

Gradient and Categorical Consonant Cluster Simplification in Persian:  
An Ultrasound and Acoustic Study

Reza Falahati Ardestani

Thesis submitted to the  
Faculty of Graduate and Postdoctoral Studies  
in partial fulfillment of the requirements  
for the PhD degree in Linguistics

Department of Linguistics  
Faculty of Arts  
University of Ottawa

## Abstract

The main goal of this thesis is to investigate the nature of an optional consonant deletion process, through an articulatory and acoustic study of word-final consonant clusters in Persian. Persian word-final coronal stops are optionally deleted when they are preceded by obstruents or the homorganic nasal /n/. For example, the final clusters in the words /næft/ “oil”, /suχt/ “burnt” and /qæsd/ “intention” are optionally simplified in fast/casual speech, resulting in: [næf], [suχ], and [qæs]. What is not clear from this traditional description is whether the coronal stop is truly deleted, or if a coronal gesture is produced, but not heard, because it is obscured by the adjacent consonants. According to Articulatory Phonology (Browman & Goldstein 1986, 1988, 1989, 1990a, 1990b, 1992, 2001), the articulatory gestures of the deleted segments can still exist even if the segments are not heard. In this dissertation, ultrasound imaging was used to determine whether coronal consonant deletion in Persian is categorical or gradient, and the acoustic consequences of cluster simplification were investigated through duration and spectral measures. This phonetic study enables an account for the optional nature of the cluster simplification process. A general phonological account is provided for the simplification of coda clusters with rising sonority, and the acoustic and articulatory investigation focuses on the simplification of clusters with coronal stops.

Ten Persian-speaking graduate students from the University of Ottawa and Carleton University, five male and five female, aged 25-38 participated in the articulatory and acoustic study. Audio and real time ultrasound video recordings were made while subjects had a guided conversation with a native speaker of Persian.

662 tokens of word-final coronal clusters were auditorily classified into unsimplified and simplified according to whether they contained an audible [t]. Singleton coda consonants and singleton /t/s were also captured as controls.

The end of the constriction plateau of  $C_1$  and beginning of constriction plateau of  $C_3$  were used to define a time interval in which to measure the coronal gesture as the vertical distance between the tongue blade and the palate. Smoothing Splines ANOVA was used in a novel way to compare tongue blade height over time across the three conditions.

The articulatory results of this study showed that the gestures of the deleted segments are often still present. More specifically, the findings showed that of the clusters that sounded simplified, some truly had no [t] gesture, some had gestural overlap, and some had reduced gestures. In order to explain the optional nature of the simplification process, it is argued that the simplified tokens are the result of two independent mechanisms. Inevitable mechanical and physiological effects generate gesturally reduced and overlapped tokens whereas planned language-specific behaviors driven by phonological rules or abstract cognitive representations result in no [t]-gesture output. The findings of this study support the main arguments presented in Articulatory Phonology regarding the underlying reasons for sound patterns and sound change. The results of this study are further used to examine different sound change models. It is argued that the simplified tokens with totally deleted [t] gesture could be the result of speakers changing their representations based on other people's gestural overlap. This would be instances of the Choice and Chance categories in Blevins' (2004) CCC sound change model. The acoustic results did not find any major cues which

could distinguish simplified tokens from controls. It is argued that articulatory data should form an integral part of phonetic studies.

## *Dedication*

این رساله دکتری را به روح نازنین مادرم که اولین آواها و نغمه های زبان فارسی را به من آموخت تقدیم می کنم

**This dissertation is dedicated to my late mom who taught me the first sounds and  
melody of Persian**

## Acknowledgements

In the journey of life we are sometimes blessed and feel privileged to have had the chance to meet some individuals who, like landmarks, bring big changes and play pivotal role in one's academic work, career, as well as personality.

First, I would like to deeply thank my dear supervisor Jeff Mielke for all he taught me. In the first seminar course that I took with Jeff, he introduced the ultrasound technology and opened a new window for research before my eyes. Jeff is an extraordinary supervisor. His vast and deep knowledge of the field as well as his experimental skills put him in a very special position. I have learnt so much from him. Thank you, Jeff, for the very high standards of academic work and the great responsibility for your students. You always encouraged me to participate and present my works at different conferences. Thank you also for your very quick replies to my emails and very constructive and thoughtful feedback on my work even on weekends. Thank you also for supporting me during the time of unexpected events and giving me the confidence that I needed. I shall never forget your full support and all you taught me during the years of my studies at the University of Ottawa. For sure, this work could not be accomplished without your supervision and guidance. I am also very grateful to Marie-Hélène Côté, my committee member. I had a phonology course and my first Comprehensive Exam with her in the department. This dissertation has benefited from the work I have done with her during the years of my program. Thank you Marie-Hélène for all the thoughtful and stimulating discussion we had in your office. You looked at my data from an angle that I never looked at and always had wonderful questions and comments for me. You were always very supportive during the whole years of my program as the chair of department as well.

I would also like to thank my other committee members: Ian MacKay, Marc Brunelle, and Shahrzad Mahootian. I took my first course in acoustic phonetics with Ian. I learnt a lot about the acoustics of the speech signal and its implications for laboratory research in that course. Ian is a wonderful teacher who has fully dedicated himself to his job. Thank you, Ian, for accepting to be on my thesis committee. I took two courses with Marc and worked for him as a research assistant. The courses I took with Marc really deepened my understanding about phonology. Marc gave me very thoughtful and careful feedback on my work. Thank you, Marc, for all your support and encouragements. I would also like to extend my thanks and appreciations to my external examiner. Shahrzad has a profound knowledge of Persian phonology and provided very constructive comments on this aspect of the thesis. Thank you, Shahrzad, for your very meticulous and encouraging comments on my thesis! I feel lucky to have you as my external examiner.

My special thanks and appreciations also go to my other professors who taught me different disciplines of linguistics during my doctoral program: Juana Muñoz-Liceras, Éric Mathieu, Laura Sabourin, Gerard Van Herk, and Catherine Anderson.

I also wish to extend my gratitude to our administrative personnel who always provided the best service and help to students in the department: Jeanne D'Arc Turpin: administrative assistant; Danielle O'Connor: academic assistant; Donna Desbiens: secretary; and Maurice Bélanger, our lab technician.

I also wish to acknowledge the major influence I received by the first courses I took in linguistics during my MA program at the University of Victoria. The first courses I took with Ewa Czaykowska-Higgins in phonology were so stimulating which sparked the earliest interest of working in this field in me. Ewa: Thank you for introducing the world of phonology to me. During my MA program, I benefited from many great teachers including

Robert Anthony who acted as my mentor during the years of my study at UVic. Robert: I deeply appreciate all your valuable advice and illuminating suggestions throughout my MA program and even after. You are always in my memory and I always remember you with great respect and appreciations.

I feel lucky and blessed to cherish the feel of love and support from my friends and colleagues in linguistics and outside linguistics: Gita Zareikar, Ahmed Abada, Ladan Hamedani, Pouneh Shabani, Nooshin Boroumand, Anousha Sedighi, Devon Woods, Mehdi Fallahpour, Sonia Pritchard, Lyra Magloughlin, Joe Roe, Saleh AlQahtani, Abdul Alamri, Jumanah Abusulayman, Tharanga Weerasooriya, Pouya Pourbeik, Yukiko Yashizumi, Moji Sedaghati, Aida Farkish, Javid Fereidoni, Mohsen Mobaraki, Abbas Akbari, Paul Melchin, Viktor Kharlamov, Farshad Khunjush, Daryl Chow, Jonathan Lim, Arezou Kashani, Pat McFadden, Zinat Goodarzi, Mitra Bahmannia, Majid Naji, Payam Sadeghi, Brandon Fry, and Brendan McDonald.

I would also like to appreciate and thank all participants in my study and ISAUO (Iranian Students' Association at the University of Ottawa) which helped me in recruiting subjects.

Last but not least, my deepest thanks go to my dear family: my parents and my sisters: Zahra, Zohreh, and Shohreh. You were the biggest support to me during the whole years of my education. I also wish to thank my brother-in-law, Ahmad, for his endless encouragement and positive energy. I should also mention Mahsa, Shayan, Pouya, and Saman, my niece and nephews, who entertained me and made me laugh by their emails and calls. This dissertation is dedicated to my late mom who was the biggest support to me in both my academic and personal life. You are always missed ..!

## Table of Contents

<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Feature-based Account of Deletion .....	3
1.2 Articulatory Phonology .....	6
1.3 Categorical versus Gradient Processes.....	11
1.4 Planned versus Mechanical Phonetic Processes .....	14
1.5 Role of the Listener and the Speaker in Sound Change .....	17
1.5.1 Ohala’s Sound Change Model .....	17
1.5.2 Blevins’ Sound Change Model .....	19
1.5.3 Lindblom’s Model.....	22
<b>Chapter 2: Background on Persian.....</b>	<b>25</b>
2.1 Phonological Structure of Persian.....	25
2.2. Consonant Clusters in Persian .....	29
2.3 Cluster Reduction and Sonority.....	36
2.4 Different Categories of Cluster Simplification in Persian.....	40
2.4.1 Coronal Stop-final Clusters with Simplification.....	41
2.4.2 Coronal Stop-final Clusters without Simplification .....	43
2.4.3 Non-coronal Stop/Fricative-final Clusters with Simplification .....	44
2.4.4 Non-coronal Stop-final Clusters without Simplification.....	45
<b>Chapter 3: Previous Studies of Consonant Articulation and Coronal Deletion .....</b>	<b>49</b>
3.1 Articulation of Consonants.....	49
3.2 Coronal Stop Deletion.....	54
3.3 Ultrasound Imaging of the Tongue .....	58
3.4 Methodological Innovations of the Study .....	59

<b>Chapter 4: Acoustic Experiment.....</b>	<b>62</b>
4.1 Methods of Acoustic Experiment .....	62
4.1.1 Subjects .....	62
4.1.2 Procedures .....	63
4.1.3 Data Analysis.....	63
4.2 Acoustic Results and Analysis .....	68
4.2.1 Statistical Analyses .....	70
4.3 Discussion for Acoustic Results .....	77
<b>Chapter 5: Articulation Experiment.....</b>	<b>79</b>
5.1 Methods of Articulation Experiment .....	80
5.1.1 Subjects .....	80
5.1.2 Procedures .....	80
5.1.3 Data Analysis.....	84
5.2 Articulatory Results and Analysis .....	102
5.3 Discussion for Articulatory Results.....	141
<b>Chapter 6: General Discussion and Conclusion .....</b>	<b>143</b>
6.1 Overall Discussion on Research Questions .....	143
6.2 Directions and Suggestions for Future Research .....	154
<b>References .....</b>	<b>158</b>

## List of Appendices

<b>Appendix 1:</b> The topics/questions used in the conversations with participants (in Persian) .....	<b>169</b>
<b>Appendix 1:</b> The topics/questions used in the conversations with participants (translated into English) .....	<b>172</b>
<b>Appendix 2:</b> The short story used for data collection (in Persian).....	<b>174</b>
<b>Appendix 2:</b> The short story used for data collection (translated into English) .....	<b>175</b>
<b>Appendix 3:</b> The list of unsimplified and simplified tokens extracted from seven subjects in the study. ....	<b>176</b>
<b>Appendix 4:</b> The list of control tokens from seven subjects in the study.....	<b>182</b>

## Chapter 1

### Introduction

This thesis is an articulatory and acoustic study of word-final consonant clusters in Persian, in order to provide a phonetically informed account of a variable cluster simplification process. Word-final coronal stops are optionally deleted when they are preceded by obstruents or the homorganic nasal /n/. For example, the final clusters in the words /ræft/ “went”, /duχt/ “sew” and /qæsd/ “intention” are optionally simplified in fast/casual speech, resulting in: [ræf], [duχ], and [qæs]. The point which is unclear is whether /t/ deletion is categorical, i.e., whether the coronal gesture is completely absent in deleted tokens, or some traces of the gesture remains either reduced in magnitude or obscured by other consonants. Since the articulatory status of final cluster simplification is in question, the term “simplification” will be used throughout this dissertation in its acoustic/perceptual sense, to describe what happens when coronal consonants are perceived to be deleted, regardless of whether there is residual articulatory evidence.

The issue of incomplete deletion has been raised and discussed in the Articulatory Phonology literature. Articulatory Phonology (Browman & Goldstein 1986, 1988, 1989, 1990a, 1990b, 1992, 2001) is a theory of phonology which considers gestures as the building units of lexical and phonological processes. This theory considers the change of magnitude and temporal duration in the gestural score as the primary reasons for phonological changes. Browman and Goldstein (1990b), for example, hypothesized that the alveolar closure gesture

for the /t/ in *must be* [mʌsbi] is still present in fast/casual speech but is completely overlapped or hidden by the labial closure gesture. This would mean that the gesture for the alveolar /t/ (raising the tongue tip/blade towards the alveolar ridge) is masked during the time that lips are closed or narrowed for labial /b/ and that this masking, not articulatory deletion, is why the /t/ is not audible.

The next goal of this research is to investigate the *optional* nature of this process. Optional consonant deletion is a complex phenomenon which is not well known yet. The old literature on phonetics makes a distinction between modifications in articulation which are the result of planned reorganization of the gestural targets versus the ones which are inevitable and mechanical (Sweet 1877; Jespersen 1904). The former are the result of features which are present in the input to speech production or gestures (Browman & Goldstein 1986), whereas the latter are due to physiological or aerodynamic constraints. For example, the controlled acoustic properties adjust to durational variations across different speech rates to keep the perceptual distance constant, whereas low-level automatic phonetic effects remain the same at different speech rates (Sole' 1992, 1995). This dissertation aims at identifying the planned versus mechanical nature of optional coronal deletion in Persian as well.

Optional cluster simplification phenomena raise important questions involving the relationship between phonetics and phonology, planned versus unplanned alternations, and the connection between gradient and categorical phenomena.

This dissertation has major and minor contributions to the field. First, it is the first experimental study which uses tongue imaging system to study the coronal stop deletion in Persian. Next, there are very few studies which have collected articulatory data from spontaneous speech using ultrasound technology and the stabilising headset. A novel application of SSANOVA for comparing between distributions of gestures is another innovation in this study.

In the following section first I will discuss the feature-based account of deletion followed by an introduction to Articulatory Phonology. Next, a number of categorical and gradient patterns will be reviewed. A discussion about planned versus mechanical phonetic processes will follow this section. The final section of Introduction deals with the role of listeners and speakers in sound change.

### **1.1 Feature-based Account of Deletion**

The feature-based account of deletion uses the distinctive feature as its smallest unit. This does not allow partial deletion. Generative Phonology, which was founded by Morris Halle in the 1950's, initially took segments as its basic unit of phonological representation. According to this approach, segments are discrete units which are characterized by binary distinctive features. Jakobson, Fant & Halle (1953) originally proposed static binary features which were defined in acoustic terms and intended primarily as a means to characterize speech sounds by their acoustic properties. Halle (1959) subsequently proposed a set of articulatorily-defined features meant to account mainly for phonological processes.

Later feature models developed more to determine the similarity or contrast among segments and were used as the baseline for defining natural classes. Some of these models

also tried to incorporate more tangible and physical constraints into phonological models. Feature geometry is a good example. The main representational advantage of this model was adding new dimensions mainly representing the vocal tract and speech production system. In this model, features are organized into tiers which themselves are organized in relation to each other in turn. The significant aspect of this model was to show which features group together. If some sets of features could spread independently of each other, they were put on separate tiers. This gives the model the power to provide a uniform description of phonological processes. The following figure adapted from Clements and Hume (1995) illustrates the feature geometry model for consonants.

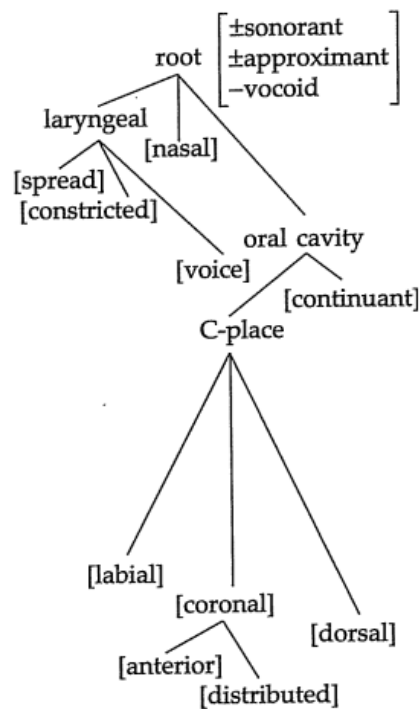


Figure 1.1 Feature geometry model (Adapted from Clements & Hume 1995, p. 292)

Assimilation happens when a segment in a sequence takes on some properties of a following or preceding segment. The nasal in the prefix *-in*, for example, becomes *-im* in

*impossible* by assimilating to the labial feature of the following consonant. Place assimilation, for example, is primarily accounted for in feature geometry by deleting the association line of a place node and spreading it to a new position on an adjacent tier (e.g., Clements 1985; McCarthy 1988). However, older models which did not use tiers to organize features had to formulate different phonological rules to explain a set of sound patterns in which the same features were involved. In Chukchi, for example, nasals assimilate to the place of articulation of the following consonants in the features [labial], [coronal], and [velar] (Clements & Hume 1995). In order to show these processes, the following rules could be formulated to introduce these three patterns informally:

$$[+nasal, -LAB] \rightarrow [+nasal, +LAB] / - [+cons, +LAB]$$

$$[+nasal, -COR] \rightarrow [+nasal, +COR] / - [+cons, +COR]$$

$$[+nasal, -DOR] \rightarrow [+nasal, +DOR] / - [+cons, +DOR]$$

The assimilation process in Chukchi, which is attested in other languages in the world, could be presented formally in this way. However, this formalism is not very appealing since these rules are not able to unify the three processes. The addition of a place tier not only does provide a uniform description of this sound alternation but also shows the active articulators involved in this sound change. This model showed a shift in incorporating general physical constraints into phonology.

Despite making the role of physical articulators more prominent, feature geometry still relied on features as the smallest phonological element. The whole physical properties of a feature are assumed to be bound and move together (see Kenstowicz, 2005, for further discussion). The transfer of features as abstract phonological units leads to categorical and

uniform alternations. In categorical processes time is only used as a reference point for segmental sequencing. There is no way to refer to particular moments within a segment or feature, if needed. This means that partial deletion of a segment cannot be captured as far as the system is bound with features as its minimal working unit.

One way to resolve the problem for describing the sound patterns which do not necessarily involve phonological manipulation of segmental feature values was to devise a unit which could identify the temporal and spatial aspects within articulators. Articulatory Phonology (Browman & Goldstein 1986, 1988, 1990a, 1990b) was proposed as a reformulation of the analytical units in phonology with gestures as the basic unit of phonological description. It is introduced in the next section.

## **1.2 Articulatory Phonology**

Articulatory Phonology (Browman & Goldstein 1986, 1988, 1989, 1990a, 1990b, 1992, 2001) takes dynamically-defined articulatory gestures as the constituent units of lexical and phonological representations. According to the computational model developed by Browman and Goldstein, a gesture is “an abstract characterization of coordinated task-directed movements of articulators within the vocal tract” (1989, p. 206). Gestures are “events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators” (1992, p. 156). An individual gesture is produced by groups of muscles that act in concert, sometimes ranging over more than one articulator. For instance, constriction of lip aperture involves the action of the upper lip, the lower lip, and the jaw; however, such a constriction is considered one gesture.

Gestures can be specified in terms of the tract variables involved in their production. In the computational model, there are 5 tract variables: Lips (L), Tongue Tip (TT), Tongue Body (TB), Velum (VEL), and Glottis (GLO). These tract variables are specified for constriction location (CL) and constriction degree (CD). Having constriction location as a variable guarantees that tract variables can make a constriction in more than one location along the vocal tract. Browman and Goldstein also have incorporated constriction degree into their model to make a distinction between different degrees of constriction made at the same location. There are five states for constriction degree in this model: [closure], [critical], [narrow], [mid], and [wide]. In this model, the tract variables for oral gestures are paired and have the two dimensions of constriction location and constriction degree. This model assumes that each of the articulators can move independently. The tract variables along with the articulators involved identified by Browman and Goldstein are shown in Figure 1.2.

Tract variable	Articulators involved
LP lip protrusion	upper & lower lips, jaw
LA lip aperture	upper & lower lips, jaw
TTCL tongue tip constrict location	tongue tip, tongue body, jaw
TTCD tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL tongue body constrict location	tongue body, jaw
TBCD tongue body constrict degree	tongue body, jaw
VEL velic aperture	velum
GLO glottal aperture	glottis

Figure 1.2 Tract variables and the articulators involved (adapted from Browman & Goldstein, 1990a, p. 301).

Gestures are abstract, discrete, and dynamic linguistic units which have an inherent spatiotemporal nature. These gestures are invariant when used in different segments; however their spatiotemporal quality allows them to overlap in time. Such an overlap between invariant gestures results in articulatory trajectories which vary based on the context. A set of gestures which shape a given utterance form a larger coordinated structure, or constellation, which is represented in a “gestural score”. The gestural score refers to a representation of a word in terms of its gestural units and its organization. A gestural score is

similar to autosegmental representation. This makes Articulatory Phonology similar to feature geometry as being a type of autosegmental representation. The minimal size of gesture and the overlap among them enables Articulatory Phonology to capture the processes which involve partial completion (e.g., partial deletion or partial assimilation). However, the feature geometry model lacks this power due to choosing a unit which does not lend itself to explaining phenomena where the unit involved in phonological processes is not a feature, node, or segment.

The innovative view of Articulatory Phonology is that the perception, production, and mental representations of speech all share the same basic primitive: articulatory gestures. According to Articulatory Phonology, physical systems impose constraints on the organization and description of phonological structures. According to this view, the output of phonological structure is an interaction between articulatory, acoustic, and purely linguistic organizations. A practical advantage of the gestural analysis of phonological patterns is that it yields a more fine-grained description of phonological phenomena than a feature-and-segment analysis does.

According to Articulatory Phonology, much of the phonological processes such as deletion and assimilation are the result of the overlap between invariant gestures. Browman and Goldstein (1990b) state that a number of patterns found in casual speech, such as deletion, insertion, and assimilation, can result from alternations in the magnitude and temporal duration in the gestural score. According to this perspective, a number of phonological phenomena can be explained by alternations in the timing relationship among gestures or alternating their magnitude.

Subsequent studies have shown that variation in gestural timing and magnitude can explain some phonological processes and sound patterns (e.g., Gick 1999; Davidson 2003; Proctor 2009). Sproat and Fujimura (1993) studied the allophonic variation in English /l/ using X-ray microbeam data. The primary articulatory gestures of light and dark variants of /l/ in English involve the tongue dorsum and the tongue blade. The results of the study show that the contrast between light and dark /l/ is quantitative and not categorical. Sproat and Fujimura explain that the variation between light and dark /l/ is due to differences in the timing and magnitude of the articulatory gestures.

As for the current study, Articulatory Phonology predicts that alveolar closure gesture for coronal stops could be present in the simplified clusters, but acoustically “hidden” by the surrounding gestures. This approach claims that the entire or partial movement of tongue tip towards alveolar ridge and then back may exist during the time of gestures for adjacent phonemes. According to this approach, overlap by following stop and fricative gestures can hide alveolar closure gestures more than liquids or glides. This approach also predicts that the following vowel may not have any overlap with the alveolar gestures of the preceding coronal stops. This approach also states that variation in phonological processes is quite systematic, i.e., the simplified and unsimplified clusters could vary along a single dimension of gesture timing.

The inclusion of physical constraints such as continuous gestural features will provide the framework with the power to show the completion or lack of completion for some phonological processes. Adopting a phonetic perspective which involves gradient and variable phenomena was considered to be indispensable for explaining some of the

phonological processes since not all events are categorical. In fact it is shown that some of these seemingly categorical patterns could have more justifiable phonetic explanations. The following section will review some of the studies in this area.

### 1.3 Categorical versus Gradient Processes

Many different studies have demonstrated that not all phonological processes can be described as categorical. Indeed, the gradient nature of some processes is well attested across different languages (Hardcastle & Roach 1979; Browman & Goldstein 1990b; Barry 1992; Byrd 1994; Hardcastle 1994; Zsiga 1995; Gow & McMurray 2007 among others). Hardcastle (1994), for example, showed that alveolar assimilation in English lacks complete stop closure on the alveolar ridge but shows some tongue blade or tongue body gestures. He used electropalatography (EPG) and combination of coronal + velar such as *Fred can*, *can go*, or *Susan can't* where assimilation potentially happens. The results of this study show that alveolar assimilation is not a total replacement of an alveolar stop by a velar stop, as a discrete phonological analysis would describe it.

Post nasal voicing of voiceless stops has also received both categorical and gradient explanation. Hayes and Stivers (1996) stated that this process is quite categorical in Ecuadorian Quechua, while the same process is shown to be gradient in English. In order to test this, Hayes and Stivers examined the amount of closure voicing in stop consonant /p/ where it was preceded by phonemes /m/ and /r/. The stop was shown to be more voiced after nasal /m/ than /r/. The speakers in the experiment showed significantly more closure voicing in some tokens like *tompa* than *tarpa*. The authors explain that there is a “phonetic tendency toward post-nasal voicing” (p. 29). It was also shown that post nasal effect was stronger in more rapid speech.

Zsiga (1995) in an electropalatographic experiment studied the articulatory behaviors of alveolar obstruents such as /t, d, s, z/ which turn into [tʃ, dʒ, ʃ, ʒ] before the palatal glide /j/ in American English. The palatalization process is obligatory at the lexical level as in *habit/habitual*, *confess/confession* pairs; however, this process is optional at the post-lexical level as in the phrases *made you* and *please yourself*. The results of this study showed that post-lexical palatalization is gradient and variable while lexical palatalization involves a categorical alternation. She argues that gradient palatalization happens mainly because of gestural overlap.

Previous studies on the deletion of consonants have provided different explanations for this process in different positions cross-linguistically. These accounts mainly differ from each other on the basis of whether they provide primarily a categorical account of this process (e.g., Clements 1985; McCarthy 1988) or whether they include substantive elements such as articulatory factors (Browman & Goldstein 1986, 1988, 1990a, 1990b, 2001) and/or acoustic/perceptual elements in their explanations (e.g., Ohala 1981a, 1995; Côté 2000, 2004a, 2004b; Mielke 2003, 2010).

Darzi (1991), for example, in the study of glottal consonant deletion in colloquial Persian assumes that the deletion of glottal /ʔ/ and /h/ is full and categorical and this results in the compensatory lengthening of the adjacent vowel. The glottal consonants /ʔ/ and /h/ are deleted in coda position in Persian. The words /ʃeʔr/ “poem”, /zohr/ “noon”, and /sobh/

“morning”, for example, turn into [ʃe:r],[zo:r], and [so:b]<sup>1</sup> in colloquial speech. Darzi uses a version of moraic theory proposed by Hayes (1989) and CV phonology to account for the deletion of these segments in Persian. However, some more recent phonetic studies have shown that the process of glottal deletion in Persian is partial and gradient. These studies have shown that glottals get weakened but not totally deleted (e.g., Sadeghi & Bijankhan 2007). Moreover, Shademan (2003) has shown that the vowel in this context is not always lengthened. Based on the standard view of moraic and X theories, the deletion of a moraic segment will trigger compensatory lengthening as the timing slot of the deleted segment is allocated to the adjacent vowel (Hayes 1989). The residue left after the glottal deletion and the disproportionate vowel length do not support the categorical account of this phenomenon in Persian.

In addition to glottal deletion, Persian shows the simplification of word-final clusters. The true nature of this phenomenon has not been investigated yet. It is not clear whether this process is categorical or gradient. If this process is not complete, we expect to find articulatory evidence of residual gestures. Previous studies have been based on acoustic data. In the absence of articulatory data, it is not possible to look for residual articulatory gestures from apparently deleted consonants and see whether deletion is gradient or categorical.

Another logical possibility is that categorical and gradient deletion patterns may coexist in the same speaker or the same population. Gradient deletion may also be an intermediate stage on the way to categorical deletion. The fact that this process may or may not be applied by speakers (i.e., this process being optional) could also help us understand

---

<sup>1</sup> I am using “:” here to show the lengthening of the vowel. This preference has no theoretical implications.

more about the nature of this process. Speakers may or may not delete the same coronal consonant in the same environment. The fact that speakers delete only a subset of tokens which potentially reduce could be interpreted in two ways. One interpretation is that speakers *randomly* delete the consonant because of physical or physiological constraints. The other approach could be that speakers *consciously* plan for such a process and its optional nature is under speaker's control. The next section elaborates more on this issue.

#### **1.4 Planned versus Mechanical Phonetic Processes**

One of the issues which has not been discussed thoroughly in the literature is the distinction between mechanical effects due to the physiological or aerodynamic systems of speech production organs and language-specific behaviors which are controlled by the speakers. The lack of references on this topic is partly due to the fact that the nature of a phonetic feature may not be the same for all language users at a given time (Sole' 2007). Different studies have shown that the status of phonetic features may change over time. This could be for boosting the saliency of speech or due to different interpretation of speakers' input features by listeners. For example, Kingston and Diehl (1994) explain that speakers may *decide* to alter acoustic cues, which used to have a physiological base, to enhance a contrast. The articulatory gestures used in producing high versus low vowels, for example, generate different acoustic cues. The high vowels have shorter duration and lower pitch while low vowels have longer duration and higher pitch. The listener who is aware of the perceptual effects of the articulatory gestures and intends to expand the vocalic contrast, deliberately adjusts the relevant acoustic cues via gestural changes to fulfill this task.

Researchers have proposed different methods in order to tease apart planned from unplanned phonetic processes. Ohala (1981b) states that manipulating syllable structure can

determine whether a phonetic process is intended by the speaker or it is the result of low-level mechanical effects. The underlying assumption is that the syllable structure has phonological effects on the duration of segments. In a phonetic process such as consonant epenthesis, if the epenthetic consonant affects the duration of the preceding segment, as usually happens to vowels before consonants, it could be concluded that this process is controlled and not due to low level phonetic constraints. For example, if the insertion of coronal stop [t] in prin[t]ce is planned, we should expect this affect the duration of the preceding nasal whereas if this is due to mechanical constraints, this should not have consequences on the temporal duration of preceding consonant. In addition to syllable structure, prosodic conditions (De Jong 2004) and speaking rate (Byrd 1994, 1996a, 1996b; Byrd & Cheng Tan 1996; Kirchner 2001) are suggested to change the temporal duration of features which are planned by the speaker. These factors are not expected to have such an effect on automatic mechanical phonetic processes.

In order to distinguish properties which are actively controlled by the speaker from properties which are the by-products of a mechanical execution, Sole' (2007) has listed a number of processes with durational variations. The status of vowel nasalization, VOT differences, and vowel duration differences before voiceless and voiced obstruents in various languages are suggested as good contexts to test the dichotomy of planned versus mechanical processes. The features under investigation in these processes are expected to behave differently if they are planned versus when they are biomechanical.

Sole' (1992, 1995) examined the oral and nasalized portions of a vowel followed by oral and nasal consonants in the production of three American and Spanish native speakers at four different speaking rates. The results of her study showed that the nasalized portion of

the vowel was longer at slower speech rates for American English speakers. Contrary to this, Spanish speakers showed a constant portion of nasalized vowel across different speaking rates. This means that the vowel nasalization process in American English is a language-specific process whereas for Spanish speakers it is an uncontrolled behavior.

This approach has also been adopted by Sole´ and Estebas (2000) on the implementation of the voicing contrast in stops in onset position in Southern British English and Catalan. These two languages differ from each other on the grounds that English uses VOT to signal voicing contrast whereas voicing contrast in Romance languages is done via presence or absence of vocal fold vibration during the constriction of consonants. Since VOT is a language-specific property of English, the ratio for the duration of aspiration to duration of segments is expected to remain the same under different speaking rates to ensure a constant perceptual contrast. The results of Sole´ and Estebas’ study showed a positive correlation between duration of aspiration and all four speech rates for the English speakers. This means that aspiration in English does not happen mechanically but that it is planned and executed by cognitive system. In contrast to English speakers, the results for Catalan speakers did not show any correlation between voicing lag and segmental duration. This indicates that duration of aspiration and speaking rate work independent of each other and this is a confirmation to its mechanical nature.

In general, the results of these studies show that variation in the temporal duration of segments triggers the temporal change of controlled effects whereas they have no impacts on mechanical effects. These effects can be interpreted as adjusting the gestural values to the perceptual demands. Considering how this process could unfold involves the role of the listener in sound change. The following section explores this topic.

## 1.5 Role of the Listener and the Speaker in Sound Change

### 1.5.1 Ohala's Sound Change Model

It is generally accepted that in human interactions, listeners are actively involved in adjusting and modifying their speech in response to what they hear from speakers. This could be in the form of normalization or (phonetic) accommodation. Normalization is the ability of listeners to correct and understand a wide range of variation which exists in speaker's production. This idea is mainly inspired by John Ohala in his listener-based theory of sound change (Ohala 1981a, 1995) and the further contribution of other scholars in the field (e.g., Mielke 2003, 2008; Babel & Johnson 2010). Ohala attributes a variety of sound changes to the listener's failure to accurately normalize the sounds produced by the speaker. The reason for such a failure on listener's part could be two: *hypocorrection* and *hypercorrection*.

Hypocorrection refers to patterns where listeners fail to correct for a contextual effect. Let's consider, for example, the underlying /VN/ where phonetic factor for changing it into a nasalized vowel [Ṽ] is coarticulatory variation. The lowering of the velum for producing a nasal consonant could overlap the vowel, causing vowel nasalization. In languages in which vowel nasalization is not phonologized, there could be variation in the gestural alignment with the oral articulators despite of the fact that the gesture itself is stable. This means that longer nasal consonants happen with shorter nasalized vowels and longer nasalized vowels co-occur with shorter nasalized consonants (Beddor 2009). If listeners attribute the intrinsic nasal features of the consonant to the vowel, this will lead to a sound system where nasal vowels will appear as phonemes.

On the other hand, hypercorrection happens when listeners wrongly attribute an intrinsic property of the input produced by speakers to contextual influence. For example,

let's assume a language where plain consonants such as /k/ are variably palatalized to /kʲ/ when followed by front vowels. The plain and palatalized consonants can variably happen in this context. However, in other contexts, the distinction between plain and palatalized consonants is phonemic. Listeners may take the intrinsic palatal feature of these phonemically palatalized consonants as a contextual effect and consider their underlying representation to be a plain consonant. The subsequent phonologization of the misparsed input by listeners amounts to a new phonological system. This is of course done through some articulatory and/or acoustic/perceptual paths. To conclude, Ohala's account of sound change considers listeners' perceptual errors as the main source of phonological change and emphasizes the active role of listeners in sound change.

Such work under the rubric of Laboratory Phonology has deepened our understanding of the major role articulatory and acoustic factors play in sound change. Laboratory phonology has facilitated experimental research into the questions that once were addressed primarily in formal linguistics. This trend has created the chance of revisiting the phonological processes which were investigated in the past. Mielke (2003), for example, showed how /h/ deletion in Turkish correlates with the low perceptibility of this segment in contexts where it deletes. Following Steriade, Mielke considers the amount of information lost resulting from the change as well as the instability of less perceptible sounds to be factors affecting deletion.

