown in Figure 4.6—that lexical vowels /v/ shorten more in Casual speech than non-lexical vowels <v>. In other words, the relative timing of the gestures for C and /r/ is more stable in underlying clusters than in underlying /CVr/ sequences.

4.3.2 Differences in F1

To test for the predicted interactions of lexical status with speech style and consonant place on vowel quality, I conducted a mixed effects linear regression analysis within each V_2 condition, which compared non-lexical vowels to the harmonizing lexical vowels predicted by previous analyses (Clements & Sezer 1982, a.o.). Non-lexical vowels were assigned to the hypothesized categories [i] before V_2 = /i/, [u] before V_2 = /o/, and [ɯ] before V_2 = /a/, and compared to their lexical counterparts, as in Ch. 3. Fixed effects were the underlying word shape (/Cr/ or /CVr/) and its interactions with the preceding consonant (/b d g/) and prosodic status (Careful focused; Casual first mention; Casual deaccented second mention). The models included random slopes and intercepts for word shape, consonant, and speech style, by subject. Models were tested against each other using maximum likelihood ratio tests. I report the most complex models justified by the data.

4.3.2.1 F1 of vowels before /i/

For vowels before /i/, the best-performing model of F1 includes a three-way interaction between preceding consonant, prosodic condition, and word shape (Table 4.2). None of the individual three-way interaction terms in the model came out significant. The only significant effect involving the vowel's lexical status was an interaction with a preceding /b/ ($\beta = 29.0$, SE = 9.18, $p < 0.005$), showing that non-lexical vowels after /b/ had a raised F1 compared to lexical vowels after /b/. Since F1 is correlated with jaw and tongue height, this suggests that the mouth does not open as much in /br/ as it does in the sequence /bir/, consistent with there being a vocalic target in /bir/ but no such target in /br/.

Table 4.2: Fixed effects table, model of F1 in vowels before /i/

```
Formula: vf1 ~ c * mention * /Cr/ + (/Cr/ + c + mention | subj)
   Data: v1i.df
                       B      SE       df  t value Pr(>|t|)
(Intercept)       332.30    25.45    5.20  13.059 3.52e-05 ***
c=/b/             -22.59    11.62    8.80  -1.944  0.08447 .
c=/g/             -23.55     7.75   27.40  -3.040  0.00515 **
Cas1               10.01    10.54   10.30   0.950  0.36390
Cas2               16.40     9.57   12.30   1.713  0.11193
/Cr/               14.93     9.59   14.70   1.556  0.14087
c=/b/:Cas1         26.11     8.99  943.10   2.904  0.00377 **
c=/g/:Cas1         -0.80     9.17  943.20  -0.087  0.93090
c=/b/:Cas2         11.61     9.00  943.20   1.290  0.19730
c=/g/:Cas2         -0.72     9.16  943.60  -0.078  0.93746
c=/b/:/Cr/         29.01     9.19  943.20   3.156  0.00165 **
c=/g/:/Cr/         13.57     9.22  943.10   1.472  0.14147
Cas1:/Cr/          -3.29     9.43  943.60  -0.349  0.72712
Cas2:/Cr/          -9.86     9.30  943.10  -1.059  0.28966
c=/b/:Cas1:/Cr/  -20.70    13.14  943.30  -1.575  0.11555
c=/g/:Cas1:/Cr/   19.59    13.15  943.30   1.490  0.13650
c=/b/:Cas2:/Cr/   -8.06    13.09  943.20  -0.616  0.53835
c=/g/:Cas2:/Cr/    5.77    13.16  943.40   0.439  0.66089
```

4.3.2.2 F1 in vowels before /o/ and /a/

In the other two vowel conditions (V_2 = /o/, V_2 = /a/) the best-performing model of F1 included an interaction between the vowel's lexical status and speech style (Table 4.3, Table 4.4). None of the individual effects had significant *p*-values in either model, however. (Note that the /d-o/ condition has been excluded due to a flaw in the experimental design; see the Materials section in Ch. 2.)

Table 4.3: Fixed effects table, model of F1 in vowels before /o/

```
Formula: vf1 ~ /Cr/ * mention + c + (/Cr/ + c + mention | subj)
   Data: v1u.df
```

	B	SE	df	t value	Pr(>|t|)	
(Intercept)	334.48	19.10	5.100	17.515	9.27e-06	***
/Cr/	10.580	9.74	9.700	1.087	0.303	
Cas1	9.909	12.25	8.100	0.809	0.442	
Cas2	11.947	13.85	7.000	0.863	0.417	
c=/g/	4.055	8.47	5.100	0.479	0.652	
/Cr/:Cas1	4.683	8.86	405.400	0.529	0.597	
/Cr/:Cas2	3.334	9.04	403.900	0.369	0.712	

Table 4.4: Fixed effects table, model of F1 in vowels before /a/

```
Formula: vf1 ~ Cr * mention + c + (Cr + c + mention | subj)
   Data: v1I.df
```

	B	SE	df	t value	Pr(>|t|)	
(Intercept)	395.511	24.692	5.1	16.018	1.51e-05	***
/Cr/	7.256	8.685	8.0	0.835	0.4277	
Cas1	5.008	12.841	6.5	0.390	0.7091	
Cas2	13.416	18.081	5.6	0.742	0.4878	
c=/b/	9.333	4.837	6.4	1.930	0.0986	.
c=/g/	-21.284	8.463	5.1	-2.515	0.0527	.
/Cr/:Cas1	-3.437	6.859	701.6	-0.501	0.6165	
/Cr/:Cas2	-1.555	6.921	702.2	-0.225	0.8223	

4.3.3 Differences in F2

Following the same procedure as the analysis of F1, I conducted a mixed effects linear regression analysis of F2. Again, only harmonic underlying vowels were considered.

In all vowel conditions, the best performing model included an interaction of lexical status with speech style. Before /i/ and /o/, consonant place also intereacted with lexical status; before /a/, it interacted with speech style. Differences between lexical and non-lexical vowels were clearest before /i/.

4.3.3.1 F2 of [i] before /i/

For the F2 of vowels before /i/, the best-performing model included two-way interactions between lexical status and preceding consonant, and lexical status and prosodic condition (Table 4.5). Just as in the Careful speech condition alone, in this model of all speech styles, there was a significant main effect of lexical status: F2 was lower in non-lexical vowels than in lexical /i/ (β = -224.75, SE = 61.14, $p < 0.01$). This result is in line with the hypothesis that non-lexical vowels lack /i/'s [-back] target, and accords with the ultrasound results from the Careful condition.

Table 4.5: Fixed effects table, model of F2 in [i] before /i/

Formula: vf2 ~ Cr * mention + c * Cr + (Cr + c + mention | subj)
Data: v1i.df

	B	SE	df	t value	Pr(>|t|)	
(Intercept)	1970.79	75.63	5.20	26.06	1.11e-06	***
/Cr/	-224.75	61.14	6.20	-3.68	0.00972	**
Cas1	-16.52	22.97	10.50	-0.72	0.48781	
Cas2	-52.35	17.54	51.60	-2.98	0.00434	**
c=/b/	128.00	55.75	5.50	2.30	0.06544	.
c=/g/	157.64	39.77	6.00	3.96	0.00745	**
/Cr/:Cas1	16.61	24.14	948.60	0.69	0.49163	
/Cr/:Cas2	50.07	24.21	947.60	2.07	0.03887	*
/Cr/:c=/b/	-196.32	24.40	950.90	-8.05	2.66e-15	***
/Cr/:c=/g/	-128.69	24.43	941.80	-5.27	1.71e-07	***

There was also a significant interaction between lexical status and preceding consonant. Lexical /i/ in the context /g_ri/ had a significantly increased F2, compared to lexical /i/ in /d_ri/ (β = 157.6, SE = 39.8, $p < 0.01$). In /giri/, Turkish phonology requires /g/ to palatalize to its allophone [ɟ], since it is the onset of a syllable with a front vowel nucleus (Göksel & Kerslake 2005). This palatalization of /g/ is probably the source of the raised F2 in lexical /i/ after /g/. In comparison, non-lexical vowels in /g_ri/ have a lower F2 than those in /d_ri/ (β = -128.69 Hz, SE = 24.43, $p < 0.001$). This suggests that /g/ does not palatalize in /gri/, where /r/ intervenes between the velar and /i/. As a result, /g/ remains velar, and has a backing effect on the non-lexical vowel, lowering its F2, unlike the palatalized [ɟ] that raises the F2 of lexical /i/. If we take Göksel & Kerslake's (2005) characterization of palatalization seriously, then the lack of palatalization in /gri/ could indicate that /g/ is not part of the onset for the syllable containing /i/, meaning that /gri/ is syllabified as [g.ri], with [g] being an extrasyllabic or consonant. However, it also seems entirely possible that /g/ only palatalizes under the influence of an immediately adjacent front vowel, and that the apparent palatalization facts here have no bearing on syllable structure, but reflect only on linear adjacency. The degree of palatalization in /gri/ vs. /giri/ could also be verified using the ultrasound images of tongue position during [g], but I leave this for future research.

Non-lexical vowels preceded by /d/ also have a higher F2 than those preceded by /b/ (β = -196.32 Hz, SE = 24.4, $p < 0.001$). This interaction between the preceding consonant and a vowel's lexical status suggests that /d/

has a much larger raising effect on the F2 of non-lexical vowels than lexical vowels, in keeping with the hypothesis that non-lexical vowels are targetless and therefore more subject to coarticulatory effects from all sides. A preceding /b/, meanwhile, significantly lowers the F2 of non-lexical vowels, as expected given the tendency of labial consonants to lower F2 in adjacent vowels.

Finally, speech style or prosodic condition had a significant main effect, with F2 being significantly lower in lexical vowels in Casual second mentions compared to Careful focused speech (β = -52.35, SE = 17.54, $p<0.005$). However, non-lexical vowels in Casual second mentions had a significantly higher F2 (β =50.07, SE = 24.21, $p<0.05$) compared to their Careful counterparts, suggesting that the deaccented, Casual speech condition does result in greater overlap of the following /i/ with the preceding ICI and its non-lexical vowel.

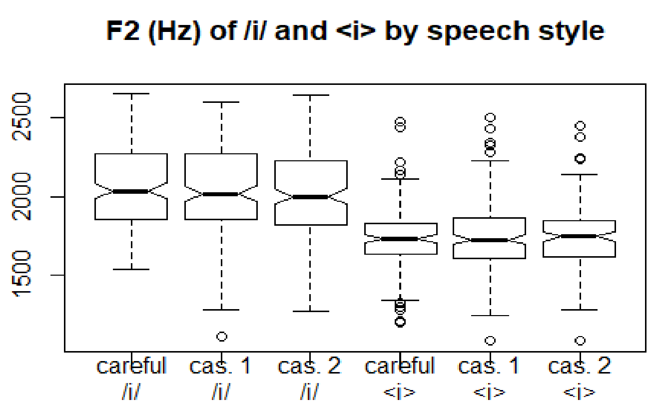


Figure 4.7: F2 of vowels before V_2=/i/

To summarize, non-lexical vowels before /i/ have a lower F2 than lexical /i/, as expected if they are targetless. Their F2 is especially lower following /g/ or /b/, and higher after /d/. This shows that the preceding consonant affects F2 in non-lexical vowels more than in lexical vowels. Furthermore, in deaccented Casual speech, F2 is diminished in lexical vowels, as predicted if the /i/ gesture undershoots in fast speech. But F2 in non-lexical vowels before /i/ actually increases in deaccented Casual speech compared to focused Careful speech, suggesting that the gesture for the following /i/ overlaps the ICI more in Casual speech, as predicted. These results support the hypothesis that non-lexical vowels in Turkish onset clusters are intrusive, and their quality is determined by the timing and overlap of adjacent gestures.

4.3.3.2 F2 of [u] before /o/

For a word containing /o/ to be fully harmonic in Turkish, any preceding high vowel must be /u/. Comparing the F2 of lexical /u/ and non-lexical <u> in the same context, the best-performing model includes significant interactions of lexical status with both speech style and the preceding consonant and speech style (Table 4.6). Only the interaction with the consonant is individually significant: F2 in non-lexical <u> is higher after /g/ than after /b/ (β = 202.7 Hz, SE = 32.2, $p < 0.0001$), reflecting /b/'s lowering effect on F2. That is, the preceding consonant has a stronger coarticulatory effect on non-lexical (targetless) <u> than on targeted /u/.

Table 4.6: Fixed effects table, model of F2 in [u] before /o/

Formula: vf2 ~ Cr * mention + c * /Cr/ + (/Cr/ + c + mention | subj)
Data: v1u.df

	B	SE	df	t value	Pr(>|t|)	
(Intercept)	1145.194	41.770	8.100	27.42	2.57e-09	***
/Cr/	-70.268	59.987	8.600	-1.17	0.273	
Cas1	65.135	48.947	9.500	1.33	0.214	
Cas2	86.614	53.550	8.400	1.62	0.143	
c=/g/	-13.469	57.979	6.500	-0.23	0.823	
/Cr/:Cas1	-8.509	38.921	409.600	-0.22	0.827	
/Cr/:Cas2	-31.464	39.657	408.200	-0.79	0.428	
/Cr/:c=/g/	202.667	32.234	403.400	6.29	8.40e-10	***

4.3.3.3 F2 of [ɯ] before /a/

Before /a/, the only harmonic high vowel possible is [ɯ]. The best-performing model of [ɯ] before /a/ includes significant interactions of speech style with lexical status and consonant (Table 4.7). No effects involving lexical status were significant within the model, although there are significant interactions between speech style and the preceding consonant. As expected, F2 is significantly lower after /b/ (β = -201.7, SE = 40.0, $p < 0.005$). In /d/ conditions, F2 is higher in Casual speech (Casual, first mention: β = 80, SE = 23, $p < 0.005$; Casual, second mention: β = 65, SE = 20.6, $p < 0.005$). But when /b/ or /g/ takes /d/'s place, this raising of F2 is negated. This suggests that a preceding /d/ leads to a stronger coarticulatory fronting effect in Casual speech.

Table 4.7: Fixed effects table, F2 of [ɯ] before /a/

```
Formula: vf2 ~ /Cr/ * mention + c * mention
      + (/Cr/ + c + mention | subj)
Data: v1I.df
                  B       SE       df t value Pr(>|t|)
(Intercept) 1527.48    63.34    5.20  24.117  1.5e-06 ***
/Cr/          17.56    13.87   26.20   1.266 0.216542
Cas1          80.60    23.05   19.50   3.496 0.002337 **
Cas2          65.05    20.61   38.30   3.156 0.003112 **
c=/b/       -201.74    39.81    5.90  -5.067 0.002429 **
c=/g/         62.76    37.22    6.00   1.686 0.142501
/Cr/:Cas1    -11.15    17.20  702.30  -0.648 0.517016
/Cr/:Cas2      0.72    17.32  702.30   0.041 0.966976
Cas1:c=/b/   -62.17    20.65  702.40  -3.011 0.002696 **
Cas2:c=/b/   -45.89    20.80  702.50  -2.206 0.027711 *
Cas1:c=/g/   -75.59    19.63  702.20  -3.851 0.000129 ***
Cas2:c=/g/   -56.05    19.86  702.30  -2.823 0.004891 **
```

4.4 Discussion

This study hypothesized that onset cluster repair in Turkish results from gestural timing, not phonological epenthesis, and predicted that a change in speech style—and therefore a change in the degree of gestural overlap—would affect non-lexical vowels differently than lexical vowels. The results bore out this prediction. Lexical vowels were shorter and showed formant differences that suggest gestural undershoot and reduction in second mentions in Casual speech, showing that deaccentuation had applied, as intended. The change in speech style caused significant differences in duration, F1, and F2 of non-lexical vowels as well. The direction and magnitude of these acoustic changes depended on the vowel's lexical status; all models performed significantly better when an interaction of lexical status and speech style was included.

With regard to duration, all vowels, regardless of lexical status, are shorter in deaccented second mentions. However, the duration of non-lexical vowels does not change as much as that of lexical vowels, suggesting that even in Careful speech, the two consonant gestures are already coordinated about as closely as Turkish gestural alignment permits (Figure 4.6, Table 4.1). The increased stability of ICI duration in /Cr/ words compared to /CVr/ words is predicted if the /Cr/ consonant cluster represents a complex onset, because the two consonants in a complex onset are expected to be coupled together with an anti-phase relation (Goldstein et al. 2008). (In Alignment terms, their timing will be governed by an alignment constraint demanding limited overlap.) This direct phasing relationship will stabilize the relative timing of C_1-C_2 in /CC/ words. But in /CVC/ words, C_1 and C_2 are simplex onsets to separate syllables, and have no direct phasing relationship. Consequently, the relative timing of C_1 and C_2 (and hence the duration of the ICI) can vary more in /CVC/ words than in /CC/ words. Thus, the duration results here support an analysis of Turkish onset cluster repair as vowel intrusion into a complex onset, which does not alter its syllabification.

This finding is surprising given previous descriptions of intrusive vowels that are absent in fast speech:

> "A third characteristic of intrusive vowels is that they are often variable in duration, and may disappear at fast speech rates or in casual speech, as reported for Saami (Bye 2001:139), Argyllshire Gaelic (Holmer 1938:32), Finnish (Harms 1976:77), Spanish (Quilis 1981:298), Hamburg German (Jannedy 1994), Moroccan Colloquial Arabic (Heath 1987; Gafos 2002), Mono (Olson 2003), and Chamicuro (Parker 1994)."
>
> Hall 2006

Of these languages, at least four (Saami, Finnish, Mono, Chamicuro) involve intrusion between the coda of one syllable and the onset of the next. If consonant gestures are not directly coordinated with each other across syllable boundaries, then it makes sense that C_{Coda}-C_{Onset} timing would not be stable across speech styles, with the result that intersyllabic intrusive vowels would be able to disappear as a result of greater gestural overlap in fast speech. Hall (2006) also notes that not all intrusive vowels are vulnerable to speech rate, citing those in Scots Gaelic, Hocank, Dutch, and Moroccan Colloquial Arabic (Gafos 2002). Intrusive vowels in Turkish onset clusters can be added to this list. I analyze their stability as a result of a grammaticized gestural alignment that specifies minimal overlap between C_1 and C_2, as previously proposed by Gafos (2002) for some intrusive vowels in Moroccan Colloquial Arabic.

In this experiment, rates of intrusion were approximately equal across speech styles. What, then, explains Clements & Sezer's (1982) claim that intrusion occurs only casual speech? One possibility is that their consultant was a fluent bilingual (similar to S3 in this study) who employed a French- or English-like gestural timing in careful speech, but a more Turkish gestural timing in casual speech. Another possibility is that a greater degree of "epenthesis" was perceived in casual speech because intrusive vowels in fast speech are proportionally longer, compared to lexical vowels, when those lexical vowels are shortened and reduced.

In this experiment, speech style also affected the formant values of lexical and non-lexical vowels differently. These differences were clearest before V_2 = /i/. Deaccented speech decreases F2 in lexical /i/, but increases F2 in non-lexical vowels. This suggests that /i/'s front tongue body target is undershot in deaccented casual speech, while targetless intrusive vowels gain a higher F2 by virtue of the following /i/'s greater overlap with the ICI.

This study also found that the preceding consonant affected non-lexical vowels more than lexical vowels. A preceding /d/ increases F2 more in <i>

than in /i/, and /b/ decreases F2 more in <u> than /u/. Intrusive <i> also has a much lower F2 after /g/. Furthermore, models of F1 in vowels before /i/ and F2 of [ɯ] before /a/ are significantly more successful when they include an interaction between lexical status and preceding consonant. This greater effect of the preceding consonant on non-lexical vowels than on lexical vowels is expected if the acoustics of non-lexical vowels in Turkish onset clusters are determined by gestural pressures from the surrounding context, as in vowel intrusion.

While intrusive vowels are often said to vary with speech style, few studies directly compare their properties in different speech styles. This study set out to do so, and found differences between careful and casual speech, even in the laboratory setting. Greater acoustic differences—reflecting a greater change in the degree of gestural overlap—occurred when target words were not only produced with a "casual" speech style, but were also deaccented, because they occurred in a given-information, second-mention context. Many of the interaction effects between lexical status and speech style in this study were only significant when Careful, focused speech was compared to Casual, deaccented speech. This illustrates the gradient, continuous nature of speech style and gestural overlap.

Appendix to Ch. 4: Models of Casual speech only

In addition to the statistical analysis across speech styles reported above, I also conducted an analysis within the Casual speech condition, paralleling the analysis reported in Ch. 3 for Careful speech. This section reports those results for Casual speech only (without the comparison to Careful speech).

Casual speech only: ICI as duration measure

The most complex model of the duration of the ICI that was justified by the data included interaction between lexical status and v_1 category, as well as between lexical status and mention, in addition to fixed effects for C place and V_2. It outperformed the epenthetic model that did not included lexical status in a maximum likelihood ratio test ($X^2(10) = 261.69$, $p < 0.0001$).

```
Formula: vdur_ici ~ c + Cr * harm_v1_yes_rounding + v2 + Cr * mention +
   (Cr + mention | subj) + (Cr + mention | target_word)

                              Estimate Std. Error       df t value Pr(>|t|)
(Intercept)                     56.3574     4.7741  16.2500  11.805 2.18e-09 ***
cb                              -6.5920     1.7186  20.5800  -3.836 0.000991 ***
cg                               9.5099     1.7030  19.4900   5.584 2.00e-05 ***
CrTRUE                          -5.0474     5.3914  13.2300  -0.936 0.365950
harm_v1_yes_roundingi            4.0043     2.2504 144.9500   1.779 0.077271 .
harm_v1_yes_roundingu           10.3805     5.3593  11.6400   1.937 0.077404 .
v2i                             13.2847     4.0314  14.6400   3.295 0.005044 **
v2o                             -2.4531     4.5330  14.7100  -0.541 0.596496
mention2                        -4.5316     0.8598  22.7600  -5.271 2.47e-05 ***
CrTRUE:harm_v1_yes_roundingi -11.7538     4.1995  17.3100  -2.799 0.012182 *
CrTRUE:harm_v1_yes_roundingu  -8.1001     5.7435  12.9000  -1.410 0.182111
CrTRUE:mention2                  2.0276     1.1453 113.7300   1.770 0.079352 .
```

As expected, the ICI was shorter after /b/ than /d/ (β = -6.6ms, SE = 1.7 ms, $p < 0.001$) and longer after /g/ than /d/ (β = 9.5 ms, SE = 1.7ms, $p <$

0.001). Also as expected, the second mention of a target within a sentence was shorter than the first (β = -4.5 ms, SE = 0.86ms, $p < 0.001$). As found for the careful condition, v_1 was also longer before V2 = /i/ than before /a/ or /o/ (β =13.3ms, SE = 4.0ms, p = 0.005). But in non-lexical vowels, this effect was cancelled out by an interaction with hypothesized v_1 category, with non-lexical <i> being 12.8ms shorter than its lexical counterpart (SE = 3.6 ms, $p < 0.005$). The difference between ICI duration for first and second mention is marginally smaller for underlying clusters compared to underlying vowels (β = 2.0 ms, SE = 1.1 ms, $p < 0.1$).

The above model with two interactions was marginally better than a model without the interaction between lexical status and mention, shown below. Like the two-interaction model, it outperforms the epenthetic model in a maximum likelihood ratio test ($X^2(9) = 258.93$, $p < 0.0001$).

```
Formula: vdur_ici ~ c + Cr * harm_v1_yes_rounding + v2 + mention + (Cr +
    mention | subj) + (Cr + mention | target_word)

                              Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)                    55.3178     4.6764  15.3400  11.829 4.10e-09 ***
cb                             -6.5987     1.6976  20.8000  -3.887 0.000863 ***
cg                              9.5114     1.6815  19.6600   5.657 1.65e-05 ***
CrTRUE                         -3.2651     5.2424  11.8700  -0.623 0.545177    
harm_v1_yes_roundingi           4.0170     2.2287 127.2500   1.802 0.073851 .  
harm_v1_yes_roundingu           9.7389     5.1798  11.4500   1.880 0.085747 .  
v2i                            13.2659     3.9097  14.2200   3.393 0.004290 ** 
v2o                            -1.8917     4.4055  14.4700  -0.429 0.673963    
mention2                       -3.5863     0.7033   9.5000  -5.099 0.000545 ***
CrTRUE:harm_v1_yes_roundingi -11.6634     4.0805  16.9400  -2.858 0.010911 *  
CrTRUE:harm_v1_yes_roundingu  -7.9273     5.5538  12.6700  -1.427 0.177663    
```

Lexical vowels are shorter after /b/ (β = -6.6, SE = 1.7, $p < 0.001$) and longer after /g/ (β = 9.5, SE = 1.7, $p < 0.001$), as expected. They are significantly longer before /i/ (β = 13.3, SE = 3.9, $p < 0.005$), and significantly shorter in deaccented second mentions (β = -3.6, SE = 0.7, $p < 0.001$). Non-lexical vowels before /i/ exempt from the lengthening of lexical vowels in this context (β = -11.7, SE = 4.1, $p < 0.05$).

Casual speech only: High amplitude periodicity (v1) as duration measure

Using the duration of high amplitude periodicity during the ICI as the duration measure, similar results are found. The most complex model justified by the data includes a main effect of lexical status, and outperforms the epenthetic model in a maximum likelihood ratio test ($X^2(7) = 181.2$, $p < 0.0001$).

```
Formula: vdur ~ Cr + mention + harm_v1_yes_rounding + v2 + c + (Cr +
mention | subj) + (Cr + mention | target_word)
                        Estimate Std. Error       df t value Pr(>|t|)
(Intercept)              49.7220     3.2494  13.5000  15.302 6.46e-10 ***
CrTRUE                  -11.2249     3.6314   7.9100  -3.091  0.01506 *
mention2                 -4.7391     0.6375  10.2800  -7.434 1.91e-05 ***
harm_v1_yes_roundingi     5.2527     2.1468 154.9100   2.447  0.01553 *
harm_v1_yes_roundingu     9.1619     3.5748  23.4700   2.563  0.01724 *
v2i                       1.1455     2.4735  96.4400   0.463  0.64433
v2o                     -10.8204     3.5029  22.2600  -3.089  0.00531 **
cb                       -2.6727     1.6566  20.2000  -1.613  0.12217
cg                        5.0493     1.6424  19.1000   3.074  0.00621 **
```

Similar to the duration of the ICI, the duration of non-lexical vowels was signficantly shorter than the duration of lexical vowels (β = -11.22 ms, SE = 3.63, $p < 0.05$). As expected, second mention vowels were also significantly shorter than first mention vowels (β = -4.7 ms, SE = 0.6ms, $p < 0.001$).

Further, /i/ and /u/ were significantly longer than /ɯ/, as expected (/i/: B = 5.3 Hz, SE = 2.15, $p < 0.05$; /u/: B = 9.16 Hz, SE = 3.57 Hz, $p < 0.05$) Vowels were significantly longer after /g/ than after /d/ (β = 5.05 ms, SE = 1.6 ms, $p < 0.01$). The following vowel also had a significant main effect on v_1 duration, with v_1 being shorter before /o/ compared to /a/ (β = -10.8 ms, SE = 3.5 ms, $p < 0.01$).

Casual speech only: Burst + VOT as duration measure

The third possible duration measure is the duration of the burst + any positive VOT. Here the most complex model justified by the data does not include any interactions, but includes a main effect for lexical status, and again outperforms the epenthetic model in a maximum likelihood ratio test ($X^2(9)=60.9$, $p<0.0001$). However, the lexical status of v_1 did not actually have a significant effect within the model, indicating that differences in the duration of the ICI between lexical and non-lexical vowels are not primarily the result of differences in consonant burst duration, but rather differences in the vowels themselves.

```
Formula: vdur_burst ~ Cr + c + mention + harm_v1_yes_rounding + v2 + (Cr +
  mention | subj) + (Cr + mention | target_word)

                      Estimate Std. Error       df t value Pr(>|t|)
(Intercept)            15.1032     2.2706   6.4500   6.652 0.000412 ***
CrTRUE                 -1.7183     1.1820   9.0000  -1.454 0.179985
cb                     -4.2296     0.5792  60.9800  -7.303 7.00e-10 ***
cg                      4.2718     0.5641  54.0400   7.573 4.81e-10 ***
mention2                1.0518     0.7337   5.4200   1.434 0.206735
harm_v1_yes_roundingi  -0.7256     0.9216 175.5900  -0.787 0.432159
harm_v1_yes_roundingu   1.4968     1.6326  30.1700   0.917 0.366508
v2i                     0.8839     0.9694 105.7700   0.912 0.363975
v2o                     0.1248     1.5905  26.4600   0.078 0.938055
```

Casual speech only: F1

For F1, the best intrusive model included an interaction between lexical status and the category of v_1, as well as an interaction between mention and V2. The intrusive model outperformed the epenthetic model ($X^2(3)=25.4$, $p<0.0001$). I used only random intercepts, not random slopes, because the epenthetic model wouldn't converge when random slopes were included.

```
Formula: vf1 ~ harm_v1_yes_rounding * Cr + v2 * mention + c + (1 | subj) +  (1
| target_word)

                             Estimate Std. Error       df t value Pr(>|t|)
(Intercept)                   376.941     23.094    5.700  16.322 5.12e-06 ***
harm_v1_yes_roundingi         -40.366      5.474   41.500  -7.374 4.56e-09 ***
harm_v1_yes_roundingu          -9.364      8.331   23.700  -1.124  0.27224
CrTRUE                          1.164      6.229   22.000   0.187  0.85343
v2i                           -18.300      7.176   30.900  -2.550  0.01595 *
v2o                           -35.407      7.911   32.800  -4.476 8.68e-05 ***
mention2                       10.301      3.971 1734.600   2.594  0.00956 **
cb                             -1.560      4.249   22.300  -0.367  0.71696
cg                            -10.365      4.222   21.500  -2.455  0.02266 *
harm_v1_yes_roundingi:CrTRUE   24.256      8.066   22.500   3.007  0.00639 **
harm_v1_yes_roundingu:CrTRUE   10.767      9.983   21.100   1.079  0.29296
v2i:mention2                  -14.714      4.967 1730.800  -2.963  0.00309 **
v2o:mention2                  -15.851      5.813 1744.100  -2.727  0.00646 **
```

As expected, F1 is significantly lower in [i] than in [ɯ] (β =-40.4Hz, SE = 5.5Hz, $p<0.001$), as well as being lower before /i/ (β = -18.3Hz, SE = 7.2Hz, $p<0.05$) and before /o/ (β = -35.4Hz, SE = 7.9Hz, $p<0.001$) than before /a/. The effect of the following vowel was particularly strong in the second mention, showing a significant interaction between V2 and mention (V2=/i/: B=-14.7Hz, SE = 5.0Hz, $p<0.005$; V2=/o/: B=-15.9Hz, SE = 5.8Hz, $p<0.01$), meaning that there is indeed more coarticulation in the second mention. In addition, the interaction between lexical status and hypothesized v1 category was significant, with non-lexical <i> having a higher F1 than lexical /i/ (β =24.3Hz, SE = 8.1, $p<0.01$). This is compatible with non-lexical <i> lacking lexical /i/'s [+high] target.