Ohala's model is primarily concerned with the initial stage of sound change (see also Hansson 2008), and it raises further questions about how this "mini sound change" (Ohala 1995) spreads throughout a speech community and how it changes into a phonologized

segment. I will resume this subject below when discussing the sound change model proposed by Blevins.

### **1.5.2 Blevins' Sound Change Model**

According to Blevins' Evolutionary Phonology model (Blevins 2004), recurrent sound patterns find their origins in phonetically motivated sound changes. According to Blevins' model, the three natural sources of sound change are Change, Chance, and Choice (this is also known as CCC sound change model). The first category Change happens due to perceptual similarities between the actual phonetic signal and the perceived one. In this sound change, the original phonetic utterance is misperceived by the listener. For example, the results of some experiments show that listeners show a biased tendency to hear /ki/ as /tʃi/ and /θ/ as /f/ (Eilers 1977; Guion 1998). The misperception on the side of the listener due to auditory reasons could sometimes lead to sound change.

The second type of sound change in this model is Chance. This sound change happens due to the intrinsic ambiguity of the original phonetic utterance. Here the listener accurately perceives the original signal but associates an underlying representation to it which is different from that of speaker's. According to Hansson (2008), this happens when the underlying representation /...k<sup>w</sup>u.../ is produced as [...k<sup>w</sup>u...] where the phonemic labial /k<sup>w</sup>/ is followed by a rounded vowel. The listener will hear this correctly as [...k<sup>w</sup>u...], but could attribute the intrinsic labial feature of the consonant to the coarticulatory effects of its adjacent vowel and take it underlyingly as /...ku.../. This is the same phenomenon described above as hypercorrection. Hansson (2008) emphasizes that this misinterpretation

need not affect the pronunciation of the listener. This type of sound change can be different from Change in three ways. First, the listener does not misperceive the acoustic signal whereas in Change such a misperception happens. Second, Change does not trigger any immediate change in the pronunciation of the listener. The last difference between these two sound change types is related to how they evolve. The probability of a sound change happening based on Change depends on the bias in human perception system; whereas, the priming effects produced by pre-existing sound patterns will boost the sound change (see Blevins, 2004, for further discussion).

The third category of sound change proposed in this model is Choice. According to Blevins (2004), “multiple phonetic signals representing variants of a single phonological form are accurately perceived by the listener, and due to this variation, the listener (a) acquires a prototype or best exemplar of a phonetic category which differs from that of a speaker; and/or (b) associates a phonological form with the set of variants which differs from the phonological form in the speaker’s grammar” (p. 33). For example, If we take the gestural overlap between adjacent consonants as discussed by Browman and Goldstein (1986, 1988, 1990a, 1990b, 2001) and assume that, due to gestural overlap, a consonant cluster like /XTK/ (X, T, and K represent uvular, coronal, and velar obstruents, respectively) can have multiple phonetic signals all of which are equally acceptable as follows:

Speaker: [X<sup>xxxxx</sup>T<sup>ttttt</sup>K<sup>kkkkk</sup>], [X<sup>xxx</sup>T<sup>txkk</sup>K<sup>kkk</sup>], [X<sup>xxxx</sup>T<sup>tttk</sup>K<sup>kkk</sup>], [X<sup>xxxx</sup>T<sup>ttkk</sup>K<sup>kkk</sup>]

Listener: [X<sup>xxxx</sup>T<sup>tttk</sup>K<sup>kkkk</sup>], [X<sup>xxxxx</sup>T<sup>tttt</sup>K<sup>kkkkk</sup>], [X<sup>xxxx</sup>T<sup>txkk</sup>K<sup>kkk</sup>], [X<sup>xxx</sup>T<sup>txkk</sup>K<sup>kkk</sup>]

The superscripts ‘x’, ‘t’, and ‘k’ show the range of gestures of the related consonants. If we could assume that there is no overlap between the consonants then their phonetic signal could look like [X<sup>xxxxx</sup>T<sup>ttttt</sup>K<sup>kkkkk</sup>]. However, studies have shown that in

fast/casual speech, there is varying degrees of gestural overlap and this could be less (e.g., [X<sup>xxxx</sup>T<sup>tttxk</sup>K<sup>kkkk</sup>]) or more (e.g., [X<sup>xxx</sup>T<sup>txxkk</sup>K<sup>kkk</sup>]) in some clusters. The change in articulation could also affect the acoustic percept of these clusters. The different degrees of overlap in clusters lead to variability in the acoustic signal. Listeners have access to the phonetic details of the ambient speech as well as information about specific lexical items. Listeners store the detailed acoustic information of the speech in their memory. Categories are represented in memory by the cloud of remembered tokens of that category. The variable nature of speech production may lead into a mismatch in “phonological form” or “best exemplar” between the speaker and the listener.

Both Ohala’s and Blevins’ sound change models take a synchronic and non-teleological approach to explaining phonological processes. According to this view, sound change does not happen in order to generate a more audible signal or an easier segment to articulate. The first two categories of sound change proposed by Blevins are very close to Ohala’s hypocorrection and hypercorrection. In both of them, the phonological representation formed by the listener differs from that intended by the speaker.

In addition to formalizing these types of sound change, Blevins contributes the Choice category where the role of speaker is more prominent. This departs from the Change category in two ways. First, in the Change category, the listener misperceives the speaker’s phonetic signal from the outset; whereas, in the Choice category, the listener accurately perceives the phonetic signal. Second, in Change category, the internal ambiguity in a one-to-one relationship between the speaker’s and the listener’s phonological form is the driving force for change whereas in Choice it is the many-to-one relationship itself which is the source of ambiguity.

The simplification of consonant clusters in Persian could be due to gestural overlap. This hypothesis corresponds to the multiple variants involved in Blevins' Choice category. Listeners hear tokens with more or less overlap and can interpret the more overlapped tokens as the intended utterance. As the target involves more overlap, the range of natural productions includes tokens where /t/ is inaudible, and listeners can either interpret that accurately as complete overlap, or interpret it as having no /t/. This could correspond to the final stage of Ohala's hypocorrection, resulting in complete deletion, and qualifying as Chance, because complete overlap is ambiguous between an intended /t/ and the absence of /t/.

A question which is equally as important here is to know whether this gestural overlap is due to mechanical effects or it is planned by the speaker. If the deletion of coronal stops is a language-specific process, we should expect variation across subjects on the magnitude of the gestural overlap. This could be triggered by varying speaking rates which affects temporal duration. From the other side, if mechanical effects are the underlying reasons for gestural overlap, no such a variation is expected. The next section introduces a third sound change model in which the speaker plays an active role in optimizing the speech signal.

### **1.5.3 Lindblom's Model**

The hyper- and hypo-articulation (H&H) theory proposed by Lindblom (1990) takes a very different approach to sound change. According to this theory, speakers are adaptive and take the active role of tuning their speech to accommodate the listeners. The principle of

economy plays a major role in this theory. Speakers have the choice of whether or not to undershoot phonetic targets. If undershoot or hypoarticulation does not impede accurate perception by the listener, then it will be applied by the speaker. In this model, speakers are aware of the factors which affect listener's perception.

This model is fundamentally different from Ohala's and Blevins' in more than one respect. This model takes a goal-oriented approach towards phonological processes. According to this, speakers are conscious about their own production and try to keep them maximally salient. At the same time, they also sometimes need to underarticulate their speech due to the economy principle and the demands of the speech setting. This model also diverges from the other two models because it takes speaker as the source of sound change. Based on this model, it is the speaker who anticipates the need for perceptability by the listener and adjusts his/her production accordingly.

Despite of the differences between this model and the ones discussed from Blevins, some commonalities could also be found. Lindblom's model considers speakers to play a major role in sound change, something which is echoed by Blevins' Choice category of sound change. This means that the listener accurately perceives the signal, and sound change happens when he/she assumes the role of a speaker. This is different from Blevins' Change and Chance categories, where the sound change is planted in the listener at the time when he/she misperceives or misparses the phonetic signal. But in Lindblom's model, the speaker consciously pursues the sound change while in Blevins' sound change model no such will exists.

The simplification of consonant clusters in these three models is accounted for differently. Ohala considers the underlying reason for consonant deletion to be hypocorrection where the acoustic cues of a consonant are hidden by the adjacent

consonants. Blevins' model involves more than one possible phonological form for the consonant clusters. Sound change happens when there is a mismatch between the "phonological form" or "best exemplar" of the listener and the speaker. In this model, listeners store the tokens they have been exposed to with detailed information. In contrast to Ohala's and Blevins' models, the deletion of coronal stops in consonant clusters in Lindblom's model would be due to the high precision that coronal stops require. If more energy is required to articulate a consonant that also has low cost on the side of the listener (because it is already difficult to recover from the signal and not critical to understanding the utterance), then deleting it is economical. According to this approach, the simplification of coronal stops at initial stage is planned by the speaker through hypo-articulation. However, both Ohala's and Blevins' models attribute it to the mechanical effects. According to these models, speakers do not necessarily plan to simplify clusters at the initial stage. This happens because of physical and physiological constraints in human's vocal tract.

The next chapter introduces Persian phonology and the different types of clusters which may or may not be simplified.

## Chapter 2

### Background on Persian

#### 2.1 Phonological Structure of Persian

Persian is a member of the Indo-European family and within that family it belongs to the Indo-Iranian branch. Its segmental inventory has twenty-three consonants and eight vocalic phonemes (six vowels and two diphthongs). The consonant and vowel inventories of Persian (more specifically, the Tehrani Farsi variety of Persian) are given below.

(1) Persian consonant inventory

	Labial	Labiodental	Alveolar	Postalveolar	Velar	Uvular	Glottal
Stop	p   b		t   d		k   g	q	
Fricative		f   v	s   z	ʃ   ʒ		χ	
Affricate				tʃ   dʒ			
Nasal	m		n				
Lateral			l r				
Glide			j				h ʔ

Where symbols appear in pairs, the one to the right represents a voiced consonant.

(2) Persian vowel and diphthong inventory

	Front	Back
High	i	u
Mid	e	o
Low	æ	ɑ
Diphthong		ow, ej

The inventory in (2) is the most typical arrangement of the vocalic segments in Persian. However, there are some studies which have used different representations for the two low vowels. For example, Rohany Rahbar (2012) used /a/ to show the front low vowel and Darzi (1991) and Mahootian (1997) used /a/ for back low vowel. The symbols for other four vowels were mainly used consistently in the literature. Quality distinction (i.e., height, backness / frontness) is considered to play the primary role in vowel inventory in Modern Persian whereas quantity is assumed to have a secondary role to play in the system as well (Darzi 1991; Mahootian 1997; Samareh 2002; Rohany Rahbar 2012). These studies mainly consider /i u ɑ / as long vowels and /e o æ/ as short vowels (see Mahootian, 1997, for more discussion). Old Persian is traditionally accepted to have only a quantity-based system for vowels with long and short pairs for /i u æ/ (Natal Khanlari 1987). It is suggested that the language has gone through the process of quantity loss for vowels from the middle to modern period (see Rohany Rahbar, 2012, for further discussion).

There is no general consensus among linguists on the status of the diphthongs. Some studies have investigated the phonemic status of the two diphthongs and claimed that /ow/ is

phonemic and /ej/ is not. This is attributed to the status of the individual segments of each diphthong in the inventory. Since consonants are the main focus of this thesis, no further discussion is presented here on vocalic segments.

As for the Persian consonant inventory, the status of glottal consonants also lacks consensus. The traditional phonetic classification describes /ʔ/ and /h/ as “glottal stop” and “glottal fricative”, respectively. The production of /ʔ/ involves “a complete closure of the glottis with no extra constriction of the larynx”, while /h/ involves “an open glottis typical of voiceless sounds but with turbulent airflow” (Durand 1990, p. 103). As for their acoustic features, the records of /ʔ/ correspond to “a short period of silence with abrupt rise time and cut-off in relation to a following and preceding vowel” while the records of /h/ correspond to broad spectrum noise from about 400 to 6500 hertz (Durand 1990, p. 103).

Kavitskaya (2002) convincingly argues that the glottal consonant /ʔ/ in Persian is an approximant. Firstly, she shows that cross-linguistically it is only the glottal consonant /ʔ/ in the stop class whose deletion causes compensatory lengthening. Compensatory lengthening is never triggered by the deletion of oral plosives. There is no other study reporting that the deletion of a plosive other than /ʔ/ triggers compensatory lengthening. Secondly, she states that the glottal consonant /ʔ/ does not share phonetic characteristics with other consonants such as fricatives and liquids, the segments which usually trigger compensatory lengthening in other languages. She considers this as another argument for not considering /ʔ/ as a glottal

stop in Persian. She further provides phonetic data from the spectrograms of two native speakers of Persian which show that the glottal consonant in question is an approximant. The vocalic nature of this segment is further supported by providing some evidence from Ket (a Siberian language) in which the glottal stop freely alternates with vowels.

Kavitskaya used Ohala's listener-oriented sound change model to develop a phonetically-based phonologization model to account for compensatory lengthening in Persian. Her results show a hypocorrection effect where the listener fails to correct for contextual effects. She argues that the deletion of glottals /ʔ/ and /h/ is due to their weak perception. Since glides have longer vocalic transitions compared to other consonants such as stops, the listener may take this longer transitional length as an intrinsic feature of the preceding vowel at the time of glide deletion.

The syllable structure in this language is considered to be (C)V(C)(C). This means that Persian can have the following six syllable types: V, CV, CVC, VC, VCC, and CVCC<sup>2</sup>. This language allows consonant clusters in the coda position. The maximum number of consonants in this position could be two; however, we may have three consonants in row if the following word starts with a consonant in its onset position. In the following section, I will present the possible word-final consonant clusters in Persian and discuss their simplification.

---

<sup>2</sup> Some linguists suggest that syllables in Persian should have a consonant in their onset position. This will reduce the number of syllable types in Persian to three: CV, CVC, and CVCC. Since the main focus of this study is consonant clusters in word-final position, this debate will not be an issue (see Samareh 2002).

## 2.2. Consonant Clusters in Persian

Persian allows various classes of consonants to combine together in the final position of words. These include stops, fricatives, affricates, nasals, liquids, and glides. Liquids have the highest number of combinations with other consonants. Stops and fricatives are also very dynamic in combining with other consonants either as C<sub>1</sub> or C<sub>2</sub> of the clusters. Persian allows all combinations of consonants except for liquid-liquid. A detailed list of attested word-final consonant clusters in Persian with some examples is provided in Table 2.1.

Table 2.1 Attested word-final clusters in Persian

Type	Combinations
stop-stop	/tb, qb, bt, qt, qd, bd, bk, tk, tq, bq, dq/
Examples:	/qotb/ “pole”, /næqb/ “burrow”, /zæbt/ “record”, /væqt/ “time”, /æqd/ “engagement”, /æbd/ “servant”, /sæbk/ “style”, /potk/ “sledge-hammer”, /notq/ “speech”, /tebq/ “according”, /sedq/ “truthfulness”
fricative-stop	/sb, zb, ft, st, χt, ʃt, ʃd, sd, zd, fq, sq, ʃq, zq, sk, ʃk/
Examples:	/næsb/ “installation”, /hezʃb/ “party”, /hæft/ “seven”, /dorost/ “right”, /loχt/ “naked”, /zeʃt/ “ugly”, /roʃd/ “growth”, /qæsd/ “intension”, /mozd/ “payment”, /bafq/ “a city”, /fesq/ “sinful act”, /eʃq/ “love”, /rezq/ “daily bread”, /mask/ “mask”, /χoʃk/ “dry”
nasal-stop	/mb, nb, mp, mt, md, nd, mq, nk, ng/
Examples:	/bomb/ “bomb”, /dʒænb/ “next”, /pomp/ “pump”, /sæmt/ “direction”, /æmd/ “intentional”, /tʃærænd/ “nonsense”, /omq/ “depth”, /bank/ “bank”, /ræng/ “color”

liquid-stop	/lp, lb, rb, lt, rt, ld, rd, lq, rq, lk, rk, rg/
Examples:	/ɒlp/ “name of a mountain”, /qælb/ “heart”, /sɔrb/ “lead”, /qælt/ “roll”, /tʃɔrt/ “nap”, /dʒeld/ “cover”, /χɔrd/ “ate”, /χæɪlq/ “creation”, /ʃærq/ “east”, /melk/ “land”, /tærk/ “depart”, /mærg/ “death”
glide-stop	/jb, ʔb, jd, hd, ʔd, ht, jt, jk/
Examples:	/qejb/ “unknown world”, /roʔb/ “fear”, /ejd/ “holiday”, /mæhd/ “cradle”, /bæʔd/ “after”, /tæht/ “under”, /bejt/ “home”, /pejk/ “messenger”
stop-fricative	/ds, ks, qs, bs, qz, bz, bʃ, qʃ, tf, qf, qv, dv, bχ/
Examples	/hæds/ “guess”, /æks/ “photo”, /ræqs/ “dance”, /hæbs/ “imprison”, /mæqz/ “mind”, /næbz/ “pulse”, /næbʃ/ “corner”, /næqʃ/ “picture”, /lotf/ “kindness”, /væqf/ “devotion”, /læqv/ “cancellation”, /bædv/ “beginning”, /tæbχ/ “cook”
fricative-fricative	/fs, χs, fz, χz, fʃ, χʃ, sf, ʃf, zf, ʃv, zv, fv, fχ, sχ/
Examples:	/næfs/ “nature”, /ʃæχs/ “person”, /hefz/ “keep”, /æχz/ “take”, /kæfʃ/ “shoe”, /pæχʃ/ “distribution”, /nesf/ “half”, /kæʃf/ “discovery”, /hæzf/ “deletion”, /næʃv/ “growth”, /ozv/ “part”, /æfv/ “forgive”, /næfχ/ “flatulence”, /fæsχ/ “abolition”
nasal-fricative	/ms, ns, mz, nz, nf, mʃ, nχ/
Examples:	/læms/ “touch”, /ʃans/ “chance”, /ræmz/ “code”, /tænz/ “joke”, /senf/ “job”, /ʃemʃ/ “bullion”, /senχ/ “kind”
liquid-fricative	/ls, rs, lf, rf, lχ, rχ, rv, rz, rʃ/
Examples:	/sɒls/ “one third”, /qɔrs/ “pill”, /serf/ “only”, /tæɪχ/ “bitter”, /sɔrχ/ “red”, /særv/ “cypress”, /dærz/ “seam”, /tɔrʃ/ “sour”

glide-fricative	/hs, js, ?s, hz, jz, ?z, jʃ, ?ʃ, hʃ, jf, ?f, hf, hv/
Examples:	/bæhs/ “discussion”, /hejs/ “direction”, /jæ?s/ “disappointment”, /mæhz/ “for”, /fejz/ “honour”, /væ?z/ “preach”, /ejʃ/ “joy”, /næ?ʃ/ “corpse”, /væhʃ/ “wild”, /hejf/ “pity”, /zæ?f/ “weakness”, /kæhʃ/ “cave”, /mæhv/ “disappear”
stop-nasal	/km, tm, qm, kn, bn, tn/
Examples:	/hokm/ “order”, /χæt̪m/ “end”, /ræqm/ “despite”, /rokn/ “principle”, /qæbn/ “loss”, /mætn/ “text”
fricative-nasal	/sm, zm, χm, ʃm, sn, ʃn, zn, fn/
Examples:	/dʒesm/ “object”, /næzm/ “order”, /ʃoχm/ “plough”, /tʃejm/ “eye”, /hosn/ “advantage”, /dʒæʃn/ “celebration”, /væzn/ “weight”, /dæfn/ “bury”
nasal-nasal	/mn/
Example:	/æmn/ “secure”
liquid-nasal	/lm, rm, rn/
Examples:	/elm/ “science”, /kerm/ “worm”, /qærn/ “century”
glide-nasal	/hm, jm, ?m, hn, jn, ?n/
Examples:	/fæhm/ “understanding”, /dejm/ “rain-watered”, /tæ?m/ “taste”, /zehn/ “mind”, /ejn/ “like”, /ʃæ?n/ “dignity”
stop-liquid	/bl, tl, dl, kl, ql, tr, dr, br, kr, qr/
Examples:	/tæbl/ “drum”, /sætl/ “bucket”, /ædl/ “justice”, /ʃekl/ “shape”, /æql/ “mind”, /ætr/ “perfume”, /qædr/ “value”, /æbr/ “cloud”, /mækr/ “trick”, /fæqr/ “poverty”

fricative-liquid	/fr, sr, zr, ʃr, ʒr, fl, sl, zl, ʎl/
Examples:	/kɒfr/ “blasphemy”, /æsr/ “afternoon”, /ɔzr/ “excuse”, /ɒʃr/ “degree”, /estæʎr/ “pool”, /qɒfl/ “luck”, /mesl/ “like”, /fæzl/ “superiority”, /dæʎl/ “income”
nasal-liquid	/ml, mr/
Examples:	/hæml/ “carry”, /omr/ “age”
liquid-liquid	...
glide-liquid	/ʔl, hl, jl, hr, jr/
Examples:	/dʒæʔl/ “forging”, /æhl/ “inhabitant”, /mejl/ “tendency”, /zæhr/ “poison”, /qejr/ “else”
stop-glide	/bh, dh, th, qh, bʔ, tʔ, qʔ/
Examples:	/sobh/ “morning”, /mædh/ “praise”, /sæth/ “surface”, /feqh/ “jurisprudence”, /robʔ/ “quarter”, /qætʔ/ “cut”, /væqʔ/ “respect”
fricative-glide	/sh, ʃj, fj, sʔ, fʔ, zʔ/
Examples:	/mæsh/ “rituals of prayer”, /mæʃj/ “method”, /næfj/ “rejection”, /vosʔ/ “ability”, /næfʔ/ “benefit”, /væzʔ/ “situation”
nasal-glide	/nh, mʔ, nʔ/
Examples:	/konh/ “nature”, /dʒæmʔ/ “addition”, /mænʔ/ “prohibition”
liquid-glide	/lh, rh, lʔ, rʔ/
Examples:	/solh/ “peace”, /ʃærh/ “explanation”, /zelʔ/ “side”, /færʔ/ “minor”
glide-glide	/hj, jʔ, ʔj/
Examples:	/væhj/ “revelation”, /bejʔ/ “sell”, /sæʔj/ “try”

The above list shows that Persian allows most of the consonant types to adjoin word finally. There are some other lists provided which lack a few clusters and are not as comprehensive as the list presented in this study. The list presented by Mahootian (1997), for example, does not include nasal-glide, liquid-liquid, and glide-glide combinations. One reason that she hasn't included them is that most of the words belonging to these clusters have Arabic origin. It should be noted that not all the attested combinations have the same frequency. The cluster types nasal-nasal, liquid-nasal, nasal-liquid, glide-liquid, fricative-glide, nasal-glide, liquid-glide, and glide-glide are found in fewer words compared to other combinations. There are also some consonants in the attested combination types which do not actively participate in making clusters. For example, the phonemes /p/, /χ/, and /v/ do not frequently adjoin to other segments to make clusters. In contrast to these phonemes, /l/ and /r/ are very dynamic in forming consonant clusters. There are also some other segments which show preferences in occupying either of the positions in clusters. Samareh (2002) states that both /h/ and /j/ show a strong tendency to occupy the first position in the clusters.

Some of the consonant clusters tend to simplify more than others. The process of cluster simplification usually happens in fast and/or colloquial speech. Pisowics (1985) confirms that the deletion of consonants in Persian is observed more often in everyday and common words. Mahootian (1997) lists the various types of consonant clusters in Persian which occur syllable finally and whether they get simplified. She defines the simplification of consonant clusters in final position as a "rule" in this language. Her study shows the predominant deletion of coronal stops /t/ and /d/ in colloquial Persian. Mahootian also

provides a limited number of words which show that /r/ is optionally deleted as the second element of a cluster.

The number of studies on consonant clusters in Persian is quite limited, and most of them have focused on the deletion of glottals. Working on the compensatory lengthening phenomenon, Shademan (2003) measured the duration of the vowel in words with the syllable template CVCC, CVGC, CVCG, CVC.CVC, and CVG.CVC, where G is used for glottal. The results of this study showed that the deletion of glottals in CVGC had no significant effect on the duration of the preceding vowel; however, it did have some effects in CVG.CVC words. This shows that the deletion of glottals does not always result in the lengthening of the vowel, particularly in word-final clusters. One could explain this by considering the phonologized nature of this process in this language and the fact that length distinction is not phonemic for Persian vowels. The initial phonetic effects of vowel lengthening could shift back to normal to respect the non-phonemic nature of length among vowels in this language.

In another study on glottal consonant deletion in Persian, Sadeghi and Bijankhan (2007) have suggested that the compensatory lengthening in Persian is a gradient process where the deletion of glottals is a matter of degree rather than category. Sadeghi and Bijankhan asked twenty native speakers of Persian to utter colloquially ten words with the CVGC template (five CVhC and five CV?C) and ten words with CVC syllabic shape which were the counterparts of the first ten words. They used pairs such as /zæhr/ “poison” vs. /zær/ “gold” and /bæ?d/ “later” vs. /bæd/ “bad” in their study. Using a spectral tilt measurement, they showed that the glottal gesture is indeed weakened in the coda position of CVGC words

and it acquires vocalic attributes. This study suggests that vowel length is dependent on the laryngeal variation of glottal consonants in coda position.

Assadi (2007) conducted an acoustic study on consonant and vowel deletion in colloquial Persian. She recorded the individual conversation of three native speakers which resulted in the production of 4038 consonants. The results of this study show that almost 8.1 % of the consonants get deleted, mainly glottals, alveolars, and velars (glottal 82%, alveolar 7.22 %, velar 7.44 %). To explain the deletion process in the language, Assadi considers the “position of the sound” and the “class of word” as the two main factors for the simplification of clusters. She explains that the final position in a word is more vulnerable to deletion because the articulatory effort in that position decreases. She further adds that the simplification of homorganic clusters is due to having a constraint based on articulatory gestures, requiring that the least possible effort be spent on producing intelligible sounds. Assadi attributes the higher rate of deletion in grammatical words rather than lexical words to their higher frequency.

In general, the number of phonetic studies on cluster simplification in Persian is low and none of them have specifically focused on coronal stops. There are a few studies which have proposed some descriptive generalizations for consonant cluster simplification including coronal stops in Persian (Mahootian 1997; Lazard 2006), but none of them have provided articulatory or perceptual analyses of this process in Persian. Moreover, these studies have not considered the clusters which are not reduced and they don't provide any explanation for why simplification does not apply to them. The current study mainly addresses one of the major categories of consonant cluster simplification in Persian: coronal stops. This study uses articulatory and acoustic data to provide an explanation for the

simplified word-final coronal stop clusters. This thesis also explores the optional nature of this process to see why the same speaker may or may not apply it. But before investigating the nature of coronal stop deletion, the simplifying consonant clusters that violate the Sonority Sequencing Principle (SSP) are discussed.

### **2.3 Cluster Reduction and Sonority**

Some consonant cluster simplification in Persian can be accounted for in terms of the Sonority Sequencing Principle (SSP). According to this principle, the occurrence of a set of consonants in the onset or coda position of a syllable is determined by the sonority hierarchy. This principle requires that “the level of sonority must not increase from the nucleus to the edges of the syllable, i.e., it must not decrease in the onset and increase in the coda” (Côté 2004b, p. 158). Clements (1990) proposes the following hierarchy for sonority:

(3) Glides (G) > Liquids (L) > Nasals (N) > Obstruents (O)

The occurrence of each class of sound in a cluster will have different effects on the violation of the SSP. The clusters which violate the SSP in Persian are formed by a combination of obstruent + glide, obstruent + liquid, obstruent + nasal, and nasal + liquid. The data given in (4) shows the words in which the last consonant in the cluster is more sonorous than the preceding segment and can get deleted; however, this does not apply to the words in (5).

- (4) a. /næfj/ → [næf] “rejection”  
 b. /mæfj/ → [mæf] “method”  
 c. /ʃekl/ → [ʃek] “shape”  
 d. /qofl/ → [gof] “lock”  
 e. /qæbl/ → [qæb] “before”  
 f. /æql/ → [æq] “wisdom”  
 g. /ætr/ → [æt] “perfume”  
 h. /ozr/ → [oz] “excuse”  
 i. /sæbr/ → [sæb] “patience”  
 j. /sefr/ → [sef] “zero”
- (5) k. /χæfm/ → [χæfm] \*[χæf] “anger”  
 l. /hæzm/ → [hæzm] \*[hæz] “digestion”  
 m. /æzm/ → [æzm] \*[æz] “will”  
 n. /væzn/ → [væzn] \*[væz] “weight”  
 o. /omr/ → [omr] \*[om] “age”  
 p. /æmr/ → [æmr] \*[æm] “order”  
 q. /hæml/ → [hæml] \*[hæm] “carry”

According to the sonority sequencing hierarchy given in (3), the first segments in the clusters given in (4) and (5) are less sonorous than the second segments. Therefore, they appear to violate SSP. However, not all clusters violating sonority will simplify. Those ending in a glide (4a-b) or a liquid (4c-j) do, but obstruent-nasal clusters remain intact (5k-q). The different behavior shown between the two patterns given in (4) and (5) can be related to

the degree of the SSP violation. If we assign a value to each class of sound and measure the sonority distance between the two adjacent consonants, then the intensity of each violation can be measured. We can then incorporate the parameters of SSP violation into a system like perception and expect that the clusters which violate the SSP more, or possibly the less perceptible clusters, are more marked and more prone to deletion. If we assign the following values to each class of sounds, we can determine the SSP violation in each cluster:

(6) Sonority Scale (Clements 1990)

<b>class</b>	<b>value</b>
Obstruent	0
Nasal	1
Liquid	2
Glide	3

The above scale shows that the clusters obstruent + glide violate the SSP more than obstruent + liquid clusters since the magnitude of the difference between the segments of the first cluster is 3, whereas it reduces to 2 for obstruent + liquid cluster. This means that the violation of SSP is relative to the magnitude of the violation between the two adjacent segments (see Côté ,2000, for further discussion).

Drawing on the notion of perceptibility, I provide an account for the two patterns in (4) and (5) here. Ohala (1990) considers perceptibility to be how easy a segment or a gesture can be identified in a sequence. Having different sources of acoustic information will naturally facilitate the perceptibility of these segments or gestures. Sonority is a way to achieve this goal. According to this principle, the sonority of the onsets rise up to nucleus, which is the peak of the sonority in a syllable, and then it will decrease towards the end of coda. The scale for sonority starts with the segments with very narrow constriction (i.e.,

stops or fricatives) and progress into segments with wider constriction (i.e., vowels or glides). Mattingly (1981) argues that the order of this scale crucially depends on the different degrees of available information within the articulatory stages of the constriction, application of the constriction, and its release. He states that “[T]he articulations of speech must be scheduled so that periods during which constrictions are released in rank order alternate with periods during which constrictions are applied in inverse rank order” (Mattingly 1981, p. 418). This is very similar to sonority scale where a narrow constriction such as that of a stop is released into a segment with more open constriction such as a glide. According to Chitoran, Goldstein and Byrd (2002), this will lead into a more overlap between segments. This in turn will allow for a simultaneous transition of a big amount of acoustic information about more than one phonetic element. According to this view, the clusters with the maximum amount of overlap which still can be recoverable are preferred and tolerated. The ones with the drastic change in their closure which block their recoverability are not tolerated and hence abandoned. According to the data presented in (4) and (5) above, the consonant cluster in [æmr] “order” is tolerated while the one in [sæbr] “patience” gets simplified. Both clusters in the coda start with a stop and are followed by a liquid; however, the stop in [æmr] is a nasal which involves velum lowering and labial closing gestures whereas the oral stop in [sæbr] only has labial closing gesture. The maximum modulations in several acoustic parameters simultaneously will raise the salience of an acoustic signal (Kawasaki 1984; Ohala 1990). The existence of two gestures for nasal stop will secure a bigger load of acoustic information compared to oral stop where only one of these gestures is present. Therefore, the first consonant in [æmr] is acoustically more salient making the cluster more

stable while the first consonant in [sæbr] is less salient. This leads into reducing the final consonant which renders the first consonant more salient. The analysis provided here for the clusters violating SSP offers a more uniform account which is in conformity with the articulatory and perceptual tenets of coronal stop deletion discussed later in the study.

The following sections present the four major categories of clusters with or without simplification with no violation of SSP.

#### **2.4 Different Categories of Cluster Simplification in Persian**

In this section I will provide different patterns of cluster simplification in Persian. This language does not have a very complicated pattern of cluster simplification when SSP is not violated. It displays a strong preference for stop deletion, similar to other languages such as English, Québec French, Hungarian, Icelandic, and Attic Greek (Browman & Goldstein 1990; Bayley 1994; Côté 2000). The stop class which is mostly targeted in Persian is limited to one place of articulation, namely coronal. Coronal stops in cluster-final position get deleted after all types of consonants except glides, liquids, and the nasal /m/. The other consonants such as fricatives, nasals, and liquids do not get deleted in the final position. There are also some non-coronal consonants which are deleted in some restricted environments, but they are still obstruents.

In order to describe the behavior of clusters which do not violate the SSP, a four way classification of the clusters is made in this study. The first category is the coronal stop-final clusters which optionally simplify. These clusters have stops, fricatives and the nasal /n/ as the first element in the cluster (i.e., C<sub>1</sub>). The second category is the coronal stop-final clusters

which never simplify. These clusters have glides, liquids and the nasal /m/ as C<sub>1</sub>. The third category is the non-coronal stop-final clusters which get simplified. The members of these clusters all share their place of articulation. They are clusters such as /mb/, /mp/, /fv/, and /ŋg/. The final category is the non-coronal stop final clusters which never show simplification. This category is the biggest among the four discussed here. The coronal stop-final clusters which simplify are presented first.

#### 2.4.1 Coronal Stop-final Clusters with Simplification

The coronal stops /t/ and /d/ show unrestricted deletion after stops, fricatives, and /n/. They comprise the sequences stop + /t/ (a-d), fricative + /t/ (e-l), stop + /d/ (m-n), fricative + /d/ (o-s), and /n/ + /d/ (t-u). List (7) below presents examples for this category.

- |     |             |           |            |
|-----|-------------|-----------|------------|
| (7) | a. /ræbt/   | → [ræb]   | “relation” |
|     | b. /zæbt/   | → [zæb]   | “record”   |
|     | c. /væqt/   | → [væχ]   | “time”     |
|     | d. /seqt/   | → [seq]   | “abortion” |
|     | e. /næft/   | → [næf]   | “oil”      |
|     | f. /mof t/  | → [mof]   | “free”     |
|     | g. /dust/   | → [dus]   | “friend”   |
|     | h. /dorost/ | → [doros] | “right”    |
|     | i. /loχt/   | → [loχ]   | “naked”    |

j. /sæχt/	→	[sæχ]	“difficult”
k. /doroʃt/	→	[doroʃ]	“large”
l. /zeʃt/	→	[zeʃ]	“ugly”
m. /æqd/	→	[æq]	“engagement”
n. /næqd/	→	[næq]	“cash”
o. /qæsd/	→	[qæs]	“intension”
p. /fasd/	→	[fas]	“phlebotomy”
q. /roʃd/	→	[roʃ]	“grow”
r. /dozd/	→	[doz]	“thief”
s. /jæzd/	→	[jæz]	“a city”
t. /bolænd/	→	[bolæn]	“tall”
u. /tʃærænd/	→	[tʃæræn]	“non sense”

Coronal stops can appear after almost all types of consonants in Persian. The final consonants of the above clusters are typically deleted in fast/casual speech. No experimental work has been reported on the simplification of these clusters yet. The few studies which have focused on the simplification of clusters in Persian are either limited to glottals (Darzi 1997; Shademan 2003; Sadeghi & Bijankhan 2007) or have taken a descriptive approach in their study (e.g., Darzi 1997; Assadi 2007). The following section introduces the coronal stop-final clusters which do not simplify.

## 2.4.2 Coronal Stop-final Clusters without Simplification

Similar to the first category, this group has coronal stops in the final position; however, they are preceded by different segments. These clusters include sequences glide + /t/ (a-b), liquid + /t/ (c-f), /m/ + /t/ (g), glide + /d/ (h-i), and liquid + /d/ (j-m). They have either liquid or glide as C<sub>1</sub> except for /m/ + /t/. Unlike the first category, C<sub>2</sub> (i.e., coronal stops) in these clusters does not get deleted. List (8) below presents some examples for the sequences attested in this category.

- (8)
- |            |   |         |                       |
|------------|---|---------|-----------------------|
| a. /bejt/  | → | [bejt]  | “verse”               |
| b. /sajt/  | → | [sajt]  | “location”            |
| c. /qælt/  | → | [qælt]  | “roll”                |
| d. /volt/  | → | [volt]  | “unit of measurement” |
| e. /ʃært/  | → | [ʃært]  | “condition”           |
| f. /tʃort/ | → | [tʃort] | “nap”                 |
| g. /sæmt/  | → | [sæmt]  | “direction”           |
| h. /ejd/   | → | [ejd]   | “holiday”             |
| i. /sejd/  | → | [sejd]  | “trap”                |
| j. /dʒeld/ | → | [dʒeld] | “cover”               |
| k. /dʒæld/ | → | [dʒæld] | “fast”                |
| l. /xord/  | → | [xord]  | “ate”                 |
| m. /bord/  | → | [bord]  | “took”                |

These clusters do not simplify despite of the fact that they all have coronal stops as C<sub>2</sub>. Darzi (1991) in his study on consonant cluster simplification in Persian has not included this category in the analysis. The non-coronal stops or fricatives which simplify are presented next.

### 2.4.3 Non-coronal Stop/Fricative-final Clusters with Simplification

In addition to coronal stops which are deleted in the final position, there are some non-coronal stops or fricatives which also get deleted as C<sub>2</sub>. These are /m/ + /p/, /m/ + /b/, /n/ + /g /, and /f/ + /v/. List (9) provides some examples of this category.

- (9)
- |             |   |         |           |
|-------------|---|---------|-----------|
| a. /pomp/   | → | [pom]   | “pump”    |
| b. /lamp/   | → | [lam]   | “bulb”    |
| c. /bomb/   | → | [bom]   | “bomb”    |
| d. /polomb/ | → | [polom] | “seal”    |
| e. /ræng/   | → | [ræŋ]   | “color”   |
| f. /zæng/   | → | [zæŋ]   | “ring”    |
| g. /æfv/    | → | [æf]    | “forgive” |

The clusters in (9) show unrestricted deletion of the final consonant. Assadi (2007) reported that labial consonants are deleted the least in Persian, but does not specify the position of these consonants in the syllable. The non-coronal stop-final clusters which do not simplify are the last category presented.

#### 2.4.4 Non-coronal Stop-final Clusters without Simplification

This is the biggest category where  $C_2$  is not a coronal stop and the clusters never simplify. These clusters are /jɫ/, /jr/, /km/, /lm/, /rm/, /jm/, /rn/, /jn/, /mn/, /bk/, /lk/, /nk/, /rk/, /sk/, /tk/, /ʃk/, /jk/, /rg/, /ds/, /fs/, /ks/, /ls/, /ms/, /ns/, /rs/, /xs/, /qs/, /bs/, /fz/, /mz/, /nz/, /rz/, /qz/, bz/, jz/, /fx/, /lx/, /rx/, /sx/, /fq/, /lq/, /rq/, /sq/, /mq/, /tq/, /ʃq/, /zq/, /bq/, /lf/, /nf/, /rf/, /sf/, /tf/, /ʃf/, /jf/, /zf/, /qf/, /rv/, /ʃv/, /zv/, /qv/, /lb/, /nb/, /rb/, /sb/, /tb/, /jb/, /zb/, qb/, /lp/, /lt/, /mt/, /rt/, /ld/, /md/, /rd/, and /jd/. The list of words given in (10) provides some examples.

- (10) a. /pæʃm/ → [pæʃm] “wool”  
b. /ræmz/ → [ræmz] “secret”  
c. /dærz/ → [dærz] “stitch”  
d. /kæʃf/ → [kæʃf] “discovery”  
e. /næsb/ → [næsb] “installation”  
f. /eʃq/ → [eʃq] “love”

The clusters presented in (10) remain unaffected by simplification. The previous studies on consonant deletion in Persian have not drawn a clear distinction between these four categories which do or do not simplify. There is no articulatory data available to see whether the simplification process is complete or incomplete. Moreover, none of these studies has addressed the optional simplification trait of the clusters which do simplify.

In general, the four categories presented above show that there are some consonants which have greater tendency to get deleted. It was shown that coronal stops are better targets for deletion compared to labials and dorsals. It seems that the simplification of a consonant cluster has a direct relationship with its adjacent segment. A summary on the status of the consonant clusters which do or do not simplify is given in Table 2.2. The groups of clusters which do simplify are five: the ones with coronal stops as  $C_2$  (i.e., the target of this dissertation) are marked with vertical shading; the homorganic  $C_1$  and  $C_2$  are marked with dark horizontal shading; the ones violating SSP are marked with dark up diagonal shading. The ones which violate SSP and lead into CL are sub-divided into two groups: the ones marked with light down diagonal shading delete  $C_2$  and the ones marked with grids delete  $C_1$ . The cells in white are attested clusters which do not simplify and the ones in gray are not attested clusters. The following legend is provided as a guide for this classification and Table 2.2.

	Non-attested clusters
	Attested clusters with no simplification
	Coronal stop-final clusters with simplification
= = = = =	Homorganic non-coronal stop clusters with simplification
/ / / / /	Clusters violating SSP with simplification
\ \ \ \ \	Clusters violating SSP with simplification & CL ( $C_2$ deletion)
# # # # #	Clusters violating SSP with simplification & CL ( $C_1$ deletion)

$C_1 \backslash C_2$	p	b	t	d	k	g	q	f	v	s	z	ʃ	ʒ	ʒ	tʃ	dʒ	m	n	l	r	j	h	?
p																	checkered						
b																	checkered						checkered
t		vertical					vertical	vertical		vertical	vertical	vertical			vertical							checkered	checkered
d		vertical					vertical	vertical		vertical	vertical	vertical						vertical				checkered	checkered
k																							
g																		horizontal					
q																		horizontal					
f																						checkered	checkered
v								horizontal														checkered	checkered
s																						checkered	checkered
z																						checkered	checkered
ʃ																						checkered	checkered
ʒ																							
ʒ																							
tʃ																							
dʒ																							
m																						checkered	checkered
n																						checkered	checkered
l		diagonal			diagonal		diagonal	diagonal														checkered	checkered
r		diagonal	diagonal				diagonal	diagonal			diagonal											checkered	checkered
j							diagonal	diagonal				diagonal											diagonal
h		diagonal	diagonal	diagonal			diagonal	diagonal			diagonal					diagonal		diagonal	diagonal	diagonal			
?		diagonal	diagonal	diagonal			diagonal	diagonal			diagonal						diagonal	diagonal	diagonal	diagonal			

Table 2.2 Possible consonant clusters and their simplification status

This chapter presented different categories of consonant clusters including coronal stop-final clusters which simplify in specific contexts. The true nature of coronal simplification is still unclear. The open question is whether the simplification process is gradient or categorical and whether there are any residual gestures left when clusters get simplified. The next question is whether this is a language-specific process and planned or whether it is a low-level phonetic effect and mechanical. Articulatory data gives us an optimal window to look at these questions. The next chapter reviews related articulatory studies. The ultrasound imaging and innovations of the study are also discussed in this chapter.

## Chapter 3

### Previous Studies of Consonant Articulation and Coronal Deletion

This chapter will review a number of studies on consonant articulation and coronal stop deletion. The first section covers the research which shows how syllable position and speech rate can affect the timing between the gestures. In the second section a review on coronal stop deletion in Persian will be presented followed by a discussion on their special status.

#### 3.1 Articulation of Consonants

The purpose of the articulatory study in this thesis is to investigate the production and deletion of word-final coronal stops in Persian fast/casual speech, specifically whether the coronal is truly deleted, or if a coronal gesture is being produced, but not heard, because it is obscured by the adjacent consonants. According to Articulatory Phonology, the gestures of the deleted segments can still exist even if those segments are not heard. Scobbie, Stuart-Smith and Lawson (2008) investigated the production and perception of Scottish English /r/ which is undergoing derhoticisation. Their articulatory results showed that even in cases where the acoustic correlates (i.e., raising of F2 and steep lowering of F3) of the canonical Scottish /r/ is missing, the gestures related to /r/ production are sometimes present.

Ultrasound tongue images showed that the lack of /r/ perception in these cases is due to the delay in tongue-tip raising movement in /r/ production rather than the total deletion of /r/-related gestures.

Previous studies have shown that language users tend to employ different strategies for the simplification of consonant clusters in different positions (e.g., Gafos 2002; Han 2006; Ramirez 2006). These strategies include vowel insertion and consonant deletion. These studies have provided various explanations for such different preferences in word positions. A number of studies have shown that the consonants in onset or prevocalic position have different gestural patterns compared to those in coda or post-vocalic positions (e.g. Ohala & Kawasaki 1984; Krakow 1989; Browman & Goldstein 1992, 1995; Sproat & Fujimura 1993; Wang 1995; Byrd 1996a, 1996b, Fougeron & Keating 1997; Gick 2003; Gick, Campbell, & Tamburri-Watt 2006; Kochetov 2006). These “syllable position” effects explain the variation caused by the relative timing between gestures and also the degree or magnitude of constriction. Sproat and Fujimura (1993), for example, in a study of light /l/, which happens syllable-initially (e.g., *lease*) versus dark /l/ which appears syllable-finally (e.g., *seal*) in English, stated that the production of /l/ involves two coordinated gestures of vocalic dorsal and consonantal coronal conducted by tongue body (TB) retraction and tongue tip (TT) closure, respectively. Acoustic and X-ray microbeam data showed that the order of these two gestures differs between onset and coda positions. The vocalic gesture preceded the consonantal gesture in codas, whereas it slightly lagged or was almost synchronous with the TT closure in onsets. Sproat and Fujimura consider gestures as inherently gradient (unlike phonological features, which are categorical) and conclude that allophonic variation is not categorical at the gestural level. They emphasize that the discussion about allophonic variations like syllable-initial and syllable-final /l/ may obscure the broader generalization relevant to all syllable-initial and syllable-final consonants.

Giles and Moll (1975), using high-speed lateral cinefluorography, reported that the magnitude or the degree of constriction of tongue tip gestures also varies syllable-initially versus syllable-finally. Tongue tip raising achieves its target consistently in the syllable-initial position whereas it lacks complete closure and is smaller in magnitude syllable-finally. Giles and Moll consider “final reduction” as an effect which only applies to syllable-final allophones.

Gick (2003) examined the pre- and post-vocalic allophones of /l/, /w/, and /j/ to identify the consonantal and vocalic gestures of these segments in different positions. Two gestures (i.e., consonantal and vocalic) were measured for /l/ and /w/ while only one gesture was considered for /j/. Lip aperture (LA) for the syllable-initial /w/ was found to precede the equally open tongue body (TB) backing gesture in the syllable onset, while the two were near-synchronous in the syllable coda. The results for other segments did not show any specific pattern.

In another study conducted by Browman and Goldstein (1995), the timing of gestures for nasals and laterals in American English was examined. Each phoneme consists of two gestures: nasals involve velic opening and oral closure gestures while laterals involve tongue tip and tongue dorsum movements. The results of MRI and X-ray for the syllable-initial (e.g., *meet*) versus syllable-final (e.g., *team*) nasal showed that the velum lowering and labial closing gestures are almost synchronous syllable-initially whereas the labial movement lags behind velum lowering gestures in syllable-final nasals. The finding of this study for laterals was similar to Sproat and Fujimura’s (1993) results. The overall results of these two studies show that the vocalic gestures tend to be closer to the nucleus or synchronous with

consonantal gestures in the initial position while gestures with narrower constrictions (i.e., tongue tip movement) tend to lag behind the gestures with wider constrictions (i.e., tongue dorsum or velum lowering gestures) in syllable-final position.

According to the Articulatory Phonology approach, gestural constellations have some intrinsic configurational properties which can identify syllable-initial and syllable-final consonants without resorting to the syllable as a prosodic constituent. In this framework, the consonants preceding a vowel or following it have a different phasing or spatial displacement with respect to the vowel. Therefore, this approach considers different gestural patterns for a consonant in onset versus coda position. According to this framework, consonants in onset and coda positions are unlikely to be the same, unless they undergo alternations. As a result, the allophones of a specific phoneme are specified by the timing of their component gestures to those of surrounding segments.

In another study, Kochetov (2006) examined the “syllable position” effect in Russian using the palatalized and non-palatalized voiceless labial stops /p<sup>j</sup>/ and /p/, and the palatal glide /j/. This study investigated the inter-gestural timing and the internal gestural properties of these consonants using an Electromagnetic Mid-sagittal Articulometry (EMMA) system. The two coordinated oral gestures in the production of the palatalized consonant /p<sup>j</sup>/ chosen in this study allows for examining the timing of the two gestures in syllable-initial and syllable-final positions and comparing it with the timing of the same gestures in /pj/ and /jp/ sequences. The results of the study show major differences in syllable position effects for the same consonants. The syllable-final consonants showed more variation in timing and

magnitude compared to syllable-initial consonants. They also showed variation in the amount of lag between more open and more closed gestures across languages. The two gestures existing in the Russian palatalized stop showed more positive lag in onset position and were more synchronous or had less negative or positive lag in coda position. This differs from the timing pattern found for English nasal and laterals (see Sproat & Fujimura 1993; Browman & Goldstein 1995); however, it is similar to the one observed for English /w/ (see Gick 2003). These studies, in general, show that the different allophones of the same phoneme are generally the result of variation in the intergestural timing and magnitude patterns of the same gestures.

In addition to the role of syllable position in orchestrating the phasing relationship between different gestures in the same segment, research has shown that other extralinguistic factors influence them as well. The rate of production, which is one of the distinguishing factors between deliberate canonical speech and connected casual speech, has proved to influence gestural orchestration and consequently the rate of deletion as well (Gay 1981; Barry 1985; Kaisse 1985; Byrd 1994, 1996a, 1996b; Byrd & Cheng Tan 1996; Kirchner 2001). Byrd and Cheng Tan (1996) investigated [s#g], [g#s], [g#d], and [d#g] sequences to see whether an increase in speech rate affects all the consonants in the same way. The results of the study for some speakers showed that consonants become shorter as speech rate increases and this reduction happens somewhat more consistently in coda than in onset position. This study also showed that stops [d] and [g] are more likely to reduce than the fricative [s], and that coronals show greater reduction than the velar [g].

In another study Zsiga (1994) investigated whether an increase in speech rate would affect the formant transition between V C<sub>1</sub> and C<sub>2</sub>. She tested this hypothesis by examining the temporal overlap of the two closure gestures in word-final alveolar stop as a function of a following word-initial /p/, /t/, or /k/ in VC<sub>1</sub># C<sub>2</sub>V environment. The results of this study showed that, taking the ratio of vowel duration to consonant closure duration as the criterion for measuring rate, C<sub>2</sub> increasingly dominates the transition as rate increases. Zsiga concluded that the manipulation of rate would result in greater temporal overlap between gestures. Giles and Moll's (1975) study also showed that the tongue tip contact with the alveolar ridge was attested for final /l/ at the conversational rate but was dropped at the faster rate. Various explanations have been proposed for the higher rate of reduction in fast/casual speech. Kirchner (2001), for example, proposed that the effort needed to achieve some constriction target at a fast speech rate is higher than the effort needed for the same constriction at a slower rate. He argues that this is the main reason for having more deletion in fast speech.

This section reviewed factors which could affect gestural timing. Since this thesis investigates the deletion of coronal stops, their simplification and the special status of coronals is discussed in the following section.

### **3.2 Coronal Stop Deletion**

Persian exhibits the deletion of final coronal stops in fast/casual speech in a sequence of consonant cluster (i.e., C<sub>1</sub>C<sub>2</sub>#) where C<sub>1</sub> is either a stop, fricative or homorganic nasal /n/ and C<sub>2</sub> is a coronal stop. The deletion of C<sub>2</sub> does not occur when the preceding consonant is a

liquid, glide, or non-homorganic nasal /m/. The following lists adapted from Falahati (2008) provide some examples for the deletion and non-deletion of coronal stops in Persian.

**(1) Deletion of coronal stops**

- |             |           |              |
|-------------|-----------|--------------|
| a. /zæbt/   | → [zæb]   | “record”     |
| b. /æqd/    | → [æq]    | “engagement” |
| c. /næft/   | → [næf]   | “oil”        |
| d. /dust/   | → [dus]   | “friend”     |
| e. /dozd/   | → [doz]   | “thief”      |
| f. /suχt/   | → [suχ]   | “burnt”      |
| g. /væqt/   | → [væχ]   | “time”       |
| h. /bolænd/ | → [bolæn] | “tall”       |

**(2) Non-deletion of coronal stops**

- |            |           |             |
|------------|-----------|-------------|
| a. /bejt/  | → [bejt]  | “verse”     |
| b. /ejd/   | → [ejd]   | “holiday”   |
| c. /qælt/  | → [qælt]  | “roll”      |
| d. /dʒæld/ | → [dʒæld] | “fast”      |
| e. /ʃært/  | → [ʃært]  | “condition” |
| f. /χord/  | → [χord]  | “ate”       |
| g. /sæmt/  | → [sæmt]  | “direction” |

Word-final coronal stops have been the focus of a variety of studies, because they frequently undergo deletion when followed by another consonant (Guy 1980; Avery & Rice 1989; Browman & Goldstein 1992, 1995; Byrd 1992, 1996; Côté 2000, 2004a, 2004b; Falahati 2008, 2011). The special status of coronals versus the other two main articulators (i.e., labial and dorsal) has been investigated in different studies (Davis 1991; Paradis & Prunet 1991 among others). Some of the studies ascribe the peculiar status of coronals to their unmarked nature. Those supporting this approach have laid their reasoning on either purely theoretical or empirical bases. The lack of specifications for place features in the underlying representation is the main argument for those holding an abstract view while high frequency (Kent 1983) as well as early acquisition in L1 (Viham, Ferguson, & Elbert 1986) are provided as empirical support. Those claiming for the unmarked nature of coronals consider their appearance as epenthetic segments, freer distribution, and their active role in assimilation to support their idea.

Despite the special behavior of coronals, McCarthy and Taub (1992) argue that even unmarked plain coronal alveolars such as /t/, /d/, /l/, /r/, and /n/ must be specified for the [coronal] feature. They provide some empirical evidence showing that both unmarked plain alveolars and marked dentals or palato-alveolars behave in the same way and should belong to the same natural class. This includes restrictions on syllable-initial cluster and syllable shape where all coronals behave in the same way. This makes the argument for coronal underspecification somewhat unfounded.

The problem with the underspecification approach and the widespread adoption of Optimality Theory gave way to an account for the asymmetry between coronals and noncoronals using hierarchies of place markedness constraints. Prince and Smolensky (1993,

2004) introduced a ranking which shows that the constraint against coronal place is intrinsically ranked lower than the constraint against labial and dorsal place features; however, there are some studies which have provided some contradictory evidence showing this is not the case. Kang (2000), for example, in her study of phonetics and phonology of coronal markedness and unmarkedness provides a long list where coronal stops act as marked segments in a large number of languages. This list shows that the markedness pattern of coronals is attested in phonological processes such as assimilation, metathesis, morpheme structure constraint, and deletion.

An examination of coronal consonant inventories across world languages has shown that there are higher ranked constraints against having these sounds. Maddieson (1984) revealed that the maximum number of coronal segments that can occur in a language is smaller than the potential places of articulation attested phonetically for them.

The unmarked nature of coronals is further contradicted by the claim that coronal gestures are intrinsically more difficult to articulate than dorsals (Hardcastle & Roach 1977). Arguing in the same direction, Browman & Goldstein (1992) have suggested that the greater tendency for the reduction of coronals in coda position may be the result of the tendency of the tongue tip to rest on the floor of the mouth during vowels. The tongue tip of coronals needs to move further compared to the tongue body or lips for dorsals and labials to form a closure. This view takes economy of articulation into view which gets close to Kohler's (1990) argument that apical gestures are physiologically marked compared to tongue dorsum and lip gestures. Kohler attributes this to the greater precision of timing and muscular coordination needed to produce apical segments. He states that the articulation of these segments is linked with more intrinsic tongue muscles. Noncoronals, on the other hand,

require a more global gesture using large extrinsic tongue muscles in conjunction with continuously present jaw movements.

There are also some other studies which have attributed the asymmetry of reduction between coronals and non-coronals to the high frequency of coronal consonants in languages (e.g., Kent 1983). According to this view, the high frequency of coronal consonants gives to listeners a better chance to decode the speech easily and allows speakers to be more sloppy. This means that articulatory precision is not at a high need. In general, studies regarding the higher rate of coronal deletion compared to other consonants attribute it either to weaker perceptual cues or to greater articulatory demands.

This section showed that the true nature of coronal stop deletion is still unclear. Articulatory data can offer very valuable information on this issue. The following section introduces the ultrasound technology which is used for collecting articulatory data in this study.

### **3.3 Ultrasound Imaging of the Tongue**

In order to address the research question of this study, ultrasound data and acoustic recordings were collected. A large number of studies on articulation have used ultrasound as a reliable means for collecting data. Ultrasound is an appealing technology which has increasingly been used in different speech production studies in recent years (Davidson & Stone 2003; Mielke, Baker & Archangeli 2005; Mielke, Baker, Archangeli, & Racy 2005; Mielke & Archangeli 2005; Davidson 2007; Baker, Mielke & Archangeli 2008; Scobbie et al. 2008; Mielke 2009; Proctor 2009; Mielke, Olson, Baker, & Archangeli 2011). This technique provides a holistic picture of tongue shape changes over a period of time. It uses ultra high-frequency sound waves to create movies of tongue motion. In most speech-related

applications of ultrasound, researchers have focused on collecting data from the midsagittal contour of the tongue, although coronal slices have also been analyzed (Slud, Stone, Smith, & Goldstein 2002). Stone (2005) considers ultrasound as having advantages over other existing technologies (e.g., X-ray, EMMA, MRI, and X-ray microbeam) used for imaging the vocal tract. She enumerates the positive points of ultrasound as being non-toxic, high resolution, inexpensive, portable, and comfortable for subjects. Moreover, due to the wide area of tongue captured by the ultrasound beam, it can provide valuable information about tongue shape and gestures during speech. Like other technologies, this technique also has some limitations. Poor imaging with some speakers, lack of transparency in some images, and the need to fix the probe relative to the head are some of the challenges of using ultrasound. In general, this technology has advantages over others which are toxic (e.g., X-ray) or not easy to carry and use due to their costs (e.g., X-ray, MRI, EMMA, X-ray microbeam). Some of these devices like X-ray microbeam and EMMA are limited to capturing only a number of points of the tongue surface. Due to the wide application of ultrasound in capturing tongue gestures and its advantages listed above, this technology is being used in the experiment of the current study. The following section introduces the methodological innovations of the study.

### **3.4 Methodological Innovations of the Study**

This study used ultrasound tongue imaging to capture spontaneous speech. To the best of my knowledge, there is only one study which has been reported to collect natural speech. Scobbie, Stuart-Smith, and Lawson (2008) used ultrasound technology to provide a sociolinguistic and phonetic analysis of Scottish English derhoticization. The first results of

this study showed that different tongue shapes are used to articulate /r/, with contextual and speaker variation. No formal analysis of the data has been published yet.

The current study also used ultrasound technology to measure gestural dynamics. This measurement has been usually done before using point-tracking methodologies, like EMMA (and before that, X-ray microbeam). Each of these methods has advantages and disadvantages. The advantages of EMMA are that it has better time resolution, flesh point data, and it is easier for data extraction. But ultrasound can measure the movement of any part of the tongue surface one can image, such as tongue root and tongue dorsum. Both X-ray microbeam and EMMA could only capture a limited number of points of the tongue surface. Due to the limitations imposed by gag reflex in EMMA, it is extremely difficult to capture the motion or shape of the tongue root. Moreover, there is always the chance of missing a significant portion of the tongue contour which happens to be between two receivers.

Derrick & Gick (2011) used ultrasound to study variation in the production of flap/tap allophone of North American English /t/ and /d/. The kinematic-articulatory patterns observed in the tongue images showed free variation of the allophones across repetitions of everyday words. The ease and efficiency of ultrasound technology has provided the opportunity to open the valuable sources of information about individual differences in speech production which have been previously missed by alternative methods of generative phonology.

In this study, I used Ultrasound Stabilization Headset which holds probe steady and allows natural head movement during the recording. Using headset ensures that the radii are equivalent across all tokens from the same session. Additionally, comparisons were made

between distributions of gestures, with a novel application of SSANOVA. In this study, the x-axis and y-axis display time and the distance between the alveolar ridge and tip/blade of the tongue, respectively. The following chapter introduces the methodology of the acoustic experiment and presents its results and analysis. A discussion section on the acoustic results will conclude this chapter.

## **Chapter 4**

### **Acoustic Experiment**

The main goal of the acoustic experiment is to see whether there are any acoustic cues left after the deletion of the word-final coronal stops in consonant clusters. Coronal stops can be optionally deleted when preceded by obstruents or the homorganic nasal /n/ in Persian. This experiment is designed to see whether the simplification of the consonant clusters is complete or incomplete. In general, this chapter pursues the following two questions:

1. Is the process of coronal stop deletion in Persian categorical or gradient?
2. Are there acoustic cues to the presence of coronal gestures in simplified clusters?

In order to address these questions, the following experiment is designed. The methodology of this experiment is introduced in section 4.1 followed by the results and analysis in 4.2.

#### **4.1 Methods of Acoustic Experiment**

##### **4.1.1 Subjects**

The subjects of the study were 10 Persian-speaking graduate students from the University of Ottawa and Carleton University, five male and five female, aged 25-38. Three (two male and one female) were excluded from the analysis due to poor imaging and other technical issues. All the participants in the study had lived in the Iranian capital city and had

Tehrani Farsi accent. All of them had finished at least their undergraduate studies in their home country. In their speech, they showed no dialectal features that would suggest some other accent other than the target of this study. None of the subjects reported any hearing problem. They were not aware of the objectives and the questions of the study.

#### **4.1.2 Procedures**

The acoustic data was collected during the articulatory experiment (the articulatory experiment is introduced in Chapter 5). Acoustic signals were captured using Shure condenser microphone mounted close to subject's mouth and USBPre2 microphone preamp and A/D convertor. A desktop computer and the Audacity software were used in audio recordings. Data were recorded at the sampling rate of 44.1 KHz/16 bits.

#### **4.1.3 Data Analysis**

The recorded speech was transferred from a digital recorder to a personal computer for analysis. At the beginning, all the sound files related to target words were extracted automatically using a Praat script<sup>3</sup>. Next, a Praat (Boersma & Weenink 2012, version 5.3.19) TextGrid was used to mark the target coronal consonants (i.e., C<sub>2</sub>) as well as the preceding and following consonants (i.e., C<sub>1</sub> & C<sub>3</sub>) and also the two vowels adjacent to the three consonants in the middle. At the end, 662 tokens of word-final coronal clusters for the seven subjects in the casual conversation were extracted. However, only 248 tokens, including 51 controls, were chosen for segmentation and acoustic analysis. They were chosen since their corresponding articulatory data was available.

---

<sup>3</sup> Mielke, J. (2010). *Praat Script for Automatic Word Extraction*.

The segmentation was done in a way that one or two segments were captured before  $V_1$  and after  $V_2$  as well. For example, in a word like ([**næft bæra**-je], “oil for”) in addition to  $V_1C_1C_2C_3V_2$  (bold segments in the transcription), the preceding and following segments (i.e., [n] and [r]) were also included in the textgrid file. This was done so that  $V_1$  and  $V_2$  could be segmented more accurately and reliably. Each stop consonant could be composed of two parts: closure and burst. The closure part of each consonant was marked as  $C_2$  or  $C_3$  and the related burst as  $B_2$  and  $B_3$ . Since  $C_1$  was always a fricative, no burst was marked for this consonant. The following criteria were used in order to decide where the segmentation needed to be done:

- The start of  $V_1$  is where the formants of  $V_1$  are visible in the spectrogram and the end of it is where the formants related to that ends.
- The start of  $C_1$  is where the vowel formants disappear and the noise for fricatives [χ] and [f] starts.
- The end of  $C_1$  was the end of noise for fricatives as seen in the spectrogram.
- The start of  $C_2$  burst, if present (marked as  $B_2$ ), was determined by an abrupt, transitory aperiodic event as seen in the waveform and spectrogram which could be followed by any subsequent aspiration.
- The start of  $C_3$  was the earliest point in time that could be  $C_3$ . This could be right after the [t] release burst.
- The end of  $C_3$  closure was where silence was finished.

- The start of C<sub>3</sub> burst (marked as B<sub>3</sub>) was determined by an abrupt, transitory aperiodic event as seen in the waveform and spectrogram which could be followed by any subsequent aspiration.
- The end of C<sub>3</sub> was where the formants for V<sub>2</sub> started.
- The start and end of V<sub>2</sub> was where the glottal pulses start and end. When V<sub>2</sub> was followed by another vocalic segment, an abrupt change in formants was used to determine the boundary.

For the segmentation of the simplified tokens the following point was specifically taken into account:

- If there was sustained silence after C<sub>1</sub> with no acoustic evidence of [t], then a point which could be the end of C<sub>1</sub> and the start of C<sub>3</sub> was chosen.

The choice of where C<sub>1</sub> ends and where C<sub>3</sub> starts was sometimes challenging since the stop consonants are sometimes spirantized. Many of the stops were shown to be spirantized, based on the presence of fricative-like acoustic features. The following figures illustrate the waveforms and spectrograms of four tokens. Figure 4.1 and Figure 4.2 show the segmentation of the unsimplified tokens ([riχt ke] “splashed to”) and ([hæft b̄ar] “seven times”) which have different V<sub>1</sub> and C<sub>1</sub>. Figure 4.3 presents the segmentation of the simplified token ([hæft b̄ar] “seven times”) and Figure 4.4 shows the waveform and spectrogram of the control token ([ælaf bazi] “wasting time”). The following abbreviations are used in the following figures: V<sub>1</sub> and V<sub>2</sub> represent vowels; C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> represent noise or closure of obstruents; B<sub>2</sub> and B<sub>3</sub> represent bursts of the related stops.

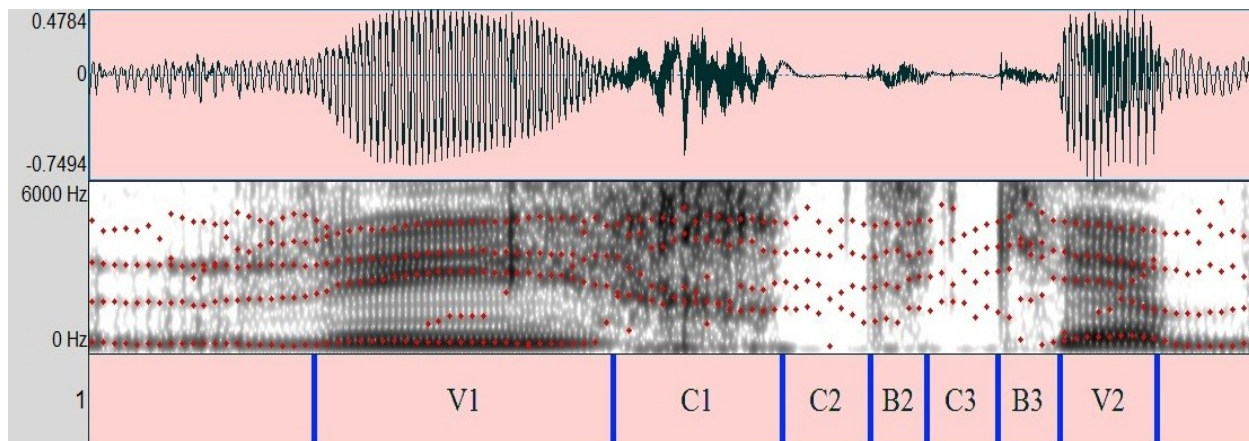


Figure 4.1 Waveform and spectrogram of unsimplified ([riχt ke] “splashed to”) for subject #2

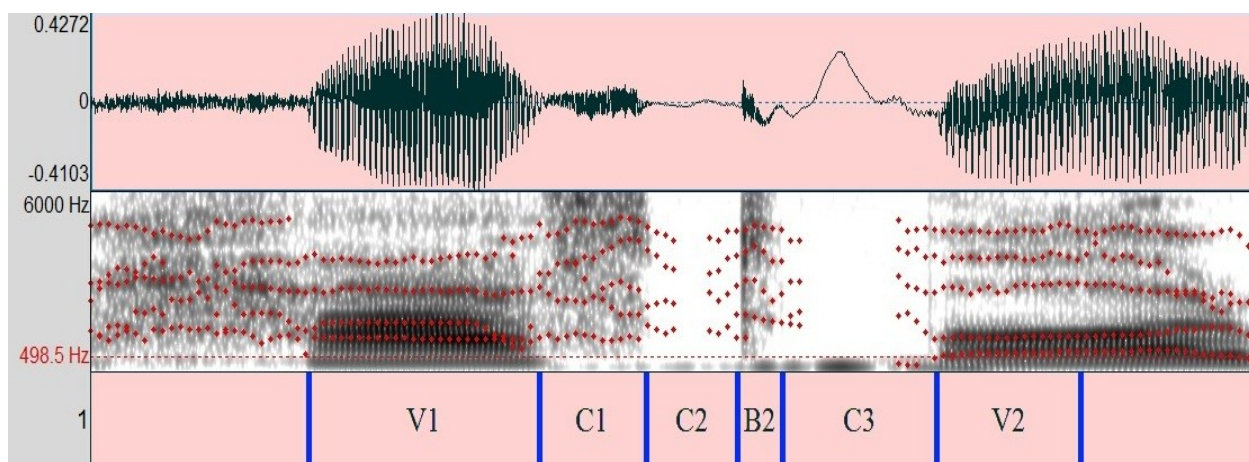


Figure 4.2 Waveform and spectrogram of unsimplified ([hæft bar] “seven times”) for subject #7

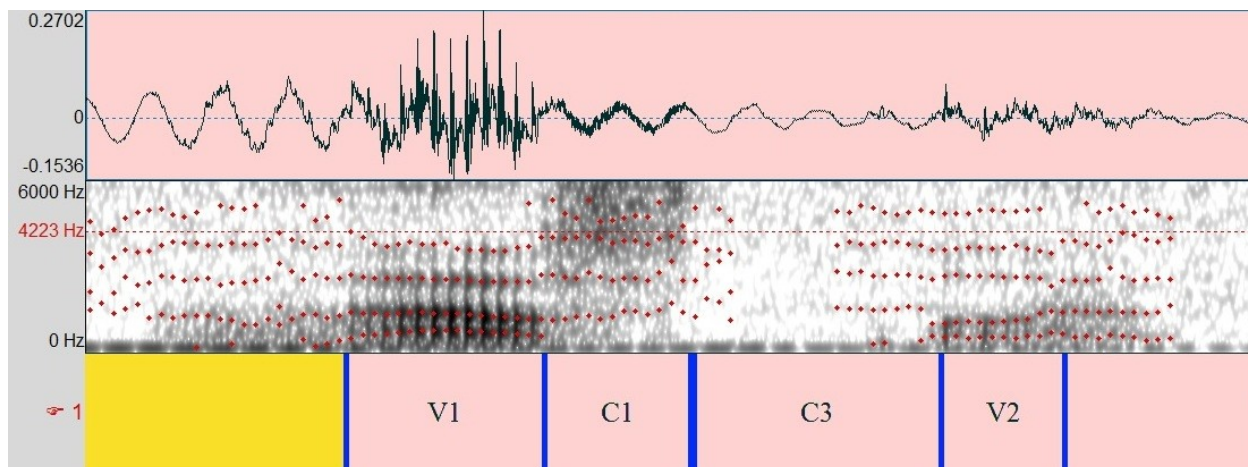


Figure 4.3 Waveform and spectrogram of simplified ([hæft bar] “seven times”) for subject #3

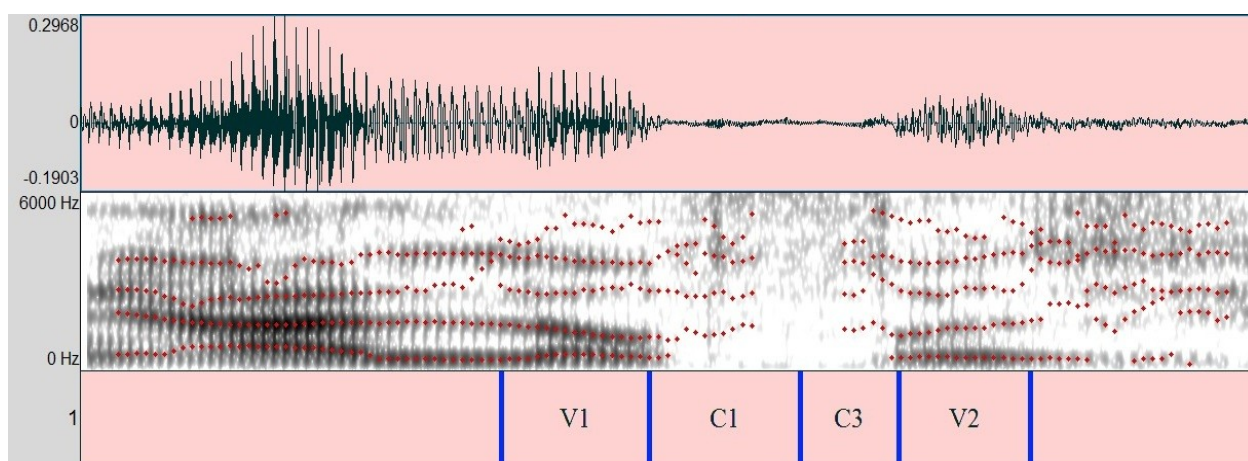


Figure 4.4 Waveform and spectrogram of control ([ælaf bazi] “wasting time”) for subject #3

After the segmentation of the acoustic data was complete, the data were analyzed.