Casual speech only: F2

For F2, the best model justified by the data included an interaction between the lexical status of v1 and the category of v1. Including mention as a fixed effect did not significantly improve the model, so it was omitted from both intrusive and epenthetic models. The intrusive model outperformed the epenthetic model in a maximum likelihood ratio test (X2(3)=44.8, $p<0.0001$). The model showed the expected main effects: when v1 is /i/, F2 is significantly higher than when v1 is /ɯ/ (β =440.6Hz, SE = 33.7Hz, $p<0.0001$), and a following /i/ also increases F2 (β =92.6Hz, SE = 53.5Hz, $p<0.1$), while a following /o/ lowers F2 (β =-122.1Hz, SE = 59.3Hz, $p<0.05$), relative to a following /a/. A preceding /b/ also lowers F2 relative to a preceding /d/ (β =-111.6Hz, SE = 35.3, $p<0.005$). Finally, there was a significant interaction between lexical status and vowel category, with non-lexical <i> being 337.7Hz lower compared to lexical /i/ (SE = 66.5, $p<0.0001$). That is, while lexical /i/ and /ɯ/ differ by 440Hz, lexical and non-lexical [i] also differ by almost as much. This suggests that non-lexical <i> is produced with the tongue farther back in the mouth, as if the non-lexical <i> lacks lexical /i/'s [-back] target.

```
Formula: vf2 ~ Cr * harm_v1_yes_rounding + v2 + c + (1 | subj) + (1 |
target_word)
                              Estimate Std. Error      df t value Pr(>|t|)    
(Intercept)                    1339.58     207.27    5.58   6.463 0.000864 ***
CrTRUE                           12.86      51.69   30.77   0.249 0.805187    
harm_v1_yes_roundingi           440.60      33.67  207.54  13.086  < 2e-16 ***
harm_v1_yes_roundingu           -83.72      68.65   28.53  -1.220 0.232593    
v2i                              92.60      53.51   36.13   1.731 0.092069 .  
v2o                            -122.09      59.26   34.53  -2.060 0.046952 *  
cb                             -111.55      35.25   30.54  -3.165 0.003504 ** 
cg                               32.17      35.15   31.16   0.915 0.367065    
CrTRUE:harm_v1_yes_roundingi  -337.57      66.48   31.07  -5.078 1.71e-05 ***
CrTRUE:harm_v1_yes_roundingu   -24.65      85.03   27.37  -0.290 0.774103    
```

Summary

As in the careful speech condition, non-lexical vowels were found to be shorter than lexical vowels. The magnitude of the difference is greater in the casual condition (12ms vs 6ms). In addition, non-lexical <i> is found to have a higher F1 and lower F2 than lexical /i/.

Chapter 5: Corpus study

5.1 Introduction

I have hypothesized that Turkish onset cluster repair results from gestural timing relations, not phonological insertion, and that apparent harmony reflects coarticulation, not phonological vowel harmony. This hypothesis predicts that the variability in onset-cluster repair (previously reported by Clements & Sezer [1982] and Yıldız [2001]) should be structured by phonetic, gestural factors that do not affect epenthesis and harmony elsewhere in Turkish.

This prediction was supported by the production study presented in the previous chapters. I found that acoustic insertion in onset clusters is more likely to occur after a dorsal C_1 (/g/ in the experiment), which is also associated with a longer ICI compared to /b/ or /d/. The formant values of acoustic inserted vowels were also affected by the place of the following vowel more than those of underlying vowels were. However, due to the necessities of experimental design, the production experiment only considered three clusters (/br/, /dr/ and /gr/), even though onset cluster repair also affects clusters with voiceless stops, stop + /l/ clusters, /s/-stop clusters, and others. The limited number of clusters also meant that it was not possible to compare different types of cluster to see, for example, whether insertion is more likely in stop + sonorant clusters than in /s/ + stop clusters (cf. Fleischhacker 2005, Broselow 1987, among many others).

In order to examine other type of clusters across a larger range of lexemes and cluster types, I analyzed onset cluster repair as transcribed in the Turkish Electronic Living Lexicon (TELL; Inkelas et al 2000). TELL consists of broad phonetic transcriptions of 30,000 lexemes pronounced by two male speakers of Istanbul Turkish in eight different morphological contexts. TELL's structure is conducive to a study of onset cluster repair because its structure makes it easy to use regular expressions to compare the orthography and the transcription to determine whether a word was a candidate for onset cluster repair, and whether it was produced with a non-lexical vowel. Since most words were produced in all five (for nouns) or four (for verbs) morphological conditions, the corpus also includes multiple repetitions of each lexeme by each speaker.

A corpus of broad phonetic transcriptions might not be the first place one would think to look for a phonetic analysis. However, this study provides an opportunity to address a methodological question: Can the variation in impressionistically transcribed data provide phonetic or gestural insights (if the data exists in sufficient quantities)? This chapter shows that phonetic factors do structure the variation in TELL, and this corpus, although not phonetically detailed, still offers insights that can supplement the understanding of Turkish onset cluster repair gained from the acoustic and ultrasound study. In addition, TELL offers the opportunity to examine factors that could not be included in the production experiment, since the corpus has a larger variety of C_1, C_2, and V combinations. Finally, the TELL data can improve the ecological validity of laboratory results, in that the TELL speakers were not burdened with an ultrasound helmet during their production of the words, and were reading the words in a variety of sentence frames to induce the different morphological contexts.

5.2 Method and Predictions

I used generalized mixed effects models and chi-squared tests to model the variation in TELL to look for effects that are predicted by the gestural account but not the epenthetic account. As noted above, TELL consists of broad phonetic transcriptions of 30,000 lexemes. Transcriptions were made by a linguistically trained native speaker of Turkish. Inserted vowels in onset clusters are not (prescriptively) represented in Turkish orthography, so all words that began orthographically with an initial consonant cluster were treated as candidates for onset cluster repair. Transcribed pronunciation was taken to indicate whether the analyst had perceived insertion and/or harmony. This yielded 1973 tokens of 306 CC lexemes, produced in 8 morphological conditions by two speakers, although some lexemes were only produced by one speaker. Of these, 1395 tokens are transcribed with an inserted non-lexical <v>: C<v>C.

Based on previous descriptions of vowel intrusion (e.g., Hall 2003, 2006) and on the results of the production experiment (Ch. 2-4), I make the following predictions for transcribed intrusive vowels:

1. *Prediction 1:* Transcribed insertion will be more likely before sonorant C_2s, which are better candidates to overlap with a vowel gesture (Hall 2003). Given the types of clusters in TELL, this is essentially a prediction that more insertion will be transcribed in obstruent+sonorant onset clusters than in s+stop clusters, which are already known to pattern differently from each other cross-linguistically (Broselow 1987, Fleischhacker 2005).
2. *Prediction 2*: Transcribed insertion will be more likely when C_1 promotes a longer ICI with more airflow and voicing.

   Prediction 2a: Transcribed insertion will be more likely after dorsal C_1s since the interconsonantal interval (ICI) will be longer after dorsal consonants (Yip 2013, Bellik 2018). Transcribed insertion may also be less likely after labial consonants, since these result in a shorter ICI, although this effect was less strong in my acoustic study.

Prediction 2b: Insertion may also be more likely after a voiced C_1, since voicing during C_1 promotes voicing during the ICI, increasing the likelihood that the ICI will sound vowel-like.

3. *Prediction 3*: Transcribed insertion—particularly, apparent harmony— will also be affected by the place and manner of C_1/C_2, not just the lexical vowel. Adjacent consonants affect the formants of vowels, resulting in different apparent vowel qualities.

On the other hand, if traditional descriptions of onset cluster repair as epenthesis are correct, then none of the above factors are predicted to contribute to the likelihood of an inserted vowel being transcribed. If anything, insertion might be predicted to be more likely when there is a flatter sonority profile between C_1 and C_2—i.e., a higher likelihood of insertion in obstruent + obstruent clusters than in obstruent + sonorant clusters, the inverse of Prediction 2 above.

5.3 Descriptive summary of results

A non-lexical <V> was transcribed in 70.7% of CC tokens, and was much more likely to be transcribed when C_2 was sonorous (Table 5.1). Of these <V>, 78.9% were transcribed as [ɯ] (Table 5.2). This echoes the finding in the production experiment that non-lexical vowels are acoustically most similar to [ɯ]. In TELL, all but 5 inserted vowel tokens were high; five tokens of *fregat* (orthographic representation; each token is from a different morphological condition) were transcribed with inserted <e>. This confirms previous claims that the vowel in Turkish onset clusters is high (Yavaş 1980, Clements & Sezer 1982, Kaun 1999, Yıldız 2010), but also suggests that the height of the vowel is not phonologically specified, as indicated by the production experiment, in which tongue position was significantly lower in non-lexical vowels before /a/.

Table 5.1: Sonority of C_2 and insertion in TELL

	C_2 [-son]	C_2 [+son]	Total
No <v>	123 (6.2%)	455 (23.1%)	122 (29.3%)
<v>	53 (2.7%)	**1342 (68.0%)**	**1395 (70.7%)**

Table 5.2: Transcribed vowel insertion and feature-spreading in TELL.

/V_2/	/CC/	Any <V>	<V> = ɯ	<V> = [+front]	<V> = [+round]
Any	1973	1395	**1100**/1395 **(78.9%)**	222/1395 (15.9%) i=179, y=38, e=5	111/1395 (8.0%) u=73, y=38
[+front]	563	563	372/563 (66.1%)	**177**/563 **(31.4%)** i=134, y=38, e=5	52/563 (9.2%) y=38, u=14
[+round]	315	315	185/315 (58.7%)	57/315 (18.1%) y=38, i=19	**111**/315 **(36%)** u=73, y=38

Only one lexeme was produced with prothesis rather than epenthesis: all five of S2's productions of *stepne* 'spare tire' were transcribed with a prothetic <i>: [istepne(+suffix)]. S1 produced this lexeme with epenthesis in all morphological conditions, however. I exclude the prothetic tokens from the following analysis.

Apparent palatal and rounding harmony both underapplied relative to previous reports (Yavaş 1980, Clements & Sezer 1982, Kaun 1999, Yıldız 2010): [+front] was only transcribed as spreading to <V> in 31% of tokens with transcribed insertion before a lexically [+front] vowel; and [+round] was transcribed as spreading in 36% of <V> before a lexically [+round] vowel

(Table 5.2). (The percentages do not sum to 100% because [y] is counted in both [+front] and [+round] percentages.) These low rates of feature-spreading are in line with the acoustic differences between <V> and harmonizing /V/ in the acoustic study. Other factors that influence the transcribed category of the non-lexical vowel will be discussed below (Sections 5.5-5.6).

5.4 Modeling insertion rate

I used R (R Core Team 2016) and the packages lme4 (Bates et al. 2015) and lmerTest (Kuznetsova et al. 2016) to perform generalized mixed effects linear modeling. The response variable was whether an inserted vowel was transcribed in TELL (<v> or no <v>). The following fixed effects were included:

- *C_2 Sonority*: Is C_2 [+son] or [–son]? Tests Prediction 1.
- *C_1 Occlusion*: Is C_1 a stop [–continuant] or a fricative [+continuant]? Tests Prediction 1 by distinguishing s-stop clusters (C_1 = [+cont]) from stop-liquid clusters (C_1 = [-cont]).
- *C_1 Place*: Is C_1 Labial, Coronal or Dorsal? Tests Prediction 2a.
- *C_1 Voicing*: Is C_1 [+voice] or [–voice]? Tests Prediction 2b.

In a model with random intercepts for speaker and morphological condition, all effects were significant in ways that bore out the predictions above. However, when random slopes by speaker were included, significance disappeared, indicating that the two speakers patterned differently. Therefore, I analyze the two speakers separately and compare them below.

5.4.1 Transcribed insertion for S1

For S1, the model with all the above fixed effects, plus random slopes by morphological condition, is shown below (6):

(6) s1: <v> ~ C_2.sonorant + C_1.stop + C_1.place + C_1.voiced +
+ (1 + C_2.sonorant + C_1.stop + C_1.place + C_1.voiced |
morph.cond)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.46174	0.17498	-2.639	0.008321	**
C_2 [+son]	2.01740	0.27836	7.248	4.24e-13	***
C_1 [-cont]	-0.64836	0.19346	-3.351	0.000804	***
C_1 [dorsal]	0.47709	0.18173	2.625	0.008660	**
C_1 [labial]	-0.52921	0.16155	-3.276	0.001053	**
C_1 [+voi]	0.07076	0.14629	0.484	0.628591	

For S1, transcribed insertion was significantly more likely when C_2 was a sonorant ($\beta = 2.02$, SE = 0.28, $p < 0.001$ – Prediction 2). Insertion was also less likely when C_1 was a stop ($\beta = -0.64$, SE = 0.19, $p < 0.001$). This result was unexpected, because the factor C_1.stop was intended to distinguish stop+liquid (TL) clusters from s+stop (sT) clusters, and insertion was expected to be more likely in TL than sT clusters. However, the dataset also contained a few s+sonorant clusters (sN) and many f+liquid clusters (fL), which the factor C_1.stop treats together with the sT clusters. Results suggest that insertion is more likely after a fricative C_1, other things being equal, perhaps because a fricative C_1

allows more airflow than a stop C_1, and greater airflow during the ICI is more likely to result in an acoustic vowel. I return to this point below (Section 5.4.3).

Transcribed insertion was also significantly more likely for S1 when C_1 was dorsal ($\beta = 0.48$, SE = 0.18, $p < 0.01$), and slightly less likely when C_1 was labial ($\beta = -0.53$, SE = 0.16, $p < 0.005$). Voicing of C_1 did not significantly predict transcribed insertion ($p > 0.5$).

5.4.2 Transcribed insertion for S2

A slightly different pattern obtains for S2. For this subject, there is much less data (515 tokens instead of S1's 1458 tokens), and the model did not converge when the sonority of C_2 was included as a fixed effect in addition to C_1 place. Therefore, I ran the model shown below that does not include C_2 sonority as a fixed effect.

```
(7) S2: <v> ~ C1.stop + C1.place + C1.voiced
          + (1 + C1.stop + C1.place + C1.voiced | morph.cond)
                 Estimate Std. Error z value Pr(>|z|)
    (Intercept) -1.2748     0.4697  -2.714  0.00665 **
    C1 [-cont]   2.6406     0.8930   2.957  0.00311 **
    C1 [dorsal] -0.2297     0.4457  -0.515  0.60626
    C1 [labial]  2.5553     0.7819   3.268  0.00108 **
    C1 [+voi]    1.1812     0.6989   1.690  0.09102 .
```

The model for S2 finds that insertion is significantly more likely to be transcribed when C_1 is a stop ($\beta = 2.64$, SE = 0.89, $p < 0.01$). The direction of the effect of C_1 occlusion is reversed for S2 compared to S1, for whom transcribed insertion was more likely with a fricative C_1. Like the place effects discussed above, this effect probably reflects the fact that the sonority of C_2 is inextricably related to C_1 occlusion, since almost all stop C_1s occur with sonorous C_2s. I disentangle these effects below using chi-squared tests.

Contrary to predictions, transcribed insertion was significantly *more* likely for S2 when C_1 is labial ($\beta = 2.56$, $SE = 0.7819$, $p = 0.001$) than when C_1 is coronal, and was not significantly affected by C_1 being dorsal. Examining a table of insertion by C_1 place for S2, we see that insertion always occurs after /b/ and is also common after /f/, whereas insertion is never transcribed after /s/ for this subject. These cells are probably driving the effect seen in the model. However, there is a confound, in that the model ignores C_2 completely. For all the labial C_1s, C_2 is a liquid, whereas in 39 of the coronal C_1 clusters, C_2 is a stop in an /s/-stop cluster. Since C_1 place and C_2 sonority are entangled here, further investigation is needed to understand these effects. Finally, transcribed insertion was marginally more likely for S2 when C_1 is voiced ($\beta = 1.18$, $SE=0.70$, $p<0.1$).

Table 5.3: Transcribed vowel insertion for S2 in TELL

S2	Labial			Coronal			Dorsal	
C_1	p	b	f	t	d	s	k	g
No <v>	16	**0**	14	9	4	40	20	4
<v>	84	51	**77**	74	15	**0**	42	65

5.4.3 Follow up: Cluster type

The models above show that insertion is less likely after a stop for S1, but more likely after a stop for S2. There is a confound, however, because the model for S2 does not include the sonority of C_2 independently, and a C_1 that is a stop guarantees a sonorous C_2. Moreover, S2 produced only one /s/+sonorant lexeme (*slovak,* transcribed without insertion), meaning that all of S2's s-initial clusters were produced without insertion. Consequently, there is also not enough data to properly distinguish the effects of C_1 place and C_1 [continuant] in a model that includes both factors.

To disentangle the effect of C_1 [-continuant] from C_2 [+sonorant] and from C_1 place, I classified all clusters into six cluster types, shown in Table 5.4 (ordered by frequency). Since the /s/+sonorant type had only 22 tokens, and I have no reason to expect s-sonorant clusters to pattern differently from f-sonorant clusters, I combined them. I also excluded 14 tokens of three lexemes that were TN or TS (Table 5.5). The analysis, then, distinguished three cluster types: TL, FN, and sT – the types that have at least 100 tokens (Table 5.6).

Table 5.4: All cluster types in TELL

Cluster type	Abbreviation	N tokens	Examples
Stop + liquid	TL	1508	brülör, gravür, plato, klik, grandi
/f/ + liquid	fL	258	frenleme, flüt, flu, frezeci, fransız
/s/ + stop	sT	171	streptokok, stand, spesifik, spor
/s/ + sonorant	sN	22	slayt, smokin, slovak, snop
Stop + nasal	TN	9	gnu, pnömatik
Stop + fricative	TS	5	psikanaliz

Table 5.5: Excluded clusters in TELL

Lexeme	Transcribed insertion	N tokens (Speaker)
gnu	<ɯx>	4 (S2)
pnömatik	<i>	5 (S1)
psikanaliz	<ɯ>	5 (S1)

Table 5.6: Cluster types in TELL (Analyzed)

Cluster type	Abbreviation	N tokens	Examples
Stop + liquid	TL	1508	brülör, gravür, plato, klik, grandi
Fricative + sonorant	FN	280	frenleme, flüt, flu, slayt, smokin, slovak, snop
/s/ + stop	sT	171	streptokok, stand, spesifik, spor

Chi-squared tests of the effect of cluster type showed that sT clusters were significantly different from TL or FN clusters (S1: $\chi^2(2) = 72.639$, $p < 0.001$; S2: χ^2 (2) = 147.37, $p < 0.001$). This is what we expect from much previous work on the differences between cluster types (Broselow 1987, Fleischhacker 2005, Yun 2016, *inter alia*). Differences in the treatment of sT clusters and other clusters are also attested in Turkish, where historically sT clusters used to be repaired with prothesis: *istasyon* < Fr. *station*, İzmir < Smyrna, *ıspanak* < Gr. *spanak* 'spinach.' This process does not seem to be synchronically productive, however (Yıldız 2010); as noted above, only one lexeme is produced with prothesis in TELL.

Since sT clusters are so different from obstruent+sonorant clusters, it is possible that the effect of C_1 occlusion in the mixed effects models above was driven by differences between cluster types rather than an independent effect of C_1. This can be tested. If C_1 being a stop matters independently of C_2's sonority, then TL and FN are predicted to be significantly different. This prediction was borne out for S1, for whom insertion was marginally more likely in FN clusters than in TL clusters (χ^2 (1) = 3.4198, $p = 0.06$). This supports Prediction 1b, that insertion is more likely when C_1 is a fricative and therefore allows more airflow during the ICI since the vocal tract is already more open compared to a stop.

The prediction was not born out for S2, however. For this speaker, TL and FN clusters were not significantly different (χ^2 (1) = 1.123, p = 0.3). This suggests the effect of [-cont] in the mixed effects model for S2 above was driven only by the difference between obstruent+sonorant clusters and sT clusters.

Chi-squared tests were also used to disentangle the effects of C_1 place and manner. In the mixed effects models, insertion was marginally less likely after a labial C_1 for S1, as predicted by the ICI duration effects found in the acoustic study. For S2, however, contrary to predictions, transcribed insertion was significantly *more* likely after a labial C_1. This effect seems to have been driven by differences between S2's f-initial clusters (lots of insertion) and his s-initial clusters (zero insertion). The apparent effect of C_1 place for S2 may be an effect of cluster type—the differences between FN and sT clusters discussed above. To test this, I excluded the sT clusters and conducted a chi-squared test of the effect of C_1 place on transcribed insertion for each subject. For S1, place still had a highly significant effect on the likelihood of insertion (χ^2 (2) = 37.0, $p < 0.0001$), providing further support for the original prediction that insertion will be more likely after a dorsal C_1 and less likely after a labial C_1 (Prediction 1a). For S2, the test found no significant effect of C_1 place (χ^2 (2) = 2.5216, p = 0.3). This is still the case if the data is subsetted to just TL clusters (χ^2 (2) = 4.1, p = 0.13). These results support the claim that apparent place effects in an unexpected direction for S2 actually reflect differences between FN clusters and sT clusters. C_1 place does not appear to have a significant independent effect for this subject, whether because of a lack of power due to the small amount of data for S2, or because the formal speech setting limited the effect of C_1 place on the duration of the ICI.

5.4.4 Summary of transcribed insertion patterns

For both TELL subjects, the likelihood of transcribed insertion of a non-lexical vowel is affected by the features of C_1 and the type of cluster (TL or FN vs. sT). For S1, a non-lexical vowel is more likely to be transcribed when C_1 is dorsal, which lengthens the ICI, and less likely to be transcribed when C_1 is labial, which shortens the ICI (supporting Prediction 1a). For S2, the place of C_1 did not significantly affect the likelihood of transcribed insertion, but a voiced C_1 did increase the likelihood of transcribed insertion (supporting Prediction 1b). For both subjects, a non-lexical vowel was much more likely be transcribed in obstruent+sonorant clusters than in s-stop clusters. This finding indicates that non-lexical vowels occurred more often when C_2 was a sonorant (Prediction 2). In addition, for S1, a non-lexical vowel was transcribed more often in FN clusters than in TL clusters, suggesting that a fricative C_1 contributes to the likelihood of a vocalic-sounding ICI more than a stop C_1. Overall, the predictions of Section 5.2 were supported, showing that Turkish onset-cluster repair is shaped by factors that do not affect epenthesis in Turkish[7], but are expected to affect gesturally-driven vowel intrusion. Furthermore, I find that the two subjects pattern differently with respect to the effect of C_1 place and voicing. This type of interspeaker variation is also not reported for epenthetic processes in Turkish, but is typically found in studies of articulation and gestural timing.

5.5 Modeling apparent palatal harmony

Previous work on onset cluster repair in Turkish describes the inserted vowel as subject to palatal harmony, consistently surfacing as [+front] when the following lexical vowel is [+front]. However, Clements & Sezer (1982) do report that all inserted vowels following velar consonants are [-front] even if the following lexical vowel is [+front], and Kabak (2011) concurs. In addition, acoustic evidence from Bokhari et al. (2016) and the production study in this dissertation (Ch. 2-4) shows that inserted vowels before lexical /i/ are more centralized than lexical /i/ in the same position. In TELL, [+front] is not transcribed as spreading to the inserted vowel in most cases. Rather, the non-lexical vowel is usually transcribed as /ɯ/ (77% of transcribed insertions, Figure 5.1), even when harmony would demand another vowel. This corresponds to the cross-linguistic tendency for intrusive vowels to be schwa-like (Hall 2003),

7 Although factors like sonority, place, and voicing in clusters can affect epenthesis in other languages, such as Irish (Carnie 1994).

since /ɯ/ is the most schwa-like Turkish phoneme. Note that since the TELL transcriptions are essentially phonemic, the use of <ɯ> does not necessarily mean that the non-lexical was fully [+back], only that it was perceptually closer to /ɯ/ than to /i/.

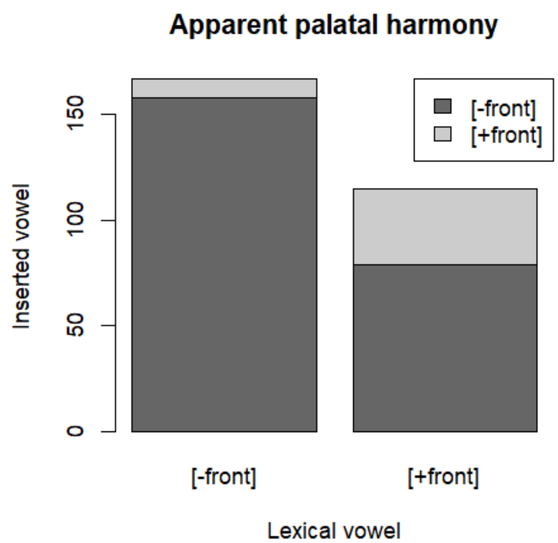


Figure 5.1: [+front] <v> in TELL

It is also of note that not every front <v> occurs before a front V_2 (Figure 5.1). A few front <v> tokens occur before a back lexical vowel, suggesting that the consonants in the cluster must play a role in determining the apparent quality of <v>. Based on Clements & Sezer (1982), we might expect that disharmonically front vowels can be inserted after /s/, ex. [s<i>por] when harmony would demand [s<u>por] or [s<ɯ>por]. However, there is no evidence of this phenomenon in TELL. Sixty-three of S1's 153 s-clusters are transcribed with insertion, but all vowels are transcribed as [+back] before

[+back] vowels. S2's 40 s-clusters are all transcribed without inserted vowels. Jonathan Washington (p.c., 2015) offers the impression that the insertion of <i> in s-clusters varies by dialect, and occurs more often in eastern Turkish[8]; if true, this may explain why it does not occur in TELL, since the TELL speakers are both natives of Istanbul. C&S do not report what part(s) of Turkey their informant(s) hail from, so there is no way to compare the possible influence of dialect here. However, I did find <i> insertion in text-setting elicitation sessions, reported in Ch. 6 (Section 6.4) and also discussed in Ch. 7. A dialectal study with more speakers could illuminate this question further.

Subsetting the data to determine where a front <v> occurred without a front V_2, I found that for S2, the non-lexical vowel can be [+front] even in the absence of a [+front] lexical vowel, if C_1 = [+labial] and C_2 = [+lateral] (8)—an effect that is completely unattested in the previous literature on onset cluster repair in Turkish. (This effect will be seen again in Ch. 6 and 7, and is discussed in Section 7.4.)

8 This does not necessarily indicate that eastern dialects of Turkish have epenthesis, since it could also reflect a slightly different gestural alignment in s-stop clusters, such that when the vocal tract opens sufficiently for a vowel to be heard, the tongue body is still fronted enough to make the intrusive vowel sound like [i].