The next section presents the acoustic results and analysis of this experiment.

## 4.2 Acoustic Results and Analysis

The main goal in this experiment is to investigate the acoustic consequences of cluster simplification in Persian. In order to show whether the deletion of the coronal stop is categorical or gradient, the acoustic traits of the tokens in the three conditions (i.e., unsimplified, simplified, and control) are examined. The tokens in the study with word-final C+/t/ clusters were categorized as unsimplified when they contained an audible [t] and as simplified when they did not. Additionally, a number of tokens such as ([jæχ kærd] “got cold”, [mæruʃ bud] “was famous”, [æləf bazi] “wasting time”) served as the control. If cluster simplification results in a complete loss of the acoustic evidence for coronal stop /t/, the simplified tokens should be indistinguishable from control tokens.

The duration of consonant clusters (i.e., C<sub>1</sub>C<sub>2</sub>C<sub>3</sub>), individual consonant (i.e., C<sub>1</sub>) and vowel (i.e., V<sub>1</sub>) and formant transitions into and out of clusters are the dependent variables in this study. Warner and Tucker (2011), in a study of stop reduction in English, found duration as a significant cue to the reduced stops. Their study showed that longer duration is an indicator of a clearer and less reduced consonant. Formant transitions were the next acoustic parameter in the analysis. It is well established that transitions of the second and third formants are cues for place of articulation in consonants (e.g., Liberman, Delattre, Cooper, & Gerstman 1954). F2 and F3 and their slopes could indicate the presence of a residual coronal gesture. In addition to F2 and F3, F4 was also included. Despite the fact that listeners are shown not to be very sensitive to F4, some studies have shown that an F4 valley has a small effect on the perception of a /t d/ in a reduced and flapping environment (Warner, Fountain,

& Tucker 2009; Warner & Tucker 2011). Since the acoustic cues pertaining to coronal gestures are the main goal of this chapter, F4 transitions were included in the analysis.

If coronal stop deletion is gradient in Persian, a prediction is that there could be some acoustic traces of a coronal stop gesture present (even weakly) in some of the simplified tokens. These properties are expected to be absent from the control tokens and fully present in the unsimplified tokens.

Figure 4.5 below illustrates the percentage of simplified tokens in different environments. According to this, when a coronal stop is followed by another stop, it gets simplified 76% of the time ( $n = 105$ ) while it only simplifies 33% of the time ( $n = 19$ ) when utterance-final or followed by a vowel.

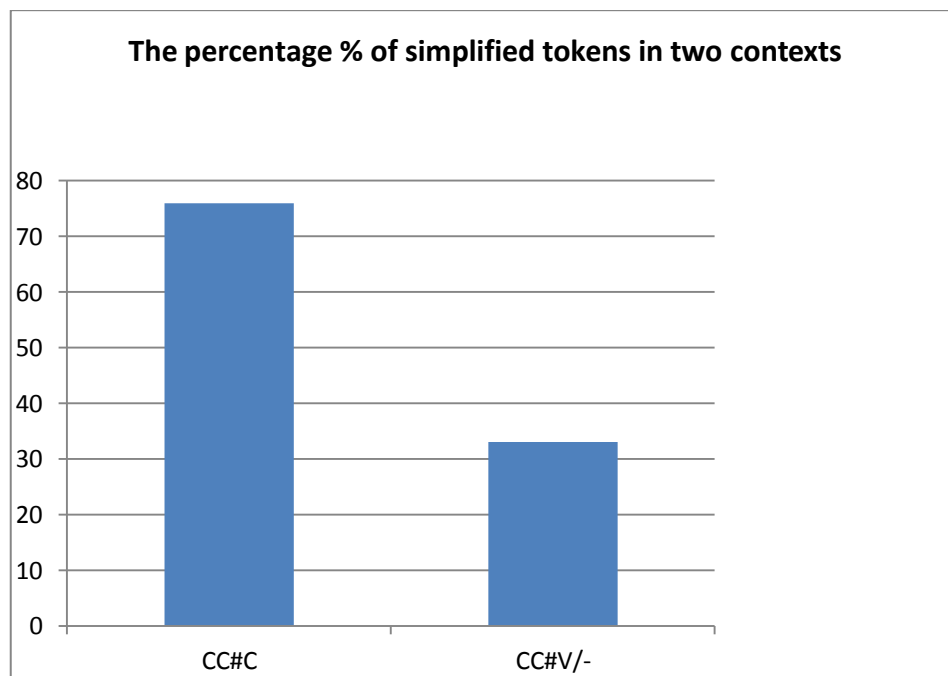


Figure 4.5 Percentage of the simplified tokens across different environments

### 4.2.1 Statistical Analyses

In order to investigate the predictions made at the beginning of this section, duration and formant transitions were considered as the acoustic parameters for measurement. As for the results of duration, all the tokens were pooled and analyzed together. Since the simplified tokens are missing all or a portion of the coronal stop /t/ gesture, duration could be considered as the most basic cue for this type of analysis. In order to conduct this analysis, the total duration of consonant clusters across simplified, unsimplified, and control conditions was measured. In order to make the tokens in the three conditions comparable, the utterance-final tokens or the C<sub>1</sub>C<sub>2</sub>s followed by a vowel were not considered in this analysis. The box plots in Figure 4.6 present the average duration of the consonant clusters in the three conditions. The x axis shows different conditions and y axis represents time in seconds.

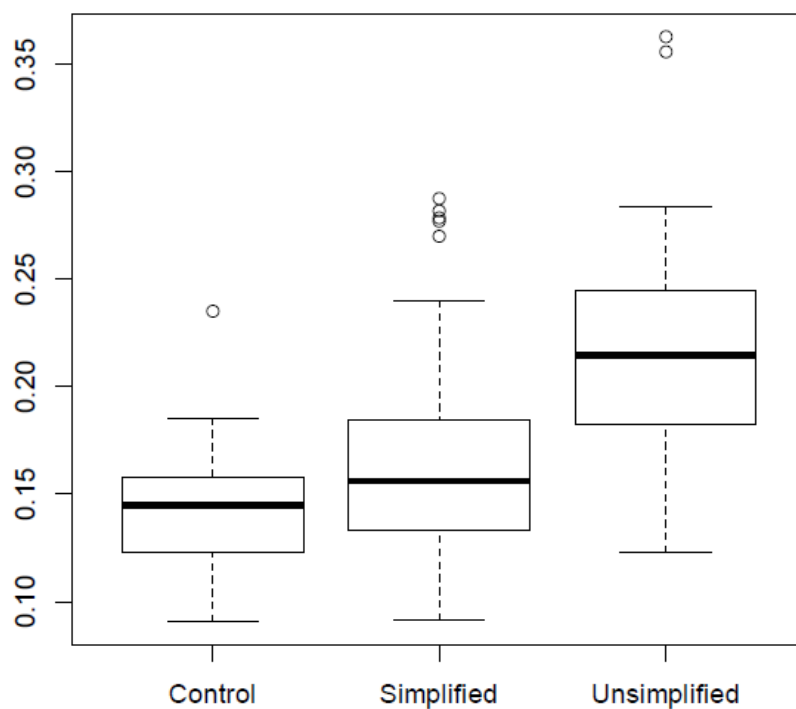


Figure 4.6 Duration of consonant clusters across three conditions

In order to test whether there is a significant difference between the duration of the clusters across the three conditions, a one-way ANOVA test was run. The null hypothesis is that there is no significant difference in the duration means across control, simplified, and unsimplified conditions. The results of the test with cluster duration as the dependent variable found significant main effect of conditions (control/simplified/unsimplified) ( $p < 0.00001$ ). Obviously, the next question to ask is where this difference lies. The pairwise comparison results of post hoc Tukey HSD test showed that duration differed significantly in unsimplified and simplified conditions at  $p < 0.00001$  level. The pairwise comparison test between simplified and control conditions, however, did not show any significant difference.

In addition to consonant cluster duration as an acoustic cue, the duration of the first consonant (i.e.,  $C_1$ ) and  $V_1$  in simplified, unsimplified, and control tokens was calculated. Figure 4.7 below presents the results for the duration of  $C_1$  in the form of box plots. This result shows that the duration of  $C_1$  across all conditions is almost the same. However, the result of the preceding vowel (i.e.,  $V_1$ ), as presented in Figure 4.8, shows a significant difference in their duration.

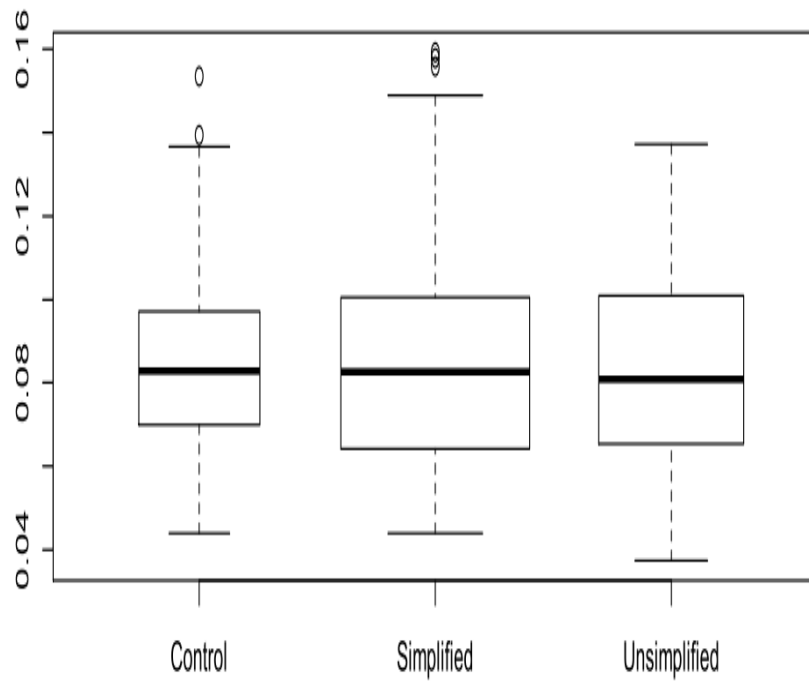


Figure 4.7 Duration of  $C_1$  across three conditions

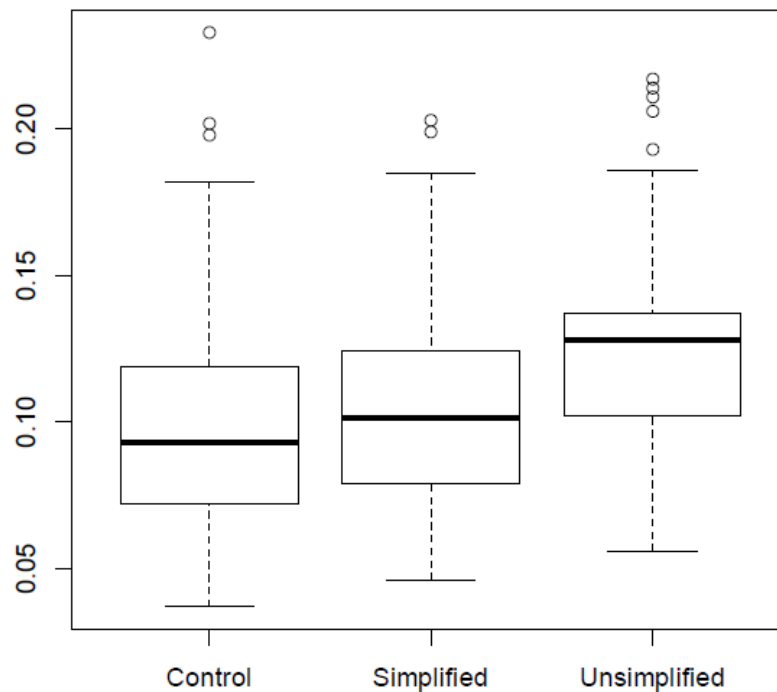


Figure 4.8 Duration of V<sub>1</sub> across three conditions

The run of a one-way ANOVA test with V<sub>1</sub> duration as the dependent variable found significant main effect of condition (control/simplified/unsimplified) ( $p < 0.00001$ ). A Tukey HSD post hoc test found significant difference in the pairs unsimplified and simplified at  $p = 0.0001$  level. The pairwise comparison test between simplified and control, however, did not show any significant difference.

In addition to duration, the formant transitions of V<sub>1</sub> at specific timing steps were also measured using a Praat script<sup>4</sup>. Two timing points in each vowel were chosen for analysis. The first one corresponds to the last glottal pulse detectable in the waveform. The second

<sup>4</sup> Mielke, J. (2013). *Praat Script for measuring formant frequencies*.

timing point was 80% of the vowel's duration (i.e., 20% before the end of the vowel). This point was chosen to make sure that the corresponding formant transition of each vowel is fully captured. The slopes between the measurements made at these two timing points were used to determine whether they differ across conditions. The slopes were calculated by subtracting the frequency of the second timing point from the first timing point.

The segments following  $V_1$  could affect formant transitions differently and this could potentially confound the results. In order to eliminate the potential problem, [f] or [χ] as  $C_1$  preceded by the most frequent  $V_1$  (i.e., [æ]) were selected and the corresponding formants data was subjected to R for statistical analyses. Figures 4.9 - 4.11 below illustrate the boxplots related to the slopes of the three formants (i.e., F2, F3, & F4) for  $V_1 = [æ]$  and  $C_1 = [f/χ]$  across simplified unsimplified, and control conditions. In the following figures the three conditions are marked with their initial letters (i.e., "S, U, & C") followed by their corresponding  $C_1$  context.

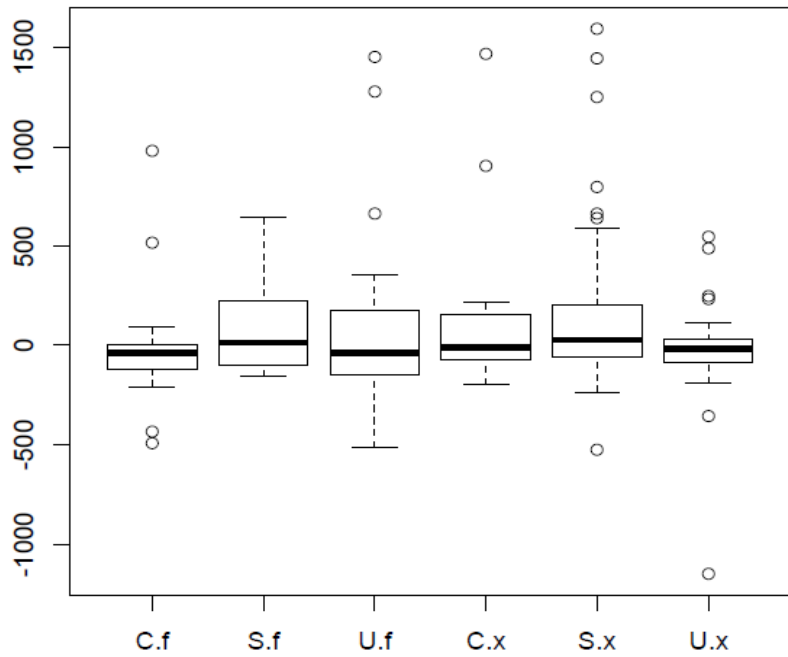


Figure 4.9 Slope of F2 transition for  $V_1=[\text{æ}]$  and  $C_1=[f/\chi]$  across three conditions

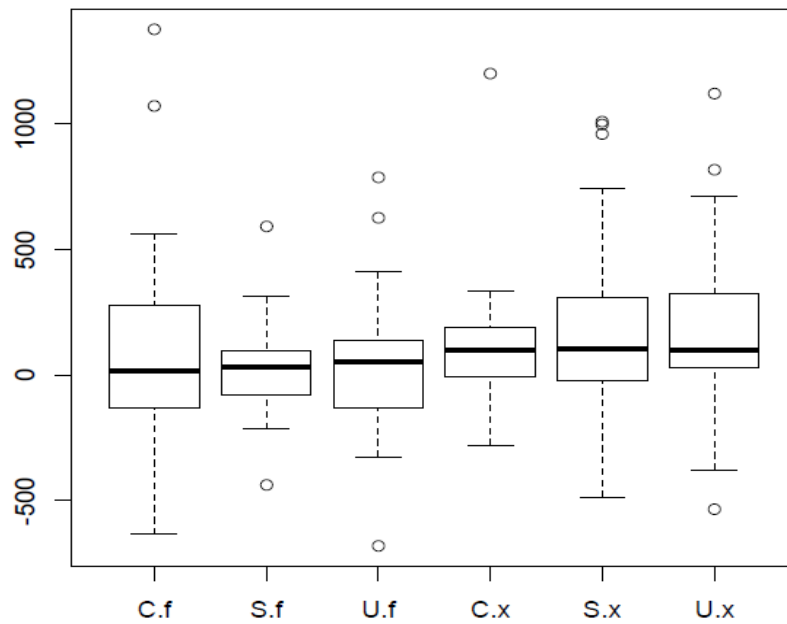


Figure 4.10 Slope of F3 transition for  $V_1=[\text{æ}]$  and  $C_1=[f/\chi]$  across three conditions

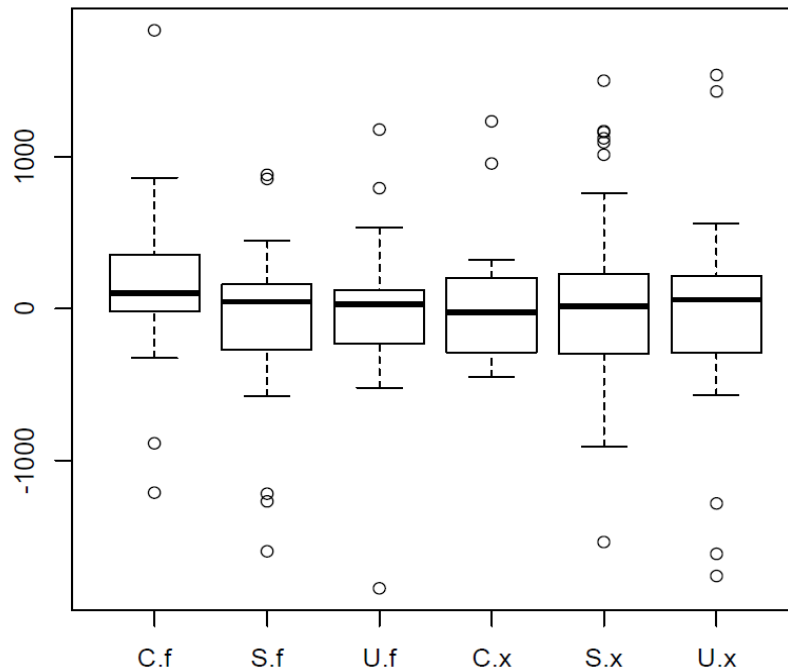


Figure 4.11 Slope of F4 transition for  $V_1 = [\text{æ}]$  and  $C_1 = [f/\chi]$  across three conditions

The results of three separate one-way ANOVA tests with formant slopes (i.e., F2, F3, & F4) as dependent variables and the three conditions (i.e., control, simplified, and unsimplified) and  $C_1$  (i.e., [f] & [χ]) as factors in  $V_1 = [\text{æ}]$  environment produced almost similar results. The results of F2 for conditions and  $C_1$  as factors did not show any significant difference ( $p = 0.303$ ;  $p = 0.701$ ). Similar to F2, the same results for F4 were not significantly different either ( $p = 0.465$ ;  $p = 0.782$ ). However, the results of the ANOVA test for F3 showed marginal significant difference for  $C_1$  ( $p = 0.041$ ), but not across the three conditions ( $p = 0.848$ ). A Tukey HSD post hoc test was run to find more about this difference. The results did not show any significant difference in the pairwise comparisons. This marginal difference could be due to different sensitivity to variability between ANOVA and pairwise tests. None of the formant slopes showed a significant interaction between conditions and  $C_1$ .

In summary, the durations of both the entire consonant cluster and the preceding vowel were significantly longer for unsimplified clusters than simplified clusters. None of the formant slopes differed significantly across conditions or consonant context.

### **4.3 Discussion for Acoustic Results**

The main goal of this chapter was to look for acoustic evidence that Persian coronal stop deletion is incomplete. One would predict that simplified tokens, if they have residual coronal gestures, could show some acoustic cues which fall between those of control and unsimplified tokens. The acoustic parameters of duration and formant transitions were used as the two dependent variables to test this prediction. The results did not show any significant difference between simplified and control conditions, even though both conditions differed from unsimplified clusters according to measures of duration and formant transitions.

One observation was that although there was not a significant difference between the control and simplified conditions, the distribution of simplified tokens is consistent with the possibility that some of them are different from controls. As the duration of consonant clusters across three conditions in Figure 4.6 above shows, the mean for simplified condition is higher than controls. This shows that some of the simplified tokens are different from controls and some of the simplified tokens are above the unsimplified mean.

The lack of any apparent acoustic differences between simplified and control tokens suggests that deletion is in fact complete. If it is not complete, then the residual articulatory gestures provide no acoustic evidence of their presence, leaving listeners no apparent way to distinguish words with simplified clusters from words that have no underlying /t/. In order to

explore this topic, an articulatory experiment was designed. The next chapter describes this experiment and presents its results and analysis. A discussion on the articulatory results will conclude this chapter.

## Chapter 5

### Articulation Experiment

The main goal of this experiment is to investigate whether there are residual gestures left after the simplification of consonant clusters in Persian. Word-final coronal stops can be optionally deleted when preceded by obstruents or the homorganic nasal /n/. Articulatory Phonology states that the simplification of consonant clusters could be due to the overlap among adjacent segments meaning that a coronal stop which is apparently deleted, and has no audible signal, could show some residual gestures. According to this approach, sound patterns and processes in phonology happen because of two underlying reasons: gestural reduction and gestural overlap. The reduction in the magnitude of individual gestures can happen both in time and space. The results presented in Chapter 4 showed that simplified and control tokens are indistinguishable acoustically, suggesting that coronal stop deletion is complete and categorical. This experiment is designed to see how articulatory data can give a fuller picture of this process. The other objective pursued in this study is to explore the optional nature of the simplification process in Persian. In general, this experiment is designed to address the following questions:

1. Is the process of coronal stop deletion in Persian categorical or gradient?
2. Are there any coronal gestures left after the simplification of clusters?
3. How do the results of this study account for the optional nature of cluster simplification?
4. Is there any inter-speaker variation regarding the magnitude of gesture overlap?

5. Do different preceding and following segments overlap differently with coronal stops?
6. Which of the sound change models receives more support based on the data and results of this study?

In order to address these research questions, we should know what the tongue tip/blade is doing at the time of the simplification process. We need a technique which allows us to record the tongue movement in real time. The articulatory experiment is designed to address all these issues. In this chapter, first the methods for the articulatory experiment are explained. The results and analysis of this experiment are presented next. A discussion section will be the last part of this chapter.

## **5.1 Methods of Articulation Experiment**

Audio and real time ultrasound video recordings were made while subjects had a guided conversation with a native speaker of Persian.

### **5.1.1 Subjects**

All the information about participants is discussed above in section 4.1.1.

### **5.1.2 Procedures**

Subjects were seated in an optometry chair located inside the sound booth in the Sound Patterns Laboratory at the University of Ottawa, Linguistics Department. Prior to recording, participants were given instructions regarding the initial stage of the experiment.

They were told that they should hold a mouthful of water for a few seconds in order to capture palate images. Then they were instructed how to produce clicks, which were used as reference points to synchronize the ultrasound images with the acoustic signal. A reading passage was also given to them so that they read and become familiar with it. They were instructed to read the passage like a story in a casual way at the end of the conversation with the researcher (see Appendix 2 for a copy of this passage). The main purpose for using this passage was to ensure that there would be words ending in single consonants, to be used as control tokens, with the idea that they would match the output of complete deletion, if it occurs. For example, words ending in /æŋ/ or /ʊf/ provide a basis for comparison with simplified tokens of words ending in /æŋt/ and /ʊft/. No attempt is made in this study to test the role of different reading styles.

Real time images of the tongue were collected using a Terason T3000 ultrasound system. Ultrasound images of the tongue were recorded with a scan rate of 60 frames per second. One of the concerns when making ultrasound recordings is being able to maintain the transducer probe in a fixed relation to the head. The transducer probe was placed under the subject's chin and was held steady by an Ultrasound Stabilization Headset from Articulate Instruments. The ultrasound headset holds the transducer probe steady and allows for a natural movement of the head. Figure 5.1 illustrates an image of the headset.



Figure 5.1 Image of the headset

The transducer was placed in a way that it could capture tongue contours midsagittally. Figure 5.2 below shows a picture of the experiment set-up. The microphone is mounted on the arm that extends in front of the subject.



Figure 5.2 Image of experiment set-up

After the subjects held water in their mouth and produced three clicks, a casual conversation with the researcher of this study was performed. The researcher chose topics and themes that were intended to elicit as many word-final coronal clusters as possible. The next objective was to lead the dialogue in a way that the coronals were produced in an environment where there was a non-coronal consonant preceding and following them. The following two sentences which are transcribed phonetically (and translated into English) are provided as examples (the target consonants are in bold):

Researcher: [be næzær-e ʃoma dær mah-ha-je əjændeh qejmæt-e næft **gerun** mi-ʃeh?]

Researcher: In your opinion, do you think that the price of **oil will rise** in the next months?

Participant: [ɑreh, mæn fek mi-kon-æm be zudi næft **gerun** mi-ʃeh]

Participant: Yeah, I think the price of **oil will rise** soon.

The above sentences show an example where the coronal stop [t] is preceded by the labial [f] and followed by the velar consonant [g]. It was possible that participants of the study could be sensitive to how the clusters were produced by the researcher, but this chance was quite small. Each casual conversation took 30-35 minutes (see Appendix 1 for the list of topics used and discussed during the casual conversation). At the end of the casual conversation, the participants of the study were asked to read the story passage which they were exposed to at the start of the experiment session.

### 5.1.3 Data Analysis

A Praat TextGrid was used to mark the target coronal consonants in the recordings. In order to get a full picture of tongue movements before and after the production of coronal stops, adjacent segments are also included. This included the consonants and vowels preceding and following the coronals. For example, in a sentence transcribed as [mæn væxt **mi**-konæm.] “I have time for it”, both the consonants and the vowels following and preceding coronal stop [t] (all shown in bold) were marked.

In addition to consonant clusters, some other types of words were marked in textgrids. The underlying assumption here is that if coronal gestural simplification is fully taking place then the remaining final consonants will be labiodental fricative /f/, alveolar fricative /s/, postalveolar /ʃ/, uvular stop /q/, and uvular fricative /χ/. These singleton coda

consonants as well as singleton /t/s were marked in textgrids. The unsimplified tokens are expected to have gestures similar to prevocalic real [t]s.

In order to make sure that the simplified and unsimplified coronal tokens are coded reliably, a native speaker of English and a native speaker of French as well as the researcher of this study (who is a native speaker of Persian) coded the initial 662 tokens independently, using a Multiple Forced Choice experiment in Praat. The two non-native speakers were trained phoneticians and familiar with the phonetic inventory of the target language. All the 662 tokens were randomly presented to the listeners as two-word phrases where the coronal stops would potentially appear at the end of the first word or as an utterance-final cluster. The task was to determine whether or not there was an audible [t] / [d] at the end of the first word. The listeners recorded their judgment by clicking either of the two boxes which appeared on the computer monitor. The rating of the listeners showed above 90% of agreement with the initial rating done in the study. As for the tokens where there was a disagreement, they were listened again by the researcher and were coded based on the majority answer.

In order to extract the corresponding ultrasound images, a python script<sup>5</sup> was run. First, the ultrasound and audio data were synchronized. In order to do so, the python script used the textgrid file and the timing (in the audio and ultrasound videos) of at least one click produced by the subject. If more than one click and ultrasound image time was used, the python script used the average as a reference point. Moreover, the python script converted all bmp images into JPEG files. The python script also inspected and checked for gaps and

---

<sup>5</sup> Mielke, J. (2010). *Python Script for Extracting Files from Ultrasound Recording*.

duplicate images in the recording. Problematic files were excluded from the analysis. Due to the challenge for separating the coronal gestures belonging to two adjacent coronals from each other, only final coronal stops in a non-coronal environment were selected for further analysis. For example, some tokens such as [mæn qæsd kærd-æm ke...] “I intended to ...” or [bæd nist ke ...] “It isn’t bad that ...” were excluded from the pool of data. This resulted in the exclusion of about 20% of the captured tokens.

The initial number of acoustic tokens collected were 662 with the duration of 300-500 ms., depending on the speech rate of the participants and whether the coronal stop was deleted or not. As for the articulatory analysis, some of them had to be discarded due to homorganicity (as discussed above) and the lack of corresponding articulatory data. At the end, 248 tokens, including 51 controls, were chosen for segmentation and articulatory analysis (see Appendices 3 & 4 for the lists of unsimplified and simplified words elicited during interviews as well as words used as controls [i.e., singleton coda consonants]).

For each token, all the frames capturing the coronal stop as well as the preceding and following environments were traced. Considering the fact that ultrasound frames were generated at 16-17 ms intervals, each token consisted of 20-32 frames. Palatoglossatron (Baker, 2005) was used to trace the tongue tip/blade as well as the tongue body in each tongue image and the alveolar ridge in each palate image. The following six figures show images of tongue and palate as well as their tracing and overlay. The images of tongue are midsagittal where the front of the tongue (i.e., tip/blade) is to the right. The palate in Figure 5.4 appears as the lighter contour at the top along with tongue trace under it. Figures 5.5 - 5.7 show the tracing of tongue and palate and how radial grids are used to calculate the tongue-

palate distance. The front radii measure the tip/blade of the tongue and they proceed into further back part of the tongue (i.e., dorsal) once we reach the radii to the left. Each radius has its own measurement and produces a single column in the output CSV measurement file. Figure 5.8 provides an overlay image of tongue and palate tracing.