(8) Disharmonic [+front] insertion, S2

	Citation form	*Orthographic form*
a.	[piłato]	plato
b.	[piłaka]	plaka
c.	[filama]	flâma
d.	[pilan]	plân
e.	[piłaj]	plaj
f.	[pixlatonik]	platonik
g.	[fiłaman]	flaman
h.	[pixlanør]	plânör
i.	[fiłoʃ]	floş
j.	[fiłaʃ]	flaş

To further explore these factors, I conducted a generalized mixed effects linear regression of the frontness of transcribed inserted v_1, and the frontness of the lexical vowel and whether C_1 was dorsal as fixed effects. When only a random intercept for speaker was included, the frontness of the lexical vowel seemed to have a very significant effect, but including random slopes for speaker eliminated this effect, showing that the two speakers again pattern differently. Therefore, I subsetted the data by speaker and conducted chi-squared tests of independence to assess the effects of the frontness of the lexical vowel, dorsal C_1, and lateral C_2 on the frontness of transcribed inserted vowels. All three factors were included for both subjects. With Bonferoni correction, the significance level was $p < 0.05/3 = 0.0167$.

For S1 (Table 5.7), when the following lexical vowel was front, the non-lexical vowel was significantly more likely to be front ($\chi^2(1) = 36.63$, $p < 0.0001$). A dorsal C_1 significantly decreased the likelihood of a front non-lexical vowel for S1 ($\chi^2(1) = 9.60$, $p = 0.0019$). A lateral C_2 did not significantly predict the frontness of S1's non-lexical vowels ($p > 0.5$).

Table 5.7: Chi-squared table for S1, palatal harmony

S1	C_1 [-dor]	C_1 [+dor]	$C_2 \neq$ /l/	C_2 = /l/	V_2 [+back]	V_2 [-back]
back <v>	110	69	133	46	120	59
front <v>	23	1	16	8	0	24

The pattern was quite different for S2 (Table 5.8), for whom the frontness of the lexical vowel did not significantly predict the frontness of the non-lexical transcribed vowel ($\chi^2(1) = 3.41$, $p = 0.065$). However, a front <v> was significantly more likely before a lateral C_2 ($\chi^2(1) = 11.81$, $p = 0.0006$), and significantly less likely after a dorsal C_1 ($\chi^2(1) = 10.33$, $p = 0.0013$).

Table 5.8: Chi-squared table for S2, palatal harmony

S2	C_1 [-dorsal]	C_1 [+dorsal]	$C_2 \neq$ /l/	C_2 = /l/	V2 [+back]	V2 [-back]
back <v>	41	24	55	10	43	22
front <v>	24	0	11	13	10	14

To summarize, the transcribed non-lexical vowel is less likely to be front after a dorsal C_1 for both speakers, in line with the claim in C&S and the results of my perceptual study. For S1 (but not S2), <v> is more likely to be front before a front lexical vowel, similar to the previous descriptions of the onset cluster repairing vowel being subject to regressive palatal harmony. Surprisingly, this effect of the following vowel was only significant for S1, not S2. For S2 (but not S1), <v> is significantly more likely to be front before a lateral, whether palatal or velar – an effect never reported before.

One possible explanation for the tendency for the ICI in S2's [pl] or [fl] clusters to sound [i]-like is a perceptual contrast effect. The transitions and bursts of labial consonants are associated with relatively low frequencies (Reetz & Jongman 2008), while voiceless labials have the strongest aspiration. Contrast effects (Rysling 2017) could lead to the high frequences inherent in aspiration being interpreted as a front vowel (high F2) since labials have low spectral weight (Amanda Rysling, p.c.). Another possible explanation is that the higher frequency of [pil]-initial words in Turkish relative to [pɯl]-initial words motivates S2 to produce the /pl/ sequence as more [pil]-like. Type counts from TELL are compatible with this hypothesis.

Table 5.9: Counts of lexemes in TELL

	pɯl[9]	pil	Cɯl*	Cil	Cɯla	Cila
Produced by S1	1	10	49	180	5	25
Produced by S2	0	4	14	112	2	5
Listed in TELL	2	10	99	312	12	35

However, this hypothesis predicts that insertion in other contexts should also be conditioned by the type frequencies of the resulting acoustic outputs, which does not seem to be the case. For example, there are no lexemes at all in TELL that begin with /Cɯro/, and yet acoustic insertion in Cro sequences usually produces an [ɯ]-like vowel. This will be discussed further in the next section.

With so few data points, any explanation of this pattern is necessarily speculative. But whatever the explanation for this pattern in S2's data, it is clear that fronting of the onset repairing vowel looks very different from regular palatal harmony in Turkish: V_2 has less of an effect and C_1 and C_2 have more of an effect. I return to the question of palatal harmony and /l/-driven <i> insertion in particular in Chapter 7.

9 Most of the [Cɯl] counts are inflated by 1 because *pılı pırtı* and *pılıpırtı* are listed as distinct lexemes.

5.6 Modeling apparent labial harmony

I now turn to labial harmony. Previous reports vary in their descriptions of when labial harmony affects the onset cluster repairing vowel. C&S (1982) indicate that it always applies, and Yıldız (2000) also does not discuss any exceptions. Yavaş (1980), however, finds that rounding of the onset cluster repairing vowel only occurs when the lexical vowel is high. Kaun (2000) also finds some support for this, along with interspeaker variation. I therefore looked for whether the rounding and height of the following lexical vowel affected the rounding of <v>. In addition, I expected rounding of <v> to be more likely when C_1 was labial, based on the finding in the acoustic study that F2 was lowered after /b/.

In analyzing the likelihood of rounding being transcribed on <v>, I initially used generalized mixed effects modeling, but abandoned this approach because there was not enough data to support random slopes or subsetting by speaker. Moreover, the chi-squared table contains zeros. So no statistical tests were conducted. Nonetheless, two tendencies are very clear. For both speakers, a round non-lexical vowel is only transcribed when the lexical vowel is round – i.e., there are zero instances where the inserted vowel is round but the lexical vowel is not round (Table 5.10). Even in the case where the lexical vowel is round, the inserted vowel is round only 35% of the time (111/315). In addition, the transcribed non-lexical vowel is much more likely to be round when the lexical vowel is high (Table 5.11), as reported by Yavaş (1980) and Kaun (1999). This fact has a potential articulatory explanation, which is that high vowels are produced with more dramatic lip-rounding than low vowels, and are therefore more likely to be perceived as round (Linker 1985, cited in Kaun 1995). Alternatively, this could reflect a phonological height agreement effect, as argued by Kaun (1999). However, no such effect occurs elsewhere in Turkish, where only the height of the target matters, not the height of the trigger.

Table 5.10: Rounding of V1

	Lexical V [-round]	Lexical V [+round]
<v> [-round]	1080	204
<v> [+round]	0	111

Table 5.11: Trigger height

	Lexical V [-high]	Lexical V [+high]
<v> [-round]	1040	244
<v> [+round]	5	106

In TELL, a labial C_1 did not increase the likelihood of rounding being transcribed on <v>. The TELL transcriber had access to the lexeme (orthographic representation of the word) and therefore knew that C_1 was was a labial, and therefore would be biased to attribute acoustic indications of rounding in <v> to C_1, and would therefore be less likely to perceive <v> as the source of any acoustic evidence of rounding (Amanda Rysling, p.c.)—another potential contrast effect (Rysling 2017).

5.7 Discussion

Modeling the variation found in TELL shows that (transcribed) insertion of a vowel is variable and appears to be subject to gestural factors like the degree of constriction (more likely when C_2 is sonorous) and the duration of VOT (more likely after dorsal C_1). A surprising result was that a non-lexical vowel is more likely after a fricative C_1 than a stop C_1. Spreading of frontness or roundness of the lexical vowel to the non-lexical vowel was highly variable, in contrast to previous reports (with the exception of Bokhari et al. 2016), but in accordance with the findings of the production study (Ch. 2-4). For one subject, the consonants in the cluster had a much stronger effect on the frontness of the non-lexical vowel than the quality of the following lexical vowel did. Factors that are irrelevant to phonological harmony elsewhere in Turkish (dorsal C_1, lateral C_2) play a significant role in predicting the frontness of the non-lexical vowel. These results support the gestural analysis of onset cluster repair, but would be difficult to account for under an epenthetic analysis.

This study was also a proof of concept that gestural factors can be inferred from corpus data. The results accord well with the acoustic and gestural studies. It is particularly striking that TELL, a relatively small corpus of largely phonemic transcriptions, could still show the influence of phonetic factors. Even more could be learned from a larger corpus, or a corpus of recordings. Studies of the variation in a corpus cannot replace careful recordings and acoustic and gestural analyses, but they can supplement experimental evidence in useful ways. A corpus offers a variety of word types that an experiment, with its carefully balanced design, may not be able to include. An experiment can inform the hypotheses to be tested in the corpus study. When, as here, experimental and corpus studies agree, we can be more certain of the interpretation of both sets of data.

Chapter 6: Syllabic and metrical status

6.1 Syllabic status of intrusive vowels

Cross-linguistically, intrusive vowels do not count as syllable nuclei for phonological processes that count syllables, such as templatic reduplication, allomorph conditioning, stress, and language games (Hall 2003: 44). For example, in Hocank, reduplication normally copies the final C(C)V syllable of the word as a suffix, or as an infix if the root is CV:C (9). (Reduplicants are bolded.)

(9) Hocank CV reduplication (data from Miner 1992, cited in Hall 2003)

a. gihú	‘swing’	gihu**hu**	‘wag tail’ M92:29
b. hit’é	‘speak, talk’	hit’é**t’é**	also ‘speak, talk’ M92:29
c. ratʃgã́	‘drink’	ratʃgã**tʃgã́**	‘drink repeatedly’ M92:29
d. No base form given.		sga:**sgá**p	‘sticky’ M79:29
e. No base form given.		ʒo:ʒók	‘slippery’ M79:29

But when the root contains an intrusive vowel, reduplication copies not just the final CV, but the whole C<v>CV sequence. This is the only time that reduplication in Hocank copies more than a CCV sequence.

(10) Hocank C<v>CV reduplication (data from Miner 1992, 1979, cited in Hall 2003)

a. -k<e>reʃ ‘make designs’ k<e>re**k<é>re**ʃ ‘spotted’

b. tʃ<i>wí̃ ‘reverberating’ tʃ<i>wí̃**tʃ<i>wĩ** ‘reverberating’

c. ʃ<a>rá ‘bald’ / ‘bare’ ʃ<a>ra**ʃ<á>ra** ‘bald in spots’

d. p<a>rás ‘flat’ p<a>ra**p<á>ra**s ‘wide’

If <v> in the C<v>CV sequence is not a phonological vowel, and the C<v>CV sequence is a monosyllable, then Hocank reduplication can be understood as always copying a monosyllable. In the ordinary case, this monosyllable contains only one vowel. But when vowel intrusion has occurred, the monosyllable contains a second acoustic vowel. This “vowel,” however, is not a phonological vowel at all, much less a syllable nucleus, but only the acoustic consequence of the gestural phasing of the surrounding C_CV. Hence, it is invisible to phonological processes, such as syllabification and reduplication.

The fact that intrusive vowels do not form syllabic nuclei also explains the fact that intrusive vowels typically do not count as syllables in poetic or musical meter. In Scots Gaelic, for example, /CVRC/ sequences are pronounced as [CVRV̲C]. Speakers of Scots Gaelic apparently can be reluctant to refer to a [CVRV̲C] sequence as a monosyllable, but they do count such sequences as a single syllable in the context of a larger word (e.g., *ʃ[œra]-ətœr’* ‘towel’ was syllabified as three syllables *ʃ[œra] – ət – œr’* by two consultants in Bogstrørm 1940). The treatment of [CVRV̲C] in poetry is also suggestive. O’Rahilly (1931, quoted in Hall 2003) reports that the [CVRV̲C] sequence counts for only one syllable in poetic meter: “in Scottish stress-verse…the epenthetic vowel is not recognized." While the intrusive vowel is invisible to syllable-counting, it does have an impact on the rhyme scheme: "the vowel preceding the consonant group rimes either with a similar vowel followed by a similar group (much as in scholastic verse) or, less commonly, with a simple long vowel" (O'Rahilly 1931:201, quoted in Hall 2003). This pattern of intrusive vowels not counting as syllabic holds not just in Scots Gaelic but also in other languages described in Hall (2003), where vowel intrusion appears to be highly regular and grammaticized.

In Turkish, however, vowel intrusion appears to be less grammaticized. By this, I mean that the phonological representation of words beginning with consonant clusters seems to vary across speakers and across lexical items. For instance, the pattern of interspeaker variation found in the production study (see Ch. 2 and 3) suggests that mastery of another language that has native complex onsets, such as English or French, may reduce the incidence of vowel intrusion in complex onsets that have been borrowed into Turkish. While each speaker's gestural phasing may be fairly stable, the fact that some speakers consistently produced shorter interconsonantal intervals than others suggests that the target gestural phasing varies across speakers. This interspeaker variation may be related to the fact that complex onsets—and the vowel intrusion that affects them—are limited to loanwords. As a result, a Turkish speaker's mental representation of these words is affected by factors such as their awareness of the foreign source of the word and their competence in the source language, as well as their attitudes toward the relationship of the source language to Turkish. Some of my consultants explicitly told me that it was “not correct” to pronounce consonant clusters with an inserted vowel. They also expressed an awareness that “some people,” “other people,” “the younger generation,” or people from elsewhere in Turkey might pronounce the words differently.

A related factor in speakers’ mental representation of loanwords with onset clusters is the prescribed spelling of the loanwords without an inserted vowel. I am told that Turkish children learning to write typically include the non-lexical vowel, and are taught not to in school. Spellings that include the inserted vowel (e.g. <kulüp> for prescriptive <klüp>, <gurup> for prescriptive <grup>) can be found sporadically in the names of businesses and in personal correspondence. However, at least among educated Turks living in the United States, the prescribed orthography definitely influences the treatment of words with complex onsets. Several of my adult consultants appealed to the orthography (“We don’t write a vowel there”), as well as explicitly referencing the words’ status as foreign borrowings, as a basis for the “correct” pronunciation, the syllable count, and for the unstable or exceptional behavior of the words.

In addition to variation among speakers, there is also variation among lexical items. While vowel intrusion is clearly an active process that applies in novel nonce words as well as familiar real words, among familiar real words, certain pronunciations are conventionalized. The production study (Ch. 2, 3) did not reveal any systematic differences between novel nonce words and familiar real words, but a survey of vowel insertion in Turkish music (this chapter) makes it clear that vowel insertion is much more strongly established in certain real words. For example, *kral* 'king' is consistently pronounced as [kɯral], while *problem* 'problem' is often produced without a vowel. Even the quality of the inserted vowel may be somewhat lexicalized: *plastik* 'plastic' is usually pronounced as [pilastik] with an <i>-like intrusive vowel, but *plan* 'plan' which offers a superficially similar environment for vowel intrusion, can be either [pilan] or [pɯlan]. This suggests a degree of lexicalization in Turkish onset cluster repair that I am not aware of in other cases of vowel intrusion.

If vowel intrusion in Turkish is not fully grammaticized, and is understood by different speakers to be at different phases in the process of phonologization, this uncertainty and variation may have consequences for the syllabic status of the intrusive vowel, particularly so if the intrusive vowel is "more present" in some words than in others. This chapter probes the syllabicity of non-lexical vowels in Turkish onset clusters. First, I report the results of a pilot syllable-counting task, which suggest that Turkish speakers' judgements of syllable counts are shaped by orthographic, rather than purely phonological, considerations. To side-step the prescriptive and orthographic influences that are inherent in an explicit syllable counting task, I then examine the text-setting of onset clusters in Turkish music. In a corpus of over forty songs containing onset clusters, the non-lexical vowel does not occupy a metrical slot in about half the onset clusters. In the remaining half, the non-lexical vowel does occupy a metrical slot. This metrical treatment contrasts markedly with the text-setting of lexical vowels in corresponding metrical positions, which overwhelmingly occupy their own metrical slots. Finally, I supplement this naturally occurring musical data with a more controlled data set, consisting of judgements of how to text-set various words into the Turkish song “Iyi ki doğdun” (“Happy Birthday To You”) by five Turkish speakers. Like the musical corpus study, the “Iyi ki doğdun” study shows that non-lexical vowels in Turkish onset clusters do not have the same metrical status as lexical vowels, and provides further evidence that different speakers have different mental representations of non-lexical vowels and onset clusters.

6.2 Syllable count judgments

In an informal data-gathering session, I asked some Turkish speakers to judge the number of syllables in a few words beginning with onset clusters.

Participants were a group of four linguistically naive Turkish speakers in the Santa Cruz community, comprised of two husband/wife pairs (M & N, S & E). Syllable count judgments were mixed and clearly affected by prescriptive notions of what constitutes a syllable. Speaker M initially said that *prens* [pirens] and *kral* [kɯral] were disyllabic. Speakers S and E then asserted that M was wrong, and *prens* and *kral* were monosyllablic. S and E argued that the definition of a syllable is a vowel, and since *prens* and *kral* are spelled with only one vowel, they were by definition monosyllabic. Faced with this argument, M lost confidence in his judgement of bisyllabicity. The fourth speaker, N, was reluctant to take a side. The group's discussion made it clear that syllable count judgements are shaped by explicit definitions of syllables that are prescriptively tied to the highly phonemic Turkish orthography, which does not represent vowels inserted into onset clusters (in most cases). Indeed, Turkish school children are taught to count the vowels (the *sesli* 'sound-ful' letters) in a word as a way to count its syllables. The rule orthographic *vowel* = *syllable* is so strong that one consultant insisted that his name *Saadeddin* [sa:.ded.din] should be syllabified as *Sa-a-ded-din,* treating orthographic <aa> as indicating two separate syllables, rather than the single long vowel it is pronounced as. The influence of prescriptive orthography biases speakers in the direction of not counting the inserted vowels as syllabic, and almost certainly contributes to maintaining the status of onset-repairing vowels as intrusive, rather than being reanalyzed as epenthetic or underlyingly present.

6.3 Text-setting in a corpus of Turkish music

This section probes the syllabic and metrical status of non-lexical vowels in Turkish onset clusters through a study of their treatment in current Turkish music. A long tradition links musical text-setting to phonological structure (Lerdahl & Jackendoff 1983, Hayes & Kaun 2001, Fabb & Halle 2008, McPherson & Ryan 2018). Musical evidence, consequently, has also been brought to bear on questions of syllabicity. For example, Ng (2017) brings evidence from text-setting of English words like "fire" and "tile" to bear on the question of the number of syllables that dipthong + liquid sequences comprise in English. Ng finds that the number of pitches assigned to these so-called sesquisyllables (Lavoie & Cohn 1999) varies across singer/song-writers, in accordance with Lavoie & Cohn's (1999) suggestion that such sequences represent unstable moraic structures, which are well-formed syllables according to general phonological principles (e.g., the Sonority Sequencing Principle) but violate language-specific moraic rules. These unstable moraic structures are expected to be assigned different mental representations by different speakers, similar to current findings for non-lexical vowels in Turkish onset clusters. Tilsen

& Cohn (2016) find that speakers who judge words as monosyllabic produce them with a shorter rime duration than those who judge them to be disyllabic. A similar finding could obtain for Turkish onset clusters, such that speakers who judge a word like *prens* [pirens] to be disyllabic would produce a longer <i>.

To the extent that vowels in onset clusters are syllabic, they are expected to be treated like lexical vowels: to consistently receive their own beat and/or pitch in music. On the other hand, if onset repairing vowels in Turkish are intrusive, they are expected to be treated differently from lexical vowels. The current study finds that non-lexical vowels in Turkish onsets may optionally receive a beat, but never receive more than one beat, unlike lexical vowels. Moreover, non-lexical vowels often do not receive a beat, while lexical vowels in equivalent metrical positions rarely fail to receive a beat. This finding demonstrates that Turkish text-setters differentiate between lexical and non-lexical vowels, ascribing less metrical "space" to the non-lexical vowels.

6.3.1 Method

Forty songs in Turkish whose lyrics contain one or more words beginning with onset clusters were collected by searching the website <https://www.xsarkisozleri.com/>, which includes a collection of Turkish song lyrics. In some cases, only song titles, not lyrics, were available on Xşarkı Sözleri, so lyrics were found on other websites, with the aid of Google Search. A full list of song titles, artists, and lyrics is available in the Appendix. The resulting corpus contains 107 tokens of onset clusters, occurring in 78 metrical lines in forty songs. The total number of tokens is so much larger than the number of lines or songs because many songs and lines contain multiple tokens of onset clusters (e.g., the rap song "Sıfır Sıkıntı" by Istanbul Trip contains *trip, speedy, stil, street, grinder, kral, bro,* and *flextedi)*. Furthermore, many clusters occur in lines that repeat multiple times in the course of one song (most notably "Bla bla bla" by Ahmet & Aslı Jackson, whose chorus "bla bla bla" repeats ten times, for thirty tokens of "bla"). Because of repetitions like this, the analysis focuses on distinct lines, rather than individual tokens. Table 6.1 shows the number of lines containing each type of cluster in the corpus. Since the corpus consists of naturally-occurring data, it is not balanced for different types of clusters, and certain words are overrepresented. Most words occur in only one song, but some occur in several songs (e.g., *kral* 'king' – six songs; *prenses* 'princess' – five songs; *problem* 'problem' – five songs).

Table 6.1: Clusters

C_2	C_1 LAB	# lines (# songs)	C_1 COR	# lines (# songs)	C_1 DOR	# lines (# songs)
r	pr	13 (10 songs)	tr	11 (5 songs)	kr	18 (9 songs)
	br	6 (6 songs)	dr	3 (1 song)	gr	5 (5 songs)
	fr	4 (4 songs)				
l	pl	7 (7 songs)	sl	1	kl	1
	bl	3 (2 songs)				
	fl	1				
other	ps	1	sp	3 (3 songs)		
			st	2		
			sn	1		
Total		33		21		24

When possible, I identified a line that metrically corresponded to the line in which the CC word occurred, in that it rhymed with the CC line, was performed to the same melody, or ideally, both. For example, the pop song "Hayat Belirtisi" by Cem Belevi contains the CC word *gram* 'gram' in the first line in (11) (underlined, onset cluster bolded). The next line rhymes with the first (both end with a progressive verb ending in *-yor*), and was sung to the same tune. Hence, I treated the second line as the metrical correspondent for the first line. For two lines, no metrically corresponding line could be identified, so correspondents were omitted.

(11) *Corresponding lines in "Hayat Belirtisi"*

gram uyku girmiyor   'not a gram of sleep enters'
kelimeler yürümüyor   'words walk'

Within each metrically corresponding line, I identified the word whose position best corresponded to that of the CC word—i.e., if the CC word was last in its line, the last word in the corresponding line was the best metrical correspondent for the CC word; if it was the third word in the line, the third word in the corresponding line was deemed to best correspond to the CC word. In "Hayat Belirtisi," the CC word *gram* is the first word in its line, so its metrical correspondent was the first word in the next, corresponding line: *kelimeler* 'words' (underlined in (11)). Since the CC-corresponding word *kelimeler* occupies more of the corresponding line than the CC word *gram* does of its line, I grouped *gram* with the following word *uyku* 'sleep' (underlined) for purposes of comparing how many vowels are in CC words vs. metrically corresponding words.

Finally, I identified what I refer to as the corresponding lexical vowel for the non-lexical vowel in the CC word. The corresponding lexical vowel is a lexical vowel that occurred in as similar a metrical position as possible to the non-lexical vowel in the corresponding CC word – namely, the first vowel in the metrically corresponding word. For the non-lexical <ɯ> in *gram*, the corresponding lexical vowel is the first /e/ in the corresponding word *kelimeler*.

Since all the target onset clusters were word-initial, non-lexical vowels were always the first vowel in the word. The first vowel of the corresponding word, then, acted as a lexical vowel control for the non-lexical vowels. Any differences in the metrical treatment of the lexical vowel controls vs. the non-lexical vowels in the same song can be assumed to come from properties local to the particular vowels in question, since their metrical surroundings are matched.

Songs were annotated in Praat using TextGrids. If the same line repeated multiple times with the same words, rhythm, and tune, then only the first instance was annotated. For one song ("Hoşgeldin" by Fatih Aydın), mp3s could not be found, so TextGrids were not made. Evaluation of meter was made on the basis of Youtube videos instead. For "Bla bla bla", only the first "bla" in each of two metrically distinct lines was counted (i.e., there are two "bla" lines included in the count, not six).

For each CC word, I annotated the number of beats falling within the non-lexical vowel v_1, the preceding and following vowels (V_0 and V_2 respectively), and the lexical control vowel in the corresponding word. The number of beats was evaluated by listening to the preceding several measures of the song and tapping along steadily to the beat. I then continued tapping or nodding to the beat while the CC line and metrically corresponding line played, and noted the number of beats that fell on the non-lexical, complex-onset repairing vowel, and on the corresponding lexical vowel. In cases of ambiguity, I consulted a native speaker of Turkish for judgements on the number of beats per vowel. In a few cases, confirmation of the relative number of beats was obtained by comparing the relative durations of corresponding lexical and non-lexical vowels.

6.3.2 Results of the musical corpus study

Comparing the number of beats received by non-lexical vowels in CC words to the number of beats given to vowels in the same position in their correspondents, we find that non-lexical vowels can be set to a beat (48/82 tokens), but often are not (34/82 tokens) (Figure 6.1). In contrast, corresponding lexical vowels are almost always set to at least one beat (79/80 tokens), and only rarely fail to receive a beat (1 token). Furthermore, it is not uncommon for lexical vowels to be set to two or more beats (12 tokens). This never occurs for non-lexical vowels in this sample. The difference between the number of beats assigned to lexical and non-lexical vowels is significant in a chi-square test ($\chi^2(2) = 46.2$, $p < 0.001$).

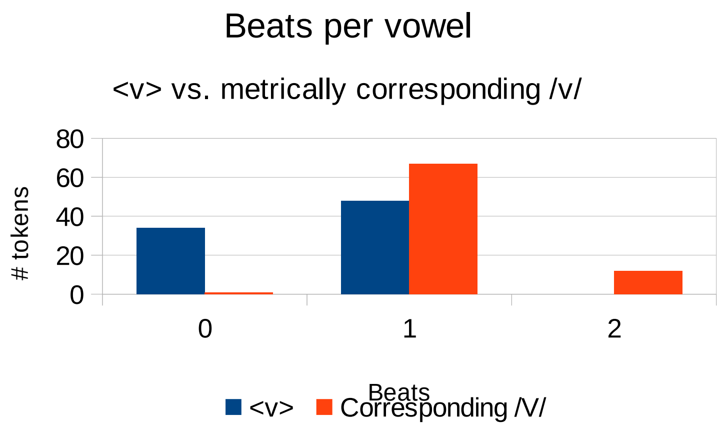


Figure 6.1: Beats per <v> vs. /v/

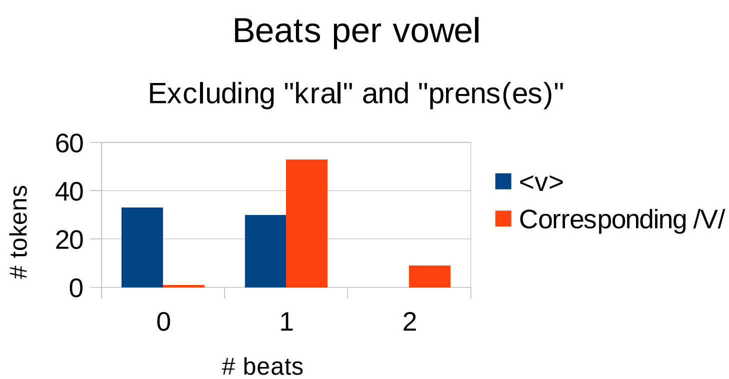


Figure 6.2: Beats per vowel (without *kral* and *prenses*)

In certain common words in this corpus, the inserted vowel is always present, and is almost always set to a beat. These words are: *kral* [k<ɯ>ral] ‘king’ (11/11 tokens are set to 1 or 2 beats), *prens* [p<i>rens] ‘prince’ (3/3 tokens set to a beat), *prenses* [p<i>renses] ‘princess’ (4/5 tokens are set to a beat). If we exclude all tokens of these three words, the tokens where <v> gets no beat slightly outnumber those where <v> gets a beat (33 zero-beat tokens vs. 30 one-beat tokens) (Figure 6.2), making the contrast between non-lexical vowels and their lexical, metrical correspondents even starker ($\chi^2(2) = 45.0$, $p < 0.001$). In tokens where <v> receives a beat, the metrically corresponding word usually contains one more lexical vowel than the CC word; where <v> does not receive a beat, the corresponding word typically has the same number of lexical vowels as the CC word. On average, the corresponding word has 0.4 more lexical vowels than the CC word, as expected given that non-lexical vowels add another metrical vowel to the CC word about half the time.

The lexical correspondent vowels discussed in the comparisons above were chosen to control for as many metrical and musical factors as possible, to provide the best comparison of lexical and non-lexical vowels. However, it may also be interesting to look at the number of beats given to the vowels that immediately precede and follow the onset cluster, V_0 and V_2 respectively (Figure 6.3), bearing in mind that these offer a much less controlled comparison. Since the onset-repairing vowel always occurs in the first syllable of the CC word, its immediate predecessor V_0 is invariably the final vowel in the word that precedes the CC word, and hence generally bears lexical stress, which accounts for the cases where it receives many beats (the maximum in this sample was 5 beats). V_2 is always the second vowel in the CC word, and like V_0, can sometimes be the nucleus of the word's final syllable, and therefore stressed. For <v>, V_0 and V_2, the most common metrical treatment is to receive one beat. The non-lexical

<v> often receives no beats, however, as noted above. This very rarely occurs for V_0 and never for V_2, just as already observed for the corresponding lexical vowels seen above. Meanwhile, V_0 and V_2 can receive two or more beats, which <v> cannot.