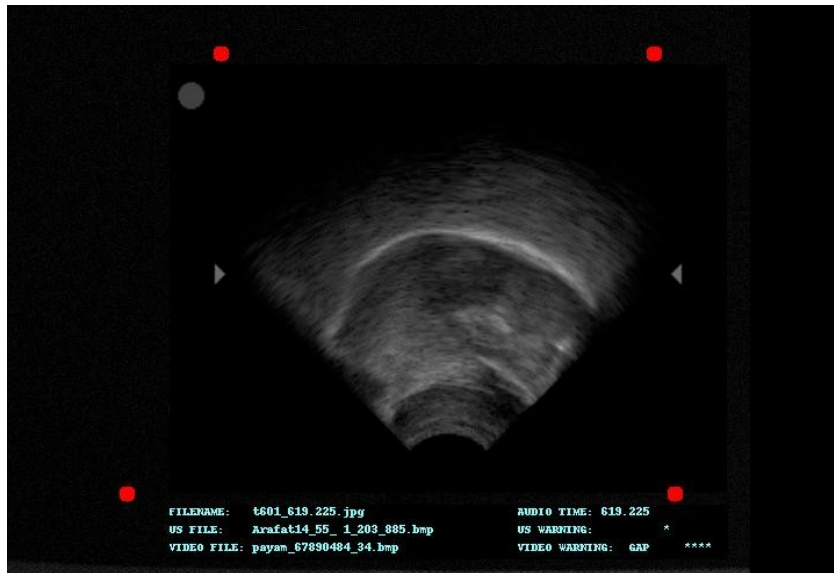


Figure 5.3 Image of tongue

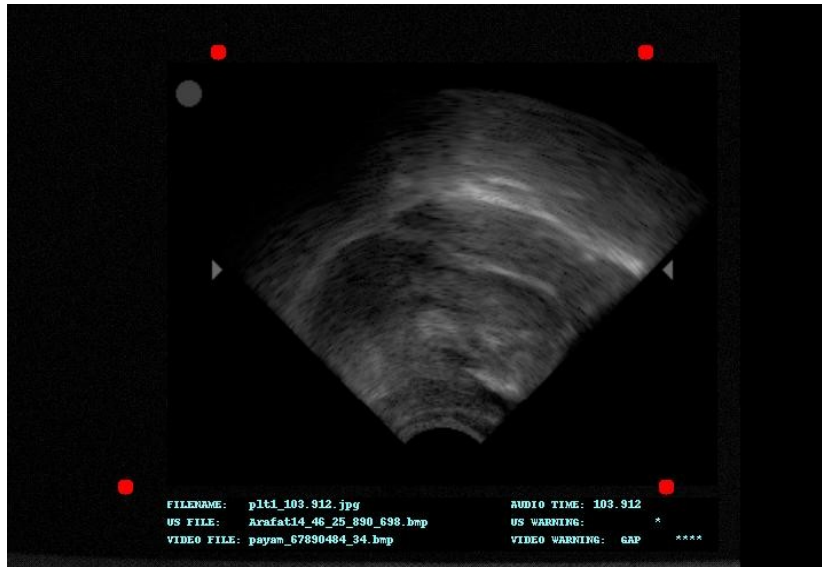


Figure 5.4 Image of palate

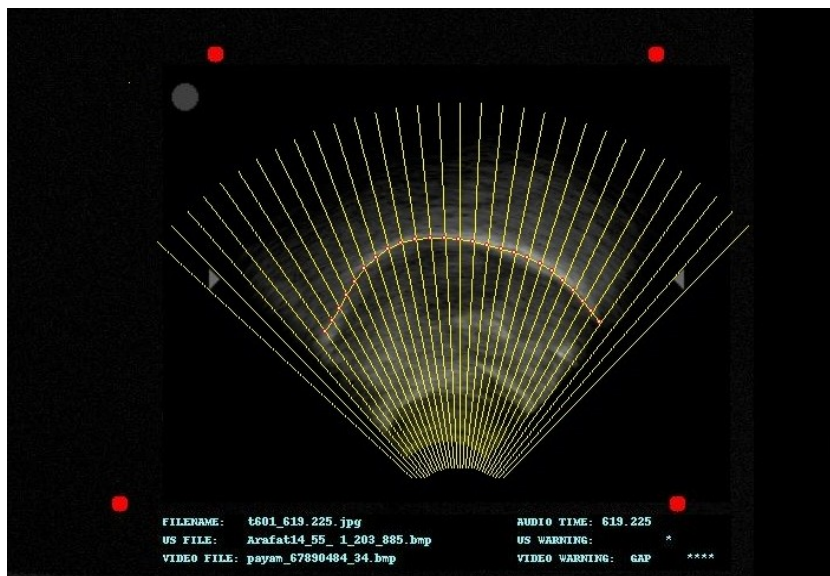


Figure 5.5 Image of radial grid over the tongue



Figure 5.6 Image of tracing of tongue

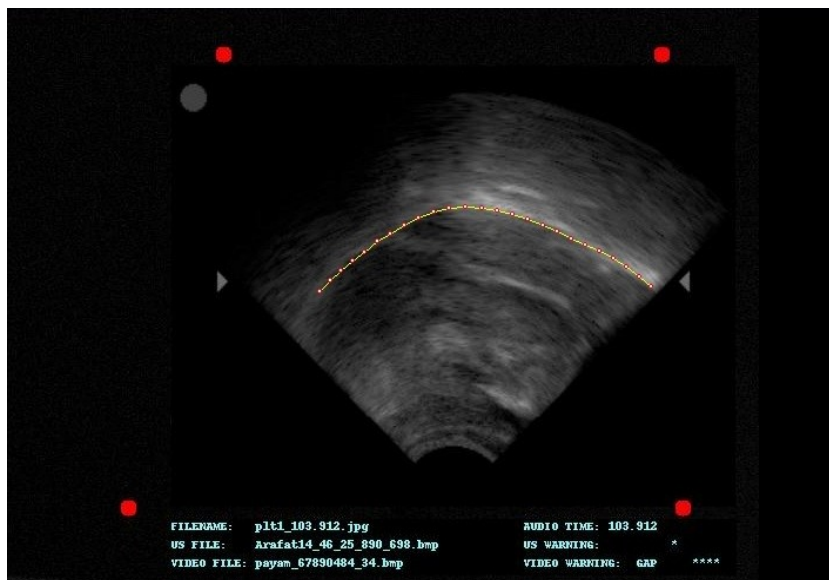


Figure 5.7 Image of tracing of palate



Figure 5.8 Image of overlay of tongue and palate tracing

The next step after tracing was to export them. The comma-separated-values (CSV) file from Palatoglossatron was fed into MATLAB, in order to plot the tongue movement over time. The MATLAB script first rescaled the data by changing distances from pixels into millimeters<sup>6</sup>. Then it used the distance between the tongue and palate to plot the tongue trajectories. Vertical velocity was calculated as the rate of time over the movement of the tongue along a particular radius. Through this, we differentiated the time function of the distance in order to obtain the velocity<sup>7</sup>. In order to get a more realistic picture of the tongue movements, some of the plots needed to be smoothed. Smoothing was necessary in order to exclude spurious peaks and zero crossings from the velocity function introduced by the quantizing ultrasound images produced with a spatial resolution of about 1mm. The smoothing process was done by averaging over a window of three frames. Some apparently erroneous measurements (which deviated by more than 1 millimeter from the average of the

---

<sup>6</sup> I appreciate the help of M. Fallahpour for this script.

<sup>7</sup> Special thanks to L. Bombien for his clarification on this point.

preceding and following measurements) were corrected by replacing them with that average value. The magnitude of the changes introduced by smoothing is small, but the consequences of smoothing are crucial for accurately identifying gesture landmarks on the basis of the velocity function.

Seven gesture landmarks were determined based on the maximum and minimum velocity in the positive and negative directions of the velocity plots: Onset of gesture (on), maximum velocity in onset (mon), beginning of constriction plateau (ps), maximum constriction (max), end of constriction plateau (pe), maximum velocity in offset (moff), and offset of gesture (off). Maximum velocity in onset (mon), maximum constriction (max), and maximum velocity in offset (moff) correspond to the kinematic events with the maximal velocity at the beginning (negative value) and end (positive value) of the gestures as well as the landmark where velocity becomes zero (Kühnert, Hoole & Mooshammer 2006).

Following Bombian (2011), the other landmarks including maximum velocity in onset (mon) and maximum velocity in offset (moff) as well as onset of gesture (on) and offset of gesture (off) are interpolated values and represent the 20% threshold of the difference between two adjacent extrema in the velocity plot. For example, the beginning of the constriction plateau is determined to be at a 20% threshold between the landmarks maximum velocity in onset and maximum constriction. The following three sets of figures show this process. In each set, the smoothed version of figures is illustrated first. The smoothed figures provide the height of tongue with arrows marking the landmarks. The corresponding non-smoothed figure is given after each smooth figure for illustration purposes. These figures could either show the tongue tip (marked as TT) or the tongue body (marked as TB). The next figure in each set shows the velocity, which is derived from the height figures. Figures 5.9 - 5.11 provide an

illustration of a non-deleted coronal stop [t] in the token [suxt giri] “getting fuel”. The next three figures 5.12 - 5.14 show a deleted coronal stop [t] in the token [sæxt gireh] “is strict”. Figures 5.15 - 5.17 show uvular stop [χ] in the same token. The x axis in all figures represent time in milliseconds. In height figures, the y axis show the distance between TT/TB and the palate.

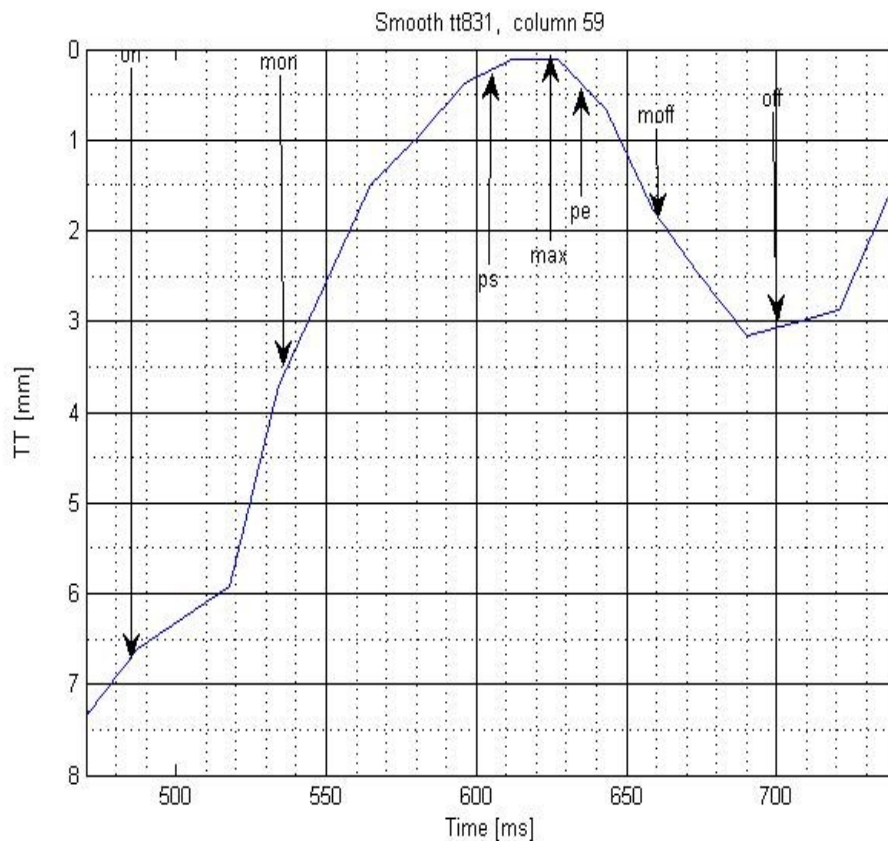


Figure 5.9 Smooth vertical tongue tip position for non-deleted coronal stop [t] in the token ([suxt giri] “fuel taking”)

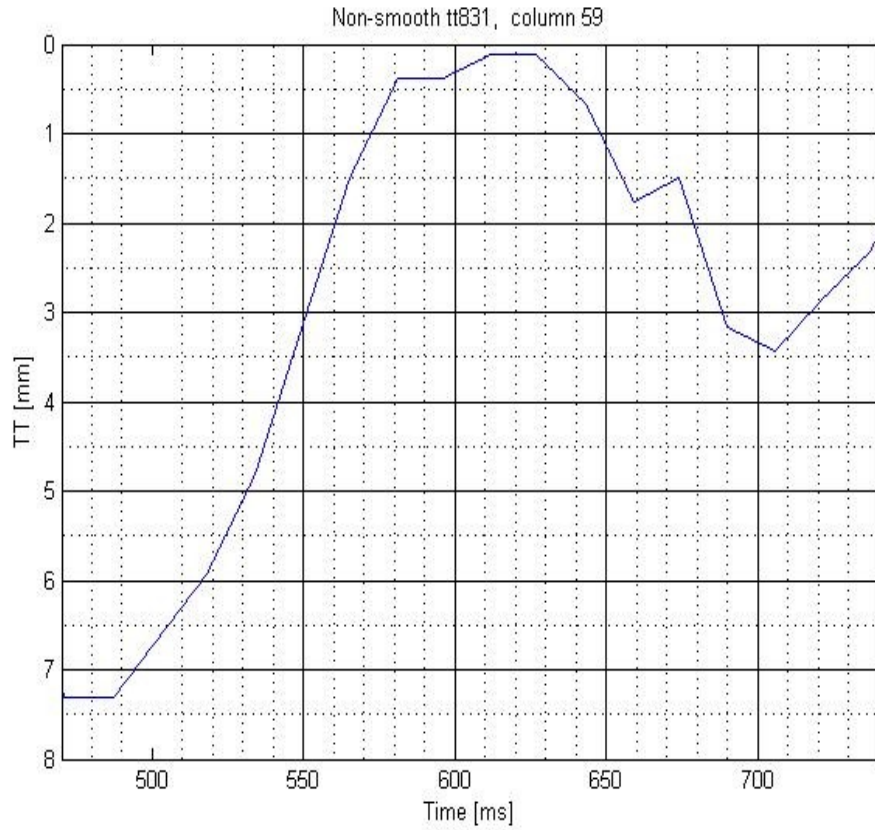


Figure 5.10 Non-smooth vertical tongue tip position for non-deleted coronal stop [t] in the token ([suxt giri] “fuel taking”)

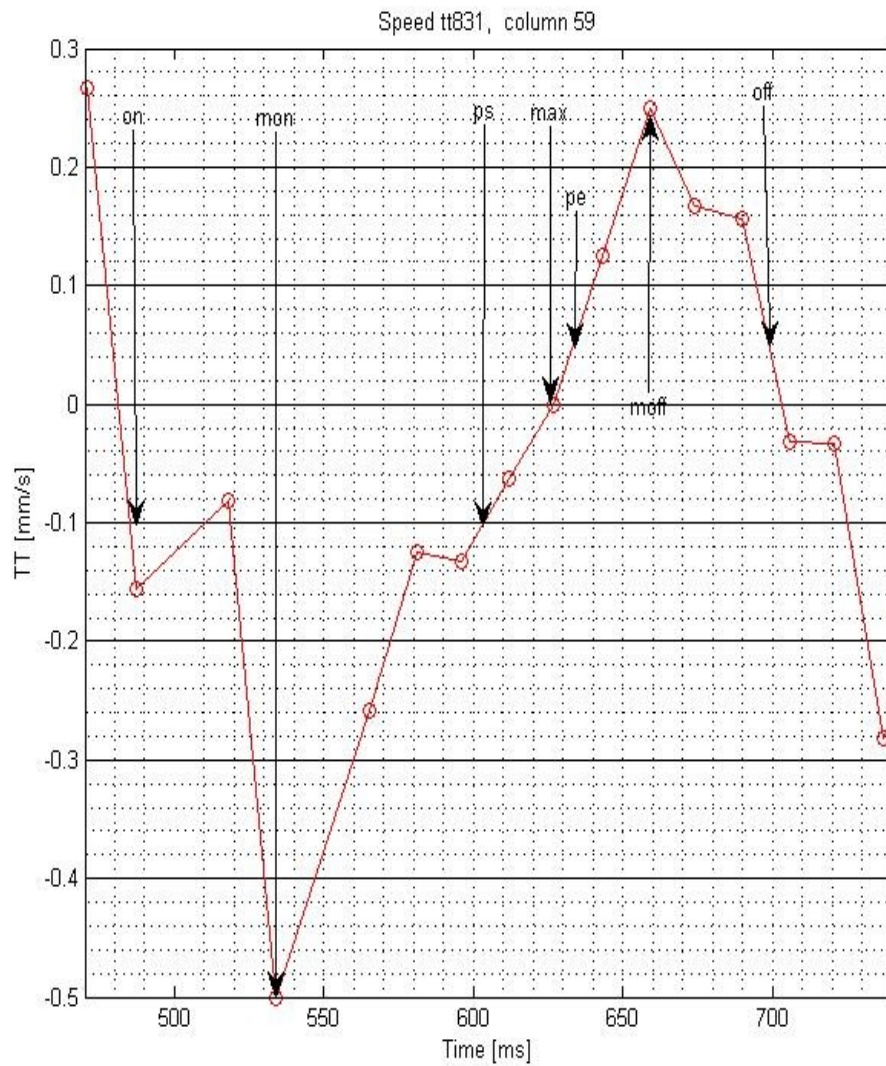


Figure 5.11 Velocity for non-deleted coronal stop [t]  
in the token ([suɣt giri] “fuel taking”)

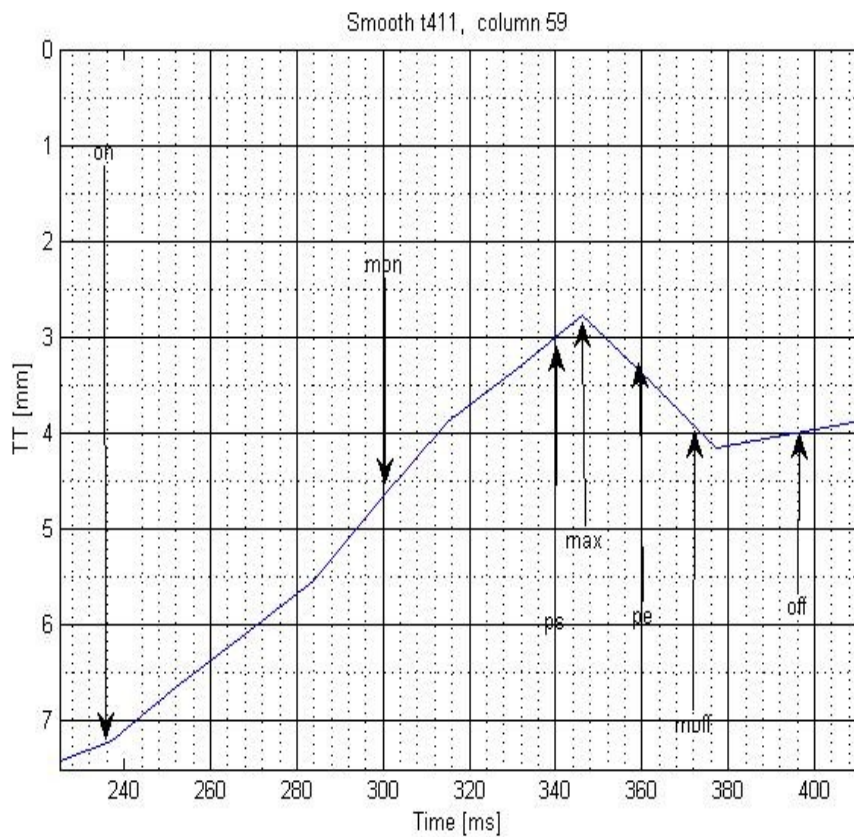


Figure 5.12 Smooth vertical tongue tip position for deleted coronal stop [t] in the token ([sæxt gireh] “is strict”)

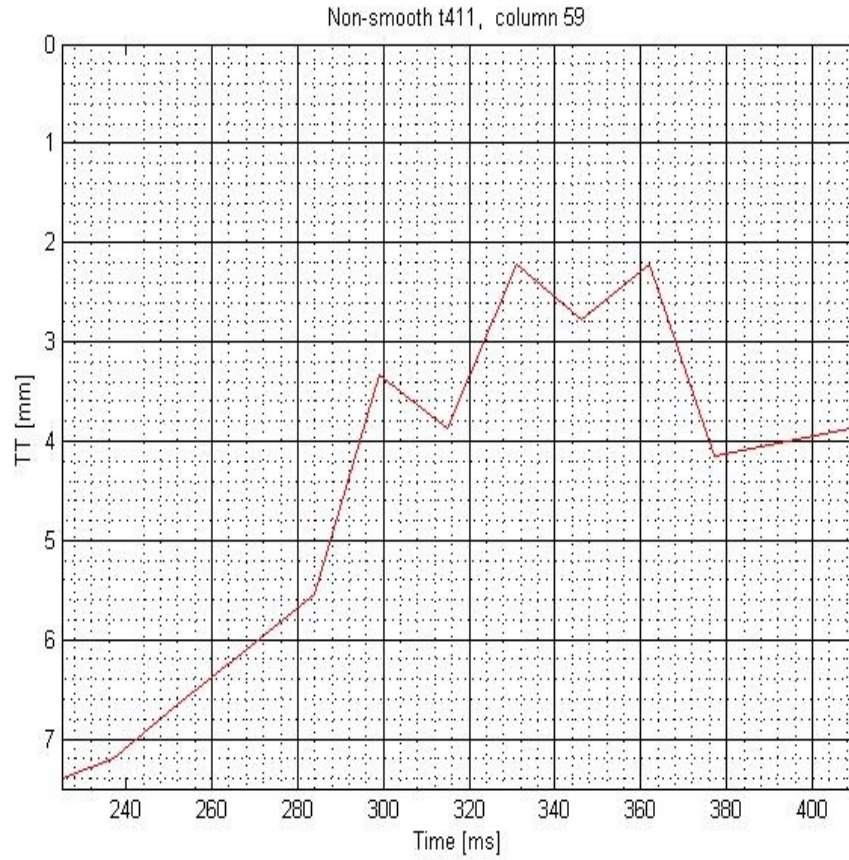


Figure 5.13 Non-smooth vertical tongue tip position for deleted coronal stop [t] in the token ([sæχt gireh] “is strict”)

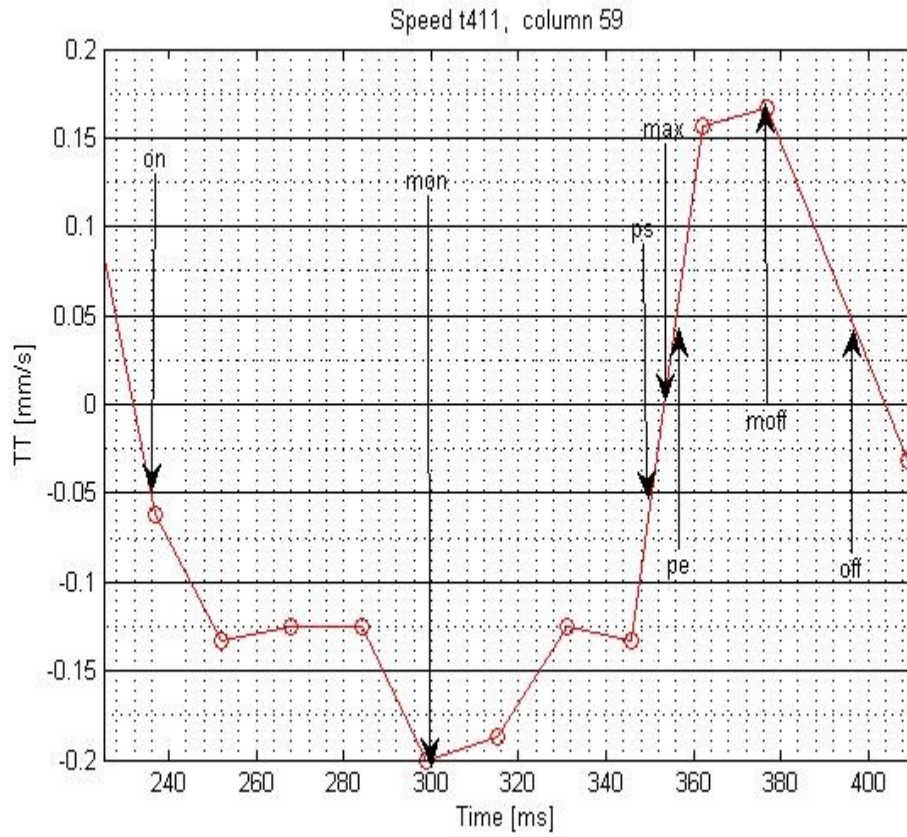


Figure 5.14 Velocity for deleted coronal stop [t]  
in the token ([sæχt gireh] “is strict”)

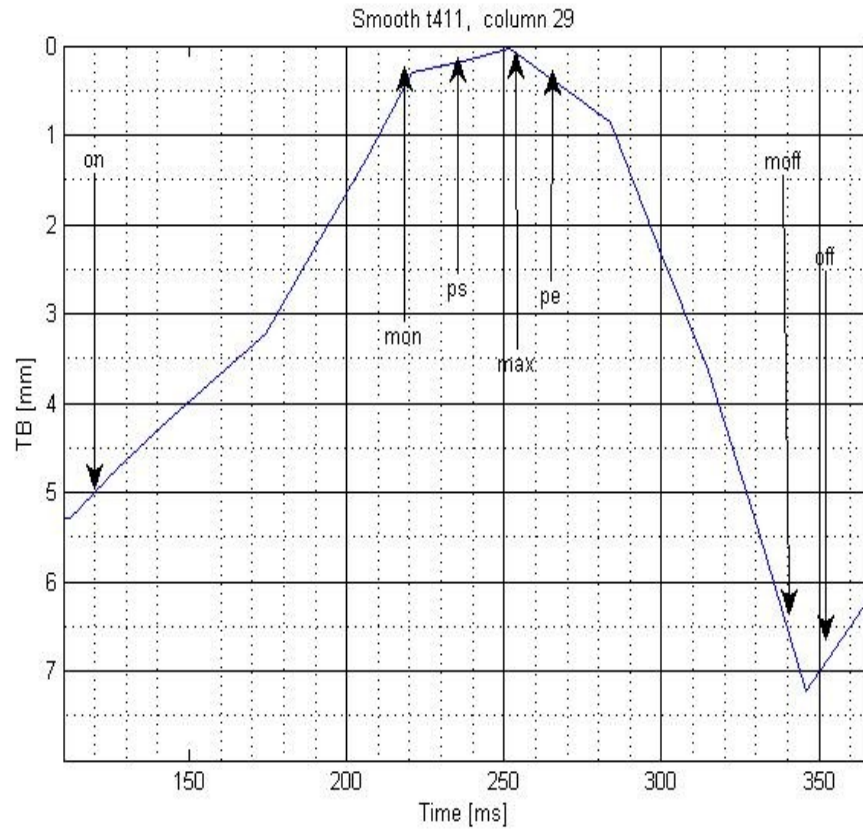


Figure 5.15 Smooth vertical tongue body position for uvular stop [χ] in the token ([sæχt gireh] “is strict”)

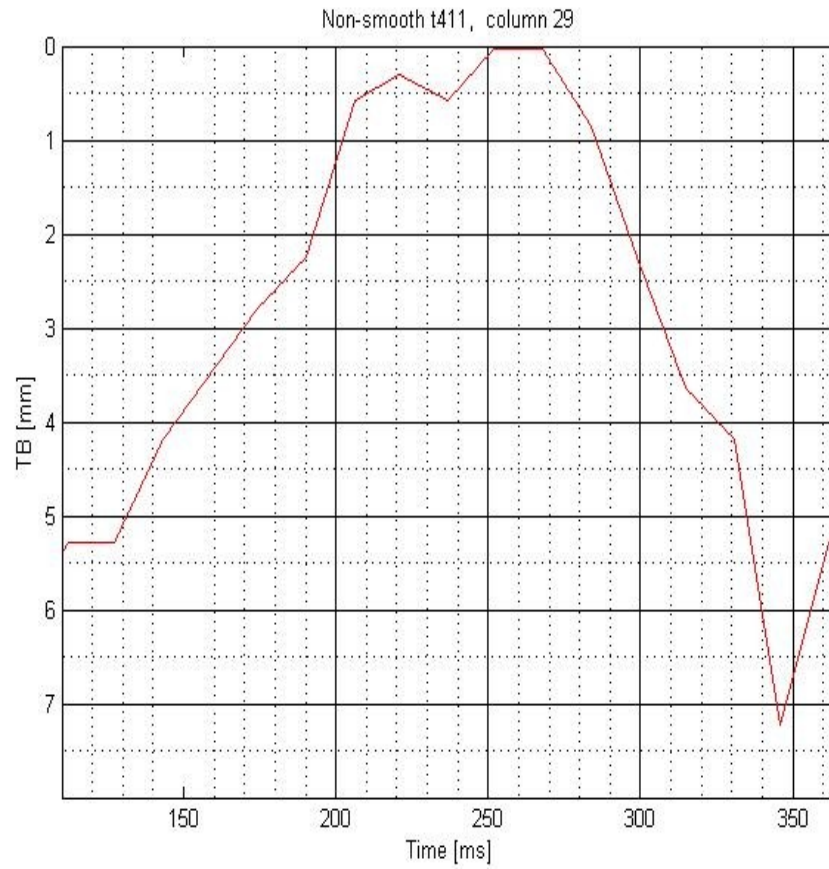


Figure 5.16 Non-smooth vertical tongue body position for uvular stop [χ] in the token ([sæχt gireh] “is strict”)

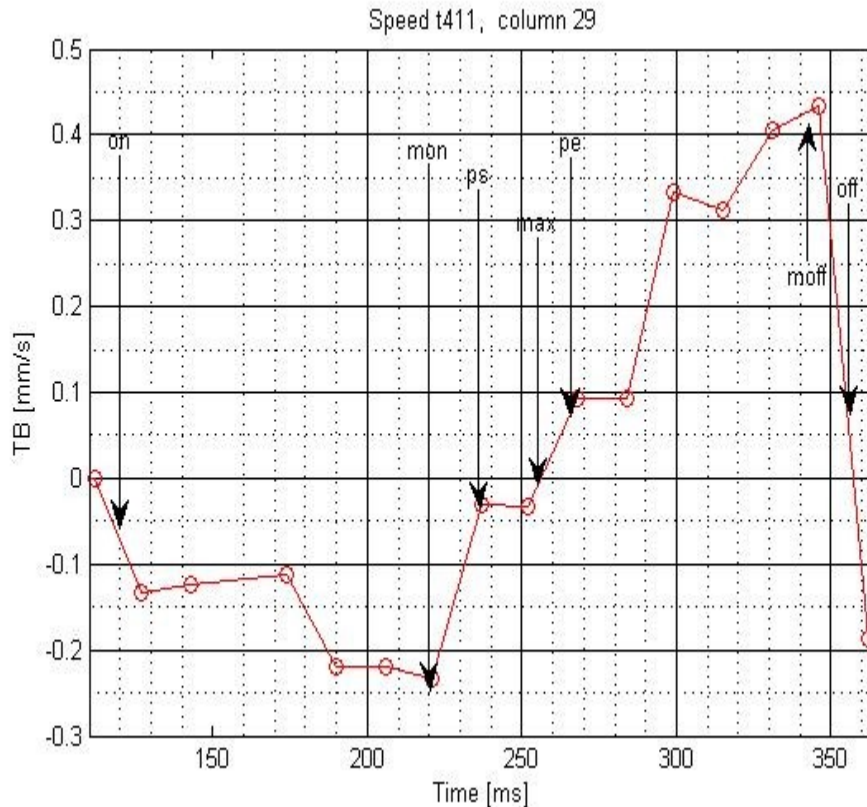


Figure 5.17 Velocity for uvular stop [χ] in the token ([sæχt gireh] “is strict”)

Researchers have used very different gesture overlap indices for many different purposes. They have mainly chosen the landmarks used in their analysis based on the purpose of their studies. For examples, those who were mainly interested in investigating the coordination among gestures, have used the latency between two or three landmarks in  $C_1$  and their counterparts in  $C_2$ . This provides information about the initiations of two or more movements in  $C_1$  and how they are similar or different from their corresponding mini-gestures in  $C_2$ .

From the other side, studies which aimed to investigate coarticulation between two consonants, mainly looked at the distance between an offset in  $C_1$  and the respective onset in  $C_2$  (e.g., Chitoran, Goldstein, & Byrd 2002; Kühnert, Hoole & Mooshammer 2006; Bombien

2011; Cunha & Harrington 2011). The main goal in these studies was to see how close the movements may “sit on each other”.

In addition to these measures, Browman and Goldstein (2000) proposed C-center as a new landmark which is calculated as the mean of the maximum constriction intervals related to individual gestures of  $C_1$  and  $C_2$ . This is mainly used when the vocalic environment preceding the consonant is being investigated as an influencing factor in the gestural scores of the consonant clusters. In the current study, following Chitoran, Goldstein, and Byrd 2002; Kühnert, Hoole and Mooshammer 2006; Bombien 2011; Cunha and Harrington 2011, I used the end of constriction plateau of  $C_1$  [pe] and the beginning of constriction plateau of  $C_3$  [ps] as the index of gestural overlap measurement. These landmarks are expected to represent the overlap between sequential consonant gestures in the best way.

For the non-lingual consonants such as bilabial (e.g., [p b]) and labiodental (e.g., [f]) consonants which were not visible in the ultrasound data, the start and the end of consonants were determined based on the spectrograms of the audio recordings. This procedure was explained in detail in Chapter 4 which dealt with the acoustic experiment.

The next step was to enter all the corresponding times to landmarks onto a CSV file. A python script<sup>8</sup> was used next to measure the vertical distance between the tongue and the palate at the different landmarks. The output of this process was a new CSV file containing all the distances between tongue and palate for each specific landmark in the designated consonants. The new CSV file was fed into the statistical software R (Version 2.15.1., Development Core Team, 2012) using a script<sup>9</sup> to compare gestures. The lingual trajectories for coronal stop [t] based on the timing of landmarks of the preceding and the following

---

<sup>8</sup> Mielke, J. (2012). *Python Script for Plotting Gesture Distance*.

<sup>9</sup> Mielke, J. (2012). *R Script for Plotting SSANOVA*.

consonants were plotted. The results and analysis of the articulatory experiment are presented in the next section.

## **5.2 Articulatory Results and Analysis**

The main objective of this study is to investigate whether coronal stop deletion in Persian is a categorical or gradient process. The results presented in Chapter 4 showed that simplified and control tokens are indistinguishable acoustically, suggesting that coronal stop deletion is complete and categorical. In this section the results of the articulation experiment of the thesis are presented and analyzed. According to Articulatory Phonology, sound patterns and processes in phonology happen because of gestural reduction and gestural overlap. The reduction in the magnitude of individual gestures can happen both in time and space. This section presents the results related to spatial gestural changes along the temporal axis.

In order to get the articulatory results, the same radii and the same part of the tongue which intersects them were tracked. This means that for the coronal stops the rightmost radius which tracks the tip/blade of the tongue was considered. For the velar and uvular consonants, the radii where the body of the tongue had the most constriction related to that consonant was chosen. This method is different from gestural scores based on X-ray microbeam and EMMA, which are point-tracking. The tracking led to tongue trajectories related to coronal stop [t], velar stops [k g] and uvular fricative [χ]. In order to determine the duration of the lingual gestures, the time corresponding to the start of plateau and the offset of the related gesture were used as the gestural landmarks (see section 5.1.3 for details regarding the procedure of choosing landmarks). For bilabial [p b] and labiodental [f]

consonants which do not have any lingual gestures, the sound files were used as the point of reference. Figures 5.19 -5.21 below present the gestural scores from individual tokens. Each gestural score shows the movement of a number of articulators which are continuously active over a given period of time. In these figures, the x-axis shows the time and the y-axis displays the distance between the tongue and the passive articulator along the selected radii of the ultrasound's field of view. The symbols C1, C2, and C3 stand for the three consonants in the clusters as they unfold in time, respectively. The two trajectories related to [χ] and [k] show the tongue body movement and the trajectory related to coronal stop [t] shows the movement of tongue tip/blade over a specific sequence of time. Each column (e.g., col 27, col 14, and col 6), highlighted in Figure 5.18 below, represents the measurements made by the 32 radii as shown in Figure 5.5 and repeated below as Figure 5.18.