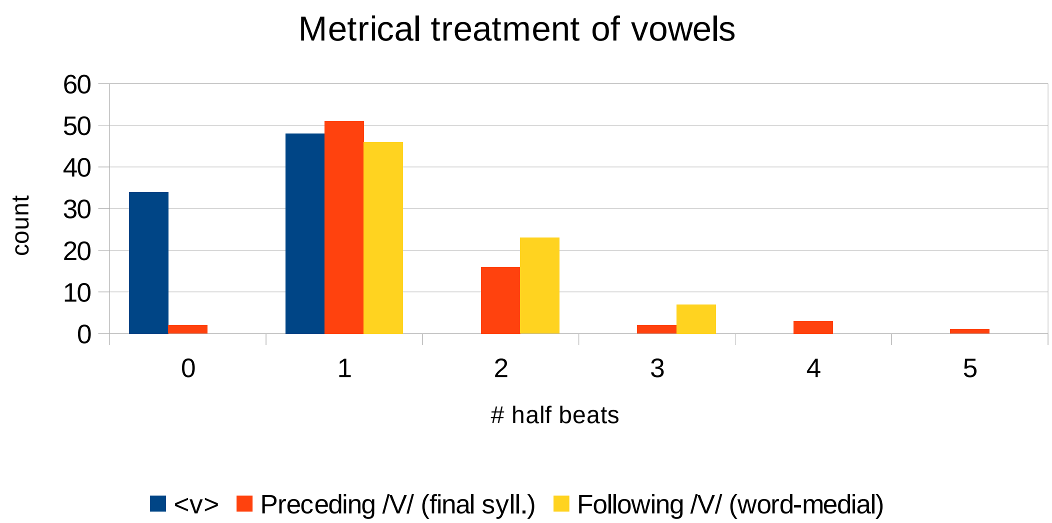


Figure 6.3: Beats for <v> vs. preceding and following lexical vowels

6.3.3 Case study: "Drama Köprüsü"

The traditional folk song "Drama Köprüsü," performed by Suavi, provides an interesting case study of the singer's ability to give the same word a different metrical treatment, depending on the demands of the remainder of the line. The song is a tragic ballad, describing the fate of a man named Hasan who is slain near the aqueduct (Turkish *köprü* 'bridge') of the town of Drama, in the present-day East Macedonia and Thrace region of Greece. Each line of the song contains eight syllables (i.e., eight vowels for lines with lexical vowels only), and is comprised of four measures of five beats each.

The CC word *Drama* appears in two different lines, shown in (12) and (13), where each beat is represented with “x”, and measures are represented by parenthesized groupings of five “x”s. According to my Turkish musical consultant, the first two beats in the measure and the last three beats in the measure form metrical sub-groupings, so this is also represented in the metrical schematization by separating them with a hyphen: (xx – xxx).

(12) (xx – x xx)$_1$ (xx - xxx)$_2$ (xx-xxx)$_3$ (xx - xxx)$_4$
(# - D<ɯ>ra)$_1$ (ma - köp)$_2$ (rü - sü)$_3$ (Ha – san)$_4$

(13) (xx – xxx)$_1$ (xx – xxx)$_2$ (xx - xxx)$_3$ (x x-xxx)$_4$
(Dra – ma)$_1$ (mah – pu)$_2$(sun – da)$_3$ (hey dost)$_4$

The first line containing *Drama* (12) contains seven lexical vowels. *Drama* is produced with a vowel between /d/ and /r/. If this non-lexical vowel counts toward the meter, then the line has eight syllables like all the others; if we do not count the non-lexical vowel for the meter, the line is unaccountably one syllable short of the pattern in the rest of the song. In fact, the non-lexical vowel in *Drama* does receive a beat, in accordance with its role as one of the line’s eight vowels. At the same time that the non-lexical vowel is counting for the meter, however, it is also counting for less than the following lexical vowel /a/, which receives two beats. This is accomplished by starting the line with a rest (represented by # in (12)), which allows the non-lexical vowel in *D<ɯ>rama* to be set to one beat, instead of taking up two beats at the start of the measure. This suggests the text-setter was reluctant to give a non-lexical vowel two beats, perhaps reflecting an awareness that these onset-repairing vowels are shorter than lexical vowels.

In contrast to the first line, where the non-lexical vowel contributes to the syllable count, the second line containing *Drama* (13) contains eight lexical vowels, meaning the non-lexical vowel in *Drama* does not contribute to the meter. Indeed, the vocalic interlude between /d_r/ is much shorter in this line, and does not receive its own beat; instead, the first two beats of the line's first measure go to the lexical /a/.

Comparing the treatment of Drama in (12) and (13), we see that the same singer, in the context of the same song, can choose to give the non-lexical vowel a beat or not. It is also interesting that in the line where the non-lexical vowel receives a beat (12), it also counts toward the syllable quota for the line (the line has eight syllables only if the non-lexical vowel counts for one), but that when the non-lexical vowel does not receive a beat (13), it also does not count toward the number of syllables (eight lexical vowels). Even when the non-lexical vowel does count for the meter, however, the singer arranges the line to keep the non-lexical quantitatively shorter than the nearby lexical vowels. This song exemplifies the pattern seen throughout the corpus where lexical vowels can be set to one beat (as with *hey* in (13)), two beats (as with *mah*), or more (*dost* gets four beats), while non-lexical vowels only get one or zero half-beats.

6.4 Text-setting in “Iyi ki doğdun”

The corpus study in the previous section has the advantage of being naturally occurring data. The downside, though, is that it does not enable a direct of comparison between the text-setting of words with different numbers of syllables in the same musical context. As a supplement to this natural but uncontrolled data, therefore, I conducted a study of the text-setting of words with different numbers of syllables, with and without onset clusters, in the same metrical slot. As the musical context, I used the line in the Turkish equivalent of “Happy Birthday To You” where the birthday-person’s name is usually inserted. The Turkish song is known as “Iyi ki doğdun,” and is sung to the same tune as “Happy Birthday To You.” The lyrics are shown in (14).

(14) Iyi ki doğdun, [name]. ‘It’s good you were born, [name]’
Iyi ki doğdun, [name].
Iyi ki doğdun, iyi ki doğdun,
Iyi ki doğdun, [name].

The slot where the birthday celebrant’s name is to be inserted corresponds to the “to you” of “Happy Birthday To You”; it is sung on two notes of equal duration.

Data was elicited from five native speakers of Turkish living in the United States (Speakers O, Sa, E, all in their thirties, and F and Se, in their early sixties). None of them participated in the production study, but Sa and E provided syllable count judgements during a separate session (discussed in Section 6.2). Speaker O is also a fluent speaker of American English, having lived in the US since the age of 9. The other four speakers came to the United States as adults (age $>$ 22). Speakers O, Sa, and E were recruited from the Santa Cruz community. F and Se reside in Miami, Florida, and were recruited through family connection with O. The five speakers represent different geographical regions of Turkey as well: Sa is from Istanbul; E is from Bolu in northwestern Turkey; and O, F, and Se are from the capital city Ankara in central Turkey.

Materials were a list of 64 Turkish words, 42 beginning with a complex onset (CC words) and 24 beginning with simplex onsets (C words). CC words contained one, two, or three lexical vowels; C words contained up to four lexical vowels. Each set of words was chosen to represent onsets produced at all major places of articulation (labial, coronal, dorsal). The list of words was initially blocked by the number of vowels in the word (Speakers O and E); for later participants (Sa, F, Se), the words were presented in alphabetical order so as to eliminate this blocking effect, after E remarked that she thought it might have influenced her to be more consistent within each block.

Participants were interviewed individually. Speaker O was interviewed in person at home, E in an office, Sa in a cafe, and F and Se over the phone while they were at home. Participants were told the researcher was interested in the way that words were sung in the song “Iyi ki doğdun,” and warned that while some of the words would be names, others would be random words that would sound silly in the song. They were then asked to sing a line or two of “Iyi ki doğdun, __,” slotting the target word into the blank. All participants were very familiar with the song before beginning the task. As they sang the song, I wrote down impressionistically how the target word was divided across the two notes of the metrical slot. In cases of uncertainty about the text-setting, participants were asked to re-sing the line, or I sang the line back to confirm it had been recorded correctly. In cases of uncertainty about the quality of the inserted vowel, participants re-said the word, hyper-articulating the inserted vowel. Due to the nature of the task, the non-lexical vowels were produced in a way that sounded impressionistically like a full vowel, in contrast to the brief and schwa-like vowels heard in the acoustic and ultrasound production experiment (Ch. 2-4).

6.4.1 Predictions

Broadly speaking, there are three possible patterns for how non-lexical vowels could be treated. These patterns are exemplified by the text-setting of CC words with one lexical vowel and potentially one non-lexical vowel, such as *prens* 'prince'. If the non-lexical vowel counts toward the meter, words like *prens* should pattern with disyllables. If it does not count toward the meter, *prens* should pattern with monosyllables. Finally, words like *prens* could optionally pattern either with monosyllables or disyllables, perhaps with different preferences based on characteristics of particular words.

Furthermore, I hypothesize that non-lexical vowels after dorsal C_1 are more likely to function metrically like a lexical vowel, since they tend to be longer (Ch. 3, 4) and are more likely to be transcribed in TELL (Ch. 5). Conversely, non-lexical vowels after a labial C_1 tend to be shorter (Ch. 3, 4), and may be less likely to function metrically like lexical vowels.

6.4.2 Results

The pattern of text-setting for non-lexical vowels varied considerably from speaker to speaker, and all three predicted patterns were represented. For Speakers E and Sa, words like *prens* consistently acted like disyllables, and could not be treated like monosyllables. For Speaker F, the opposite was true: *prens* could only pattern as a monosyllable. Finally, Speakers O and Se could optionally treat *prens* and its ilk as either monosyllables or disyllables. These results are explored in greater detail below.

6.4.2.1 Text-setting of simplex onset words

For words containing no onset clusters, there was near-total consensus among subjects. When a one syllable word, such as *pul* 'stamp', is inserted into the [name] slot of "Iyi ki doğdun," the sole vowel in the word is lengthened to stretch the word across the two metrical slots (15)a. For a two syllable word like *barış* 'peace', each syllable is allocated to one note / metrical slot (15)b. When there are three lexical vowels, the first two together occupy the first note, and the final stressed syllable occupies the second note (15)c. For four lexical vowels, two vowels can fall on each note, or three can fall on the first note and the stressed syllable may occupy the second note (15)d.

(15) Text-setting of lexical vowels in "Iyi ki doğdun"

a. 1 σ:	pul	'stamp'	pu – ul
b. 2 σ:	barış	'peace'	ba – rış
c. 3 σ:	Batuhan	(name)	Batu - han
d. 4 σ:	Saadeddin	(name)	Saaded – din
	Semiramis	(name)	Semi – ramis ~ Semira – mis

The one exception to this consensus was that Speaker F said a monosyllable could not be lengthened to fill out the slot in the song. Instead, F sang the monosyllable on the first note, and left a silent rest (#) for the second note (16). F’s treatment of polysyllabic words did not diverge from the majority’s treatment, however.

(16) Speaker F’s text-setting of monosyllables

a. pul ‘stamp’ → pul – #

b. kol ‘arm’ → kol – #

c. su ‘water’ → su – #

In short, the survey revealed that the basic pattern is for the final syllable to fill the second note, and for remaining syllables to land on the first note. To answer the question of whether inserted vowels are text-set as separate syllables or as a sub-part of another syllable, then, CC words with only one lexical vowel provide the best data. This is because in longer words, whether the non-lexical vowel is its own syllable or not, all the non-final lexical vowels will generally be grouped with it on the first note. The remainder of the discussion, therefore, focuses on these CC words that contain one lexical vowel, comparing them to unambiguous monosyllables and unambiguous disyllables.

6.4.2.2 Text-setting of complex onset words

For words beginning with onset clusters, more variation emerges. Speakers E and Sa almost always insert a vowel, and text-set it just like a lexical vowel. That is, the non-lexical vowel occupies the first metrical slot, either by itself (if there is one other vowel in the word), or along with the other non-ultimate vowels in the word (if there are two or more lexical vowels).

(17) Text-setting of non-lexical vowels – /V/-like treatment by E and Sa

a. prens	‘prince’	pi – rens,	* pre – ens
b. fren	‘brake’	fi – ren,	* fre – en
c. stil	‘style’	si – til,	* s(i)ti – il
d. spor	‘sport’	si - por, sɯ - por,	* spo-or
e. gram	‘gram’	gɯ – ram,	* gra – am
f. gri	‘gray’	gɯ – ri,	*/? gri – i
g. bre		E: * bi – re,	bre – e
		Sa: bi – re,	* bre – e
h. bro	‘bro’	E: ? bu – ro,	* bro – o
		Sa: bu – ro	? bro – o

For almost all the words, E and Sa had the same judgements, although for Speaker Sa, alternative text-settings where the non-lexical vowel did not occupy a metrical slot were slightly more acceptable than for E. Sa also made a point of saying that he would not necessarily pronounce these words with the inserted vowel in regular speech, but that he was giving me the best or correct way to “fill out the song.” For E, giving the inserted vowel a metrical slot was almost entirely obligatory; the only CC word with one lexical vowel that she was able to text-set without an inserted vowel was *bre* (17)g. It is of note that this exception involves a labial C_1, since evidence from other studies reported in previous chapters suggests that clusters with labial C_1 were less prone to vowel intrusion in Turkish. The /br/ clusters had shorter interconsonantal intervals in the production experiment, and were less likely to have $>$20 ms of vocalic material than /gr/ clusters. Sa and E’s text-setting diverged for *bre*, where they had the opposite judgements about whether vowel insertion should occur. Their judgements also diverged somewhat for *bro*, which several consultants (including E) said was "not Turkish"—which implies that they do view the other loanwords as assimilated into the Turkish language. E's reaction to seeing *bro* on the list was, "It's not Turkish, but I can do it." In the end, however, she said that there was no good way to fill the name slot in the song with *bro.* Sa, though, judged two text-settings of *bro* as acceptable (17)h.

To sum up, for two speakers out of five, an inserted vowel in a CC word in the name slot of "Iyi ki doğdun" acted like a lexical vowel. In a few cases, it was acceptable not to insert the vowel, but in general, insertion was deemed obligatory, and the optimal text-setting for E and Sa gave the non-lexical vowel a claim to a piece of the meter.

On the opposite end of the spectrum, we find Speaker F, for whom it is not possible to give the non-lexical vowel a metrical slot in rising sonority clusters. Words like *prens* must pattern like monosyllablic *pas* 'rust'. This is true even for *kral* 'king,' which is standardly pronounced with a fairly long non-lexical [ɯ] breaking up the /kr/ cluster. *Kral* is such an old borrowing that its intrusive vowel always receives a beat in my corpus of music (Section 6.3). Another particularly striking example is *plan* 'plan' (18)b, where F allows insertion but still requires the whole C<v>CV sequence to fall on the first note.

Since this is how F treats monosyllables, and not how she treats disyllables, this text-setting provides evidence that F does not consider the intrusive vowel to be syllabic. This is exactly the kind of text-setting behavior that is predicted for intrusive, non-syllabic vowels. Furthermore, F spontaneously said that words like those in (18) are "one thing"—atomic, impossible to split up. This metalinguistic remark suggests that F considers such sequences to be monosyllables.

(18) Text-setting of non-lexical vowels by Speaker F

a. prens	'prince'	prens – #,	* pi – rens
b. plan	'plan'	plan – #, pilan – #,	* pi – lan, * pla – an
c. gri	'gray'	gri – #,	* gɯ – ri, * gri – i
d. kral	'king'	kral – #,	* kɯ – ral
e. stil	'style'	si – til	
f. spor	'sport'	spor – #, sɯ – por,	* si – por

Intriguingly, however, F does allow the non-lexical vowel to occupy the first metrical slot when the vowel breaks up a /s/+stop cluster (18)e-f. I speculate that the vowels in words like *spor* and *stil* might not be intrusive vowels for F, but could be epenthetic instead. Further research is required here.

Intermediate between E/Sa who treat *prens* as a disyllable, and F who treats *prens* as a monosyllable, the other two speakers, Se and O, could optionally text-set words like *prens* either as monosyllables or as disyllables (19).

Where multiple pronunciations are acceptable, they are listed in the order the speakers provided them.

(19) Text-setting of non-lexical vowels by Speakers Se and O: optionality.

		O	Se
a. prens	'prince'	pi – rens, pre – ens	pi – rens, ?pre – ens
b. fren	'brake'	fi – ren, ?fre – en	
c. plaj	'beach'	pi – laj, ?pla – aj	pla – aj, ??pi – laj
d. bre	(title)	bre – e, * bi – re	bi – re, bre – e
e. bro	'bro'	bro – o, *bu – ro	bu – ro, bro – o
f. stil	'style'	sti – il, si – til	sti – il, *si – til
g. spor	'sport'	sɯ – por, si - por, spo-or	si – por, ??spo-or
h. gram	'gram'	gɯ – ram, ??gra – am	
i. gri	'gray'	gɯ – ri, */? g(ɯ)ri – i	
j. kral	'king'	kɯ – ral, */? k(ɯ)ra – al	

Speakers O and Se could either insert a vowel to break up the onset cluster, and slot the non-lexical vowel into the first metrical position, or they could lengthen the lexical vowel to span both metrical positions. Which option was preferred depended on the particular lexical item, and appears to be tied to the identity of the initial consonant: for all word with a dorsal C_1, there was a strong preference to give the non-lexical vowel its own beat (19)h-j. Preferences were divided for the labial C_1s, and varied both among lexical items and between speakers: *fren* is produced with <i> in the first metrical slot, *prens* is acceptable either way; *bre* cannot be produced as [bi – re] for O, but could be for Se; *plaj* cannot be [pi – laj] for Se, but can be for O.

This optionality, where the non-lexical vowel and the lexical vowel can co-occupy the first metrical slot, is not seen in words with two lexical vowels. For instance, *Dilar* can only be set as [di – lar], not as *[dila – ar].

When a CC word contains two lexical vowels, inserting a vowel becomes less preferable (although it is still an option), and the inserted vowel cannot occupy an entire metrical slot by itself, since that would require the two lexical vowels to be compressed into one beat. Instead, the final stressed syllable takes up the entire second metrical slot, and the first lexical vowel, along with the non-lexical vowel if there is one, occupy the first metrical slot: *prenses* 'princess' can be [pren-ses] or [piren-ses], but not *[pi-renses]. The asymmetry between places of articulation observed for the one-vowel CC words vanishes here; all two-vowel CC words follow the same pattern regardless of the identity of C_1.

6.4.2.3 Summary of "Iyi ki doğdun" results

Eliciting text-setting judgements from five Turkish speakers in the well-known song "Iyi ki doğdun" revealed that speakers vary a great deal in their text-setting of non-lexical vowels. Even parents and children, or husband and wife, could have systematic disagreements. The full range of predicted possible patterns was instantiated in this small sample of speakers. Two speakers treated the inserted vowels like lexical vowels, one ignored them for syllabification and text-setting, and two allowed both options. These differences may reflect differences in the mapping between phonology and text-setting for different speakers. But they also correlate to some extent with meta-linguistic remarks on the prosodic size of the words. F, who ignores the non-lexical vowels in text-setting, considers a word like *prens* to be metrically atomic. O, who has optionality in his text-setting of non-lexical vowels, judges *prens* to have "one and half syllables." And E requires the non-lexical vowels to occupy a metrical slot, suggesting that, contrary to her appeals to orthography, the non-lexical vowel may act as syllabic for her. I interpret this to mean that the variation in text-setting does reflect differences in phonological representations.

In addition, the differences in the treatment of individual words by the same speaker provide further evidence that vowel intrusion in complex onsets in Turkish is subject to lexically specific variation. This variation may be partly conditioned by historical factors, with older borrowings like *kral* being more assimilated into Turkish and hence more likely to be produced with an inserted vowel, and newer ones like *bro* being less assimilated, hence less likely to be produced without a vowel. However, it seems clearer that articulatory factors play a role: clusters with a dorsal C_1, which leads to a longer ICI, are more likely to be produced with an inserted vowel that behaves in text-setting like a lexical vowel; clusters with a labial C_1, which yields a shorter ICI, are less likely to be text-set as if they had an additional lexical vowel.

The variation among and within speakers in this study echoes the findings in Tilsen & Cohn (2016), who elicited syllable-count judgement from 28 English speakers for words like *pier, veal, pyre,* and *vile*, and also obtained audio recordings of those same speakers producing those words. They found inter-speaker variation, but also variation within speakers; 19% of syllable count judgements changed between their two syllable count judgement tasks. Furthermore, speakers who judged a word to have more syllables also produced its rime with greater durations. Tilsen & Cohn (2016) interpret this as evidence that "syllable count judgements and articulatory control utilize the same representations" (p. 31). Like Tilsen & Cohn (2016), this study is concerned with a structure of indeterminate syllabicity, and probes its representation using both metalinguistic tasks and a production experiment. Unfortunately, I do not have production data and text-setting data from the same speakers; a good direction for further research would be to elicit both types of data from the same participants in order to establish whether there is a correlation between them, as in Tilsen & Cohn (2016). For now, if both types of data reflect the same representation, then this predicts that speakers who did not distinguish the duration of <v> from /V/ (S2 and S4 from the production experiment) would also not distinguish them in their text-setting, much like E and Sa in the "İyi ki doğdun” study here. Speakers like S3 who produced relatively few non-lexical vowels might pattern with F in their text-setting and ignore <v>, and speakers whose production falls in between these extremes might, like O and Se here, display optionality.

Finally, this study is the first in this dissertation to support the previous descriptions of the quality of the vowels in onset clusters in Turkish. Participants had clear judgements about the “correct” quality of the inserted vowel. They required inserted vowels to be [-back] when the next lexical vowel was [-back] (e.g., [t<i>ren]), and often used [+round] non-lexical vowels before [+round] lexical vowels. The claim that inserted vowels are always [+back] after dorsal consonants (C&S, Kabak 2011) is supported here as well, though: /gri/ ‘gray’ → [g<ɯ>ri], *[g<i>ri]; /kredi/ ‘credit’ → [k<ɯ>redi], *[k<i>redi]. However, this study also confirmed the finding from Ch. 5 that non-lexical <i> often occurs in /Cla/ sequences, e.g., [p<i>laj], [p<i>lak], which is previously unreported and has no plausible explanation under vowel harmony. (See Section 7.4 for further discussion.) Furthermore, Clements & Sezer’s (1982) claim that <i> can be inserted in /s/-stop clusters even when the lexical vowel is [+back] is vindicated here, in contrast to the lack of support for this claim in the study of TELL (Ch. 5). This apparent discrepancy may be because of the greater dialectal diversity of the participants here (as opposed to the two native Istanbul dwellers who contributed to TELL), or alternately may be due to methodological difference. Participants in this study were encouraged to provide all the possible ways they could produce a word, whereas TELL only records one pronunciation per word. TELL does not record any optionality, so if (e.g.) [s<i>por] ‘sport’ was an acceptable option for the TELL speakers, but was not their preferred pronunciation, this was not recorded.

I interpret the patterns of quality of the non-lexical vowel in this study, to mean that Turkish speakers are able to introspect about the acoustic characteristics of the intrusive vowel and provide a categorical judgement about how they perceive its quality. This is compatible with the non-lexical vowels being intrusive and therefore not subject to phonological vowel harmony, but rather being affected by coarticulation with both surrounding vowels and consonants.

6.5 General conclusion

The studies in this chapter indicate that the non-lexical vowels in Turkish onset clusters have an intermediate metrical status, analogous to the intermediate syllabic status of words like *fire* in English. Turkish speakers are inclined to say these vowels are not syllabic, sometimes on the basis of orthography. The non-lexical vowels can receive a beat in music, or not; in the corpus of music examined here, they received a beat about half the time. When they do receive a beat, it is always one beat only, unlike lexical vowels, which can sometimes receive two or more beats, even when they are word-initial and hence unstressed. Certain words like *p<i>renses* ‘princess’ are more likely to receive a beat for the non-lexical vowel; others like *bre* and *problem* ‘problem’ are less likely to contain a non-lexical vowel at all. But the same singer can choose to give the non-lexical vowel a beat or not, depending on the metrical demands of the remainder of the line. Finally, the study of text-setting in “Iyi ki doğdun” revealed a remarkable degree of interspeaker variability, forming a spectrum from never giving the inserted vowel a beat, to optionally giving it a beat, to always giving it one. This study also hinted at the articulatory roots of this phenomenon, with dorsal consonants being associated with a stronger preference for inserting a vowel and giving it metrical space.

While the fact that non-lexical vowels sometimes receive a beat in music might at first blush suggest that they are syllabic after all, this interpretation is not supported by the text-setting results as a whole. The non-lexical vowels are definitely not being treated as metrically equivalent to lexical vowels. They receive less metrical space, much less consistently. Furthermore, their treatment varies between speakers in ways that the metrical treatment of lexical vowels does not. I believe these results are consistent with non-lexical vowels in Turkish onset clusters being non-syllabic, as would be expected if they are intrusive vowels produced by gestural phasing relations, rather than phonological epenthesis. In contrast, there does not seem to be any pressure against assigning beats to coda-repairing epenthetic vowels in Turkish, which can receive any number of beats—they can be elongated to dramatic effect, just like underlying vowels. For example, in Ece Mumay's "Bir vedayla bir ömür" ('one life with one farewell'), the word *ömür* 'life' occurs at the end of a line, and is sung to five beats (https://www.xsarkisozleri.com/ece-mumay-bir-vedayla-bir-omur-sarki-sozleri/). In a sample of half a dozen songs containing coda-repairing vowels, every coda-repairing epenthetic vowel received at least one beat, and most received multiple beats. This is unsurprising, since coda-repairing vowels tend occur in the final syllable and receive stress, and stressed vowels are more likely to be prominent in text-setting as well. Indeed, it would be peculiar for a vowel that prototypically bears stress to avoid occupying a metrical slot in music. A systematic study of the text-setting of coda-repairing vowels could establish more decisively whether they tend to receive fewer beats than lexical vowels, but I strongly suspect that true epenthetic vowels in Turkish are no less eligible to occupy metrical space than lexical vowels are.

Overall, the results of this chapter point to Turkish speakers being aware that onset clusters are typically produced with non-lexical vowels that do take up time, while also being aware that these non-lexical vowels are optional and short, and that they vary a great deal from speaker to speaker—a fact that several consultants alluded to during the elicitation sessions. This variability, the prescriptions of orthography, and knowledge of source languages all conspire to promote Turkish speakers' awareness that these vowels in onset clusters are not *true* vowels, and do not have the same syllabic and metrical status as lexical vowels.

Appendix A: Lyrics to "Drama Köprüsü"

Drama köprüsü Hasan
Dardır geçilmez bre Hasan
Dardır geçilmez
Soğuktur suları Hasan
Bir tas içilmez
Soğuktur suları Hasan
Bir tas içilmez

Anadan geçilir Hasan
Yardan geçilmez bre Hasan
Yardan geçilmez
At martini Debreli Hasan
Dağlar inlesin
Drama mahpusunda he dost
Canlar dinlesin

Mezar taşlarını Hasan
Koyunmu sandın bre Hasan
Koyunmu sandın
Adam öldürmeyi Hasan
Oyunmu sandın
Adam öldürmeyi Hasan
Oyunmu sandın

Drama mahpusunu Hasan
Evinmi sandın bre Hasan
Evinmi sandın

At martini Debreli Hasan
Dağlar inlesin
Drama mahpusunda he dost
Canlar dinlesin

At martini Debreli Hasan
Dağlar inlesin
Drama mahpusunda he dost
Canlar dinlesin

Appendix B: Selected results from “Iyi ki doğdun” study

# /V/	word	gloss	O x1	O x2	F x1	F x2	Se x1	Se x2	E x1	E x2	Sa X1	Sa X2
1, /C/	kap	container	ka	ap	kap	x	ka	ap	ka	ap	ka	ap
	pas	rust	pa	as			pa	as	pa	as	pa	as
	pul	scale, stamp	pu	ul	pul	x	pu	ul	pu	ul	pu	ul
	su	water	su	u	su	x	su	u	-	-	su	u
	top	ball	to	op					-	-	to	op
1, /CC/	bre	(Eastern?)	*bi	re	*bi	re	bi	re	*bi	re	bi	re
		bro	bre	e	bre		bre	e	bre	e	? bre	e
	bro	bro	bro	o	?		bro	o	?bu	ro	bu	ro
			*bɯ	ro			bu	ro	*bro	o	? bro	o
	fren	brake	fi	ren			fi	ren	fi	ren	fi	ren
			?fre	en	fren	x	? fre	en				
	gram	gram	gɯ	ram	gram	x	gI	ram	gɯ	ram	gɯ	ram
			??gra	am			? gra	am	*gra	am	*gra	am
	gri	gray	gɯ	ri	gri	x	gri	I	gɯ	ri	gɯ	ri
			*gri	I			? gri	I	*gri	ri	? gri	i
	kral	king	kɯ	ral	kral	x	kɯ	ral	kɯ	ral	kɯ	ral
			?kra	al			* kra	al				
	plaj	beach	pi	laj	plaj	x	pla	aj	pi	laj	pi	laj
			? pla	aj			?? pi	laj	*pla	aj		
	plak	record	pla	ak	plak	x	pla	ak	pi	lak	pi	lak
			pi	lak			? pi	lak				
	plan	plan	pi	lan	plan	x	pi	lan	pi	lan	pɯ	lan
			? pla	an	pilan	x	?? pla	an				
	prens	prince	pre	ens	prens	x	pi	rens	pi	rens	pi	rens
			pi	rens	* pi	rens	? pre	ens	*pre	ens		
	spor	sport	spo	or	spor	x	si	por	si	por	sɯ	por
			sɯ	por	sI	por	?? spo	or	sɯ	por		
			si	por	*si	por						
	stil	style	sti	il	si	til	sti	il	si	til	si	til
			si	til			* si	til	*siti	il		
	tren	train	ti	ren	tren	x	ti	ren	ti	ren	ti	ren
			tre	en	? tre	“the newer g		en	*tire	en		
2, /C/	Barış	name	ba	ış	ba	ris	ba	ris	Ba	ris	Ba	ris
			* barı	ış	* bari	is	* bari	is	*Bari	is		
	Dilar	name	di	lar	?		di	lar	Di	lar	Di	lar
			* dila	ar								
	Guray	name	* gura	ay								
			gu	ray	gu	ray	Gu	ray	Gu	ray	gu	ray
	Gürsel	name	Gür	sel					Gür	sel	Gür	sel
	Kıvanç	Pride, joy			kɯ	vanc	kɯ	vanc	Kɯ	vanc	kɯ	vanc

Chapter 7: Phonological representation

7.1 Distinguishing phonetics and phonology

Traditionally, phonetics and phonology are distinct branches of linguistics, representing distinct modules in the linguistic system. In this model, phonology is understood as a domain of categorical phenomena, and performs symbolic operations on discrete, non-overlapping segments, while phonetics handles physical implementation and gradient processes. This distinction is affirmed by a range of work on the phonetics-phonology interface (e.g., Cohn 1990, 1993; Keating 1979; Myers 2000; Strycharczuk 2012; Solé 1992; Bermúdez-Otero 2013; Hale & Reiss 2008). At the same time, phonetics and phonology govern such closely related phenomena that it can be hard to distinguish which domain a process belongs to. This close relationship has caused some researchers to propose the two are really integrated into a single domain (e.g., Ohala 1990, Flemming 2001).

A particularly influential challenge to the traditional conception of phonology comes from Articulatory Phonology (Browman & Goldstein 1986, 1990, et seq.; Gafos 2002; *inter alia*), which brings gestures into the phonological representation. Work in Articulatory Phonology revealed that many processes that had been thought to involve symbolic manipulations by rules were actually better analyzed as the gradient consequences of gestural timing. For example, the variation between light and dark /l/ was traditionally characterized as an allophonic alternation between two categorical variants [l] and [ɫ]. But Sproat & Fujimura (1993) provide evidence that the variation is actually gradient, and reflects the relative timing of the two gestures that /l/ is composed of, an apical gesture which has an affinity for the syllable margin, and a dorsal gesture that has an affinity for the nucleus. The syllable position of the /l/, then, determines whether the apical gesture precedes or follows the dorsal gesture, thus determining the perceived lightness of the /l/. Another example is the apparent deletion and/or assimilation of consonants in casual speech, as in the pronunciation of *perfect memory* as [pɚfɛkmɛmoɹi] with no audible [t], or *ground pressure* as [graumpɹɛʃɚ]. An articulatory study showed that the seemingly deleted /t/ and /d/ gestures were actually still present, but their acoustic consequences had been "masked" by overlap with the following labial consonant (Browman & Goldstein 1990). When it is possible to ground an explanation in bodily gestures, many patterns can be accounted for in principled ways without the need for abstract, symbolic rules of, e.g., deletion or feature change. Browman and Goldstein proposed to eliminate purely symbolic segments and features from the representation entirely, and make gestures the primitives of both phonetics and phonology.

Other researchers, however, have drawn on the gestural insights of Articulatory Phonology while maintaining the importance of features and segments for representing categorical phenomena. For example, Zsiga (1997) shows that vowel assimilation in Igbo is gradient, and therefore best represented in gestural terms, but that vowel harmony is categorical, and therefore is better modeled by feature-spreading. In the same vein, Hall (2003, 2006) contrasts vowel epenthesis, which adds a segment and categorically changes the phonological representation, with vowel intrusion, which does not insert a segment and is better modeled as the consequence of a particular gestural timing. A range of other work on the phonetics-phonology interface (e.g., Cohn, 1990, 1993; Keating, 1984; Myers, 2000; Strycharczuk, 2012; Solé 1992; Bermúdez-Otero 2013; Hale & Reiss 2000) likewise affirms that a distinction between gradient phonetics and categorical phonology is still necessary.

In this dissertation, I follow Hall (2003, 2006) in accounting for the properties of intrusive vowels using gestural timing, and therefore adopt an Articulatory Phonology framework. I assume, following Zsiga (1997) and many others, that phonology must be categorical (on some level). Hence, the gradience of vowels in Turkish onset clusters indicates that they are ultimately "added" during phonetic implementation. Phonology specifies the gestural score that produces these vowels as a side-effect (cf. Gafos 2002; Hall 2003, 2006; Bradley 2004), and this intrusion-producing gestural score differs only gradiently from a close-transition [CC] gestural score, but it differs categorically from the gestural score for a lexical vowel.

This chapter addresses the phonological representation of Turkish consonant clusters and the vowels that intrude into them. Section 7.2 introduces two existing approaches to modeling gestural coordination within the syllable. The first is a grammar of gestural coordination couched as constraints of Optimality Theory (Gafos 2002, Hall 2003), and the second is the coupled oscillator model (Saltzman & Byrd 2000, Goldstein et al. 2007, Goldstein et al. 2008). In Section 7.3, I combine these frameworks and propose that vowel intrusion in Turkish onset clusters occurs because the gestures for the two consonants are in an anti-phase timing relation, while the first consonant is coupled to the syllable nucleus only weakly or not at all. This set of coupling relations creates an open transition between the consonants, and pushes the first consonant away from the vowel in time. The resulting intrusive vowel is not a phonological object (segment) that would be subject to processes like vowel harmony and syllabification, as discussed in Section 7.4. This accounts for the ways that the quality of the intrusive vowel – as evaluated acoustically and according to Turkish speakers' intuitions – differs from the quality predicted by a regressive vowel harmony process posited by previous work on onset clusters. In Section 7.5, I argue that vowel intrusion in Turkish onset clusters is phonologized to different degrees by different speakers, and that historical reanalysis of intrusive vowels in onset clusters has been arrested in present-day Turkish by orthographic pressures and language-contact. Section 7.6 concludes.

7.2 Gestures in phonology

Experimental work on gestural timing has established that the gestures for onsets and codas are coordinated differently with respect to the gesture for the syllable nucleus. First, the gesture for a vowel begins simultaneously with the gesture for a simplex onset (Löfqvist & Gracco 1999). When multiple onset consonants are present, the gesture for the vowel begins at the midpoint of all the consonantal closures[10]. A coda consonant, in contrast to an onset, does not begin simultaneously with the vowel. Instead, it begins at the midpoint of the vocalic gesture, thus overlapping the vowel much less than an onset consonant. Subsequent coda consonants do not appear to be coordinated with the nucleus, but are only coordinated locally with the first consonant in the coda (Browman & Goldstein 1988). Theorists modeling these phasing relations, or gestural coordination relations, have done so both with constraints (in Optimality Theory

10 At least for some languages; I discuss to this point more fully in Section 7.2.2.

approaches, 7.2.1) and with mathematical equations (in Task Dynamic approaches, 7.2.2).

7.2.1 Gestural timing in Optimality Theoretic approaches

In an influential paper on gestural timing in Moroccan Colloquial Arabic, Gafos (2002) argues compellingly that the phonological grammar must be able to refer to the timing of gestures. The traditional view that the phonological representation refers only to the linear sequencing of features and segments, and not to any overlap in time, cannot account for the distribution of transitional schwas or the means by which templatic morphology is satisfied in Moroccan Colloquial Arabic. Gafos (2002) proposes that the grammar governs gestural timing through Optimality Theoretic constraints on the alignment of gestures, where a gesture is defined as "a spatio-temporal unit, consisting of the attainment of some constriction at some location in the vocal tract." In order to abstract away from the continuous, dynamic nature of gestures, constraints refer to the alignment of *landmarks* in each gesture. The basic template for a gestural Align constraint is shown in (20); landmarks are drawn from the list in (21).

(20) ALIGN(GESTURE$_1$, LANDMARK$_1$, GESTURE$_2$, LANDMARK$_2$): Assign a violation for every Gesture$_1$ whose Landmark$_1$ is not aligned with Landmark$_2$ of Gesture$_2$.

(21) Gestural landmarks: ONSET, TARGET, CENTER, RELEASE, OFFSET

Gafos (2002) draws on work by Browman & Goldstein (1988, 2001) to establish which gestural coordination relations need to be implemented as Align constraints; these are the coordination relations C-V (onset-nucleus), V-C (nucleus-coda), and C-C (between two consonants of a complex onset), implemented as in (22)-(24) (paraphrased slightly from Gafos 2002).

(22) C-V COORDINATION = ALIGN(C, C-CENTER, V, ONSET): Assign a violation if the C-center of an onset C is not synchronous with the Onset of the nucleus V.

(23) V-C COORDINATION = ALIGN(V, RELEASE, C, TARGET): Assign a violation if the Release of the nucleus V is not synchronous with the Target of a coda consonant C.

(24) C-C COORDINATION = ALIGN(C1, C-CENTER, C2, ONSET): Assign a violation if the C-center of C1 is not synchronous with the onset of C2, when C1 and C2 are adjacent consonants in the same syllable.

Such OT constraints on gestural alignment have since been employed in many phonological analyses (Hall 2003, Benus et al. 2004, Delforge 2008, Schmeiser 2009, Russell Webb & Bradley 2009, Halpert 2012, Bradley 2012, *inter alia*). Here, I will focus on Hall (2003)'s Timing Augmented Surface Phonology (TASP) framework, which employs the ALIGN constraints and gestural landmarks from Gafos (2002) to account for vowel intrusion, with particular attention to copy vowel intrusion. Hall (2003) is representative of various similar OT approaches to gestural coordination, all drawing on the fundamental proposal in Gafos (2002) that there must be a grammar of gestural coordination. I focus on TASP as a convenient example because it deals specifically with gestural configurations that produce different varieties of vowel intrusion, as well as those that would produce close transitions. Hall (2003) also spells out the various possible rankings and their consequences more explicitly than the other adaptations of Gafos (2002) that I have seen.

In TASP, the input to phonology consists of segments, and as phonology happens, the segmental information is supplemented by gestural alignment information, so that the output of phonology contains information about gestural timing relations. Thus, gestural timing information is part of the *output* of phonology, but not part of the *input*, and therefore never subject to faithfulness constraints. This model captures the fact that gestural timing varies from language to language, but is never contrastive within a language. Hall (2006), however, points out that gestural timing could be part of the input without being contrastive, as long as there are no faithfulness constraints that refer to gestural timing. This is the solution to the non-contrastiveness of syllabification proposed by McCarthy (2003): "It is, however, more in keeping with OT's thesis of richness of the base (Prince & Smolensky 1993) to assume that underlying representations may be syllabified [or gesturally aligned] or not in diverse ways – freely but also pointlessly, since no constraints of UG lobby for conservation of underlying syllabification [or gestural timing]." In the discussion that follows, I assume a representation in which the input consists of segments, which in turn comprise of one or more tightly coordinated gestures. (To anticipate the language of the coupled oscillator model, a segment might be modeled as an oscillator, with the oscillators for the gestures that form it being tightly coupled to each other and to the segmental oscillator.) The underlying form may or may not contain gestural timing information. Phonology contains markedness constraints on the phasing or coupling of gestures, and the output of phonology consists of a coupling graph that indicates the phasing relations of all the segments involved. Crucially, phonology does not include faithfulness constraints referring to the phasing of gestures, since the details of gestural timing are never contrastive.

Vowel intrusion occurs when the gestures for adjacent consonants are in a phasing relation that produces an open transition between them. Hall (2003) adopts the phasing relation employed in Gafos (2002) to produce intrusive schwas: $\textsc{Center}_{C1} = \textsc{Onset}_{C2}$. If the Align constraint enforcing this phasing is ranked above any constraints that favor a closer alignment (e.g., $\textsc{Release}_{C1} = \textsc{Target}_{C2}$), then an open transition will result, producing an intrusive vowel which enhances the acoustic cues to the consonants.

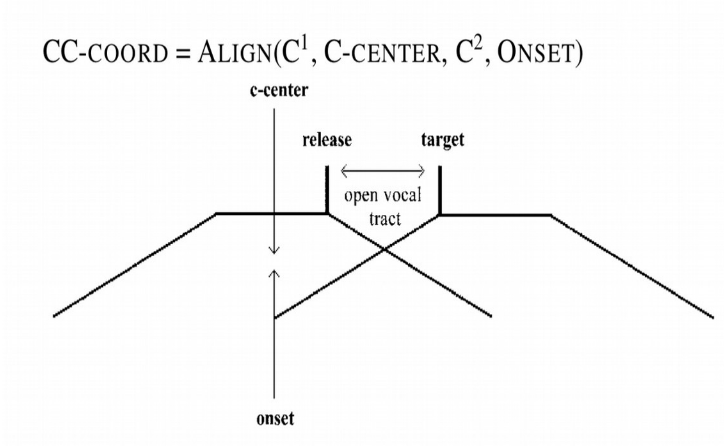


Figure 7.1: CC-Coord -- (7) in Gafos (2002)

For Hall (2003), to produce an intrusive schwa, this phasing of the consonants is all that is necessary. For an intrusive vowel that shares the acoustic quality of an adjacent lexical vowel (an intrusive copy vowel), constraints requiring the lexical vowel to span the entire syllable from edge to edge, such as those shown in (25)-(26), must also be ranked above constraints on overlapping gestures, such as the constraint in (27). (All constraints taken from Hall 2003:19.)

(25) ALIGN(V, OFFSET, SYLL, OFFSET): "The offset of every vowel is aligned with the offset of the rightmost segment that belongs to the same syllable as that vowel."

(26) ALIGN(V, ONSET, SYLL, ONSET): "The onset of every vowel is aligned with the onset of the leftmost segment that belongs to the same syllable as the vowel."

(27) *C in V: “A vowel articulation does not fully surround a consonant articulation.”

For example, Hall (2003) models the gestural alignment that produces intrusive copy vowels in Oscan with the constraint ranking shown in (28).

(28) Copy vowel intrusion in Oscan *puk<e>le* ‘son’ – (23) in Hall 2003:20.

/pukle/	ALIGN (C_1, CENTER, C_2, ONSET)	ALIGN (V, ONSET, SYLL, ONSET)	*C IN V	ALIGN (C_1, RELEASE, C_2, TARGET)
a.  [pukəle]		*!		*
b. →  [pukele]			*	*
c.  [pukle]	*!	*		

In the tableau in (28), the constraint demanding an open transition between consonants, ALIGN(C_1, CENTER, C_2, ONSET), outranks the constraint ALIGN(C_1, RELEASE, C_2, TARGET), which prefers a close transition for faster articulation. Therefore, candidate c, which has a close transition between /k/ and /l/, is ruled out. Candidate a with a schwa-like intrusive vowel is ruled out by the high ranking of ALIGN(V, ONSET, SYLL, ONSET), which requires the vowel to begin at the same time that the syllable begins. The winning candidate has both an open CC transition and a vowel gesture spanning the entire syllable, at the cost of violating the constraint against completely surrounding a consonant gesture with a vowel gesture.

To summarize, in TASP (Hall 2003), the intrusive vowel does not correspond to a segment in the phonological input. Syllabification operates on segments only, and organizes C_1, C_2, and V into a syllable. Then gestural alignment constraints (adapted from Gafos 2002) on markedness require C_1's release to occur before C_2 achieves closure, and require the gesture for V to span the entire syllable. When V spans the whole C_1C_2V syllable, V necessarily overlaps C_2, as well as the entire interconsonantal interval. The result is that C_2 is surrounded by V, and V can be heard in two pieces, before and after C_2.

7.2.2 Task Dynamics

The TASP framework crucially makes use of segments as phonological primitives. Other work on Articulatory Phonology, however, treats gestures as the phonological primitives instead (Browman & Goldstein 1986, 1988, 1990, *et seq.*). Segments are bundles of gestures that are in a more stable timing relation, and syllables are constellations of gestures that are coordinated with respect to each other.

For both the segment-based TASP and the gesture-based representations of phonology, some stage of the speech planning operates on a set of gestures and desired relative timings. In Gafos (2002) and Hall (2003), the relative timings of gestures are stated in terms of the alignment or coordination of different gestural landmarks. These landmarks include the gestural onset, achievement of closure (or other target), release of closure, and gestural offset. In Browman & Goldstein's (1986 et seq.) model of Articulatory Phonology, however, landmarks are not involved. The alignment of gestures in the gestural score is expressed in terms of their relative phasing in degrees. This gestural score is sent to the task dynamic model (Figure 7.2).

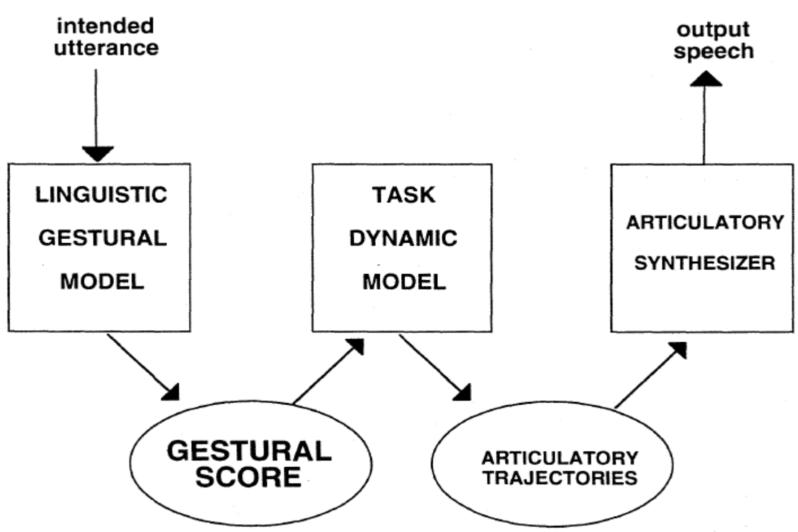


Figure 7.2: Browman & Goldstein's (1992) Gestural computation model

Task dynamics is a general model of gestural coordination that has also been used to represent manual coordination (Saltzman & Kelso 1987), which has been applied to model speech events as coupled oscillators (O'Dell & Nieminen 2009, Saltzman & Byrd 2000). Saltzman & Byrd (2000) divide the task dynamic model into two components (Figure 7.3): the Task Space, where target relative phasings are set; and the Articulator Component Space, where factors such as the initial state of the relevant articulators and their intrinsic properties (e.g., the different masses of the tongue tip vs. tongue body) come into play. The Task Space may be conceived of as essentially phonological, since it consists of target alignments between gestures that are (as yet) abstract, and produces a gestural score, as does the Linguistic Gestural Model in Browman & Goldstein (1992) (Figure 7.2). Indeed, Goldstein, Chitoran, & Selkirk (2007) refer to the coupling graph for a given word as part of the speaker's phonological knowledge. The Articulator Component Space of Saltzman & Byrd's model can then be understood as belonging to the phonetic implementation phase.

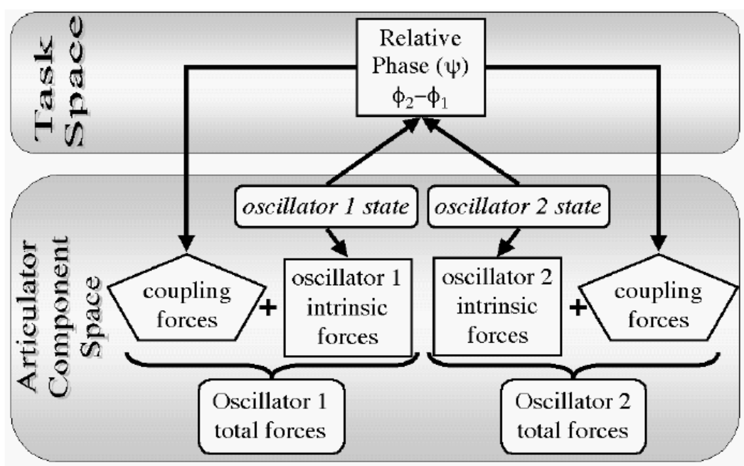


Figure 7.3: A graphical schema of the task dynamics model (Figure 3 in Saltzman & Byrd 2000)

The task dynamic model does not employ a finite set of landmarks within each gesture. Instead, each gesture is modeled as an oscillator: a process which tends to repeat regularly. The repeating, dynamic nature of an oscillator can be mathematically modeled with equations based on sine waves. Gestural coordination is represented by the coupling of oscillators, indicating that the timing of one gesture (oscillator) is affected by (coupled to) the timing of another gesture (oscillator). Coupling of two gestures G_1 and G_2, then, is directly equivalent to the presence of an active constraint ALIGN(G_1, LANDMARK$_1$, G_2, LANDMARK$_2$). With anti-phase coupling, the equivalent alignment constraint aligns different landmarks from each gesture (e.g., LANDMARK$_1$=ONSET, LANDMARK$_2$=CENTER); with in-phase coupling, the equivalent alignment constraint aligns the same landmark from each gesture (i.e., LANDMARK$_1$ and LANDMARK$_2$ are the same).

In the task dynamic framework, the collection of five gestural landmarks employed in, e.g., Gafos (2002) and Hall (2003), is not employed. Instead, the relative phasing of gestures may be expressed in terms that are either more specific than these gestural landmarks, or more abstract. In the more specific, numerical approach, gestural phasing is expressed in terms of the degrees of offset between the activation of the first gesture and the activation of the second gesture (Saltzman & Byrd 2000, Nam & Saltzman 2003). For instance, if the gestures begin simultaneously, the target relative phasing is 0 degrees: if the target phasing is attained, then the first gesture will be 0 degrees into its oscillation when the second gesture begins. Or if the target phasing is 60 degrees, when the first gesture has completed a sixth of its cycle (60 degrees / 360 degrees = 1/6), the second gesture will be activated and begin its oscillation. Even as they introduce this model, however, Saltzman & Byrd (2000) point out that it overgenerates possible gestural phasings, since the target phasing is expressed in terms of a continuous numeric variable, meaning that an infinite number of target coordinations can be expressed.

In the second, more abstract approach, the number of possible gestural coordinations is severely curtailed. The possible gestural coordinations were decided based on research on the coordination of movements of two limbs, such as bimanual gestural coordination. Bimanual coordination can be observed in tasks where participants are asked, e.g., to tap their right hand on the table in time with a metronome, while tapping their left hand once for every two taps of the right hand. This line of research has shown that only a limited set of bimanual gestural coordinations are possible. In fact, there are only two modes of coordination available to subjects without any training: *in-phase* and *anti-phase*. With in-phase coupling, both gestures begin simultaneously, maximizing overlap; with anti-phase coupling, one gesture begins as the other completes 180 degrees of its oscillation (Nam, Goldstein & Saltzman 2009), reducing overlap. Since these modes are available without any special learning, they are considered *intrinsically stable*. It has been proposed that the timing of speech gestures should also be governed by intrinsically stable modes of gestural coordination (Goldstein, Chitoran, & Selkirk 2007; Goldstein, Nam, Saltzman & Chitoran 2009; Nam, Goldstein & Saltzman 2009). This reduces the target relative phasings of two speech gestures to just two (in-phase or anti-phase)—in dramatic contrast to the infinite number of possible phasings as expressed in degrees of offset, or even compared to the 5 landmarks x 5 landmarks = 25 possible ALIGN constraints for the coordination of just two segments.

While both in-phase coordination and anti-phase coordination are intrinsically available, in-phase coordination is more stable than anti-phase coordination. This was determined based on the observation that subjects who are oscillating two limbs in an anti-phase pattern spontaneously shift to oscillating in an in-phase pattern when the frequency of the oscillation is increased, whereas no such shift from in-phase to anti-phase occurs as frequency increases (Turvey 1990, cited in Nam, Goldstein & Saltzman 2009). The coupled oscillator model of syllable structure exploits this asymmetry between the intrinsic modes of coupling to explain the asymmetric behavior of onsets and codas. It is hypothesized that onsets are coupled in-phase with the nucleus, while codas are coupled anti-phase with the nucleus (Goldstein, Chitoran & Selkirk 2007). These coupling relations reflect findings from articulatory studies that gestures for onsets and vowels in CV syllables begin simultaneously, whereas coda gestures do not begin simultaneously with the nucleus (Löfqvist &

Gracco 1999). The greater stability and availability of in-phase coordination, then, can explain the cross-linguistic preference for onsets over codas.

When an onset contains multiple consonants, all onset consonants are hypothesized to be coupled to the nucleus in-phase, while also being coupled to each other in an anti-phase relationship (29). I follow the convention in Goldstein, Nam, Saltzman & Chitoran (2008) of annotating in-phase coupling relations with a line (—), and anti-phase coupling relations with an arrow (→) whose direction indicates the order in which gestures are activated.

(29) $C_1 \rightarrow C_2$—V
|______|

Given the coupling graph in (29), not all target phasing relationships can be achieved. Initiating C_1 simultaneously with V and initiating C_2 simultaneously with V would entail initiating C_1 and C_2 simultaneously as well, in direct contradiction to the anti-phase relationship specified in the coupling graph. There is, therefore, competition between the $C_1 \rightarrow C_2$ coupling target and the C_1—V and C_2—V coupling targets. This competition is hypothesized to be the cause of the C-center effect found in numerous studies of the timing of gestures in complex onsets (Nam & Saltzman 2003, Goldstein et al. 2007). The C-center effect essentially reflects a compromise between these competing phasing constraints, in which C_1 begins earlier than V, while C_2 begins later, so that the timing between the center of the beginnings of all the onset consonant gestures (the C-center) and the beginning of the vowel gesture remains stable as the number of consonants in the onset increases.

Table 7.1: Possible coupling graphs for one vowel and two consonant gestures

<table>
<tr><td>CCV</td><td>CVC</td><td>VCC</td></tr>
<tr><td>C → C — V
|_____|
Competitive #CC cluster
(ex. English)</td><td rowspan="2">C—V → C
Non-competitive CVC
(onset + coda)</td><td>V → C → C
|_______↑
Competitive CC#
(hypothesized for Malayalam)</td></tr>
<tr><td>C → C—V
Non-competitive #CC cluster
(ex. Slovak, Moroccan Arabic)</td><td>V → C → C
Non-competitive CC#
(ex. English, Romanian)</td></tr>
</table>

The C-center effect, believed to reflect the competitive coupling discussed above, has been found in complex onsets in English (Browman & Goldstein 1988; Honorof & Browman 1995), Georgian (Goldstein, Chitoran, & Selkirk 2007), German (Pouplier 2012), and some complex onsets in Italian (/pr/ and /pl/, but not /s/ clusters; Hermes et al. 2008). Word-initial consonant clusters in Slovak (Pouplier & Beňuš 2011), Moroccan Arabic (Shaw et al. 2009), and Tashlhiyt Berber (Goldstein, Chitoran & Selkirk 2007), on the other hand, do not show the C-center effect, nor do /s/-clusters in Italian (Hermes et al. 2008). For Moroccan Arabic and Berber, this apparent right-edge anchoring has been argued to indicate that only the vowel-initial consonant is syllabified as an onset, while the peripheral consonants are either syllabic consonants or appendices to the syllable. For Slovak, on the other hand, all clusters have been considered to be complex onsets, perhaps due to the prevalence of syllabic consonants /l/ and /r/ in the language, causing Pouplier & Beňuš (2011) to raise the possibility that the c-center effect may not be a cross-linguistic universal after all. Results for Romanian (Marin 2013) also raise questions about the universality of the competitive coupling of onset clusters, since the c-center effect was found for /s/ initial clusters /sp-, sk-, sm-/, but not for stop initial clusters /ps-, ks-, kt-, kn-/. In cases where the C-center effect was not found, the phasing of #CCV appears to be more like the one schematized in the bottom lefthand cell of Table 7.1.

The discussion above concerns situations where two consonant gestures and a vowel gesture are organized such that both consonants precede the vowel. A second possible organization for a vowel and two consonants puts the vowel between the two consonants, producing a CVC closed syllable, with a simplex onset and simplex coda. Here, the onset consonant would be coupled in-phase with the vowel, accounting for the near-simultaneous onsets of onset and vowel, while the coda consonant would be coupled anti-phase with the vowel, accounting for the sequential production of nucleus and coda consonant (Browman & Goldstein 1995, Krakow 1993, Sproat & Fujimura 1993, Löfqvist &

Gracco 1999, among others). This is schematized in the middle column of Table 7.1.