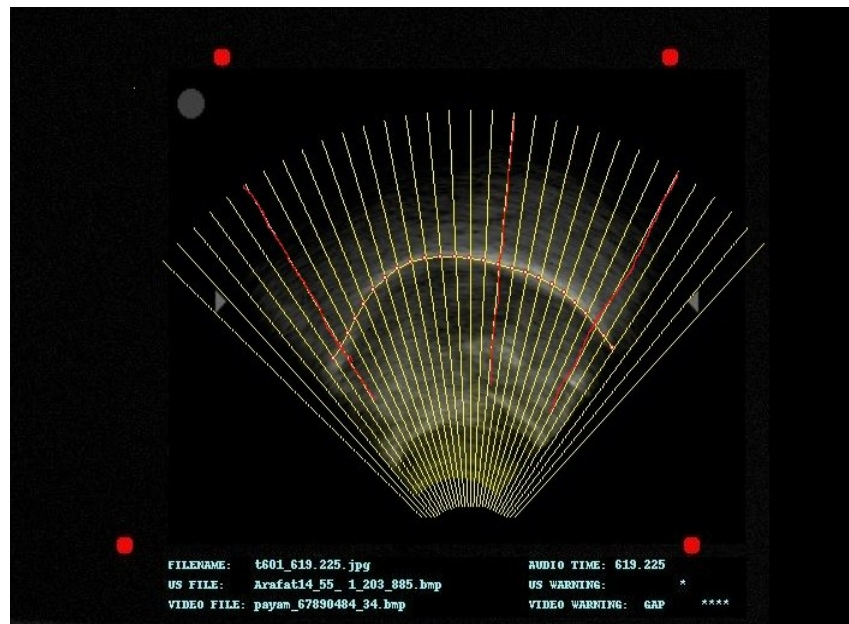


Figure 5.18 Image of radial grid over the tongue

Column 1 shows the measurement for the most posterior and column 32 for the most anterior radii. The measurement points were selected to match the place of articulation for particular consonants for particular subjects. Each rectangle represents the duration of the gestures related to segments. The beginning and the end of duration was determined based on the start of plateau and the offset of the related gesture. The three figures below show how gestures produced by different articulators unfold and overlap over time. For example, Figure 5.19 shows the gestural score of the unsimplified token ([duxt ke] “sew that”) for subject #6. In this figure, C<sub>3</sub> (col 14) is the trajectory imaged in column 14, and this column is determined to be the constriction location for C<sub>3</sub> (i.e., velar stop). It should be emphasized that in these figures the trajectory of each segment is shown for the entire time interval which includes more than just segmental duration. Figure 5.20 below taken from subject #6 illustrate the gestural scores of the simplified tokens ([duxt ke] “sew that”).

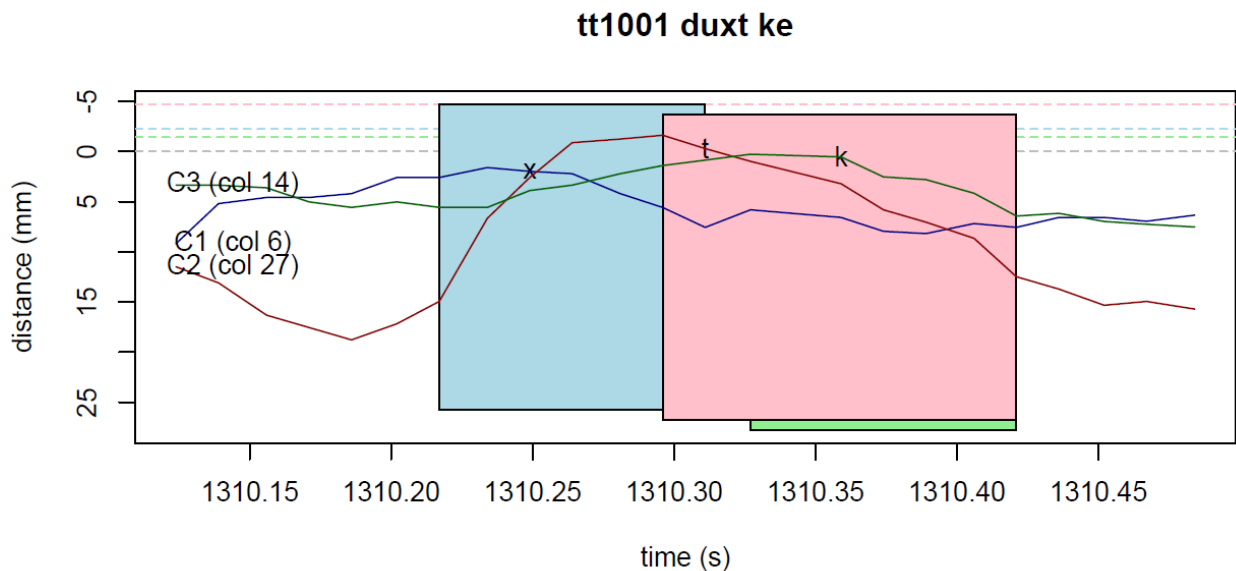


Figure 5.19 Gestural score of the unsimplified token ([duxt ke] “sew that”) for subject #6

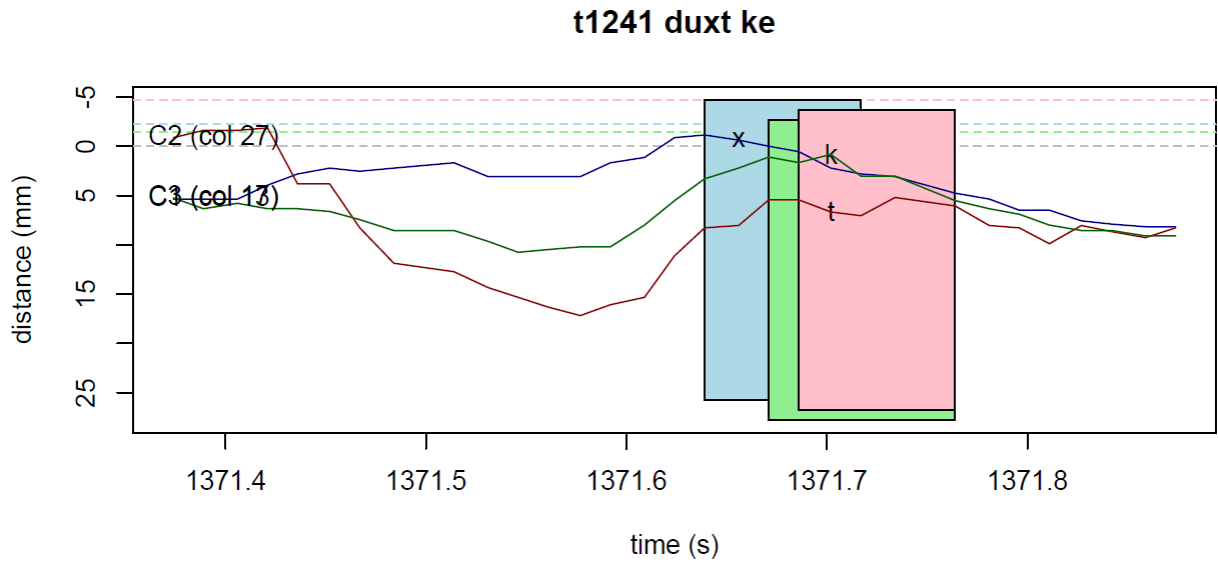


Figure 5.20 Gestural score of the simplified token ([**dux**t ke ] “sew that”) for subject #6

Figure 5.21 below illustrates the gestural score of the simplified token ([**mof**t geruneh] “is of little value”). Since labiodental [f] has no lingual gesture, only the tongue trajectories of the two other consonants are traced.

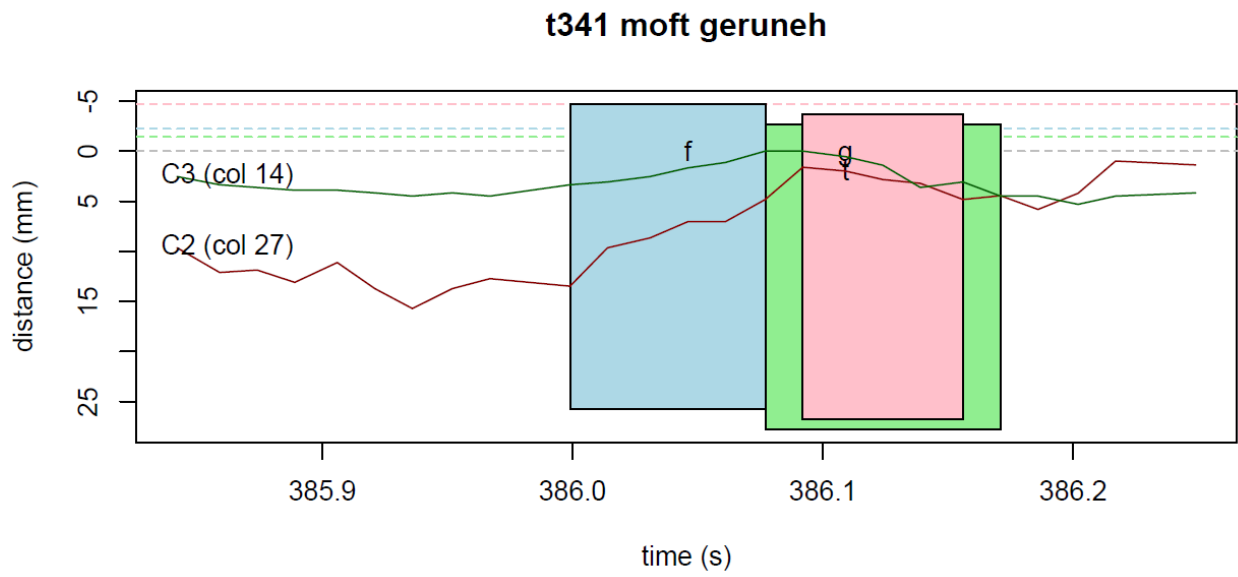


Figure 5.21 Gestural score of the simplified token ([moft geruneh] “is of little value”) for subject #6

The similar tokens were grouped together in order to get a broad picture of how simplified tokens relate to unsimplified tokens and controls. Next, the best fit (i.e., smoothing splines) for the related tongue trajectories of coronal stop [t] was found. After finding the best fit, 95% Bayesian confidence intervals around the smoothing splines were constructed. Smoothing Splines ANOVA [SSANOVA (Gu, 2002)] was applied to data to determine whether there is a significant difference between two curves. Accordingly, when the confidence intervals of the main effect curves overlap, the differences between the two curves are not significant. This technique was first used in other fields such as medical sciences where similarities and differences between two curves were assessed. This procedure was first introduced to the field of linguistics by Davidson (2006) and was widely used afterwards. For example, in a study on rhotic vowels in Canadian French, Mielke (2011) used ultrasound to see whether Canadian French rhotic-sound vowels are

articulatorily similar to English [ɹ]. The SSANOVA technique has the advantage of capturing the dynamic and overlapping nature of the tongue movement and curve. Moreover, this technique illustrates the tongue trajectories more naturally. In this study, I am using SSANOVA in a novel way. It compares gestures over time, with time on the x-axis but with tongue position on the y-axis. Previous studies have mainly used SSANOVA to compare tongue contours rather than trajectories of parts of the tongue surface, so the axes display the horizontal and vertical position of the tongue surface. Since this study explores the gestures of coronal stops, the x-axis displays time, and the y-axis displays the distance between the alveolar ridge and the tip/blade of the tongue. Figure 5.22 below shows a SSANOVA graph for illustration purpose.

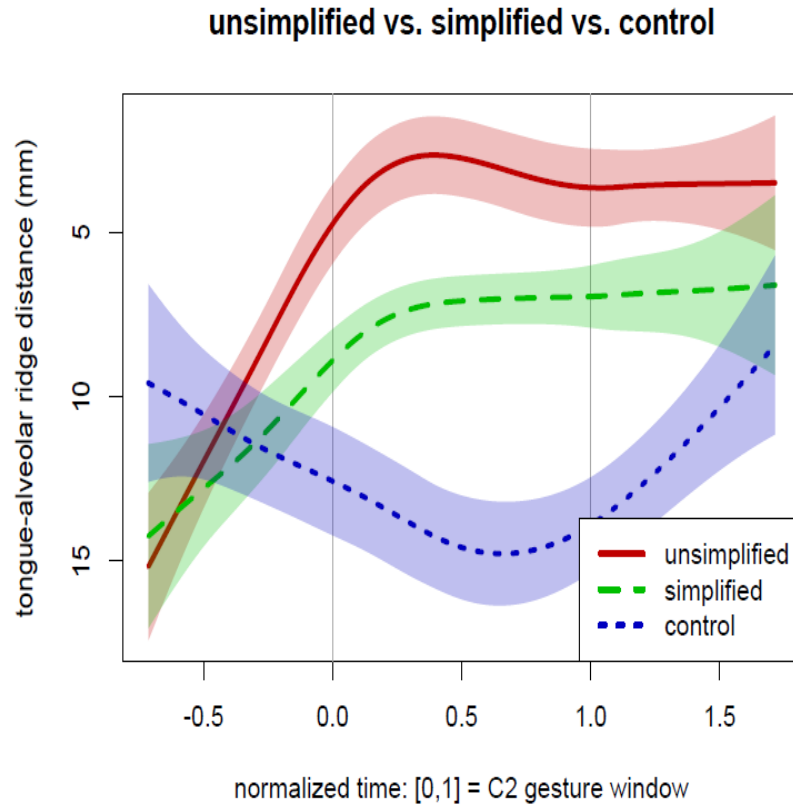


Figure 5.22 Illustrative SSANOVA graph

Figure 5.22 above illustrates the position of tip/blade of the tongue contours in the three conditions of the study (i.e., unsimplified, simplified, and control) through using different colors and patterns. Since the duration of the interval between the gestural landmarks used for comparison vary across subjects and conditions, tokens were aligned on the basis of  $C_1$  and  $C_3$  gesture landmarks as follows: the time interval  $[0,1]$  is defined as the interval between the  $C_1$  gesture maximum and the  $C_3$  gesture maximum. If the time of the gesture maxima were missing, the time of plateau start or end was substituted (favoring the end for  $C_1$  and the start for  $C_3$ ). If no  $C_3$  landmark was available, time 1.0 was arbitrarily set to 8 frames (133ms) after the  $C_1$  landmark, and if  $C_1$  and  $C_3$  landmarks were simultaneous,  $[0,1]$  was set to the preceding and following frames (a time interval of 33 ms). Trajectories

were represented by the [0,1] interval plus a pad of five preceding and five following frames. In the SSANOVA figures used in this study, the normalized time on the x axis indicates the C<sub>2</sub> gesture window in the interval [0,1], and times outside this interval corresponds to the 5-frame pads. In each graph, the mean tongue tip/blade trajectory is represented through thick lines and the shaded regions represent confidence intervals (95%) around the main effect line. It should be emphasized here that each contour represents tongue movement over time along a single radius of the ultrasound image which was selected because it is in the tongue blade/alveolar ridge region.

The following SSANOVA graphs illustrate the pooled data of each subject. The pooled data of each subject is categorized based on the three conditions. The prediction is that the contours for the unsimplified condition should show the least distance between the tongue and the palate while the contours for the control condition are expected to show the largest distance. If deletion is complete, the contours illustrating the tongue-palate distance for the simplified condition should be similar to the controls. On the other hand, if the apparent simplification is completely due to gestural overlap, the simplified contours should fall somewhere between unsimplified and control conditions.

Figure 5.23 below shows the SSANOVA graph of tongue-alveolar ridge distance for subject #1. According to the C<sub>2</sub> gesture window in this figure, the contours representing the unsimplified, simplified, and control conditions all show different tongue-palate distance. As expected, the unsimplified condition shows the least distance whereas the control condition shows the largest distance between the tongue and alveolar ridge. According to the confidence intervals in the three conditions, the unsimplified condition shows a significant difference with simplified condition. The simplified and control conditions show significant differences throughout most of the C<sub>2</sub> gesture window. The y-axis shows the distance in

millimeters between the tongue blade and the position of the alveolar ridge as measured in a palate image. The actual value of this distance is affected by vagaries in tongue and palate imaging, and differences between closure location and the selected radius. For this reason, it is important to use the unsimplified category (rather than 0) as the reference for evaluating the simplified tokens.

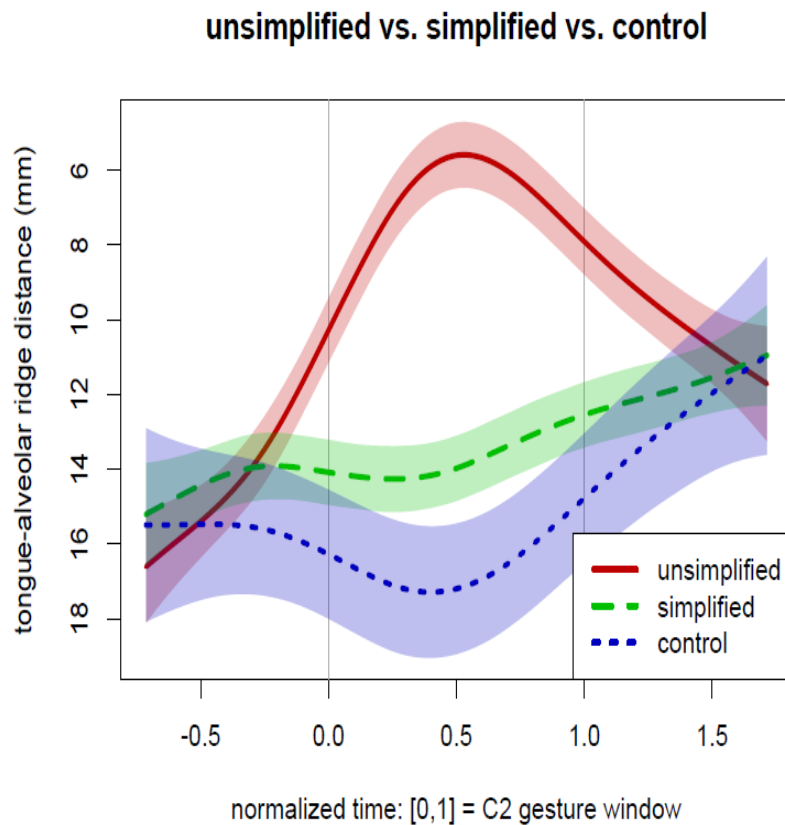


Figure 5.23 SSANOVA graph of tongue-alveolar ridge distance for subject #1

Figure 5.24 below shows the tongue-alveolar ridge distance for subject #2. Similar to the results for subject #1, the three conditions in the  $C_2$  gesture window show differences in their distance contours. The unsimplified condition shows the least distance while the

contour belonging to control shows the largest distance. The simplified contour falls between unsimplified and control contours. According to this figure, the contour for the simplified condition is significantly different from both the unsimplified and control conditions.

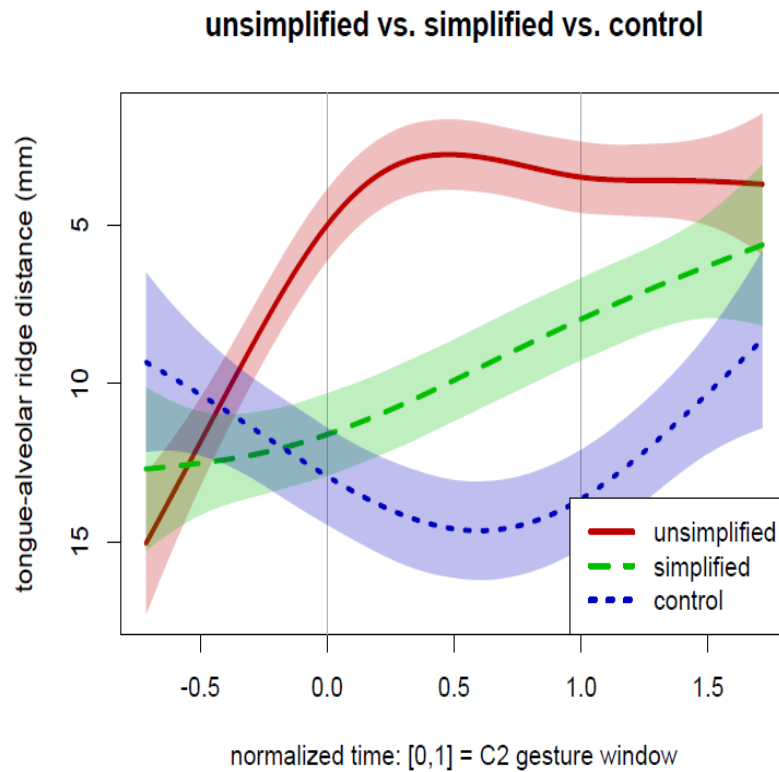


Figure 5.24 SSANOVA graph of tongue-alveolar ridge distance for subject #2

Figure 5.25 below illustrates the results for subject #3. According to this, the contour showing the tongue-palate distance for the simplified condition is very close to the control contour. However, the contours related to unsimplified and simplified conditions in the C<sub>2</sub> gesture window show significant difference from each other.

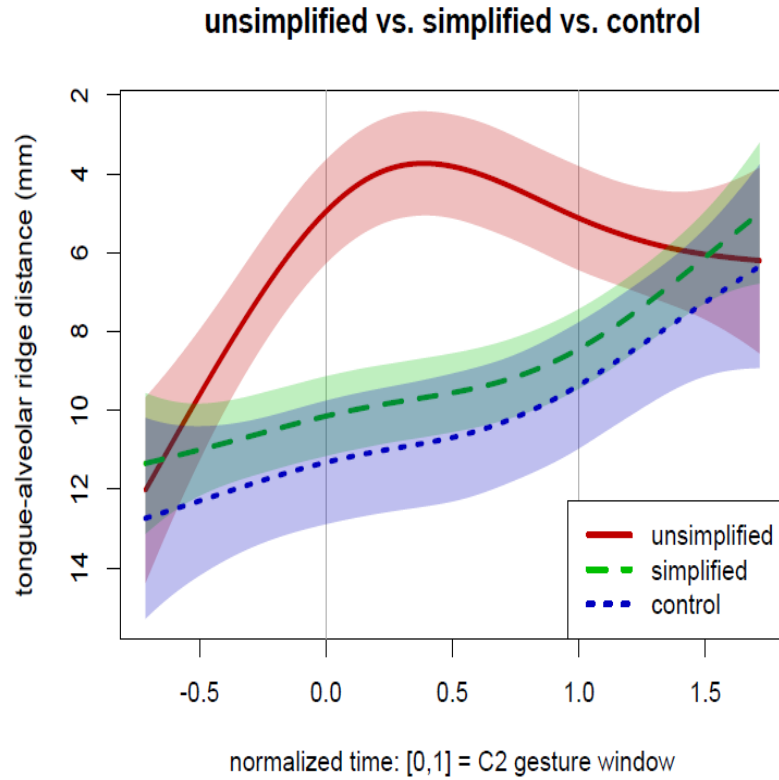


Figure 5.25 SSANOVA graph of tongue-alveolar ridge distance for subject #3

Figure 5.26 below presents the SSANOVA graphs for subject #4. The contours of unsimplified and simplified conditions show significant difference for this subject; however, only a part of the simplified contour is actually significantly more distant than the control contour. This could be attributed to different contexts in which tokens were produced. The discussion regarding how the effect of context is controlled will be taken up later in this chapter.

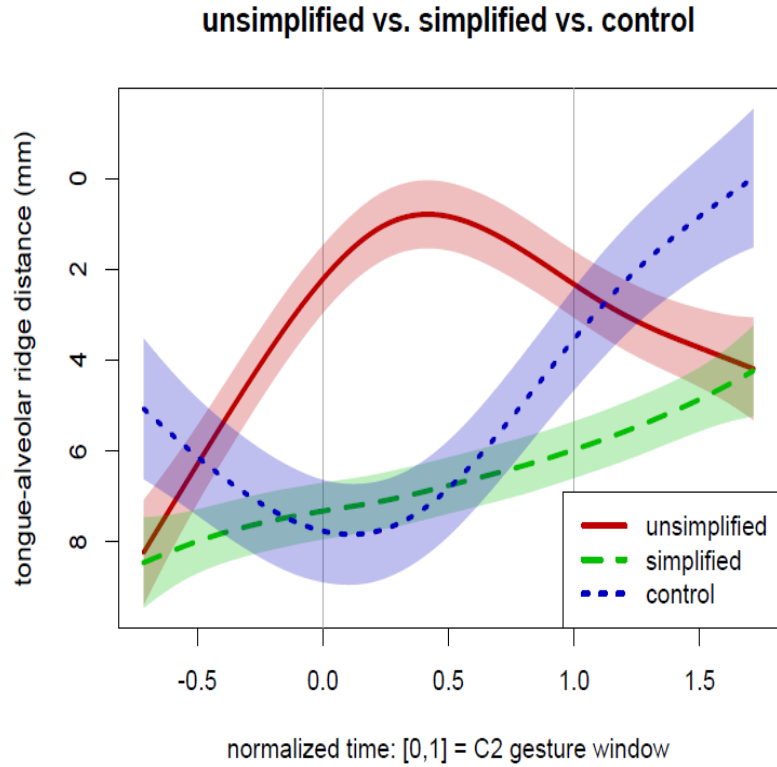


Figure 5.26 SSANOVA graph of tongue-alveolar ridge distance for subject #4

Figure 5.27 illustrates the graph of tongue-alveolar ridge distance for subject #5.

The unsimplified SSANOVA curve for this subject shows significant difference in the C<sub>2</sub> gesture window with the simplified contour. Moreover, most of the simplified contour shows significant difference with the control contour.

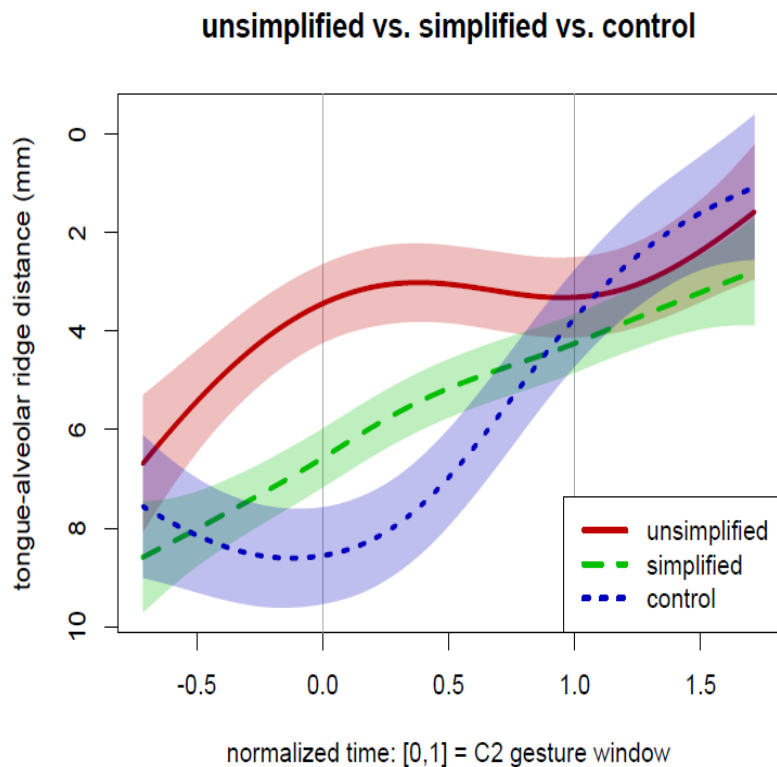


Figure 5.27 SSANOVA graph of tongue-alveolar ridge distance for subject #5

The last two subjects in the study (i.e., subject #6 and subject #7) simplified the consonant clusters far less than the other participants. In fact, one of them simplified less than 7% of the potential tokens in the interview. Due to the small number of simplified tokens for these two subjects compared to unsimplified tokens, their results are not presented in this section.

In summary, the subjects based on these results fell into 4 categories: For two out of seven subjects (i.e., subjects #3 and #4) the contours of the unsimplified condition in the  $C_2$  gesture window showed significant difference with that of simplified condition. However, the SSANOVA curves of the simplified and control conditions for these subjects did not show such a difference. The results of subject #1 and subject #2 displaying the tongue-palate

distance in the  $C_2$  gesture window showed significant difference across unsimplified, simplified, and control conditions. Subject #5 showed significant difference between simplified and unsimplified contours; however, this was only partially true between control and simplified contours. The last two subjects (i.e., subjects #6 and #7) simplified the clusters the least and did not generate the expected patterns. One can notice that in all subjects, the SSANOVA curve of the simplified condition falls between the unsimplified and control conditions. However, it patterns similar to controls in some subjects (e.g., subjects #3 and #4) and dissimilar to both unsimplified and control contours in other subjects (e.g., subject #1 and subject #2). This could raise the question whether pooling all simplified tokens together hides some facts about their internal distribution. In fact, a closer inspection of the simplified tokens shows that some of them pattern similar to unsimplified tokens, some are similar to controls, and some fall in between. Figures 5.28 - 5.32 below present a representative number of tokens which illustrate this fact. These figures demonstrate the tokens in the same environment and belong to the same subjects. They show the trace of simplified tokens on the same plot with the unsimplified and control SSANOVA curves.

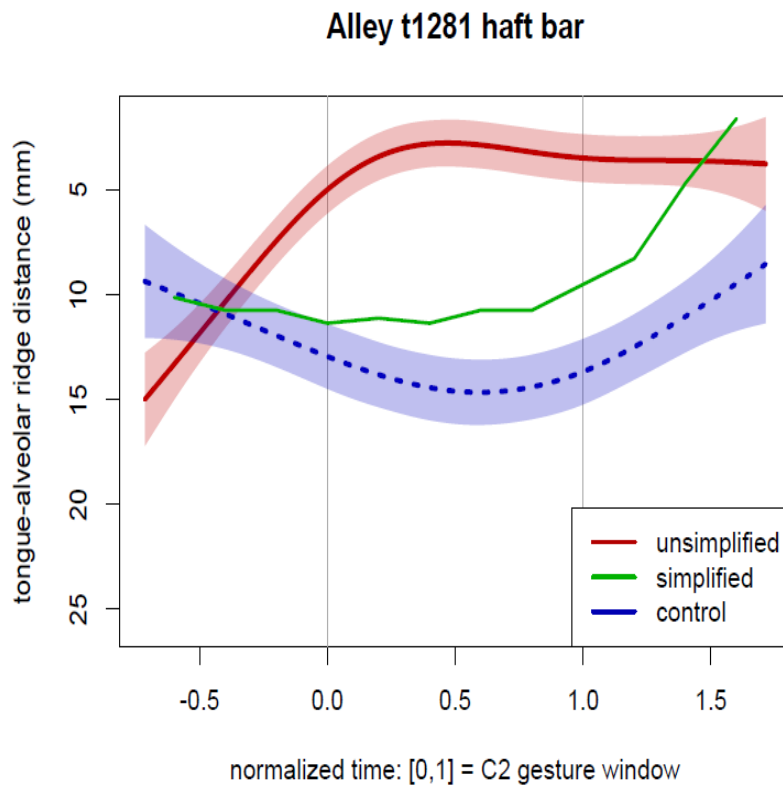


Figure 5.28 Simplified token ([hæft bɑr] “seven times”) for subject #2

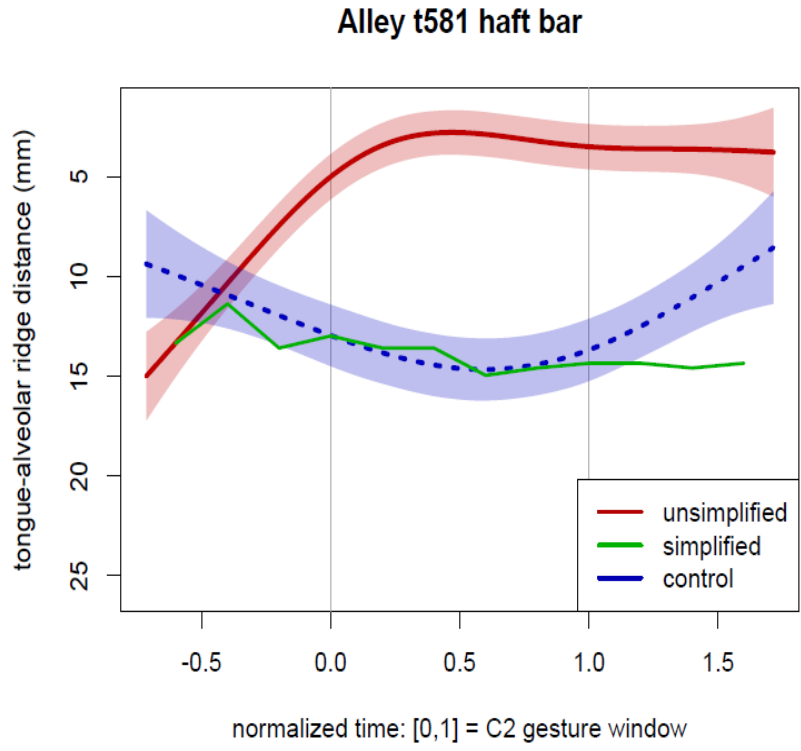


Figure 5.29 Simplified token ([hæft bar] “seven times”) for subject #2

### Alley t121 naft baraaye

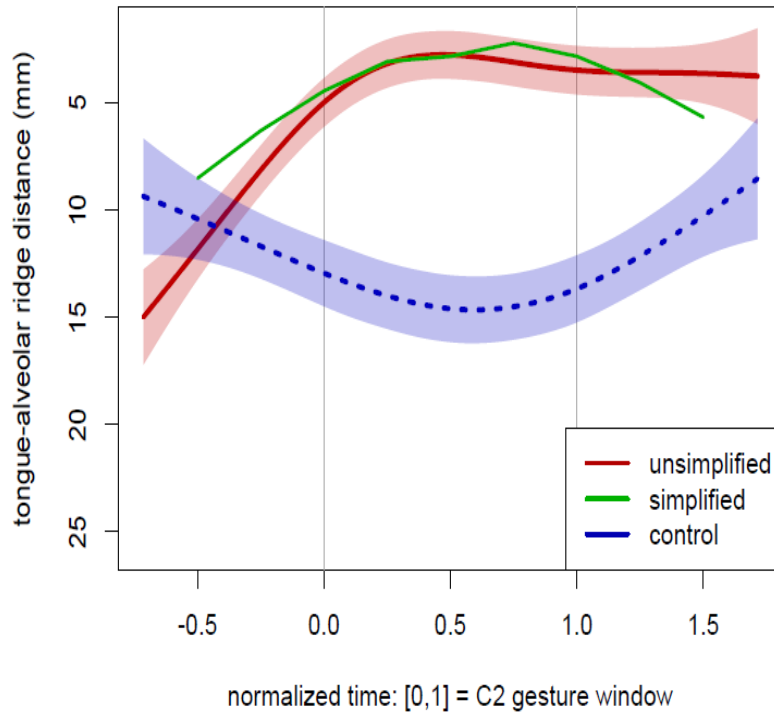


Figure 5.30 Simplified token ([næft bæraɟ-e] “oil for”) for subject #2

### Roshi2 t1071 loxt kardan

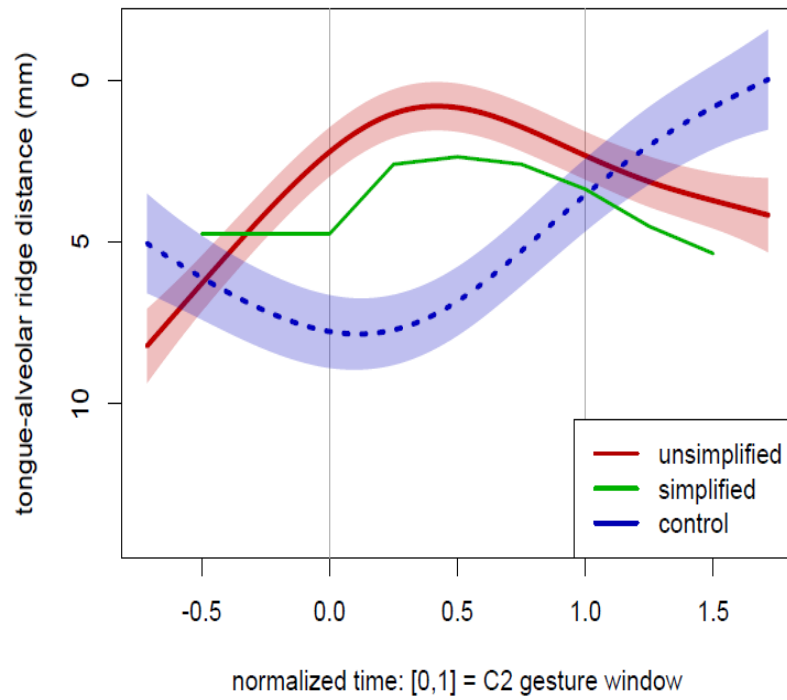


Figure 5.31 Simplified token ([loxt kærdæn] “to rob”) for subject #4

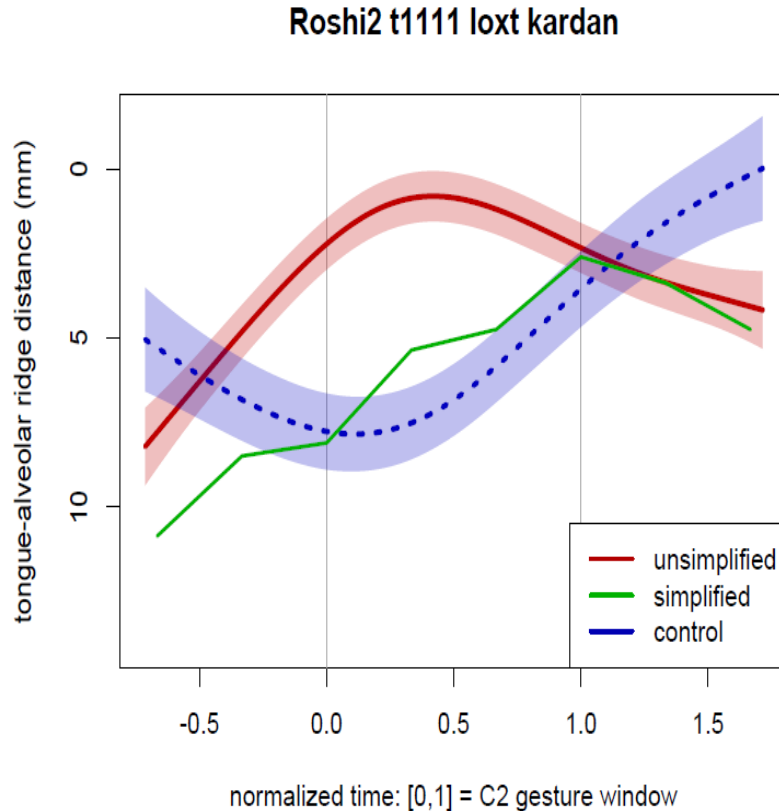


Figure 5.32 Simplified token ([loxt kærdæn] “to rob”) for subject #4

Figures 5.28 -5.29 above show the simplified token ([hæft bar] “seven times”) for subject #2. The trace of the simplified token in the C<sub>2</sub> gesture window in the first figure falls in the middle whereas in Figure 5.29 it is similar to the controls. The variation across simplified tokens is illustrated further through Figure 5.30 for the same subject where the trace of the simplified token ([næft bærqj-e] “oil for”) in the same context is similar to unsimplified tokens. Figures 5.31 and 5.32 show the traces of the simplified token ([loxt kærdæn] “to rob”) for subject #4. Figure 5.31 shows that the trace of the simplified token is between the two conditions whereas Figure 5.32 displays it as a token similar to control. In

general, the simplified tokens are shown to have variation along the tongue-alveolar ridge distance axis.

The distribution of the simplified tokens along the y axis in this way could have different interpretations. One scenario could be that all simplified tokens are intermediate and fall between unsimplified and control conditions. The other scenario is that some of the simplified tokens are the same as unsimplified and the others are the same as controls. In order to test these two scenarios, all simplified tokens of each subject are plotted individually. In order to get a better picture of their distribution, the traces of the individual simplified tokens are superimposed on the unsimplified and control SSANOVA curves. Figures 5.33 -5.37 below present the tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves. In these figures, we need to notice that the 95% confidence intervals are strictly in reference to the means of the categories, and that the tokens in all categories vary widely from the mean. Ordinarily, two categories are considered to be significantly different at x-axis values where there is no overlap between their confidence intervals. For the purpose of classifying individual simplified tokens with respect to the unsimplified and control categories, a reasonable approach is to make a one-tailed comparison with each of the other two categories (using the lower/more distant side of the confidence interval for the simplified category and the upper/less distant side of the confidence interval for the control category). For example, Figure 5.33 shows several simplified tokens that are distinct from both unsimplified and control categories at time 0.5 (overlapping neither confidence interval) and several tokens that are the same as controls (overlapping the control confidence interval or exceeding it). Figure 5.34 shows four tokens that are the same as the unsimplified category at least somewhere along the C<sub>2</sub> gesture

interval [0,1]. Figures 5.33 - 5.37 show the plots for the five subjects who produced enough simplified tokens to analyze.

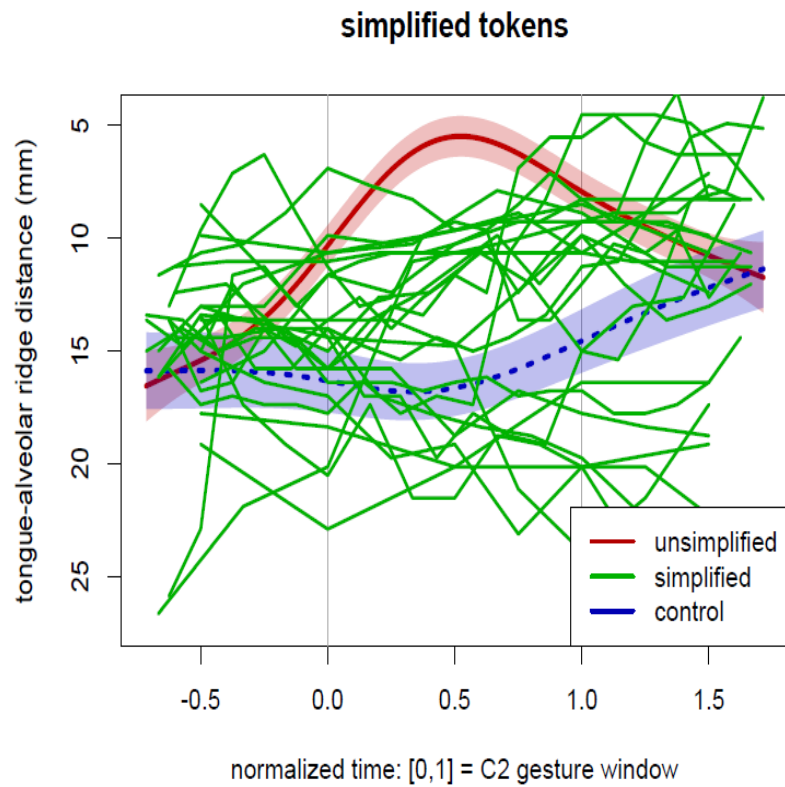


Figure 5.33 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves for subject #1

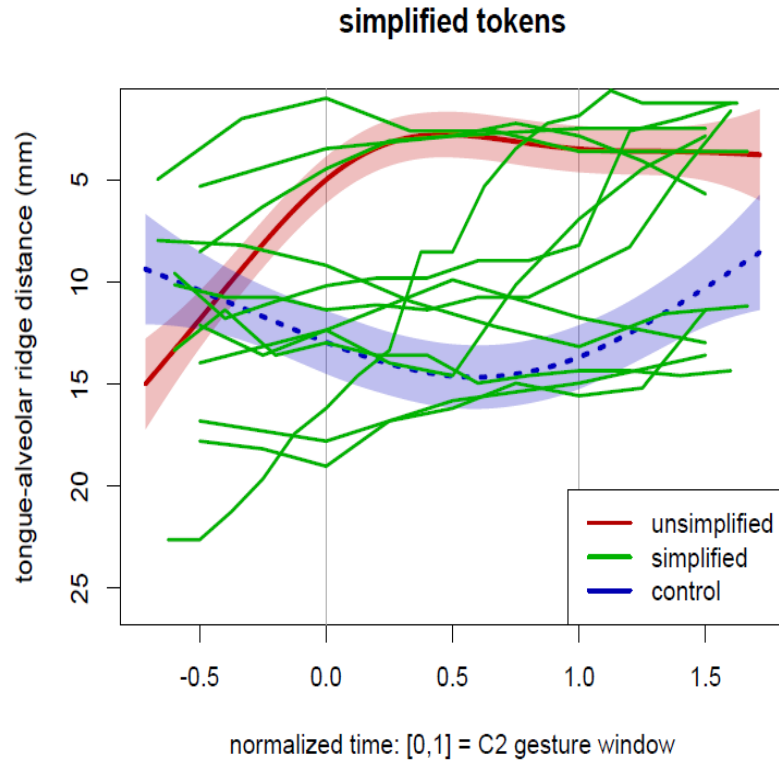


Figure 5.34 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves for subject #2

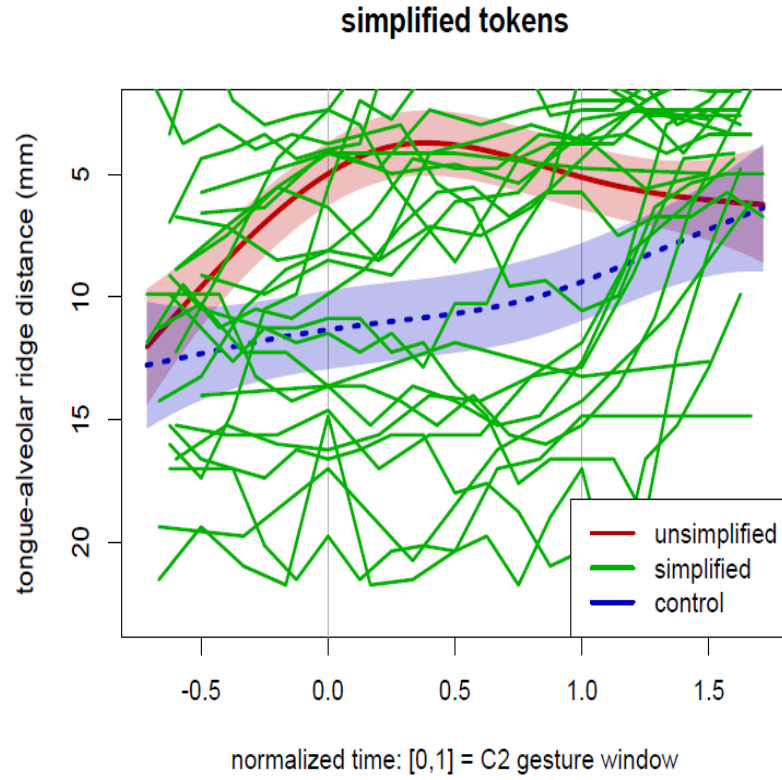


Figure 5.35 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves for subject #3

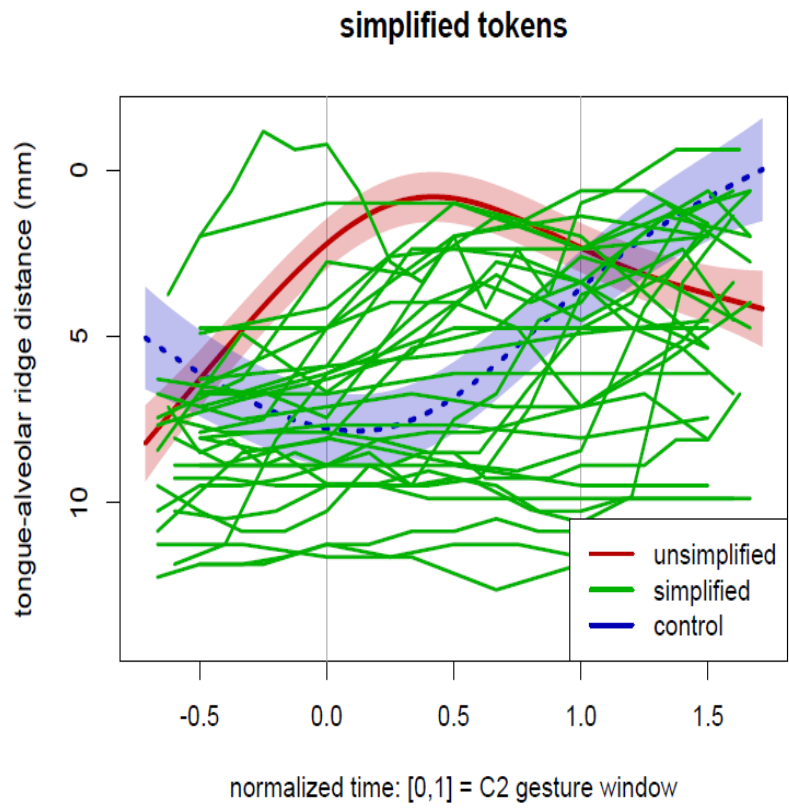


Figure 5.36 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves for subject #4

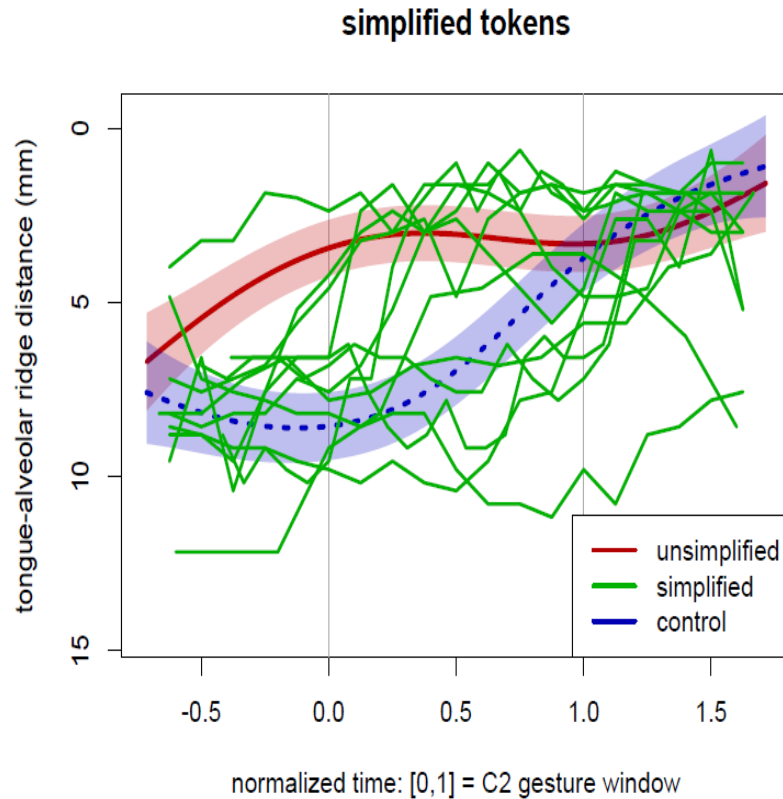


Figure 5.37 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves for subject #5

As Figures 5.33 - 5.37 above show, the traces of the simplified tokens for all of the subjects in the  $C_2$  gesture window vary widely in the height of the tongue blade. The two scenarios proposed above, one considering an intermediate position for the simplified tokens and the other taking some as unsimplified and the others as control cannot explain the distribution of the simplified tokens as presented above independently. However, combining the two scenarios together gives us a comprehensive picture of the distribution of the simplified tokens. The traces of the simplified tokens presented in the five figures above showed that some of the simplified tokens are similar to controls, some are similar to unsimplified, and there are some simplified tokens which fall between unsimplified and control conditions. Examining the simplified tokens more closely, one would notice that the

magnitude of reduction across subjects varies. For example the simplified tokens for subjects #2 and #4 as illustrated in Figures 5.34 and 5.36 fall mainly at the lower part of the plot whereas these tokens for subject #1, as shown in Figure 5.33, mainly cluster in the middle part. Moreover, one may argue that the distribution of the simplified tokens as illustrated and described above could be an artifact of the surrounding segments. In fact, different studies have shown the strong effects of the neighboring segments on their adjacent consonant. These studies have shown that the articulatory gestures of a segment could be affected by its preceding and following segments. In order to control this effect, all the simplified tokens from all subjects were pooled together and plotted based on the preceding and following segments. This will show whether the distribution of individual simplified tokens as shown above is an internal property of them or the effect of the context. Breaking down the tokens by subject and context simultaneously will result in a small number of tokens in each category. This means that one needs to take one or the other method. In this study, my approach is to try both and compare the results.

The preceding segments in this study were Lab and Dor consonants and the following segments were Lab, Dor, and V/#. Similar to Figures 5.33 - 5.37 above, the traces of the individual simplified tokens are superimposed on the unsimplified and control SSANOVA curves, but each figure includes only the subset of tokens that match a particular segmental context, and tokens are pooled across subjects. Since the figures include data from multiple subjects, the y-axis shows standardized tongue-alveolar ridge distances (standard deviations from the subject's overall mean tongue-alveolar ridge distance).

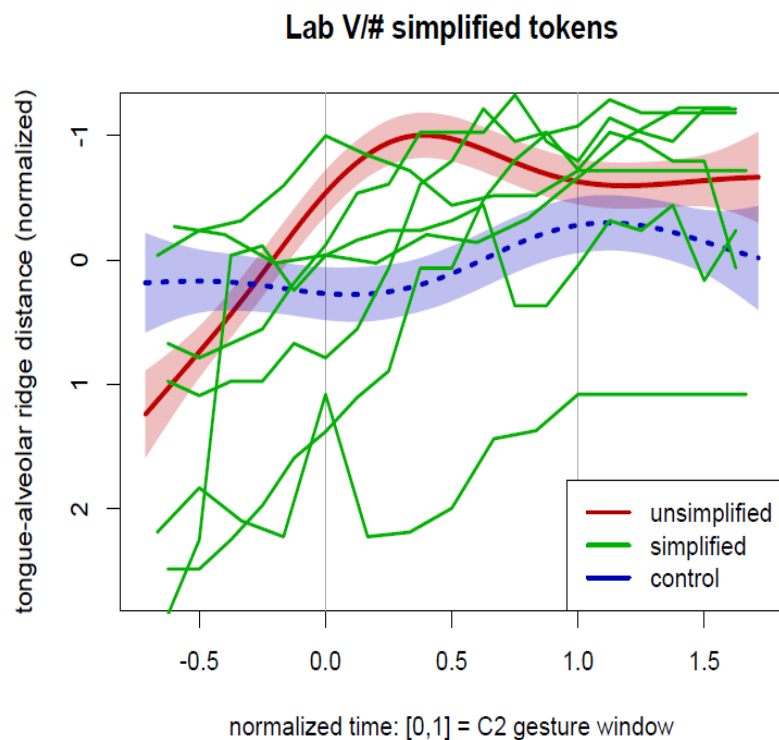


Figure 5.38 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves in Lab-V/# context

Figure 5.38 demonstrates the traces of all the simplified tokens superimposed on the unsimplified and control SSANOVA curves in Lab-V/# context. As this figure shows, the simplified tokens vary widely in tongue blade height in the  $C_2$  gesture window. Some of the simplified tokens pattern similarly to unsimplified and some have a behavior similar to controls. There are some simplified tokens which are intermediate and fall between the two conditions. According to this figure, the vocalic segments affect the lingual gestures less than labial consonants. This is shown through having more convergence between the two conditions in the left environment (i.e., Lab) versus the right environment (i.e., V/#).

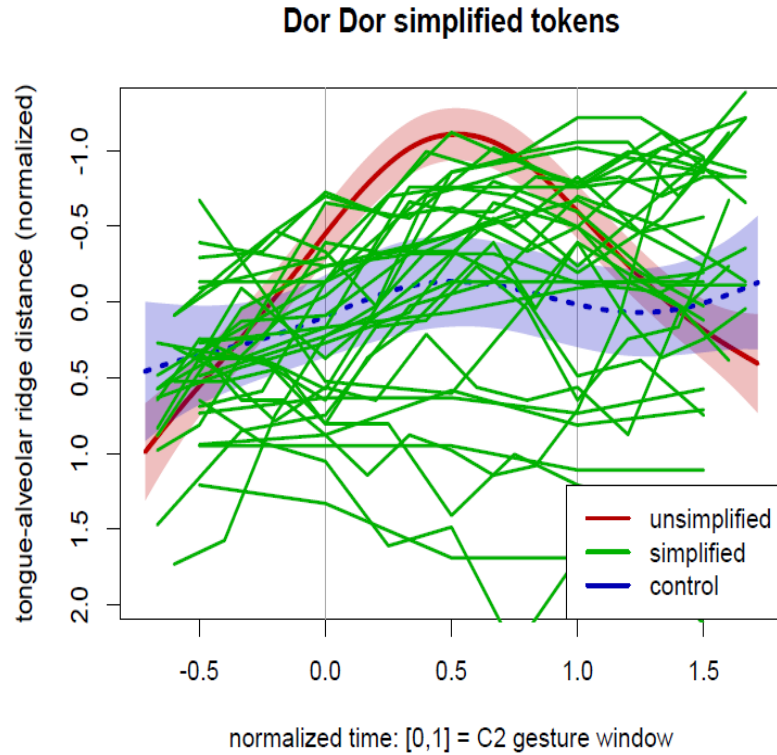


Figure 5.39 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves in Dor-Dor context

Figure 5.39 demonstrates the traces of all simplified tokens in Dor-Dor as well as the unsimplified and control SSANOVA contours. As the  $C_2$  gesture window shows, it is clear that some of the simplified tokens are similar to controls while some are similar to the unsimplified tokens. This figure also shows that preceding and following dorsal consonants affect the lingual gestures to a similar extent.

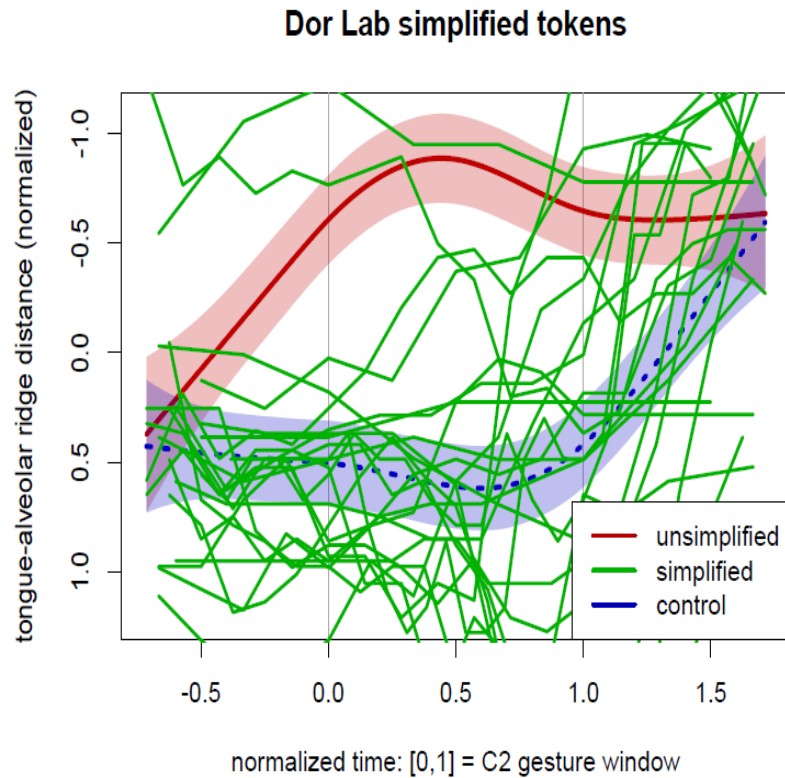


Figure 5.40 Tongue traces of the simplified token superimposed on unsimplified and control SSANOVA curves in Dor-Lab context

Figure 5.40 above shows the traces of all the simplified tokens superimposed on the unsimplified and control SSANOVA curves in Dor-Lab context. Similar to the results in the previous figures, the simplified tokens show three distinct patterns: some of them are similar to unsimplified, some are similar to controls, and the rest fall in between. Most of the simplified tokens shown in the  $C_2$  gesture window are similar to controls in this context.

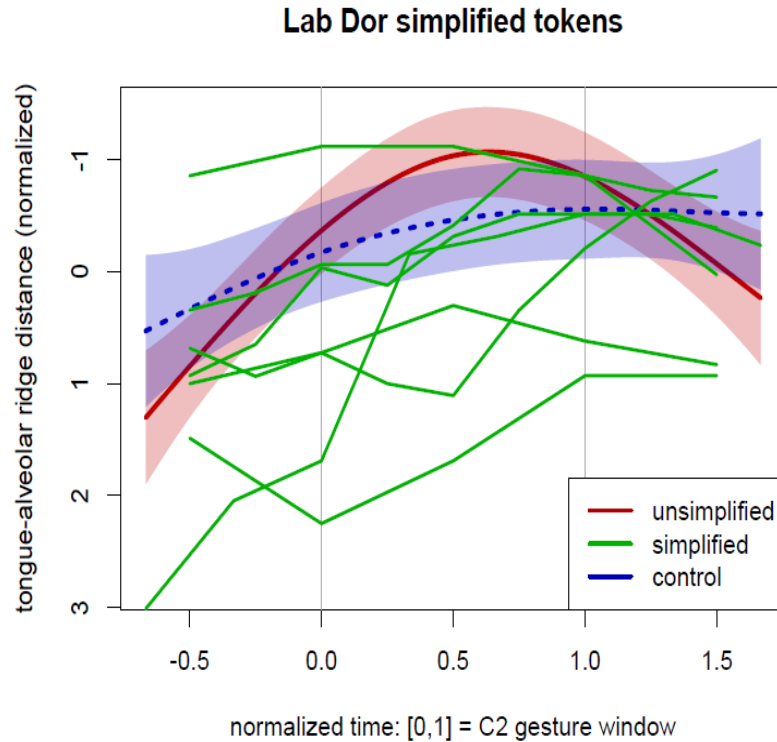


Figure 5.41 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves in Lab-Dor context

Figure 5.41 presents the traces of the simplified tokens along with unsimplified and control SSANOVA curves in Lab-Dor context. As it is shown in the  $C_2$  gesture window of the figure, the simplified tokens still tend to vary widely in tongue blade height; however, they mainly pattern similar to the controls in this context.

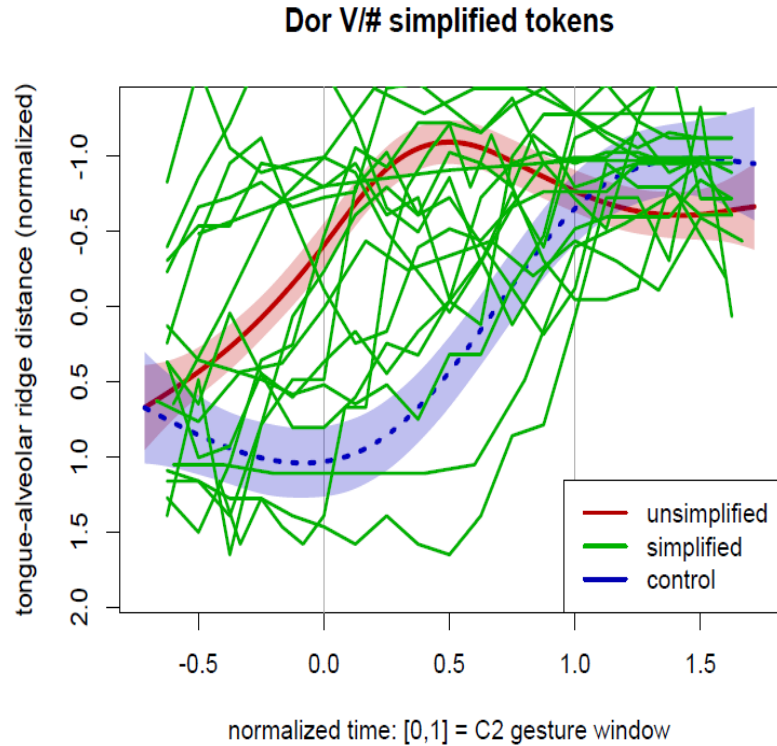


Figure 5.42 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves in Dor-V/# context

Figure 5.42 illustrates the traces of the simplified tokens superimposed on the unsimplified and control to SSANOVA contours in Dor-V/# context. Similar to the results of the previous figures, the simplified tokens vary widely in tongue blade height in the  $C_2$  gesture window. This means that some of them have a position similar to control tokens and others are similar to unsimplified tokens. There are some other simplified tokens which are in between and can be considered as intermediate. This context is similar to Lab-V/# in having vocalic segments to the right of coronal segment. In both of these two contexts, the simplified tokens tend to cluster around the unsimplified SSANOVA contour. This means that this context favors simplified tokens with full gestures; however, the general tendency of the simplified tokens to spread along the vertical axis still exists.

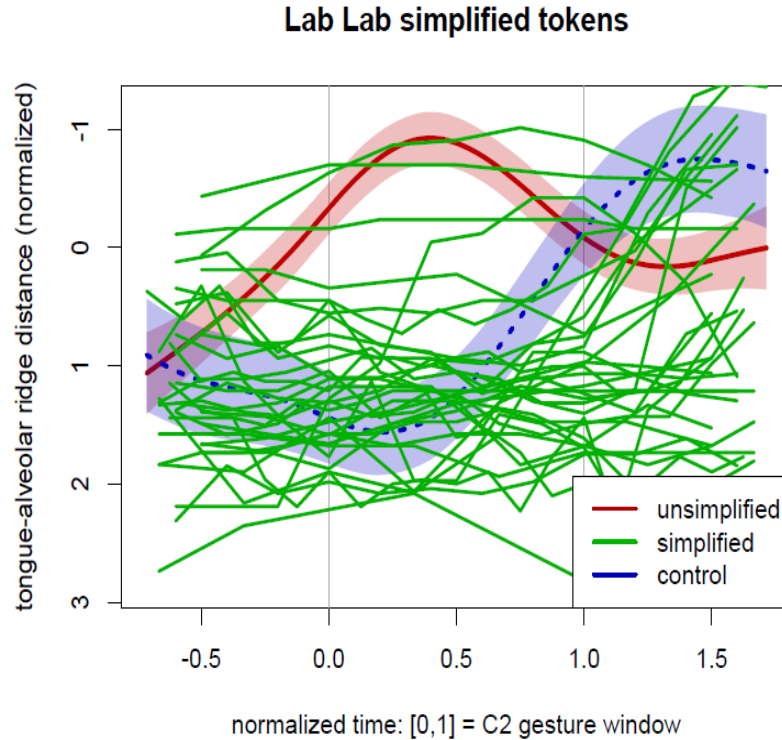


Figure 5.43 Tongue traces of the simplified tokens superimposed on unsimplified and control SSANOVA curves in Lab-Lab context

Figure 5.43 shows the traces of the simplified tokens in Lab-Lab context. As seen in the figure, a number of these tokens show similar behavior as controls and some other similar to unsimplified tokens in the  $C_2$  gesture window. Another group of the simplified tokens fall in the mid position where they show gestural position which is different from other two groups. Similar to the results presented for Dor-Lab context, Figure 5.43 shows that the simplified tokens are reduced to different degrees but they, similar to controls, mainly pattern at the lower part of the plot.

In general, Figures 5.38 - 5.43 showed that the appearance of intermediate tokens is not simply an artifact of pooling across contexts. The patterns shown in these figures are similar to the ones in Figures 5.33 - 5.37 for all subjects where simplified tokens scattered between the control and unsimplified categories. These results showed that the gradient

nature of cluster simplification in the individual simplified tokens as shown through Figures 5.33 - 5.37 for each subject did not change as they were classified based on preceding and following contexts. Different contexts could have varying effects on the tongue position; however, this did not change the gradient patterns attested for the individual simplified tokens.

In order to explore this issue further, a scatter plot was made based on the degree of gesture size (1= full gesture and 0 = no gesture). For every simplified token, every frame during the C<sub>2</sub> gesture window was given a score from 0 to 1, where 0 means the same as control and 1 means the same as unsimplified. For measurements falling in between the confidence intervals for control and simplified, the score indicates the percentage, with 0.5 being halfway between the two distributions. Scores for tokens were calculated by averaging across the scores for the frames. Thus, a token with a score of 0 has no sign of a tongue blade gesture anywhere in the C<sub>2</sub> gesture window, and a token with a score of 1 is indistinguishable from unsimplified, throughout the C<sub>2</sub> gesture window. An intermediate score indicates the degree to which the token matches either category, with 0.5 being halfway in between. The x axis in this plot represents the percentage of similarity between the simplified [t] tokens and the unsimplified ones within speakers and the y axis represents the same information within context. Figure 5.44 below shows the scatter plot which compares the similarity of simplified tokens to unsimplified tokens from the same speaker versus tokens in the same context.

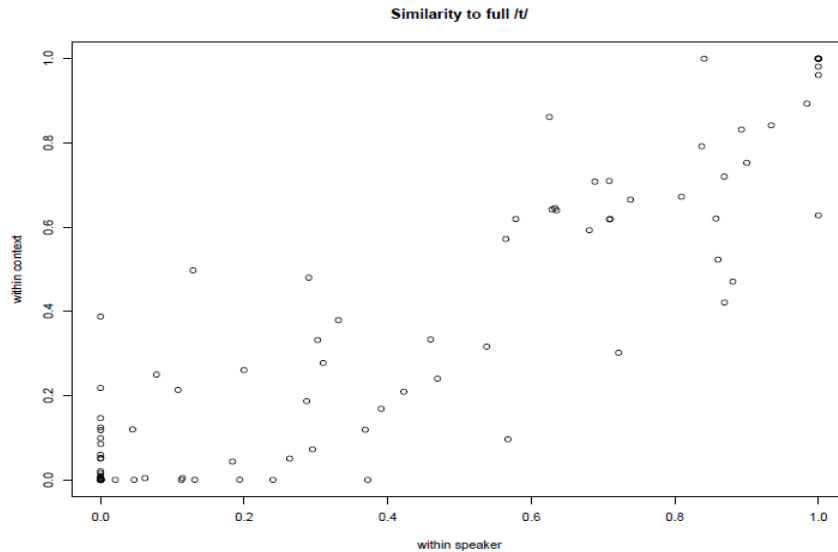


Figure 5.44 Scatter plot of simplified tokens within speaker and context

The scatter plot in Figure 5.44 shows a positive correlation between the simplified tokens within subject and context, meaning that the results are stable across the two techniques, i.e., whether tokens are combined by subject or by context. Most of the tokens in this figure are piled on top of each other at the point (0,0), and the next most frequent one which is (1,1). A linear regression was run next to quantify how the two variables are correlated. The results of the test showed the estimated coefficient of 0.847, R-squared = 0.849 at  $p < 0.00001$  level.