Finally, a vowel and two consonants may also be syllabified as a nucleus followed by a complex coda. In contrast to onset clusters, coda clusters have been characterized as exhibiting a non-competitive, local gestural organization, as shown in the lower righthand cell of Table 7.1. This organization has been found for complex codas in English (Byrd 1996, Honorof & Browman 1995, Marin & Pouplier 2010), German (Pouplier 2012), and Romanian (Pouplier 2013). However, competitive coupling for a complex coda is logically possible as well, and may well be necessary to model the intrusive copy vowels seen in RC codas in Scots Gaelic; one such possible competitive coupling is schematized in the upper righthand cell of Table 7.1. Indeed, Nam (2004) suggests that competitive coupling occurs in complex codas in Malayalam (although the particular competition in Nam's [2004] model involves decomposing the consonant into distinct closure and release gestures, which are not represented in Table 7.1).

7.3 Combining frameworks

As observed by Gafos (2002), the competition between mutually incompatible target phasing relations, as in Error: Reference source not found (repeated from (29)), is directly analogous to the competition between different constraints in an Optimality Theory system. In an OT analysis of gestural timing, the competition between ALIGN constraints results in the selection of some particular alignment of gestures as the output of phonology. (For examples of output candidates in TASP, see (28).) In the task dynamic approach, the competition between incompatible target phasing relations must be resolved during the computations of the Task Space (see Figure 7.3).

(30) $C_1 \rightarrow C_2$—V
|________|

Every edge in the coupling graph is assigned a coupling strength, and these coupling strengths are not necessarily the same for all edges. When a gestural coupling has a greater coupling force, or coupling strength, then its target phasing is more likely to be achieved in the actual implementation. The relative strengths of different edges in the coupling graph, then, are analogous to the rankings of the different Alignment constraints that in turn correspond to the edges themselves.

Table 7.2: Correspondences between OT / Align and Task Dynamics

OT systems	**Task Dynamics**
Align constraints	Edges in coupling graph
Ranking of constraints	Coupling strengths

Given these clear parallels between OT and Task Dynamics, it seems desirable to integrate the two systems, as in Gafos (2002). Task Dynamics has the virtues of being grounded in research on the coordination of other body parts, and of having a mathematical model (cf. Task Dynamics Articulations [TADA] from Haskins Laboratories) for testing its predictions. Optimality Theory offers a formal framework to connect articulatory factors with other factors like perceptibility, and to make typological predictions. Combining the two approaches may offer a more embodied model of gestural phonology. An important divergence between the two models, however, is that current uses of Task Dynamics to model syllable timing employ only two gestural coordination relations, in-phase and anti-phase (Goldstein, Chitoran & Selkirk 2007), as discussed above. The landmarks put forth in Gafos (2002) and employed in subsequent analyses by various authors (Hall 2003, Benus et al. 2004, Delforge 2008, Schmeiser 2009, Russell Webb & Bradley 2009, Halpert 2012, Bradley 2012, *inter alia*) are not represented in the coordination relations of the couple oscillator model. In the remainder of this section, I propose a revised Optimality Theoretic implementation of the coupled oscillator model which restricts the possible alignments to the intrinsically stable modes of gestural coordination, as suggested by Goldstein, Chitoran & Selkirk (2007). This restriction reduces the number of constraints, and therefore offers a more constrained and more testable theory. For convenience, I will continue to refer to the linearly first consonant in a complex onset as C_1, and the linearly second consonant as C_2, but

C_1 should be taken to represent the onset consonant that is most string-remote[11] from the nucleus, and C_2 to represent the onset consonant that is string-adjacent to the nucleus. Onset clusters of more than two consonants are not discussed here, but could be modeled using additional constraints.

7.3.1 Coupling strengths in the phonological representation

In landmark-based Align systems like Hall (2003), the choice between a closed C-C transition and an open C-C transition is determined by the relative ranking of two competing constraints on the alignment of C1 and C2, ALIGN(C_1, CENTER, C_2, ONSET) and ALIGN(C_1, RELEASE, C_2, TARGET). Neither of these constraints demands a total overlap of C1 and C2, and yet they still compete directly with each other, and this is possible because there are five different landmarks for each gesture.

11 The linear string of segments is a convenient way to represent the order in which the segments are produced. If the coupled oscillator model is given the widest scope, the linear order of the segments might be understood as being determined by the coupling graph, particularly the direction of the anti-phase coupling relations which require one segment to begin before another, rather than requiring them to begin simultaneously. If this is the case in the underlying representation of phonology, then underlying representations would be understood as coupling graphs, and faithfulness constraints on linear order would have to refer to coupling relations—which might incorrectly predict that syllabification and gestural timing could be contrastive. This might be preventable if such faithfulness constraints were restricted to only refer to the edges in the coupling graph and not to coupling strengths. However, this issue is beyond the scope of this discussion, and I will assume that the phonological representation contains segments in a linear order, following Hall (2003, 2006), along with Zsiga (1997).

However, if we want the phonological constraints to better reflect the proposals of the coupled oscillator model, then they need to refer to in-phase and anti-phase coordination, rather than gestural landmarks. Without landmarks, it is no longer possible to have two directly competing C-C alignment constraints that do not demand total overlap of the consonants, because the only coordination relation that is not totally overlapped is the anti-phase coordination. This creates a challenge for modeling the full typology of C-C coordinations, in which we need to distinguish three degrees of closeness (a closed C-C transition, a short open C-C transition with a schwa-like intrusion, and a long open C-C transition with copy intrusion). Since there are only two phasing relations available, the constraints in the new system cannot refer solely to the *existence* of phasing relations. They must also be able to refer to the *strength* of the phasing relation: the coupling strength.

In Saltzman & Byrd's (2000) model (Figure 7.3), coupling forces are the output of the Task Space where the coupling graph resides, which contribute to the total forces influencing each articulator in the Articulator Component Space. If the Task Space corresponds to phonology and the Articulator Component Space corresponds to the phonetic implementation, then the coupling strengths would be part of phonology's output, sent to phonetic implementation to contribute to the actual physical articulation. For Gafos (2002), the output candidates appear to be unweighted coupling graphs, and the coupling strengths are represented by the ranking of constraints. Implicit in the model is that at some point in the progression from phonological representation to acoustic output, each edge in the coupling graph is labeled with a coupling strength. However, Gafos (2002), Hall (2003), and other ALIGN approaches that include landmarks have not directly represented coupling strengths in the phonological output candidates (or at least, not in those that are included in tableaux), since coupling strengths are supposed to be derived from the constraint ranking only.

Here, I pursue a slightly different implementation of the same concept, in which some abstract representation of coupling strength is included in the phonological output candidates created by the GEN function, and the ranked constraints *select* a candidate based on the relative strengths of edges in its coupling graph. This becomes necessary for capturing the full range of acoustic outputs, when the range of possible coordination relations is restricted to two (in-phase / anti-phase), instead of the twenty-five relations that are possible with five landmarks.

7.3.2 GEN

For the OT system considered here, I begin by assuming that syllabification determines the actual coupling graph (e.g., what it means to be syllabified as an onset is to be coupled in-phase to the nucleus), and that syllabification/coupling is determined by familiar constraints like ONSET, NOCODA, *COMPLEXONSET, etc., and that the constraints that determine syllabification are sufficiently high-ranked that they will not be affected by gestural alignment constraints. I define the candidate-generating function GEN as a function that takes as its input a coupling graph (31) (representing a string of segments whose syllabification is already determined), and produces as its output (33) the set of all coupling graphs where each edge (= phasing relation) is assigned a coupling strength drawn from *s*, some language/system-specific finite set of possible values. Here, I arbitrarily set $s = \{2, 5\}$ (32), meaning the edges in the coupling graph can have coupling strengths of either 2 or 5; larger values of *s* would lead to larger candidate sets. The winning candidate is then determined by the relative rankings of the constraints that require in-phase or anti-phase coupling. In this system, as in previous Align + landmark OT systems, differences in coupling strengths and consequent gestural alignments across languages will reflect differences in constraint rankings. The difference is that competing candidates include coupling strengths as well as phase relations in their coupling graphs.

(31) g = a coupling graph = a set of edges $\{e_1, e_2, e_3, \ldots, e_i\}$. For an open syllable beginning with a CC complex onset, there will be three edges

in the coupling graph: $g = \{C_1—V, C_2—V, C_1 \rightarrow C_2\}$

(32) s = the set of possible coupling strengths in the system = {2, 5}

(33) $GEN(g_i) = \{g_o \mid \text{for every } e_x \text{ in } g_o, e_x.s \text{ is in } s\} = \{ \{e_1=2, e_2=2, e_3=2\},$
$\{e_1=2, e_2=2, e_3=5\}, \{e_1=2, e_2=5, e_3=2\}, \{e_1=5, e_2=2, e_3=2\},$
$\{e_1=5, e_2=5, e_3=2\}, \{e_1=5, e_2=2, e_3=5\}, \{e_1=2, e_2=5, e_3=5\},$
$\{e_1=5, e_2=5, e_3=5\}\}$

Note that since every constraint demands that its coupling relation have a higher coupling strength than all competing coupling relations, candidates where $e_1=e_2=e_3$ {2, 2, 2} and {5, 5, 5} will be harmonically bounded. They are therefore excluded from the tableaux that follow. Furthermore, in this system, since *s* contains only two possible values for coupling strengths, if all edges do not have the same coupling strength, then two will have the same coupling strength and one will have a different coupling strength (either greater [= 5] or smaller [= 2]). In this particular system, it is not possible for a given edge e_1 in the coupling graph to be stronger than e_2 but weaker than e_3. If *s* were set to contain more than two possible values, then a more complex system would result, since GEN would create many more candidates ($|GEN(g)| = |g|^{|s|}$), and the set of constraints proposed in the following section would no longer select unique winners.

7.3.3 A landmark-free CON

The next component of an OT system is its constraint set. For this system, I replace landmark-based Align constraints with landmark-free constraints on the relative strengths of different phasing relations. Recall that in Hall's (2003) Align + landmarks analysis of vowel intrusion, intrusive vowels result from the interaction of constraints aligning the center of C1 with the onset of C2 (34)a, plus constraints requiring the vowel to span the whole syllable (34)c. These constraints favoring intrusion are opposed by constraints that prefer a close transition between consonants and constraints that penalize one gesture "surrounding" another (34)b, (34)d.

(34) Align with landmarks (Gafos 2002, Hall 2003):

a. Pressure for an open transition: ALIGN(C_1, CENTER, C_2, ONSET)

b. Pressure for a close transition: ALIGN(C_1, RELEASE, C_2, TARGET)

c. Pressure for V to span the whole syllable (Hall 2003): ALIGN(V, OFFSET, SYLL, OFFSET), ALIGN(V, ONSET, SYLL, ONSET)

d. Pressure to prevent C from being imposed on V (Hall 2003): *C IN V

In the coupled oscillator model of syllable structure, the pressure for an open transition between consonants comes from an anti-phase coupling between the consonants (35)a, while the pressure for a close transition comes not from a competing coupling relation between the consonants themselves, but from the two consonants' separate demands to be produced in-phase with the vowel (35)b. The in-phase coupling between the first consonant and the vowel in the coupled oscillator model also promotes an articulation in which the vowel spans the syllable. This coupling relation corresponds to Hall's ALIGN(V, ONSET, SYLL, ONSET) only, however; typical coupling graphs do not show any coupling relation that would force the vowel to extend to the right edge of the syllable[12]. The landmark-free constraints appear below (35).

(35) Coupling constraints, no landmarks:

a. ANTI-PHASE($C_1 \rightarrow C_2$): Couple C_1 and C_2 anti-phase, with C_1 being activated first, where C_1 and C_2 are onset consonants in the same syllable. Assign a violation if the coupling relation $C_1 \rightarrow C_2$ has a smaller coupling strength than a competing coupling relation. Two coupling relations are competing if they cannot both be satisfied simultaneously because they contain mutually incompatible gestural specifications.

b. IN-PHASE(C_1–V), IN-PHASE(C_2–V): Couple an onset C_N in-phase with a nucleus V. Assign a violation if the coupling relation C_N—V has a smaller coupling strength than a competing coupling relation.

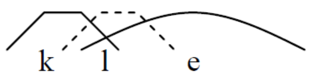
12 The alignment of the right edge of the syllable with the end of the vowel might be represented by including the release of the vowel as a distinct gesture to be coordinated. This has been proposed for the release phase of consonants in complex onsets in Georgian (Goldstein et al. 2008).

Hall's (2003) constraint *C IN V is problematic to translate into in/anti-phase coupling relations. This constraint prohibits a consonant articulation being "surrounded" by a vowel. Hence, it appears to oppose an articulation where a consonant gesture is superimposed on a vowel gesture (cf. Öhman 1966). This constraint might even penalize an in-phase coupling relationship between (onset) consonants and vowels. But we know from a large body of articulatory work that C and V do begin simultaneously when C is the onset for V, so it seems problematic to prohibit exactly that unmarked timing relation. It is not clear what exactly Hall (2003) means by one gesture completely overlapping another – her justification for this constraint is simply that there must be constraints that oppose the total overlap in order to account for languages without vowel intrusion. But *GESTURE-IN-GESTURE might not be the right constraint, particularly since Hall (2003) appeals to the problem of having [-son] inside a vowel to explain the lack of copy intrusion across obstruents, whereas other work (Ch. 5; Davidson 2003, 2006) suggests that vowel intrusion can occur across obstruents, not just sonorants (although the intrusive vowel in these cases may not sound like a copy vowel). Fortunately, the same work that *C IN V does can be done by an already necessary constraint, IN-PHASE(C_2, V), which requires the onset consonant to begin simultaneously with V. When C_2 only begins midway through V's articulation, IN-PHASE(C_2, V) is violated, since they are not beginning simultaneously. But the canonical onset-vowel timing where C and V begin together fulfills the constraint, so that this in-phase coordination is unmarked, as desired.

7.3.4 Distinguishing schwa intrusion and copy intrusion

For Hall (2003), whether an intrusive vowel is schwa-like or copy-like is determined by the relative rankings of *C IN V and ALIGN(V, ONSET, SYLL, ONSET): ranking *C IN V higher prevents the vowel from fully overlapping the ICI, resulting in a schwa-like intrusive vowel, while ranking ALIGN(V, SYLL) higher causes V to already have attained its target during the ICI, and forces the closer C inside the vowel's articulation. That is, Hall's (2003) proposed gestural alignment for a schwa-like intrusive vowel would translate into a coupling graph where C_1 is not coupled to V at all (36)a, (37). In contrast, the coupling graph for copy-vowel intrusion does require coordination between C_1 and V. For Hall (2003), copy intrusion results when the pressure to produce C_1 in-phase with V overrides the pressure to produce C_2 in-phase with V, so that C_1 and V begin simultaneously, V reaches its target before C_1 is released, and C_2 begins after V, so that the ICI is longer and V is already in its target position during this open transition (36)b.

(36) /pukle/ → [puk<e>le] in Hall 2003:20.

/pukle/	ALIGN (C_1, CENTER, C_2, ONSET)	ALIGN (V, ONSET, SYLL, ONSET)	*C IN V	ALIGN (C_1, RELEASE, C_2, TARGET)
a.  [pukə̲le]		*!		*
b. →  [puke̲le]			*	*
c.  [pukle]	*!	*		

(37) Translating (36)a into a coupling graph: C→C—V

In the OT system considered here, however, the three possibilities of schwa-intrusion all reflect the same coupling graph, but with differently weighted edges. This weighting of the edges represents the coupling strengths from Task Dynamics, and allows the system to include the C1→C2 anti-phase coordination while still differentiating the different timing relations that produce close transitions, schwa intrusion, and copy intrusion. If the C_1—V in-phase coupling and the pressure for the C_1→C_2 anti-phase coupling are stronger than the C_2—V in-phase coupling, then the pressure for C_1 to begin simultaneously with V will be greater than the pressure for C_2 to begin simultaneously with V, with the result that C_2 begins after V, producing a long, open ICI (copy intrusion). But if the C_1—V coupling and the C_2—V coupling are equally strong, then C_1 and C_2 will both imperfectly fulfill their in-phase coupling relations, producing the C-center effect as a compromise between two equally strong coupling relations. These relative coupling strengths are schematized in the table in (38).

(38) Possible coupling strengths

	Coupling strength		
Acoustic result	$C_1 \rightarrow C_2$	C_1—V	C_2—V
a. k → l — e |______| [k<e>le]	High	High	Low
b. k → l — e |______| [k<ə>le]	High	Low (to none)	High
c. k → l — e |______| [kle]	Low	High	High

In (38), the three acoustic outputs all correspond to the same coupling graph, but with different coupling strengths. The Oscan output (38)a, [puk<e>le] with an intrusive copy vowel, results from $C_1 \rightarrow C_2$ and C_1—V having high coupling strengths, while C_2—V has a low coupling strength. A similar competition occurs in (38)c, but the coupling strengths for C_1—V and C_2

—V are equal, while the $C_1 \rightarrow C_2$ anti-phase coupling is weaker, so that the coupling graph is implemented with a close transition as [pukle] with no intrusion, as in English. This indicates that the difference between a canonical complex onset and a complex onset with an intrusive copy vowel is a gradient distinction between the magnitudes of the relevant coupling relations, not a categorical distinction between difference coupling graphs.

Finally, the intermediate case of schwa-like intrusion, as in Turkish, could result from assigning a low coupling strength to the C_1—V coupling relation (38)b. Since this coupling relation is weaker than the competing C_2—V and $C_1 \rightarrow C_2$ relations, it violates the constraint requiring C_1—V to have a higher coupling strength, and it will be produced with C_1 beginning before V. However, the coupling graph still has the same edges as the graphs for a long intrusive vowel. Thus, the distinction between different types of intrusion can be gradient, similar to the more acoustically salient distinction between copy vowel intrusion and an English-style complex onset. Schwa-intrusion reflects the same coupling graph as the other two productions of complex onsets above, just with a weakened C_1—V relation.

Another possibility, not included in the OT system explored here, is that the grey coupling relation C_1—V in (38)b is not included in the graph at all. If this relation is not included (i.e., the coupling graph only includes the lines shown in black), then the difference between schwa-intrusion and copy-intrusion reflects a categorical difference in the coupling graph, similar to Hall's (2003) proposal. In that case, [puk<ə>le] with a schwa-like intrusive vowel reflects a competition-free coupling graph. The $C_1 \rightarrow C_2$ coupling and the C_2—V coupling are very strong, and C_1—V has no coupling relation or strength at all.

Both coupling graphs can reasonably account for the acoustic qualities of intrusive schwa. But when the larger landscape of possible realizations of complex onsets is considered, it seems preferable to propose that schwa-intrusion lies on a continuum of gestural alignments between copy intrusion and closed complex onsets (39).

(39)

Copy <V>		**Schwa**		**Close, with release**		**Close, no release**
[puk<e>le]	→	[puk<ə>le]	→	[puk^{ə}le]	→	[pukle]
Oscan		Turkish		French		English

7.3.5 Copy intrusion

The remainder of this section demonstrates how the proposed OT system can model copy intrusion, a closed CC transition, and finally schwa-like intrusion as in Turkish onset clusters.

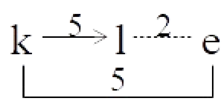
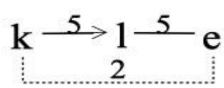
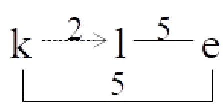
To produce copy vowel intrusion, as in Oscan, Scots Gaelic, and Hocank, I proposed above that $C_1 \rightarrow C_2$ anti-phasing and C_1—V phasing are stronger than the C_2—V phasing. In OT terms, this translates to a constraint ranking as in (40), where the constraint requiring the C_2—V coupling strength to be greater than its competitors is outranked by those that maximize the coupling strength of the other two phasing relations.

(40) *Constraint ranking for copy-vowel intrusion (Oscan):*

$C_1 \rightarrow C_2$, C_1—V >> C_2—V

The consequence of this constraint ranking is that, in the winning candidate, the coupling strengths of relations favored by the high-ranked constraints must be greater than the coupling strength of the phasing relation whose constraint is low-ranked, as illustrated in the tableau in (41).

(41) Tableau of copy vowel intrusion, as in Oscan /pukle/ [puk<e>le] 'son'

/kle/	$C_1 \rightarrow C_2$	C_1—V	C_2—V
→ a. [k<e>le] 			*
b. [k<ə>le] 		*	
c. [kle] 	*		

The winning candidate (41)a avoids violations of $C_1 \rightarrow C_2$ and C_1—V by giving both of those phasing relations a coupling strength of 5. Numbers assigned to coupling strengths are intended to express relative strengths, without making any quantitative claims about the exact magnitudes of the coupling forces involved. This candidate violates C_2—V, because this phasing is assigned coupling strength of only 2, which is less than 5; however, this constraint is outranked by the other two, so under this constraint ranking, the copy-vowel candidate wins over (41)b and (41)c, since these candidates violate the requirements for C_1—V to have the strongest coupling, and for $C_1 \rightarrow C_2$ to have the strongest coupling, respectively.

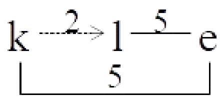
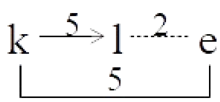
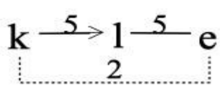
7.3.6 Close C-C transition

In contrast, in languages like English and German that require a close transition between onset consonants and exhibit the C-center effect, the constraint $C_1 \rightarrow C_2$ is lower-ranked, and the open transition between the consonants is lost in the effort to maintain in-phase coupling between both onset consonants and the vowel.

(42) *Constraint ranking for close CC transition (English):* C_1—V, C_2—V >> $C_1 \rightarrow C_2$

The winning candidate (43)a violates only $C_1 \rightarrow C_2$, since this phasing relations does not have as high a coupling strength as the two in-phase relations (C_1—V, C_2—V). Candidate (43)b satisfies $C_1 \rightarrow C_2$ at the expense of beginning C_2 after V, thus violating C_2—V, while candidate (43)c satisfies $C_1 \rightarrow C_2$ at the expense of the in-phase relationship of C_1—V.

(43) Tableau of a close CC transition, as in English *clementine* [klementajn]

/kle/	C_1—V	C_2—V	$C_1 \rightarrow C_2$
→ a. [kle] 			*
b. [k<e>le] 		*	
c. [k<ə>le] 	*		

The next section explores the constraint ranking that produces the intermediate case of schwa-like intrusion, as seen in Turkish consonant clusters.

7.3.7 Phasing in Turkish onset clusters

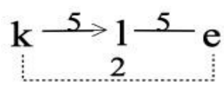
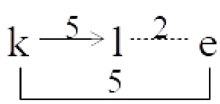
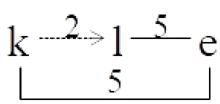
To account for the presence and quality of the intrusive vowels in Turkish complex onsets, I propose that the constraint ranking in (44) obtains in Turkish.

(44) *Constraint ranking for schwa-intrusion (Turkish):*

$C_1 \rightarrow C_2$, C_2—V >> C_1—V

Here, the requirements for an open $C_1 \rightarrow C_2$ transition and for C_2 to be produced in-phase with V (C_2—V) outrank the requirement for C_1 to be produced in-phase with V. As a result, the winning candidate will give the C_1—V phasing a low coupling strength compared to the $C_1 \rightarrow C_2$ and C_2—V phasings. This is shown in the tableau in (45), where the winner (45)a assigns the C_1—V phasing a low coupling strength of 2, while the other two phasings whose constraints are higher-ranked receive coupling strengths of 5. In the losing candidates (45)b and (45)c, the C_1—V coupling has a higher coupling strength, but one of the other phasing relations has a lower coupling strength, thus violating one of the high-ranked phasing constraints instead of the lowly C_1—V.

(45) Tableau of intrusion in Turkish *kleptoman* [k<ɯ>łeptoman] 'kleptomaniac'

/kle/	$C_1 \rightarrow C_2$	C_2—V	C_1—V
→ a. [k<ə>le] ~ [k<ɯ>le] 			*
b. [k<e>le] 		*	
c. [kle] 	*		

In this coupling graph, as usual for complex onsets, C_1 is in an anti-phase coupling relationship with C_2, meaning that C_1 and C_2 should *not* begin simultaneously (i.e., should not be phased 0 degrees apart). Rather, the gesture for C_1 should already be half-way done when C_2 begins (i.e., C_2 is phased to begin 180 degrees into C_1's oscillation). As a result, C_1's closure will tend to be released before C_2's closure is achieved, creating an open transition between C_1 and C_2, as in Turkish *s<i>por* ‘sport.’

Since V is coupled in-phase to C_2, the gesture for V will also begin at the same time that C_2's gesture begins. So both C_2 and V’s gestures will be in progress during the interconsonantal interval, but neither gesture will have achieved its target yet during this open transition (although it may be possible for V to already be at or at least near its target position during the ICI in very overlapped speech). For example, during the ICI for *prens* ‘prince’, the gesture for /p/ is being released, while the gestures for /r/ and /e/ are in their onset phases. So the tongue tip is approaching /r/’s brief coronal closure, while the tongue body is approaching /e/’s front, mid position. The result is an open transition where the tongue body’s frontness is in between that for /i/ and for /ɯ/, so that there is a brief vowel of intermediate quality: [p<i>rens] ~ [p<ɯ>rens].

Finally, and crucially, the in-phase coupling between C_1 and V is present, but weak (or conceivably even absent). This is the most important contrast between the proposed gestural score for Turkish consonant clusters, and the gestural score for a canonical complex onset with a closed transition between the consonants, as in English *sport*. As discussed above, in such complex onsets, C_1 and C_2 are also in an anti-phase relationship, but there is a stronger in-phase coupling relationship between C_1 and the nucleus V, which demands that C_1 and V begin simultaneously (Nam & Saltzman 2003). Since it is not possible for C_1 and C_2 to both begin simultaneously with V without beginning simultaneously with each other, it is not possible to satisfy all of these target timing relations. Instead, the different coupling relations compete with each other, creating the famous C-center effect.

In contrast, in a coupling graph where the coupling between C_1 and V is weak or even absent, the stronger C_2—V and $C_1 \rightarrow C_2$ coupling relations dominate the phasing competition. Both the anti-phase coupling between C_1 and C_2, and the in-phase coupling between C_2 and V, can be achieved simultaneously.

7.3.8 Modeling interspeaker variation

Speakers in the production study reported in Ch. 2-3 varied in their phonetic implementation of onset clusters. This variation does not necessarily mean that their grammars select differently shaped coupling graphs. Instead, it could reflect either differences in *s* (the parameter of the GEN function that determines possible coupling strengths), or differences in the phonetic implementation of the coupling graphs. In articulation, speakers vary in just how weak their C_1—V phasing relation is. For example, speakers who are more experienced with complex onsets due to early exposure to languages like English or French may produce C_1—V timing. If the grammar is to account for this, it could be ascribed to a higher coupling strength for the C_1—V phasing relation. In the production study reported in this dissertation, the speaker who was an early Turkish-French bilingual (S3) produced relatively few vocalic ICIs; I hypothesize that this reflects a phonetic implementation in which the C_1—V phasing relation receives a stronger coupling force. Conversely, speakers who have minimal exposure to the source languages for these loanwords, or who are familiar with languages like Moroccan Arabic that employ syllabic consonants, may assign a particularly low coupling strength to the C_1—V phasing, or even fail to assign it any coupling strength at all. This may be the case for the one monolingual Turkish speaker in the study (S4) who did not differentiate durationally between lexical and non-lexical vowels.

The tableau in (46) shows some examples of possible variations in coupling strengths for C_1—V that are all equally optimal within the proposed Turkish constraint ranking. These candidates would not all be produced by the same GEN function, since they require different sets of coupling strength values. The constraints do not differentiate among these coupling graphs because all of them assign C_1—V a lower coupling strength than C_2—V, thus satisfying the highly ranked constraint C_2—V which requires the coupling strength of C_2—V to be greater than any competing coupling strength. For this coupling graph, C_1—V and C_2—V compete, since they cannot both be satisfied simultaneously. But $C_1 \rightarrow C_2$ and C_2—V do not compete, since it is possible to satisfy them both at once. Candidate (46)a represents a coupling graph where C_1 and V are not coupled at all, such that C_1 is a sort of appendix to the syllable. Monolingual S4 may employ this sort of coupling graph. Candidate (46)b represents schwa-like intrusion as described above, although I have transcribed the intrusive vowel as <ɯ> because the preceding /k/ in this context would give it a particularly [+back] articulation. Finally, candidate (46)c shows a coupling graph where C_1 /k/ is treated more like a true part of the onset, since it has a larger C_1—V coupling strength compared to the other candidates in this tableau. A speaker who is more experienced with complex onsets may use this sort of production.

(46) Tableau of variation in acceptable coupling strengths within Turkish

/kle/	$C_1 \rightarrow C_2$	C_2—V	C_1—V
a. [k$^{\text{ɯ}}$le] → $\underbrace{k \xrightarrow{6} l \overset{6}{\text{—}} e}_{0}$			*
b. [k<ɯ>le] → $\underbrace{k \xrightarrow{6} l \overset{6}{\text{—}} e}_{0.5}$			*
c. [k<ə>le] → $\underbrace{k \xrightarrow{6} l \overset{4}{\text{—}} e}_{2}$			*

The coupling graphs above would each be generated by different GEN functions which differ in the set *s* of possible coupling strengths employed. (46)a would result from s = {0, 6}; (46)b from s = {6, 0.5}; and (46)c from s = {6, 2, 4}. Thus, s is a possible source of variation among speakers. It is also entirely possible that the interspeaker variation seen in previous chapters of this dissertation could reflect differences in phonetic implementation, rather than grammar, and indeed, may be more desirable from a theoretical standpoint. I leave it to future research to address this issue, however.

7.3.9 Appendix or complex onset?

Since the C_1—V coupling is weak or absent in Turkish, little to no C-center effect is predicted. This is similar to the phonological representations that have been proposed for initial consonant clusters in Moroccan Arabic (Shaw et al. 2009) and Berber (Goldstein et al. 2009). In these languages, the rightmost consonant does not shift toward the vowel as more consonants are added on the left side of the cluster, indicating that the additional consonants are not syllabified as part of a complex onset. Instead, only the rightmost consonant has the onset relation to the vowel. Other consonants in these languages have been proposed to be syllabic consonants (Dell & Elmedlaoui 1985, 2002) or perhaps appendices to the syllable.

In the coupled oscillator model, only segments within a syllable are expected to be coupled directly to each other. In addition to the *intralevel* coupling of segments within a syllable, there is also *interlevel* coupling of oscillators at different levels of the prosodic hierarchy (O'Dell & Nieminen 2009), such as syllable oscillators being coupled to a foot oscillator. Thus, a coupling graph that depicts interlevel coordination as well as intralevel coordination of oscillators will closely resemble a prosodic tree depicting the phonological organization of segments into syllables, feet, words, and so forth. In this model, segments within a syllable are coupled to a syllable oscillator, and coordination of segments across syllables is determined indirectly by intralevel coordination between the syllable oscillators. Hence, if the first consonant in a #CCV sequence forms a separate syllable from the vowel, then the gesture for this first consonant is not predicted to be coupled directly to any of the segments in the CV syllable. On the other hand, if the initial, non-onset consonant is an appendix and still part of the syllable, then it can still be coordinated directly with other segments in the syllable, such as C_2. This predicts that C_1-C_2 timing should be more stable when C_1 is an appendix than when C_1 is a syllabic consonant.

(47) Possible coupling graphs given a simplex onset

a. *Syllabic consonants:* $[\ C_1\ ]_\sigma\ [\ C_2 — V\ ]_\sigma$

b. *Appendices:* $[\ C_1 \rightarrow C_2 — V\ ]_\sigma$

For Turkish, I propose that C_1 in a #CCV sequence is either weakly part of the complex onset, or is an appendix, but is not a syllabic consonant. This accounts for the fact that C_1 does not appear to count for a syllable in musical meter, and certainly not in explicit syllable count judgements. In cases where C_1 occupies a different beat in music than C_2, C_1 is always accompanied by an intrusive vowel (see Ch. 6, particularly Section 3.3). In this, Turkish text-setting contrasts with that for Berber or Moroccan Arabic (Dell & Elmedaoui 2002). This coupling graph also accounts for the greater temporal stability of the interconsonantal interval in /Cr/ words compared to /CVr/ words (Ch. 4). In /Cr/ words, /C/ and /r/ are part of the same syllable, and have a direct anti-phase coupling relation governing their relative timing. But in /CVr/ words, /C/ and /r/ are each simplex onsets to different syllables, so they are not directly coupled to each other, and their relative timing is more free to vary.

7.4 Harmony and the intrusive vowel

Whether the consonants in Turkish consonant clusters are both part of a complex onset, or consist of an appendix and a simplex onset, there is no distinct vocalic gesture between them. This can be seen in the shortness of the intrusive vowel compared to lexical vowels, its gradient duration, and the clear contrast between text-setting of the intrusive vowel and lexical vowels. I conclude, therefore, that there is no vowel segment between the two consonants—another way of saying that the vowels in Turkish onset clusters are intrusive, not epenthetic.

Since phonology operates on segments, phonology cannot directly manipulate the interconsonantal interval, no matter how vowel-like it may sound. More specifically, in the absence of a vowel segment, there is nothing in the consonant cluster for vowel harmony to operate on. This conclusion conflicts with previous accounts of the non-lexical vowels in Turkish onset clusters. These accounts assume that these vowels are epenthetic, and state that regressive vowel harmony determines the backness and rounding of the non-lexical vowel (Clements & Sezer 1982, Yavaş 1980, Kaun 2000, Yıldız 2010). Each of these accounts has built theoretical conclusions on the assumption that the inserted vowel is the result of (optional) epenthesis. Yıldız (2010) uses the supposed epenthesis as evidence for an indeterminate ranking between *COMPLEXONSET and DEPV. All authors use the supposed vowel harmony as evidence that harmony in Turkish can be regressive, as well as progressive. Kaun (2000) argues that Turkish rounding harmony is sensitive to the height of the harmony trigger as well as the target, thus making a case that Turkish speakers can access the UG constraint on gestural uniformity GESTUNI([ROUND]) that requires harmonizing round vowels to share a height specification.

Despite these claims, data within the previous literature, particularly Clements & Sezer (1982), indicates that whatever process influences the quality of the non-lexical vowel in Turkish onset clusters does not behave the same way as the progressive vowel harmony that determines the backness and rounding of vowels in suffixes, as well as the true epenthetic vowels that repair illicit coda clusters. The previous literature agrees that the inserted vowels are variable—either subject to speech style (Clements & Sezer 1982) or entirely optional (Yıldız 2010). There are hints that the quality of the vowel is affected by consonant place (Clements & Sezer 1982). And of course the rounding harmony is described as sensitive to the trigger's height instead of just the target vowel's height as in the standard progressive rounding harmony in Turkish (Yavaş 1980, Kaun 2000). Kabak (2011) summarizes these differences and concludes that the harmony that affects vowels in onset clusters is not fully phonologized, unlike the harmony that affects epenthetic vowels in coda clusters. Pilot phonetic support for this comes from a study conducted by Bokhari et al. (2016), who found that /i/-like vowels in onset clusters had a lower F2 than lexical /i/ or epenthetic /i/, suggesting they are more central, and not phonologically specified as [-back].

The various studies in this dissertation confirmed that the non-lexical vowel in onset clusters does not behave like a harmonizing, epenthetic vowel. The ultrasound study shows that the position of the tongue body during the non-lexical vowel does reflect the backness of the following lexical vowel (i.e., fronter before /i/ than before /a/ or /o/), but that it is much less front than a lexical /i/. Furthermore, following /g/, the tongue body is not significantly fronter than lexical /ɯ/, even when the following lexical vowel is /i/. This gestural finding indicates that tongue body position in these intrusive vowels is being determined by coarticulation with both the following vowel and the preceding consonant, not by categorical backness harmony. In addition, the gestural study found evidence of lowering in intrusive vowels before /a/, further demonstrating coarticulation and also indicating that the intrusive vowel is not categorically [+high] as claimed in the epenthetic account.

The acoustic and corpus studies also confirmed that categorical vowel harmony does not determine the quality of the vowel in onset clusters. Acoustically, non-lexical vowels are more subject to coarticulation than lexical vowels, and even at their front-est, they are more central than lexical /i/. In the TELL corpus, the non-lexical vowels are usually transcribed as <ɯ>, even in contexts where harmony would demand [i]. TELL also confirms that a preceding velar consonant blocks fronting of the intrusive vowel; there is only one instance of a front vowel being transcribed in a cluster where the first consonant is dorsal. Rounding harmony also underapplies in TELL, and this is true whether the following lexical round vowel is high or low. It is especially stark for low vowels, however, in keeping with previous findings that some speakers only produce a round non-lexical vowel if the potential trigger is [+high] (Kaun 2000, Yavaş 1980).

The TELL study also revealed another peculiarity of the quality of non-lexical vowels in Turkish onset clusters, which is that <i> is often (but not always) transcribed in P+/l/ words, where P stands for a labial consonant.

(48) TELL /pł-/ and /fl/ words – Speaker 2 only

a. [piłato]	'Plato'	
b. [pixlatonik] (*sic*)	'platonic'	
c. [pilan]	'plan'	
d. [piłaʒ]	'beach'	
e. [piłaka]	'plaque, (license) plate'	
f. [pixlanøɾ] (*sic*)	'planner'	plânör
g. [pɯlasenta]	'placenta'	
h. [fiłaman]	'filament'	flaman
i. [fiłoş]	'floss'	floş
j. [fiłaş]	'flash'	flaş

A phonological harmony analysis might take some hope from the fact that a palatal /l/ can be a harmony trigger when it occurs at the end of the word; [l] induces [-back] suffixes even across another consonant (49).

(49) /kalp + E/ 'heart + DAT' [kalp – e], *[kalp – a]

Indeed, this may be occurring in (48)f, where the orthography indicates a lexically palatal /l/ with the caret on top of <a>. Some other light /l/s are transcribed as well ((48)b, c). However, a close look at the transcriptions shows that there is often no [-back] harmony trigger. The lexical vowel is either /a/ or /o/, which are both [+back], and the /l/ in (48)a, d, e, and h - j also surfaces as its [+back] allophone [ɫ], in keeping with the regular allophonic pattern in Turkish where onset /l/ matches its nucleus in backness (Göksel & Kerslake 2005). We should take TELL's transcriptions seriously since they are based on the considered judgements of a phonetically trained annotator who had access to audio recordings and spectrograms. Furthermore, while /l/ triggering front suffixes is extremely consistent across speakers and lexical items, /l/ triggering instrusive <i> varies by speaker and lexical item. E.g. in TELL, Speaker 1 never has <i> in pla- words, but Speaker 2 often does. However, even Speaker 2 does not always have <i> in these words – sometimes he has <ɯ> (48)g. This degree of variability is unexpected for phonological harmony in Turkish. But if the quality of the vowel is being gradiently shaped by gestural timing and coarticulation, this sort of variability is expected.

From the perspective that the vowels in onset clusters are intrusive, we might speculate that the apparent <i> in these examples is a result of a perceptual contrast effect, as discussed in Ch. 5 (Section 5). Alternatively or additionally, articulatory factors may help explain the occasional frontness of the vowel in labial+/l/ clusters. Since a labial consonant does not make any demands on the tongue body, articulation of /l/ is able to start sooner in these clusters than it would in clusters where C_1 involves the tongue. The articulation of a lateral involves two gestures, an apical gesture and a tongue body gesture; these gestures are timed differently depending on the lateral's position in the syllable, resulting in different degrees of "lightness" or "darkness" (Sproat & Fujimura 1993). Sproat & Fujimura suggest that the apical gesture gravitates toward the syllable margin, which in this case would be toward C_1, while the dorsal gesture gravitates toward the nucleus. In a labial + /l/ cluster, then, /l/'s apical gesture may have already moved the tongue toward an /i/-like position by the time the labial consonant is released. The release of the C_1 closure begins the open transition into C_2, which allows an [i]-like vocalic interlude in the ICI. As the ICI ends, the dorsal gesture of the /l/ takes place, closer to the nucleus than the apical gesture was. Since the "light," apical part of the /l/ occurred or at least began earlier and was transcribed as [i], the remainder of the /l/ tends to be perceived and transcribed as a dark /l/.

The fact that this unexpected <i> intrusion only occurs in TELL with labial + /l/ clusters may reflect imbalances in the place of C in the set of C + /l/ clusters, as much or more than it reflects a meaningful contribution by the labial consonant. Since /tl/ and /dl/ clusters are illegal in the main donor languages French and English, and TELL contains only one /sl/ cluster produced by Speaker 2, there is no real chance for /l/-driven <i>-intrusion to occur with a coronal C_1 in TELL. (Zimmerer & Kabak [2010] conducted a production study of epenthesis in onset clusters that included /tl/ clusters in pseudo-Russian words, but unfortunately their poster does not break down the inserted vowels by their consonant context, only by vowel.) TELL does include quite a few /gl/ and /kl/ clusters, but here the dorsal gesture from C_1 has such a large coarticulatory influence that the intrusive vowel always sounds [+back]. This was seen gesturally in the ultrasound results (Ch. 2), and acoustically in the formant values (Ch. 4), both of which confirmed previous descriptions saying that inserted vowels in clusters with a velar C_1 are always [+back] (Clements & Sezer 1982, Kabak 2011, Walter 2017).

TELL does not contain any instances of <i> occurring in /bl/ clusters, but since the corpus only contains one /bl/ word produced by S2 (*blöf* [bɯłøf] 'bluff'), this may not mean anything. I tested whether the voicing of the labial consonant matters by collecting some acceptability judgements from two Turkish speakers, based on orthography. Both speakers (Sa and E, who also participated in the "Happy Birthday" study, Ch. 6) said that front vowels could occur in both /pl/ and /bl/ (Table 7.3). Intriguingly, Sa also reported that the actual quality of the vowel is *between* /i/ and /ɯ/, which shows that at least some Turkish speakers can notice and label the gradient quality of the vowel in these words.

Table 7.3: Acceptability of <i> and <y> in /pl/ and /bl/ clusters (Speakers E and Sa)

		E	Sa
a. p<i>lastik	‘plastic’	✓	✓
b. p<i>latin	‘platinum’	✓	✓
c. p<i>lonjon	‘dive’	–	✓
d. p<ü>lonjon			*
e. b<i>lazer	‘blazer’	b<i>leyzir	✓
f. b<ü>lucin	‘blue jeans’	*	✓
g. b<i>lok	‘block’	✓	✓
h. b<ü>lok		✓	*
i. b<ü>luz	‘blouse’	*	✓

The elicitation above and the results of the “Happy Birthday” study show that Turkish speakers do have intuitions about the quality of the vowel. Furthermore, these intuitions can be gradient. Like Sa, Speaker O has volunteered gradient judgments about the quality of the intrusive vowel, as well as saying that it seems to contribute “half a syllable.” Speaker Se has also said that when producing a word with an onset cluster, “you put your tongue in the position for the vowel, but you don’t actually say the vowel” between the consonants. Despite this awareness of the indeterminate quality of the intrusive vowel, speakers are also able to assign it to a category when asked, as in these studies and Kaun (2000). The chosen phoneme category may be the one demanded by harmony, but is not consistently so. Variation occurs within and across speakers. Front vowels may fail to occur where they would be predicted (underapplication), or may occur where they are not predicted (as in labial + /l/ clusters and /s/ + stop clusters; overapplication?). Round vowels do not seem to occur in the absence of a round trigger (no overapplication), but may fail to appear (underapplication). Speakers often disagree about the quality of the vowel, even speakers from the same family or region. I interpret these judgements about the quality of the intrusive vowel as a manifestation of Turkish speakers' ability to draw on their memories of the acoustic, non-lexical vowels that they have produced and perceived in the course of their lifetimes, and extract generalizations about which phonemes they are most perceptually similar to. I do not believe they reflect categorical harmony judgements, considering: how variable the judgements are; the fact that speakers can access gradient judgments about the quality (Sa), syllabicity (O), and gestural properties (Se) of the intrusive vowel; and the evidence from the various production and corpus studies reported in this dissertation.

7.5 Historical development

Native Turkish phonology does not contain any complex onsets, nor does the native vocabulary of Persian or Arabic, two languages that Turkish has historically borrowed many words from (Yıldız 2010). The word-initial consonant clusters that are produced with intrusive vowels all come from relatively recent borrowings from European languages, predominantly French and English. The fact that the intrusive vowels all occur in loanwords raises the question of how the current gestural configuration came about, and whether it has been stable across time. As suggested in Ch. 3 (Section 3.4.1), the initial analysis of the vowels in these loanwords when they first entered the language probably varied among speakers, depending on their knowledge of the source language. Bilinguals would have been aware that the source language did not contain vowels in the consonant clusters, and the vowels bilinguals produced would likely have been intrusive, arising from their recruitment of familiar or intrinsic modes of gestural coordination to deal with this novel sound structure. The gestural coordination that produces a closed CC transition, along with the C-center effect, must be learned; it is a compromise between three coupling relations, each of which is in an intrinsic mode (in-phase or anti-phase), but which cannot simultaneously be satisfied. Hence, Turkish speakers producing these loanwords would not be able to replicate the French gestural timing with its C-center effect (Kühnert et al. 2006), but would have needed to employ an intrinsic mode of gestural coordination to produce them. The two intrinsic modes are in-phase (beginning simultaneously) and anti-phase (beginning the second gesture midway through the first). Beginning both consonant gestures simultaneously (coordinating them in-phase) would obscure the formant transitions in and out of at least one consonant closure, since both closures would be occurring at the same time. This would drastically reduce the perceptibility of each consonant. Consequently, these Turkish speakers employed anti-phase coupling between C_1 and C_2. As seen in Section 7.3.7 above, this leads to an open transition between the consonants, and an intrusive vowel.

Monolingual Turkish listeners, hearing these intrusive vowels, may not have noticed that they were not underlying vowels. Without knowing the source language, they would have had no compelling reason to think the intrusive vowels were non-lexical. Since Turkish has no prefixes, the initial C in a word initial consonant cluster never has the chance to be syllabified as a coda. Therefore, there are no syllable-structure changes at the left edge of the word to produce C.CV ~ C<v>CV alternations that would serve as cues to listeners that <v> is non-lexical. Without this kind of evidence, they would have no reason to think that the inserted vowel is intrusive or even epenthetic. Many monolingual Turkish speakers, then, must have rapidly re-analyzed [C<v>CV] as /CVCV/.

Other listeners, however, may have been attuned to the subtle, gradient differences in formant values and duration that signal the targetlessness of the intrusive vowels. This is not entirely far-fetched, given the astute observation by my consultant Sa that the vowels sound in words like [p<i>lastik] *plastik* is intermediate between /i/ and /ɯ/. Sensitive listeners like Sa, then, may have recognized that these vowels were not lexical vowels, and analyzed them as epenthetic (due to the native Turkish constraint ranking *COMPLEXONS >> DEPV), or even detected that these vowels were intrusive.

In the absence of exposure to the source language via speech or spelling, however, over time the reanalysis of intrusive vowels as underlyingly present would tend to prevail. We have evidence that this occurred in Ottoman-era Turkish, in the form of /CC/ loanwords from European languages, which were written down in Ottoman-era texts with an inserted vowel (Walter 2017). Some of these words are well-integrated into present-day Turkish, and are always spelled with a lexicalized intrusive vowel, reanalyzed as an underlying vowel (e.g., *tulumba* (50)a, *pırasa* (50)e). Others have retained their status as containing a consonant cluster, and may be pronounced with or without an intrusive vowel (e.g., *istop~stop* (50)h).

(50) Non-lexical vowels reanalyzed as lexical vowels in Ottoman texts (data from Walter 2017)

	Source word	*Turkish borrowing*	*Gloss (for Turkish)*
a.	tromba	t**u**lumba	'pump'
b.	pflug	p**u**lluk	'plow'
c.	Crimea	k**ı**rım	'Crimea'
d.	quartz	k**u**vars	'quartz'
e.	Prason (Gr.)	p**ı**rasa	'leek'

f. sgombro	**u**skumru	'mackeral'
g. ?	**i**skota ~ **ı**skota	'sheet, clew line'
h. stop	**i**stop	'stoppage'
i. ?	**i**skarpela	'carpenter's chisel'
j. scala (It.)	**ı**skala	'gamut'
k. screw	**u**skur	'screw'
l. stuppi (Gr.)	**ü**stüpü	'cotton tow, waste'

Walter (2017) analyzes the vowels in these words as epenthetic, and points to the preponderance of inserted [i]s as a strategy for marking loanwords. These inserted [i]s can surface even where they are disharmonic for backness and rounding ((50)g-i). Interestingly, Walter's (2017) corpus contains a great many /s/+stop clusters that are repaired with prothesis. The prothetic vowel sometimes harmonizes in backness and/or rounding ((50)j-l), but is often a disharmonic <i>. In fact, all the disharmonic <i>s in Walter's corpus are prothetic vowels occurring before /s/+stop clusters, which suggests that their frontness might be driven by perceptual or gestural reasons having to do with /s/. Clements & Sezer (1982) report that disharmonic [i] can be inserted in /s/ clusters, citing words like [s<i>por ~ s<ɯ>por] 'sport.' This seems to be restricted to certain dialects in present-day Turkish—neither of the two TELL speakers follows this pattern, but all of my consultants in the "Iyi ki doğdun" study allowed non-lexical [i] in at least one of *s<i>til* and *s<i>por*. Apparently it was a more widespread phenomenon in Ottoman Turkish, however; <i> appears in 90/188 forms examined by Walter (2017), and 41 of those instances are disharmonic (<i-ı> or <i-o>; <i-u> never occurs). Likewise, while prothesis was a very common repair strategy in Ottoman Turkish, it is not active in present-day Turkish. It is an open question whether intrusive prothetic vowels are even possible. A prothetic vowel does not reflect an open C_1-C_2 transition, but rather an open transition into C_1. This could be modeled with a gestural coordination in which V spans the whole syllable, and begins before or simultaneously with C_1. The prothetic vowel would correspond to the onset phase of both V and C_1, and if C_1 is /s/, the high front tongue position required for /s/ could lead to the intrusive vowel sounding like [I]. It is unclear what gestural factors would favor this alignment, however, since it does not yield the open CC transition demanded by the phasing relation $C_1 \rightarrow C_2$. One possibility is that /s/+stop clusters that are repaired by prothesis have different phasing relations, which would be a natural consequence if they are actually complex segments, as argued by Broselow (1987, 1988) and LaMontagne (1993).

This evidence from spelling in Ottoman Turkish, in conjunction with the lexicalization of many (but not all) of the inserted vowels and their standardization in spelling, supports the proposal that intrusive, non-lexical vowels have historically been reanalyzed as either epenthetic vowels, or as underlying vowels. Historically, before the average Turkish speaker had so much knowledge of the source languages for these complex onsets, and before the writing system was reformed to its present-day transparency, these intrusive vowels were reanalyzed as lexical vowels as /CC/ loanwords were increasingly integrated into Turkish and used by speakers who had no connection to the source languages.

Today, however, changes to the Turkish writing system and educational system stalled this reanalysis. The adoption of the Latin alphabet as a highly phonemic writing system, combined with the prescriptive spelling of most /CC/ loanwords without an inserted vowel, provides an on-going reminder to Turkish speakers who encounter the written forms of /CC/ loanwords that these words are “not supposed to” have a vowel. Furthermore, changes to the educational system in the last few decades have greatly increased the average Turk’s exposure to English (Yıldız 2010), such that younger generations of Turks are likely to be more aware of the intrusion-less source versions of loanwords typically produced with an intrusive vowel in Turkish. The current gestural phasing represents Turkish speakers’ adaptation of existing gestural phasing relations to accommodate onset clusters in foreign loanwords. This phonological representation with a consonant cluster has been maintained in present day Turkish by everyday Turkish speakers’ knowledge of prescriptive spelling and the source languages, as well as their experience of language-internal variation in the production of these onset clusters. This variation comes in from speech rate (intraspeaker variation) and also dialectal differences (interspeaker variation).

Since studying English is compulsory in Turkish schools today, I suspect that the intrusive status of vowels in loans from English is likely to be maintained more robustly than that of intrusive vowels in loans from less-studied languages like French and especially Italian. Intrusive vowels in loans from French may go the way of vowels in loans from Greek. Greek was once commonly spoken in Turkey but has very few speakers there now, and formerly non-lexical vowels in loans from Greek like *İstanbul* and *pırasa* 'leek' are fully lexicalized today. French is still commonly taught in Turkish schools as an elective, but it is not as pervasive as it once was. However, French does show signs of becoming more popular as Turkey's economy and government have been shaken in recent years (Rega 2018, Cafe Babel), so perhaps French loanwords will continue to resist reanalysis as well.

7.6 Discussion

In this chapter, I have summarized how vowel intrusion in onset clusters in loanwords in Turkish is conditioned by gestural factors. Adopting a framework in which phonology can refer to both segments and gestures, and restricting the possible gestural coordinations to the intrinsic modes argued for by Goldstein, Chitoran & Selkirk (2007), I have shown that vowel intrusion can be represented phonologically as the result of a coupling graph where the anti-phase relationship between $C_1 \rightarrow C_2$ competes with the in-phase relationships between C_1—V and C_2—V, just like in the C-center effect in languages like English. The difference between languages with intrusion and those without is the relative ranking of constraints demanding that particular phasing relations be given maximal coupling strengths. An open transition between the consonants occurs when $C_1 \rightarrow C_2$ outranks one of the in-phase constraints. The Turkish case of schwa-like intrusion is argued to reflect a constraint ranking in which the C_1—V phasing relation has a weak coupling strength, perhaps so weak that this phasing relation might as well not be present in the graph at all.

This representation of Turkish onsets means that the vowels in Turkish consonant clusters are not phonological segments, and therefore cannot be targets for vowel harmony. This analysis fits well with the observed facts, that the quality of Turkish intrusive vowels is variable, gradient, and determined by coarticulation. Some puzzles, such as /l/-driven insertion of <i>, and what kind of gestural coordination might motivate prothesis, deserve further investigation. But overall the evidence from the production and corpus studies makes a strong case that the quality of vowels in Turkish onset clusters is not determined by phonological vowel harmony.

Finally, I discussed the diachronic trajectory of intrusive vowels in Turkish. Data from Ottoman Turkish suggests that intrusive vowels have often been reanalyzed as lexical vowels, at least historically. At the present time, however, it is likely that this nativization and reanalysis is being slowed down by young Turkish speakers' high degree of exposure to English, the source language for many of these loans. This knowledge, in conjunction with the prescriptive spelling without a vowel, as well as language internal evidence in the form of variation and gradience in the production of vowels in onset clusters, may provide Turkish speakers with sufficient evidence of intrusive vowels' different status, to prevent them from being reanalyzed as lexical vowels. Turkish speakers' awareness of the non-lexical status of these vowels can be seen in their metalinguistic commentary and syllable-count judgements (often based on orthography), as well as (more convincingly) their text-setting of intrusive vowels, which do not get the same metrical treatment as lexical vowels. Overall, this evidence suggests that Turkish speakers are indeed aware that the intrusive vowel is not a full vowel—or in more technical terms, they are aware on some level that the vowel is intrusive and does not belong as an object in the phonology.

Chapter 8: Conclusion

Although they produce similar acoustic consequences, from a representational standpoint, vowel intrusion and vowel epenthesis are fundamentally different. Epenthesis is a categorical, phonological process that adds a segment. Intrusion is a gradient, phonetic process that does not add a segment, but merely implements the gestural plan and acoustic consequences entailed by the gestural score or coupling graph provided as phonology's output. To put a finer point on it, epenthesis adds a symbol, and intrusion does not. This dissertation has explored the regularities and the variability of vowel intrusion in one particular language, Turkish.

Previous descriptive and even experimental evidence (Yavaş 1980, Clements & Sezer 1982, Kaun 1999, Yıldız 2010) characterizes vowel insertion in complex onsets in Turkish as epenthesis. However, these descriptions still hinted at a greater variability in the repair of onset clusters that in coda clusters, indicating that they are affected by the preceding consonant and vary with speech style in ways that true epenthetic vowels in Turkish do not. These previous works also reported differences in the harmony process that was supposed to determine the quality of the vowel in onset clusters, to the point that Kabak (2011) concludes that whatever this process is, it is not the same harmony at work elsewhere in the word.

This dissertation breaks with previous descriptions to pursue a gestural account of the non-lexical vowels in Turkish onset clusters. I account for their seemingly harmonic behavior as the result of coarticulation, not phonological vowel harmony. Gesturally, ultrasound imaging revealed that the non-lexical vowels have tongue positions that are lower (before /a/) and more central than lexical high vowels /i/ or /ɯ/ (Ch. 2). Their duration was gradient, indicating that this vowel insertion process is gradient and therefore phonetic (Ch. 3). Their formant values, like their tongue position, were also intermediate between those of lexical vowels (Ch. 3), even in casual speech (Ch. 4) where they became less central (e.g., higher F2 before /i/), reflecting greater gestural overlap, in contrast to the formant values of the lexical vowels themselves which were reduced in casual speech (e.g., lower F2 in lexical /i/). Furthermore, the duration of the ICI in clusters proved more stable across speech styles than that of underlying vowels, indicating that there is a direct C-C coordination relation in clusters (produced with non-lexical, intrusive vowels), unlike the indirect coordination between onsets of adjacent syllables (for /CVCV/ words) (Ch. 4).

Looking beyond the voiced stop + /r/ clusters of the experiment, similar results were found through studies of the Turkish Electronic Living Lexicon (Ch.

5) and text-setting in Turkish pop, rap, and folk music (Ch. 6). TELL showed that insertion is common across cluster types and is affected by factors that make sense from a gestural standpoint but would be surprising if the inserted vowels are supposed to repair the syllable structure. Vowels are more likely to be transcribed after a voiced, fricative, or dorsal C_1—although the particular effects differed between the two speakers. S1 inserts in all types of clusters, included /s/ + stop, whereas S2 only inserts in clusters where C_2 is a sonorant. S1 never has inserted front vowels if the following lexical vowel is not front, but S2 consistently inserts a front vowel in labial + /l/ clusters. For both speakers, non-lexical round vowels occurred almost exclusively before high round lexical vowels, in line with Kaun (1999) and Yavaş (1980). Most inserted vowels were transcribed as /ɯ/. These results show that onset cluster repair is extremely variable in ways that are structured by phonetic factors, and that the influence of these phonetic factors can be seen even in a corpus of broad transcriptions.

Examining text-setting of the non-lexical vowels in Turkish onset clusters (Ch. 6) revealed that Turkish speakers are aware on some level of the durational or syllabic differences between lexical and non-lexical vowels. Whereas lexical vowels are reliably set to one or two beats in pop, folk, and rap music, non-lexical vowels could be set to maximally one beat. They were equally likely to receive no beat at all, which was very rarely an option with lexical vowels. The same singer, with the same word, in the same song, could set a lexical vowel to a beat in one line, and set it to no beats in another line, based on what worked for the meter of the rest of the line. I also looked at text-setting in "İyi ki doğdun" ("Happy Birthday To You"), where two speakers treated non-lexical vowels like lexical vowels, one ignored non-lexical vowels, and two exhibited optionality, with preferences shaped by the individual word and by the consonant it began with. (Non-lexical vowels after velar consonants were more likely to be text-set like lexical vowels, presumably reflecting their greater duration). This is further evidence that non-lexical vowels do not behave like underlying, phonologically present vowels for metrical purposes, and suggests that they are not syllabic, at least not in the same way that lexical vowels are.

In Chapter 7, I proposed that non-lexical vowels in Turkish result from a gestural score that puts a much greater weight (coupling strength) on the anti-phase coupling of the two initial consonants than on the in-phase coupling between the peripheral consonant C_1 and the nucleic vowel. This accounts for the relative stability of the C-C timing, as well as the intermediate, somewhat centralized quality of the intrusive vowel in Turkish. To represent this, I adopted the assumption that phonology can refer to both segments and gestures (following Zsiga 1997, Hall 2003, among others), and integrated the coupled oscillator model of syllable structure with Optimality Theory to model the range of gestural scores that could result from different rankings of constraints on the strengths of different phasing relationships. I also summed up the evidence from previous chapters to argue that the quality of non-lexical vowels in Turkish onset clusters reflects gradient coarticulation on the intrusive vowel, not phonological vowel harmony. Lastly, I argued that intrusive vowels tend to be reanalyzed as lexical vowels, but that this reanalysis can be stalled or prevented by knowledge of the source language (via second language acquisition or prescriptive orthography), or by language-internal evidence in the form of variation across speakers and speech styles, as well as subtle acoustic differences between underlying and intrusive vowels.

The analysis of Turkish onset clusters advanced here contradicts previous descriptions that characterize Turkish onset clusters as being repaired by phonological epenthesis, accompanied by regressive vowel harmony (Yavaş 1980, Clements & Sezer 1982, Kaun 1999, Yıldız 2010). This raises the question of why previous analyses did not come to the same conclusion. One factor is that Articulatory Phonology (Browman & Goldstein 1986 *et seq.*) was not part of the theoretical landscape when Clements & Sezer (1982) and Yavaş (1980) were writing. Furthermore, Articulatory Phonology remained divorced from formal Optimality Theory until Gafos (2002)—well after Kaun (1999) proposed an OT account of the variation in vowel quality in Turkish onset clusters. Furthermore, without an instrumental study, the acoustic difference between the non-lexical and lexical vowels can easily be obscured by categorical perception on the part of the linguist who is transcribing what they hear with discrete, categorical symbols. Indeed, for some speakers (like S4 in the production study), there is complete acoustic neutralization between /Cr/ and /Cɯr/, and only a gestural study reveals the articulatory differences between them. Epenthesis is a familiar process to most phonologists, while intrusion is less familiar, so it may have come to mind more readily, particularly to phonologists interested in syllable structure and repairs to syllable structure. The familiar process of epenthesis did a reasonably good job describing the data in broad strokes. In addition, the non-lexical vowel also displays alternations in vowel quality that, upon closer reflection, are better explained by coarticulation, but given that vowel harmony was independently necessary in Turkish phonology, it is understandable that phonologists used the theoretical tools that came readily to hand in their analyses of Turkish onset clusters. This dissertation, then, can be taken as a cautionary tale of how impressionistic transcriptions may not be adequate for a phonological analysis. They are too vulnerable to the biases of the transcriber. Rather, phonology would benefit if impressionistic studies were more often supplemented with instrumental studies and meta-linguistic evidence such as syllable counts, templatic morphology, or language games that can shed light on syllable structure.

In addition to adding to the knowledge base for vowel intrusion and for Turkish phonology, this case study also has a contribution to make to the study of vowel harmony. While the quality of intrusive vowels is not determined by vowel harmony but by coarticulation, the effects are impressionistically similar enough that previous authors adopted the opposite analysis. Turkish onset clusters and intrusive vowels, then, have something to say about the gestural or even perceptual roots of vowel harmony. One characteristic of harmony systems cross-linguistically is that vowel harmony is much more common than consonant harmony (Hanson 2001). This echoes the finding in this dissertation that the following vowel matters more to the quality of the intrusive vowel than the preceding consonant does, although both contribute. This is a logical consequence of the proposed gestural alignment, where the onset phase of the following vowel overlaps the entirety of the ICI, while the release phase of the preceding consonant only overlaps the first portion of the ICI. Greater overlap of the vowel gesture will lead to the vowel having a greater articulatory influence. This influence of the vowel, however, can be largely blocked when the preceding consonant is dorsal. In this, the coarticulation seen here differs from what is seen in phonological vowel harmony, where blockers or opaque segments are those that are specified for the spreading feature (Rose and Walker 2011), and must intervene between target and trigger to block harmony. (When /k/ blocks coarticulatory spreading in Turkish, the sequence is $[\mathrm{k}<\mathrm{v}>\mathrm{C}_2\mathrm{V}]$ = blocker-target-C_2-trigger.) These differences are a reflection of the fact that phonological harmony systems are not driven solely by gestural factors, unlike the intrusion and coarticulation seen here. Perceptual enhancement also plays a role in phonological harmony (Kimper 2017), as with rounding harmony where sounds whose rounding is phonetically weaker are more likely to be perceptually supported by triggering rounding harmony on other segments (Kaun 1995). In vowel intrusion, if any enhancement is occurring, it is enhancement of the cues to the consonants, not the vowels.

Fundamentally, spreading and blocking in phonological harmony are categorical, whereas in the vowel intrusion and coarticulation seen here, they are gradient. Nonetheless, one can imagine that the type of coarticulation seen here could be reanalyzed by speakers (not just linguists) as epenthesis combined with phonological vowel harmony. Something similar may have occurred historically with the harmonizing epenthetic vowels now seen in Turkish coda-clusters. These clusters derive from Arabic loanwords, where they are produced with epenthesis or intrusion, depending on the dialect. If the vowels were intrusive in the source dialect, Turkish speakers would have had to reanalyze them for them so that they could become epenthetic, as they are today. (See Ch. 1 for reasons to believe these vowels are epenthetic, not intrusive.)

To sum up, this dissertation has presented evidence from a variety of methodologies (ultrasound, acoustic, corpus, text-setting) that shows that vowels in Turkish word-initial onset clusters are intrusive, not epenthetic. This means that the consonant clusters are complex onsets, not simplex onsets to adjacent syllables, and it also means that vowel harmony in Turkish is never shown to proceed regressively. This study provides yet another example of phonology's ability to refer to gestural timing, following previous work on Articulatory Phonology, and also suggests a new way to represent articulation and gestural coordination as OT constraints, taking insights from the coupled oscillator model. Future work could test the proposed coupling graphs for Turkish and for other languages. Perceptual studies of these intrusive vowels could shed further light on Turkish speakers' mental representations of these clusters, and how they may interact with their knowledge of vowel harmony and of prescriptive orthography. Also, dialectal or language acquisition studies with many speakers could illuminate patterns in the interspeaker variation seen here, which could not be fully explored due to the small number of participants; this could contribute to our understanding of variation in phonology and its role in the phonetics-phonology interface, particularly as it relates to phonologically governed gestural timing.

References

Alfonso, P. J., and T. Baer. (1982). Dynamics of vowel articulation. *Language and Speech*, 25(2).

Anderson, A. H., M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert (1991). "The HCRC map task corpus." *Language and Speech* 34(4): 351-366.

Baković, E. (2000). *Harmony, Dominance and Control* (Doctoral dissertation). Rutgers University. Retrieved from https://doi.org/doi:10.7282/T3TQ60BJ

Barbosa, A.V. & Vatikiotis-Bateson, E. (2014). Optical flow analysis for measuring tongue-motion. Paper presented at the 168th Meeting of the Acoustical Society of America, Indianapolis, IN.

Bates, D., M. Maechler, B. Bolker, S. Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01

Bellik, J. (2017). Turkish onset cluster repair: An ultrasound study. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 43.

Bermúdez-Otero, R. (2013). Amphichronic explanation and the life cycle of phonological processes. In: Honeybone, Patrick and Joseph C. Salmons (eds). *The Oxford handbook of historical phonology*. Oxford: Oxford University Press.

Bingöl, A. S., & Karaca, N. (1995). Spectral Analysis of Turkish Vowels and a Comparison of Vowel Normalization Algorithms. *Speech Recognition and Coding,* Vol. 147 of the NATO ASI Series, pp. 317-320.

Benus, S., I. Smorodinsky, A. Gafos (2004). Gestural coordination and the distribution of English 'geminates.' *University of Pennsylvania Working Papers in Linguistics, 10(1).* http://repository.upenn.edu/pwpl/vol10/iss1/4

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08. Retrieved January 2015 from http://www.praat.org/

Bokhari, H., M. Durmaz, & J. Washington (2016). An acoustic analysis of vowel insertion at syllable edges in Turkish. Slides presented at the *2nd Conference on Central Asian Languages and Linguistics*, University of Indiana, Bloomington.

Bradley, T. (2004). Gestural timing and rhotic variation in Spanish codas. In: T. L. Face (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 197-224). Berlin: Mouton de Gruyter.

Bradley, T. (2012). Labialization and palatalization in Judeo-Spanish phonology. In: J. Smith, T. Ihsane (eds.), *Romance Linguistics 2012: Selected papers from the 42nd Linguistic Symposium on Romance Languages (LSRL), Cedar City, Utah, 20-22 April 2012.* Philadelphia, PA: John Benjamins Publishing Company, 2015.

Broselow, E. (1987). An investigation of transfer in second language phonology. In: Ioup, G. & S. Weinberger, eds. *Interlanguage phonology: the acquisition of a second language sound system.*

– (2015). The typology of position-quality interactions in loanword vowel insertion. Retrieved from https://linguistics.stonybrook.edu/faculty/ellen.broselow/selected.publications

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

– (1987). Tiers in Articulatory Phonology, with some implications for casual speech. *Haskins Laboratories: Status Report on Speech Research* SR-92.

– (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45: 140-155.

– (1993). Dynamics and Articulatory Phonology. *Haskins Laboratories Status Report on Speech Research,* SR-113, 51-62.

Browman, C. (1994). Lip aperture and consonant releases. In P. Keating (ed.) *Phonological structure and phonetic form: Papers in laboratory phonology III*, pp. 331-353. Cambridge: Cambridge University Press.

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

Carnie, Andrew. (1994). "Whence sonority? Evidence from epenthesis in Modern Irish." In Andrew Carnie, Heidi Harley, and Tony Bures, eds. *Papers on phonology and morphology: The MIT Working Papers in Linguistics*, volume 21. Cambridge, Massachusetts: MITWPL. pp. 81-108.

Cheng, A., E. Remirez, S. Lin. (2017). An Ultrasound Investigation of Covert Articulation in Rapid Speech. Northwestern Phon{etics;ology} 3, University of British Columbia, May 2017.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207-229.

Clements, G. and E. Sezer (1982). Vowel and Consonant Disharmony in Turkish. In H. van der Hulst and N. Smith (eds.), *The Structure of Phonological Representations*. Dodrecht: Foris.

Cohn, A. C. (1990). WPP, No. 76: Phonetic and Phonological Rules of Nasalization. *UCLA: Department of Linguistics, UCLA*. https://escholarship.org/uc/item/1xq3d5hr

– (1993). Nasalization in English: Phonology or phonetics. *Phonology* 10(1): 43-81. https://doi.org/10.1017/S0952675700001731

Davidson, L. & M. Stone (2003). Epenthesis Versus Gestural Mistiming in Consonant Cluster Production: An Ultrasound Study. *WCCFL 22 Proceedings*, eds. G. Garding and M. Tsujimura, pp. 165-178. Somerville, MA: Cascadilla Press.

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics* 19(6-7):619-33.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics,* 34(1), 104-137.

Davidson, L. (2007). The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology,* 24, 261-286.

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, 38(2), 272-288.

Davidson, L. and J. Shaw. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics,* 40, 234-248.

Delforge, A. (2008). Gestural Alignment constraints and unstressed vowel devoicing in Andean Spanish. In: C. Chang & H. J. Haynie, eds. *Proceedings of the 26th West Coast Conference on Formal Linguistics.* Somerville, MA: Cascadilla Proceedings Project.

Dell, F. & M. Elmedlaoui (1985). Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber, *Journal of African Languages and Linguistics* 7, 105-130.

Dell, F. & M. Elmedlaoui (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic.* Dordrecht: Kluwer.

Fabb, N. & M. Halle (2008). *Meter in Poetry: a New Theory.* Cambridge: Cambridge University Press .

Fleischhacker, H. (2005). *Similarity in phonology: Evidence from reduplication and loan adaptation*. (Doctoral dissertation), University of California, Los Angelos. (Order No. 3208408). Available from Dissertations & Theses @ University of California; ProQuest Dissertations & Theses A&I. (305033127). Retrieved from https://search.proquest.com/docview/305033127?accountid=14523

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology,* 18, 7-44. Retrieved from http://www.jstor.org/stable/4420187

Gay, T. (1981). Mechanisms in the Control of Speech Rate. *Phonetica* 38:148–158. doi:10.1159/000260020

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20(2), 269-337. https://doi.org/10.1023/A:1014942312445

Göksel, A. & C. Kerslake (2005). *Turkish: a comprehensive grammar.* New York, NY: Routledge.

Goldstein, L., I. Chitoran, and E. Selkirk. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. *Proceedings of the International Conference on Phonetic Science.* Saarbrucken.

Goldstein, L., H. Nam, E. Saltzman, and I. Chitoran. (2008). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in Phonetics and Speech Science: Festschrift for Wu Zongji.*

Gouskova, M., & Hall, N. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation* (pp. 203-225). London: Equinox Publishing.

Gu, Chong. 2002. *Smoothing spline ANOVA models*. Springer Series in Statistics. New York: Springer-Verlag.

Hale, M. & C. Reiss. (2008). *The phonological enterprise*. Oxford & New York: Oxford University Press.

Hall, K.C., Allen, C., McMullin, K., Letawsky, V., & Turner, A. (2015). Measuring magnitude of tongue movement for vowel height and backness. *International Congress of Phonetic Sciences.*

Hall, N. (2003). *Gestures and segments: vowel intrusion as overlap* (Doctoral dissertation). University of Massachusetts, Amherst. Available from Proquest. AAI3110499. https://scholarworks.umass.edu/dissertations/AAI3110499

– (2006). Cross-linguistic patterns of vowel intrusion. *Phonology* 23: 387–429. Cambridge University Press. doi:10.1017/S0952675706000996.

– (2011). Vowel Epenthesis. In The Blackwell Companion to Phonology (eds M. Oostendorp, C. J. Ewen, E. Hume and K. Rice). doi:10.1002/9781444335262.wbctp0067

– (2013). Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *Journal of Phonetics*, 41(2), 133-143. https://doi.org/10.1016/j.wocn.2012.12.001

Hall, N. & E. Sue (2018). *Hocank (Winnebago) vowel epenthesis: A phonological re-examination in light of phonetic data.* 15th Old World Conference on Phonology, London, United Kingdom. Abstract retrieved from http://www.ucl.ac.uk/~ucjtcwh/OCP15/abstracts/OCP_15_paper_96.pdf

Halpert, C. (2012). Overlap driven consequences of Nasal place assimilation. In: P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, B. Kühnert, eds. *Consonant Clusters and Structural Complexity.* Walter de Gruyter, Berlin/Boston.

Hansson, G. 2001. *Theoretical and typological issues in consonant harmony*. Dissertation. UC Berkeley.

Hayes, B. & A. Kaun (2001). *The Role of Phonological Phrasing in Sung and Chanted Verse. The Linguistic Review*. 13. 10.1515/tlir.1996.13.3-4.243.

Hermes, A., M.Grice, D. Mücke, H. Niemann (2012). Articulatory coordination and the syllabification of word initial consonant clusters in Italian. In: P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, B. Kühnert, eds. *Consonant Clusters and Structural Complexity* (pp. 157–176). https://doi.org/10.1515/9781614510772.157

Honorof, D.N. & Browman, C.P. (1995). The center or edge: how are consonant clusters organised with respect to the vowel? In: K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (vol.3, pp. 552-555). Stockholm, Sweden: Congress Organisers at KTH and Stockholm University.

Inkelas, S., G. Ö. Hansson, A. Küntay, O. Orgun (2001). Labial attraction in Turkish: an empirical perspective. *Turkic languages* (5), 169-97.

Inkelas, S., A. Küntay, R. Sprouse & O. Orghun (2000). Turkish Electronic Living Lexicon (TELL). *Turkic Languages* 4, 253-275.

Ito, J., & A. Mester (1995). The core-periphery structure of the lexicon and constraints on reranking. In: J. Beckman, S. Urbanczyk & L. Walsh, eds. *Papers in Optimality Theory*,181-210. Amherst: GLSA.

Jannedy, S. (1994). High vowel devoicing in Turkish. *Journal of the Acoustical Society of America*. 96(5). doi: 10.1121/1.410727.

Kaun, A. (1995). *The typology of rounding harmony: an Optimality Theoretic approach* (Doctoral dissertation). University of California, Los Angelos. (Order No. 9530008). Available from ProQuest Dissertations & Theses A&I. (304173330). Retrieved from https://search.proquest.com/docview/304173330?accountid=14523

– (1999). Epenthesis-Driven Harmony in Turkish. *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society: Special Session on Caucasian, Dravidian, and Turkic Linguistics (2000)*, 95-106.

Keating, P. A. (1979). *A phonetic study of a voicing contrast in Polish.* Doctoral dissertation, Brown University.

– (1990). The window model of coarticulation: articulatory evidence. In J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology, volume 1*. Cambridge University Press.

Kiliç, M. H., & F. Öğüt (2004). A high unrounded vowel in Turkish: Is it a central or back vowel? *Speech Communication,* 43, 143–154.

Kimper, W. (2017). Not crazy after all these years? Perceptual grounding for long-distance vowel harmony. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 19. DOI: http://doi.org/10.5334/labphon.47

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Hearing and Language Research*. 18(4): 686-706.

Kühnert, B., P. Hoole, C. Mooshammer (2006). Gestural overlap and C-center in selected French consonant clusters. 7th International Seminar on Speech Production (ISSP), Dec 2006, Ubatuba, Brazil. pp. 327-334, 2006. <halshs-00185737>

Kuznetsova, A., P. B. Brockhoff, & R. H. B. Christensen (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. Retrieved from https://CRAN.R-project.org/package=lmerTest

Lamontagne, Gregory A. (1993). "Syllabification and consonant cooccurrence conditions" . Doctoral Dissertations Available from Proquest. AAI9316687. https://scholarworks.umass.edu/dissertations/AAI9316687

Lavoie, L.M. & A. C. Cohn (1999). Sesquisyllables of English: the structure of vowel-liquid syllables. *Proceedings of the International Conference of Phonetic Science*. San Franscico.

Lees, R. B. (1966). On the interpretation of a Turkish vowel alternation. *Anthropological Linguistics* 8(9), 32-39.

Lerdahl & Jackendoff (1983). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Lewis, G. (1967). *Turkish Grammar*. Oxford: Oxford University Press.

Li, M.; C. Kambhamettu; and Maureen Stone. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7), 545-554.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773. doi: 10.1121/1.1918816.

Lisker, L. (1989). On the articulatory interpretation of vowel "quality": The dimension of rounding. *Journal of the International Phonetic Association*, 19, 24-30.

Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. Journal of the Acoustical Society of America, 105(3), 1864-1876. http://dx.doi.org/10.1121/1.426723

Marin, S. (2013). The Temporal Organization of Complex Onsets and Codas in Romanian: A Gestural Approach. *Journal of Phonetics*, 41 (3-4), 211-227, doi:10.1016/j.wocn.2013.02.001

Marin, S. & Pouplier, M. (2010). Temporal Organization of Complex Onsets and Codas in American English: Testing the Predictions of a Gestural Coupling Model. *Motor Control*, 14, 380-407

– (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics* 42: 24–36.

McCarthy, John J., "Sympathy, cumulativity, and the Duke-of-York gambit" (2003). *The Syllable in Optimality Theory*. 79. Retrieved from https://scholarworks.umass.edu/linguist_faculty_pubs/79

McPherson, L. & K. M. Ryan (2018). Tone-tune associations in Tommo So (Dogon) folk songs. *Language* 94(1).

Mücke, D., A. Hermes, L. Goldstein. (2012) Coupling of tone and constriction gestures in pitch accents. In: P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, B. Kühnert, eds. *Consonant Clusters and Structural Complexity*, pp.205-230. https://doi.org/10.1515/9781614510772.205

Murdoch, D. & E. D. Chow (2018). ellipse: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions. R package version 0.4.1. Retrieved from https://CRAN.R-project.org/package=ellipse

Myers, S. (2000). Boundary disputes: the distinction between phonetic and phonological sound patterns. In: *Phonological Knowledge: Conceptual and Empirical Issues*, ed. by Burton-Roberts, N., P. Carr and G. Docherty, 245-272. Oxford: Oxford University Press

Nam, H. (2004). A Competitve, Coupled Oscillator Model of Moraic Structure. Abstract. *9th Conference on Laboratory Phonology*, June 24-26, 2004. Retrieved from https://labphon.org/sites/default/files/previous_conferences/LabPhon9.pdf

Nam, H., L. Goldstein, & E. Saltzman (2009). Self organization of syllable structure: a coupled oscillator model. In: eds. François Pellegrino, Ioana Chitoran, Egidio Marsico, Christophe Coupe (2009). *Approaches to Phonological Complexity*. Walter de Gruyter.

Nam, H. & E. Saltzman (2003). A competitive, coupled oscillator of syllable structure. *Proceedings of the XIIth International Congress of Phonetic Sciences* (3): 2253-2256.

Ng, S. (2017). Music & syllables. Undergraduate thesis, Utah.

O'Dell, M. & T. Nieminen (2009). "Coupled Oscillator Model for Speech Timing: Overview and Examples." Nordic Prosody X, At Helsinki, Nordic Prosody: Proceedings of the Xth Conference, Helsinki 2008, p. 179-189

Öhman, S.E.G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *Journal of the Acoustic Society of America,* 39, 151. doi:10.1121/1.1909864

Padgett, Jaye (2011). Consonant-vowel place feature interactions. In Van Oostendorp, M., C. Ewen, E. Hume & K. Rice (eds.) *The Blackwell companion to phonology*, vol. 3. Malden, MA: Wiley-Blackwell, 1761-1786.

Patel, A. D. (2008). *Music, language, and the brain.* New York, NY, US: Oxford University Press.

Pöchtrager, M. (2016). Is there phonological vowel reduction in Turkish? In M. Güven, D. Akar, B. Öztürk, M. Kelepir (eds.) *Exploring the Turkish Linguistic Landscape: Essays in honor of Eser Erguvanlu-Taylan.* Amsterdam: John Benjamins Publishing Company. 21-39.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In: S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 63–96). Frankfurt am Main: Peter Lang.

Pouplier, M., & Beňuš, Š. (2011). On the phonetics status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2, 243–273

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Reetz, H. & A. Jongman (2008). *Phonetics*. Malden, MA: Wiley-Blackwell.

Rega, Lawrence. (2018) "Will learning French bring Turks more freedom?" Cafe Babel, June 11, 2018. Accessed March 26, 2019. https://cafebabel.com/en/article/turkey-will-learning-french-bring-turks-more-freedom-5b0ef527f723b37a833af2b7/.

Rose, S. & R. Walker. (2011). Harmony Systems. In J. Goldsmith, J. Riggle, A.C.L. Yu, eds. The Handbook of Phonological Theory, Second Edition. Blackwell Publishing. https://doi.org/10.1002/9781444343069.ch8

Russell Webb, E., & T. Bradley. 2009. Rhotic Metathesis Asymmetries in Romance: Formalizing the Effects of Articulation and Perception on Sound Change. In: *Romance Linguistics 2007: Selected Papers from the 37th Linguistic Symposium on Romance Languages (LSRL)*. ed. by P. J. Masullo, E. O'Rourke, & C. Huang, 321-337. Amsterdam: John Benjamins.

Rysling, A. (2017). *Preferential early attribution in segmental parsing* (Doctoral dissertation 1122). University of Massachusetts, Amherst. https://scholarworks.umass.edu/dissertations_2/1122

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. Psychological Review, 94(1), 84-106. http://dx.doi.org/10.1037/0033-295X.94.1.84

Saltzman, E. & D. Byrd (2000). Task-dynamics of gestural timing: Phase windows and multi-frequency rhythms. *Human Movement Science*, 19, 499-526.

Schmeiser, B. (2009). An acoustic analysis of intrusive vowels in Guatemalan Spanish /rC/clusters. *University of Pennsylvania Working Papers in Linguistics*, 15(1), 193-202.

– (2009). On the Current State of Vowel Intrusion Analysis in Spanish within Optimality Theory. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, ed. Joseph Collentine et al., 207-218. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2214.

Schreuder, M. & D. Gilbers (2004). The Influence of Speech Rate on Rhythm Patterns. In: Gilbers, D., Schreuder, M., and N. Knevel (eds.). *On the Boundaries of Phonology and Phonetics*. (pp. 183-202). Rijksuniversiteit Groningen.

Shaw, J., A. Gafos, P. Hoole. (2009). Syllabification in Morrocan Arabic: evidence from patterns of temporal stability in articulation. *Phonology* 26: 187–215. doi:10.1017/S0952675709001754

Shaw, J. & L. Davidson. (2011). Perceptual similarity in input-output mappings: A computational/experimental study of non-native speech production. *Lingua,* 121, 1344-1358.

Shaw, J. & S. Kawahara (2018). The lingual gesture of devoiced [u] in Japanese. *Journal of Phonetics* 66, 100–119.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in Gradient Symbol Systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science,* 38, 1102-1138.

Smolensky, P. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. Ms. http://roa.rutgers.edu/content/article/files/1552_smolensky_1.pdf

Sole, M.-J. (1992). Phonetic and phonological processes: the case of nasalization. *Language and speech*, 35(1,2), 29-43.

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

Steriade, D. (1990). Gestures and autosegments: Comments on Browman's and Goldstein's paper. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology II: Between the Grammar and Physics of Speech*, 382-397. Cambridge: Cambridge University Press.

Strycharczuk, P. (2012). *Phonetics-phonology interactions in pre-sonorant voicing.* Doctoral dissertation, University of Manchester.

Tilsen, S., & Cohn, A. C. (2016). Shared Representations Underlie Metaphonological Judgments and Speech Motor Control. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 14. DOI: http://doi.org/10.5334/labphon.52

Traunmüller, H. (1990). "Analytical expressions for the tonotopic sensory scale." *Journal of the Acoustic Society of America,* 88, 97-100.

Underhill, R. (1986). Turkish. *Studies in Turkish Linguistics*, eds. D. Slobin and K. Zimmer. Philedelphia: John Benjamin Publishing Company.

Walter, M. (2018). The significance of vowel insertion in Ottoman Turkish. In: Akıncı, Mehmet Ali & Kutlay Yagmur, *The Rouen Meeting: Studies on Turkic Structures and Language Contacts.* Turkologica 114. Wiesbaden, Germany: Harrassowitz Verlag.

Warner, N., Jongman, A., Cutler, A., & Mücke, D. (2002). The phonological status of Dutch epenthetic schwa. *Phonology,* 18, 387-420.

Wrench, Alan. (2008). *Articulate Assistant user guide*, version 1.17.

Yavaş, M. (1980) Some pilot experiments on Turkish vowel harmony. *Linguistics*, 13(3), 543-562. doi:10.1080/08351818009370510

Yıldız, Y. (2010). *Age Effects in the Acquisition of English Onset Clusters by Turkish Learners: An Optimality-Theoretic Approach*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Yip, J. (2013). *Phonetic Effects on the Timing of Gestural Coordination in Modern Greek Consonant Clusters* (Doctoral dissertation). University of Michigan, Ann Arbor. http://hdl.handle.net/2027.42/102475

Yun, S. (2016). *A theory of consonant cluster perception and vowel epenthesis* (Doctoral dissertation). Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/107089

Zimmerer, F. & B. Kabak (2010). Epenthetic vowels: Mere gestural mistiming? Poster presented at LabPhon 2010. DOI: 10.13140/RG.2.2.23724.97921

Zsiga, E. (1997). Features, gestures, and Igbo vowels: An approach to the phonology/phonetics interface. *Language* 73(2): 227 – 274.