In order to examine how the simplified tokens within subject and within context would pattern independent of each other, the degree of similarity for the simplified tokens to unsimplified tokens (in percentages) within subjects and within context on x axis and the duration of clusters (as measured from the spectrogram) on y axis was plotted. Figures 5.45 and 5.46 below illustrate the distribution of the simplified tokens in these two situations.

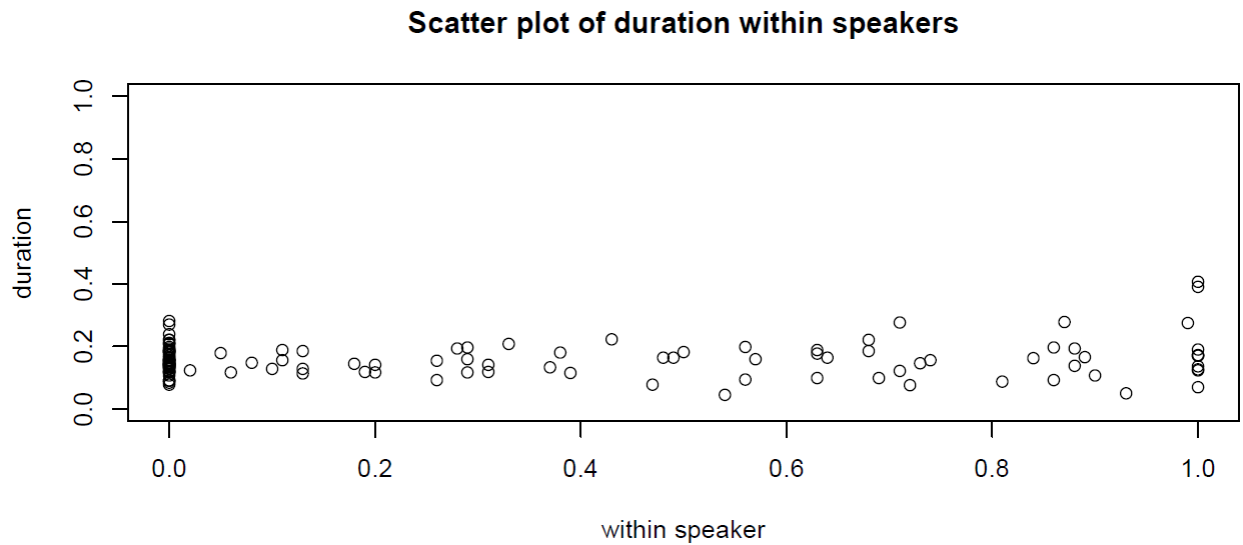


Figure 5.45 Scatter plot of simplified tokens within speakers

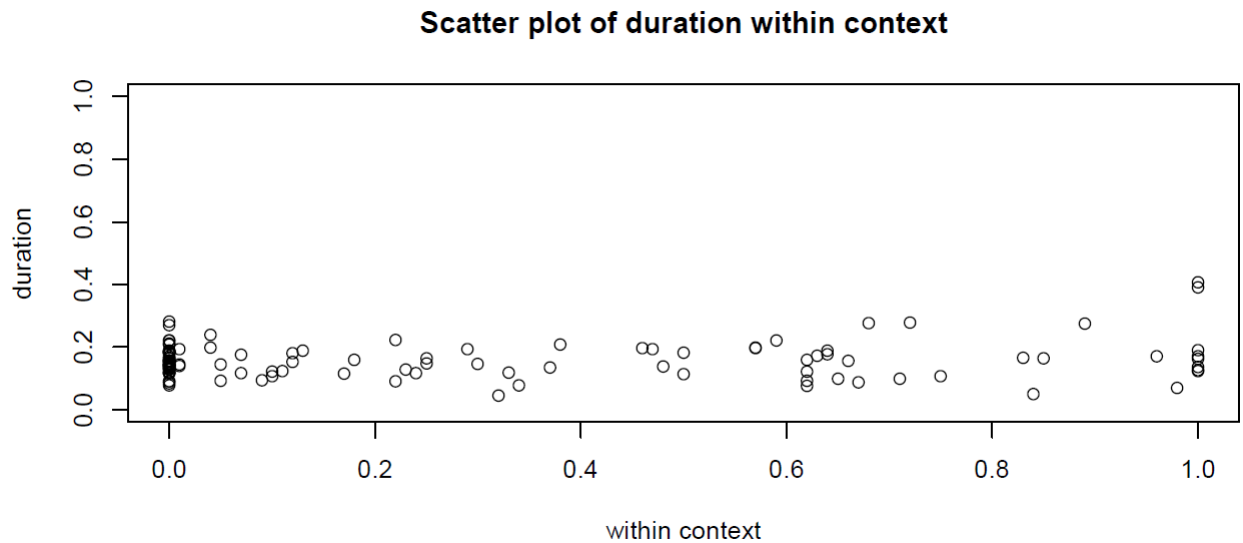


Figure 5.46 Scatter plot of simplified tokens within context

The x axis in Figures 5.45 and 5.46 shows the percentage of the similarity for the simplified tokens to unsimplified tokens and the y axis shows the cluster duration. Figure 5.45 shows the simplified tokens within subject and Figure 5.46 shows the same tokens

within context. The two figures show very similar patterns for the simplified tokens across subject and context. This means that the variation shown for the simplified tokens in this study cannot be attributed to the context.

The general results based on SSANOVA plots showed that the simplified tokens can be classified into distinct groups based on their gestural position and whether they fall above the lower bound of the unsimplified confidence interval (i.e., full [t] gesture), below the upper bound of the control confidence interval (i.e., no [t] gesture) or in between (i.e., partial [t] gesture).

Table 5.1 below reports the number of tokens for all three categories in the simplified condition in addition to unsimplified condition for each subject. This table is based on individual checking of all the simplified and unsimplified tokens from all subjects. The information for the two subjects (i.e., subjects #6 & #7) who had low rate of simplification is also presented.

	Unsimplified	Full [t] gesture	Partial [t] gesture	No [t] gesture
Subject #1	17	9	7	12
Subject #2	9	2	8	2
Subject #3	9	8	6	10
Subject #4	11	3	13	15
Subject #5	4	4	3	4
Subject #6	12	0	1	3
Subject #7	14	0	2	9

Table 5.1 Number of tokens in all simplified and unsimplified categories across subjects

The questions which still remain are how the simplified tokens with full [t] gesture, are different from unsimplified tokens, and how the other two categories of the simplified tokens relate to each other and other categories. The prediction is that the simplified tokens with full [t] gestures should have short interval between C<sub>1</sub> and C<sub>3</sub> landmarks, because the release of the [t] gesture is apparently obscured. The other two categories of the simplified tokens could have short or long intervals since they don't have complete [t] gestures, and there is therefore no chance of an audible [t] release. In order to address this issue, the time interval between the end of plateau for C<sub>1</sub> [pe] and the start of plateau [ps] for C<sub>3</sub> (in the cluster of C<sub>1</sub>C<sub>2</sub>C<sub>3</sub>, the middle consonant is a coronal stop) in all three categories of the simplified tokens is calculated. As the point of reference, the same measurement is also made for the unsimplified and control tokens.

Figures 5.47 - 5.51 below illustrate box plots which show the time intervals between the designated landmarks for  $C_1$  and  $C_3$  for the three categories in the simplified condition along with unsimplified and controls as reference points for each subject.

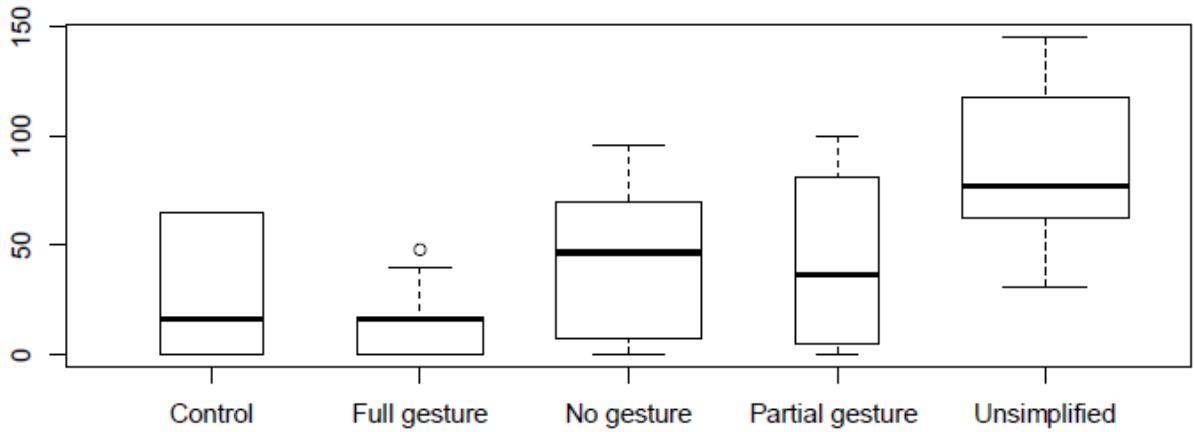


Figure 5.47 Time intervals between [pe] and [ps] landmarks across five categories for subject #1

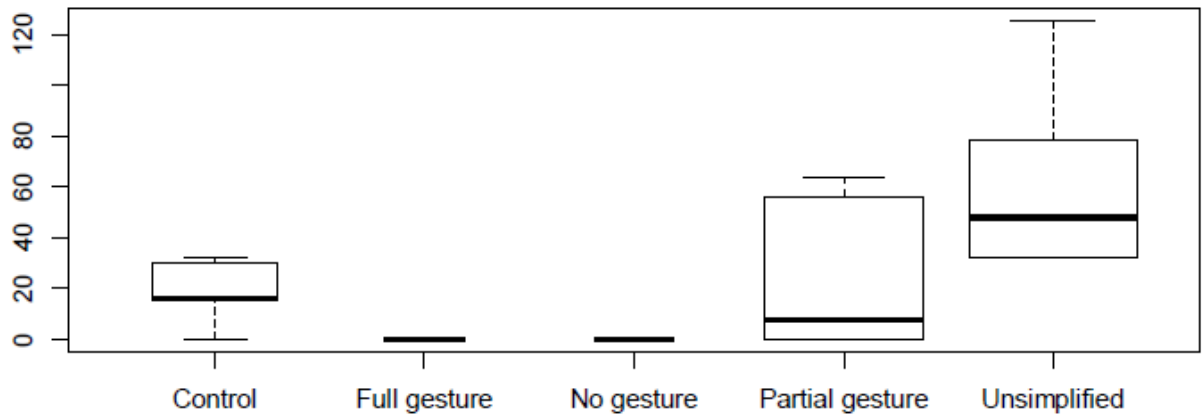


Figure 5.48 Time intervals between [pe] and [ps] landmarks across five categories for subject #2



Figure 5.49 Time intervals between [pe] and [ps] landmarks across five categories for subject #3

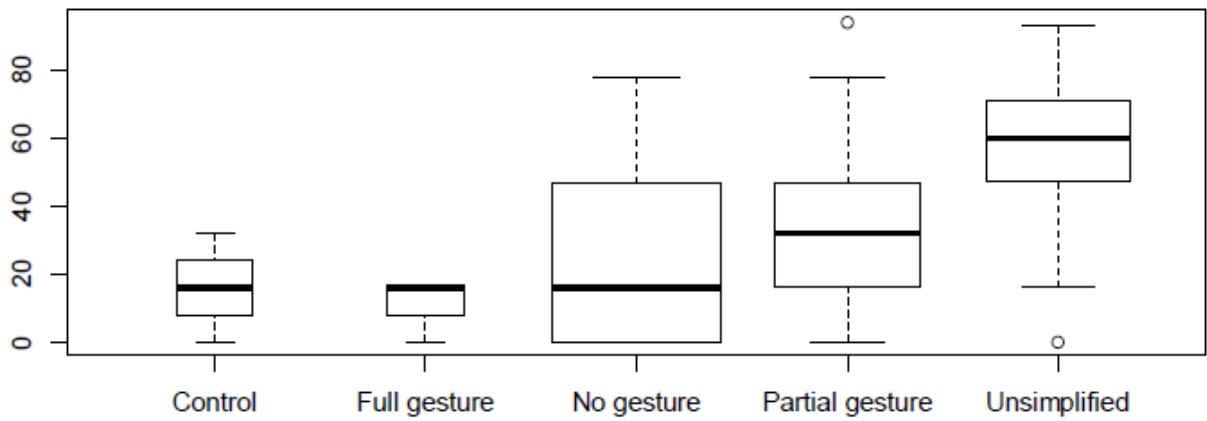


Figure 5.50 Time intervals between [pe] and [ps] landmarks across five categories for subject #4

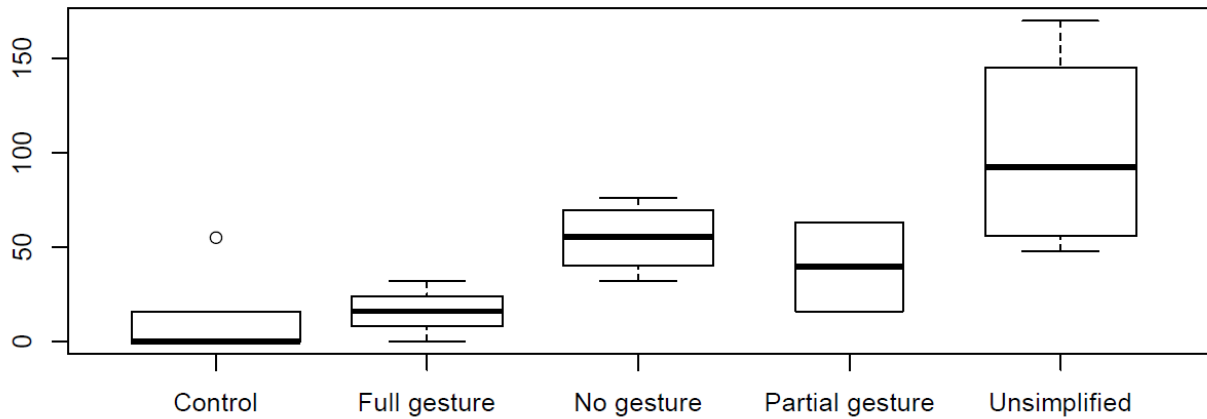


Figure 5.51 Time intervals between [pe] and [ps] landmarks across five categories for subject #5

The results of the five categories presented above showed that the simplified tokens with full [t] gesture have very short time intervals. This is very similar to the results for control tokens. The tokens in the unsimplified condition show the longest time intervals between the two landmarks. The other two categories of the simplified condition, namely partial [t] gesture and no [t] gesture, show time intervals which are longer than full [t] gesture category and shorter than unsimplified category. The following section presents the discussion related to articulatory results.

### 5.3 Discussion for Articulatory Results

In general, the results of this chapter showed that the deletion of coronal stops in Persian is a gradient process. In addition to this, the traces of the simplified tokens showed that they make a mixture of tokens that are completely deleted, unreduced but overlapped, and in between. The articulatory results in this chapter showed that speakers can produce a simplified [t] in three distinct ways: (1) by temporal reduction (obscuring the release of the

[t] gesture by reducing the interval between C<sub>1</sub> and C<sub>3</sub>) or (2) by spatial reduction (reducing the magnitude of the gesture, so that there is no release), or (3) by producing no [t] gesture at all. Tokens produced with a full gesture and a long interval between C<sub>1</sub> and C<sub>3</sub> would have an audible [t] release, and be classified as unsimplified. Based on the results, different subjects could show different patterns for simplification. This could be due to different reasons. It is also possible that participants of the study are reacting differently to the experiment setting, and speaking more carefully. The similar pattern for simplified and control contours for some subjects is an indication of phonologization. The next chapter will consider how these different patterns of simplification fit into the phonologization process. In general, the results of this chapter showed that articulatory data should be an integral part of phonetic studies. Without this information, there have been no reasons even to suspect that deletion was gradient. The following chapter presents the general discussion and conclusions of the thesis. A section devoted to suggestions for further research will conclude this chapter.

## Chapter 6

### General Discussion and Conclusion

This dissertation investigated the deletion of the word-final coronal stops in Persian. This topic is important for different reasons. First, it is not clear whether this process is gradient or categorical. There hasn't been any experimental work in Persian to show whether the application of this process is complete or incomplete. Second, the optionality nature of this process has not received enough attention in the academic sphere yet. We still don't know much about why the same speaker may or may not simplify the same cluster. A better understanding about optionally deleted clusters could shed light on the distinctions between phonetics and phonology. Next, using different types of data in this study (i.e., acoustic and articulatory) created an environment where their relative contribution to the major findings in the study was examined. In this thesis, acoustic data were examined first to see how reliably they could attest the presence of coronal gestures in the simplified clusters. The articulation data proved to be quite significant in revealing some facts about this process and addressing the main research questions in this thesis. The current study also created an optimal context where different sound change models could be evaluated.

#### 6.1 Overall Discussion on Research Questions

According to Articulatory Phonology, variation in gestural magnitude and gestural overlap (in both time and space) can account for many phonological processes across languages. In this approach, what sounds like deletion may sometimes really be gestural overlap or reduction. The articulatory results of this study showed that the gestures of the deleted segments are often still present. For clusters that sounded simplified, some truly had

no [t] gesture, some had reduced gestures, and some had gestural overlap. This supports the main arguments presented in Articulatory Phonology regarding the underlying reasons for sound patterns and sound change. The dynamic nature of gestures allows them to be active over a period of time. This will create the environment where the final stage of a gesture related to a segment could overlap with the initial stage of the next gesture. This happens in our everyday speech and under normal circumstances. The magnitude of reduction or the amount of overlap likely varies when the speaking rate or speech register changes. Some studies have shown that there is more overlap among adjacent gestures in fast/casual speech. The various reactions of the participants to the experimental settings could be also another factor. As gestures shorten, their likelihood of being completely overlapped and hence acoustically hidden increases.

The gradient nature of coronal stop deletion was shown both through SSANOVA curves and the traces of individual simplified tokens in this study. The results of the simplified tokens presented through Figures 5.33 - 5.37 showed a gradient reduction of coronal stop gestures for each subject. The individual simplified tokens for all subjects showed that they span a wide range of tongue-alveolar ridge distances. However, the SSANOVA results presented in Figures 5.23 - 5.27 showed that subjects exhibit different rates of cluster simplification and different rates of complete versus incomplete deletion. For example, the position of simplified and control contours in subjects #3 and #4 presented in Figures 5.25 and 5.26 was similar. This means that a greater number of simplified tokens for these two subjects are truly deleted. The results for subjects #1 and #2 presented in Figures 5.23 and 5.24 showed a large number of simplified tokens with partial gestures. The contours of the simplified tokens for these subjects fell between those of unsimplified and controls in

the C<sub>2</sub> gesture window. The pattern for these subjects was different from subjects #3 and #4. The contour related to simplified tokens for subject #5 presented in Figure 5.27 was only partially similar to controls. This means that they had a pattern which was different from the other subjects and was in between subjects #1/#2 and subjects #3 / #4. The seven subjects in the study showed different range of simplified tokens (i.e., full [t] gesture, partial [t] gesture, and no [t] gesture), as seen in Table 5.1. The fact that there is a continuum for the observed simplified clusters conforms to the findings of the studies on the residual gestures found for other phonological processes (e.g., Ohala 1992; Holst & Nolan 1995). The findings that the same process (i.e., cluster simplification) involved full, partial and zero degrees of gestural reduction could be attributed to different sources. According to Articulatory Phonology (as it was discussed in detail earlier in Chapter 1), the phonological processes in fast/ casual speech happen because of either gestural overlap or gestural reduction. The speakers may intend to facilitate their fluency and speed which result in a higher gestural overlap and reduction. The different magnitudes of reduction and overlap result in producing a continuum of coronal stops with a wide range of residual gestures, as shown in Table 5.1. The SSANOVA graphs as well as the time interval figures presented in Chapter 5 showed the intermediate stage of the simplified tokens with partial and full [t] gestures. The intermediate status of these tokens was established through articulatory and acoustic results. The tokens with partial [t] gesture could be explained to be the result of decrease in gestural magnitude and the ones with full [t] gesture result from increasing gestural overlap in time. In other words, the former is being changed spatially and the latter temporally. The fact that

two of the subjects in the study (i.e., subjects # 6 and #7) did not produce any simplified tokens with full [t] gesture supports the independent status of these two processes.

Following the same line of argument for the simplified tokens with full [t] gesture, one may expect the third category of simplified tokens (i.e., simplified tokens with no [t] gesture) to result from increase in gestural overlap in time. If this argument is valid, we should expect this category to be shorter than the one with full [t] gesture. The results of the time intervals between [pe] and [ps] landmarks across five categories presented through Figures 5.47 - 5.51 do not show that the simplified tokens with no [t] gesture are shorter. The equal or longer duration of this category compared to the one with full [t] gesture does not support the idea of gestural overlapping for this category.

The explanation offered by Articulatory Phonology regarding gestural overlap and reduction for casual speech phonological processes fails to account convincingly for the simplified tokens with no [t] gesture. According to this approach, the reduction and overlapping of gestures for the simplified tokens with full [t] gesture and partial [t] gesture could correspond to the mechanical and physiological factors. The speakers neither intend to reduce nor plan to overlap gestures. These are the natural by-products of human's articulation system when engaged in a fast/casual conversation. In order to reach such a goal, no phonological rule or operation is needed or activated in the cognitive system. This makes the tokens with no [t] gesture distinct from the other two categories in the simplified tokens. According to this analysis, the three categories of the simplified tokens involve two independent strata: gradient gestural reduction or overlapping which are the results of low-

level phonetic mechanisms versus the categorical deletion caused by cognitive system. The former is not planned whereas the latter is intended and therefore controlled. This classification is similar to the one proposed for sibilant assimilation in English (Holst & Nolan 1995). The existence of these two separate mechanisms and the gradient nature of overlap and/or reduction provide the context for variation in the speech production.

The production process described above along with the perceptual reinterpretation results in sound change. Ohala (1990) has emphasized that analyzing phonological processes in terms of articulatory gestures does not rule out a perceptual/acoustic component to these processes. Any change in gestures or their magnitude or timing will result in perceptual/acoustic changes. In order to explain the sound change process, we could posit that the articulatorily reduced or overlapped tokens could be heard accurately and categorized into a phonological group with zero phonetic cues by some listeners. The simplified tokens with no [t] gesture could be the result of speakers changing their representations based on other people's gestural overlap. Over time, the existence of tokens with no [t] gesture suggests that subjects have phonologized this process to some degree. The lack of simplified tokens with full [t] gesture and the minimal number of tokens with partial [t] gesture by subjects #6 and #7 suggests that they either do not simplify clusters or do it completely. It seems that for them simplification happens only as a fully deleted process should they apply this phonological process at all.

This approach to sound change assumes that the gradient reduction of articulatory gestures does not necessarily map directly to the auditory system. This means that the partial coronal gestures in some simplified tokens were not perceived in the auditory system based

on a gradient scale. The result of more than 90 % agreement reached on categorizing the tokens of this study as unsimplified or simplified by two trained phoneticians and the researcher of this study supports this idea. Despite of the fact that the simplified tokens had a wide range of residual gestures, the three raters categorized them similarly more than 90 % of the time. Moreover, the acoustic results of this study showed that the simplified and control tokens mainly shared similar acoustic cues. The listeners, who are exposed to the reduced tokens, may use a myriad of reduced tokens in their production intending to be zero when they assume the role of a speaker. These are heard as a zero segment by the listeners. Once this cycle gets repeated over time, it becomes a dominant pattern and a norm within the speech community. According to this, the sound change happens within a listener; however, it comes into existence when the listener assumes the role of a speaker. Putting that in another way, the speaker is mainly responsible for gradual increase in gestural overlap to the point that [t] is not heard; however, it is the listener who would abruptly interpret a word as having no [t].

The results of the coronal stop reduction and its subsequent total deletion presented here matches with the sound change model proposed by Blevins. According to the Choice category in her model, the listener perceives the “multiple phonetic signals” by the speaker accurately; however, selecting a prototype/best exemplar or assigning a phonological form to the set of variants which is different from that of the speaker will trigger the change. Putting this within the context of the current study, Figures 5.28 - 5.32 showed that the same speaker may reduce the coronal stops to different degrees. This could extend from partial [t] gesture, as in Figure 5.28, to full overlapped [t] gesture as in Figure 5.30. These simplified tokens are

all heard accurately by the listener. However, they may be categorized as fully reduced/deleted since the coronal stop acoustic cues for these reduced tokens are very weak. The results of this study showed that acoustic cues could not clearly distinguish the simplified tokens from controls. Therefore, these two categories could be classified as similar acoustic tokens and merge later on. A new experiment should be developed to target the perceptual aspect of the simplified tokens. This is further discussed in the next section where suggestions for future studies are provided.

The results of this study also showed that the preceding and following contexts to coronal stop affected the tongue position. This had an impact on the measures used to assess coronal stop gestures. The SSANOVA comparisons of unsimplified and control conditions showed that the vocalic context seemed to have the least effect on the tongue position whereas the consonantal context displayed a major impact on it. The SSANOVA comparisons of unsimplified and control conditions in Lab-V/# context in Figure 5.38 showed less effects of vocalic environment than consonants. The simplified tokens plotted based on the preceding and following contexts also showed a clear role of the consonant order. The simplified tokens in the contexts of Dor-Lab and Lab-Lab, as shown in Figures 5.40 and 5.43, tend to mainly cluster around controls. Some studies have shown that the front-to-back order for consonants shows more overlap compared to back-to-front (Byrd 1992; Wright 1996; Chitoran 1998). The patterns related to the simplified tokens in this study do not seem to strongly support this prediction.

The results of this study also showed that coronals readily simplify, if SSP is not violated.. The question will be why coronals show such a behavior. In order to answer this question, the articulatory characteristics of these segments could be taken into account. The velocity plots in this study, as shown in Figures 5.11, 5.14, and 5.17, depicting how fast

different parts of the tongue travel in the mouth, should be compared to see if the contours for tongue tip movements in general show sharper angles than those for tongue body. If the gestures for coronal segments are more rapid than dorsal segments, the greater tendency of coronal gestures for reduction could be partly explained. Other studies have previously shown the more rapid nature of coronal gestures (Browman & Goldstein 1990). Browman and Goldstein (1990, p. 362) state that because coronal gestures are very quick, they are the most likely gestures to be hidden articulatorily when overlapped by adjacent gestures. Greater gestural overlap leads into more acoustic muffling.

According to the findings of this study, stops are potentially good targets for deletion in the absence of SSP violation. This is in accordance with the findings of other studies which have emphasized the special status of stops compared to other consonants in phonological processes such as deletion (Côté 2000) and assimilation (Hura, Lindblom & Diehl. 1992). Why is it that stops tend to get deleted more than other consonants? The peculiar articulatory movements of stops, compared to other consonants, could shed some light on why they are a good potential target for deletion. Stops usually have three different phases of production which are: “obstruction for occlusion”, “holding the occlusion”, and “the release of the closure”. From these three stages in stop production, it is the application and release of the obstruction which demystifies the nature of place of articulation (Blumstein 1981; Dorman, Studdert-Kennedy, & Raphael 1977). Fricatives, on the contrary, are very distinct acoustically and auditorially throughout the whole stages of their production. This allows their place of articulation to be internally salient during the constriction, especially for sibilants. Any articulatory reduction of any sort does not distort the place of articulation identity for fricatives.

Under equal conditions, liquids are also expected to have better perceptual cues than stops. For example, it is well known that /l/ in English has two allophones. The light /l/ in onset and dark /l/ in coda positions. Research has shown that the timing between the gestures of the two articulators for /l/ (i.e., tip of the tongue raising and tongue dorsum retraction) determines the nature of these two allophones (Sproat & Fujimura 1993; Browman & Goldstein 1995; Gick, Campbell & Tamburri-Watt 2006). The reduction in either of these two gestures could change the gestural timing of these allophones, hence their perceptibility. However, similar to fricatives, the information regarding the place of articulations for liquids is available throughout the constriction period. Wright (1996) states that the strongest place of articulation cues for some of the laterals is distributed over an entire syllable. This is very different from the case in stops where the lack of release could lead into hiding the place of articulation identity. The lack of audible release cues for stops could trigger their misperception. The articulatory results of this study showed that for some of the simplified coronal stop clusters, the release of the [t] gesture was obscured by reducing the interval between C<sub>1</sub> and C<sub>3</sub>, and for some there was no actual release due to the spatial reduction.

Sound change could be partially due to the limitations of the human auditory system. Listeners do not pay attention to all specific points in an acoustic signal belonging to a segment. This could be due to either redundancy of those points or the resort of listeners to other informative sources to distinguish a segment. Many phonemes could relate to many acoustic cues signalling to that phoneme. The acoustic results of this study showed that both control and simplified tokens had similar acoustic cues. This suggests that these two categories will merge once the minimal distinctive acoustic cues between them fade away.

As for the role of position, there are two reasons which render consonant clusters in utterance-initial position perceptually different from their counterparts in final position. First, it is well established that onset positions play a major role in lexical access (Marslen-Wilson 1987). The less gestural overlap or reduction in the onset will boost their perceptibility level. Kohler (1990) explains that speakers are aware of the high signaling values of the syllable- or word-initial position and they respect it by providing more articulate production for listeners. Second, it is well known that formant transitions into or out of vowels are major acoustic cues for stops (Stevens & Blumstein 1981; Ohala 1992; Steriade 1997; Côté 2000). In the case of consonants in the onset position like # C<sub>1</sub>C<sub>2</sub>V, the second consonant (i.e., C<sub>2</sub>) is acoustically quite salient due to its release into the following vowel; however, C<sub>1</sub> lacks this property. In order for C<sub>1</sub> to be perceptually salient, its release should be audible. Less overlap between C<sub>1</sub> and C<sub>2</sub> serves this purpose. This is not the case for a consonant cluster in coda position (i.e., VC<sub>1</sub>C<sub>2</sub>#). The consonant after vowel gets its formant transitions out of vowel and the second consonant can be released in an utterance-final position. The fact that C<sub>1</sub> is perceptually more salient in this position compared to utterance-initial position gives it the chance of having more overlap.

If we merge the idea of less overlap in the onset position with the explanation for consonant cluster perceptual saliency presented above, a legitimate question could be born out. If a survey shows that the maximum number of consonants in the onset position across languages, in general, is less than the maximum number of consonants in the coda position, could it be explained by resorting to maximal modulations/recoverability as well as gestural overlap concepts? If we find that the maximum number of consonants in onset position is less than coda position across world languages, could it be due to the less gestural overlap allowed in the onset position? In fact, the literature on syllable structure and OT has shown

such a tendency cross-linguistically (Maddieson 1992; Kager 1999). As it was mentioned above, languages prefer an optimal amount of gestural overlap since non-optimal or lesser gestural overlap will lower the saliency of the acoustic signal in terms of its amplitude, frequency, spectral shape and so forth. I leave this as an open question. This requires searching the data bases to see the maximum number of consonants and the type of consonants in onset and coda positions. The magnitude of constriction for each consonant (narrow versus wide) as well as its gestural components and their corresponding acoustic information are the variables which need to be taken into account.

In conclusion, the results of this study showed that the acoustically-similar simplified tokens are composed of three articulatory distinct groups. The first group are the temporally reduced segments in which the release of the [t] gesture is obscured by the adjacent segments. The second group of the simplified tokens are the spatially reduced tokens which have no release. Finally, the last group of the simplified tokens are the ones which are produced by no [t] gesture at all. In order to account for the optional nature of this process, it was discussed that the first two groups of simplified tokens are the result of mechanical or physiological constraints. The production of these tokens is not controlled by the speaker. In contrast to these two groups, the last group is intended by the speaker and gets executed via a phonological rule or abstract cognitive representations. The deletion of coronal stop starts as a fully optional and gradient phonetic effect, triggered by some articulatory and perceptual factors, but it undergoes phonologization later on. This process could progress towards the obligatory and categorical end of the continuum, and eventually acquire full phonological status. It was shown that Blevins' CCC sound change model best describes the results of the

consonant cluster simplification in Persian. The last section of this thesis provides suggestions for further studies.

## **6.2 Directions and Suggestions for Future Research**

One of the interesting research goals to follow is to investigate the perception of the three categories of the simplified clusters which auditorily sounded similar but gesturally distinct. These categories were the simplified tokens with reduced [t] gestures, overlapped [t] gestures, and the ones with no [t] gestures. In general, the mainstream theories of speech perception either consider speaker's gestures as the object of perception or acoustic/auditory signals. In the gestural model, perceivers use structure in the acoustic signal as information about the speaker's articulations. According to this view, the object of perception is speaker's gestures. Beddor, McGowan, Boland, Coetzee, & Brasher (2013), for example, used eye-tracking to assess listeners' use of coarticulatory vowel nasalization as that information unfolded in real time. In the experiment, subjects heard the nasalized vowels with two different time latencies. The prediction was that subjects will fixate on the related image sooner when they hear the nasalized vowel earlier. The results showed that listeners use the relevant acoustic cues, which is argued, to allow listeners to track the gestural information. It is also possible that different listeners utilize various cues.

In contrast to gestural theories, auditory theories assume that speech sounds are perceived via general cognitive and learning mechanisms. In this view, speech is not special and listeners do not perceive gestures. This means that there is no special module for perceiving speech in terms of gestures. The proponents of this approach mainly take categorical perception of non-speech sounds or categorical-like perception of non-human

animals as an evidence for their argument. They also consider some of the cross-linguistic sound patterns and the “maximal auditory dispersion” in the vowel systems as further support for their claim.

An experiment could be set up to use the available data to further test these views regarding speech perception. One prediction is that if subjects are gesturally-sensitive, their reaction times should reflect the degree to which relevant gestures are present in the stimuli. The stimuli set with no coronal gestures is expected to have the shortest reaction time and the set with only overlap (i.e., full coronal gesture) should have the longest reaction time. The stimuli set with only reduced coronal gesture is expected to fall in between. Both identification and discrimination tasks could be used to collect data on the perception of subjects. The reason for using two tasks is to make sure that the gestural cues are really tapped. Research has shown the major role of experimental task in the outcomes of different studies (see, for example, Côté & Kharlamov 2008).

In addition to perception experiment, the phonological status of the simplified tokens with no [t] gesture can be further investigated. If the argument made in the discussion section is valid, the phonological rule (i.e., the simplification with no [t] gesture or the categorical deletion of coronal stops), similar to other phonological rules, should be sensitive to higher levels of linguistic structures. This means that changing a variable like speaking rate should not affect this category whereas the other two categories which belong to low-level phonetic mechanism (i.e., full [t] gesture and partial [t] gesture) will be variably affected at different speaking rates.

Another line of research could test the role of suprasegmental features. Prosodic boundaries and domains are shown to delimit segmental and tonal processes (Keating 2006).

The edges of these boundaries are marked by pitch reset and lengthening. This means that the durations of segments could vary depending on their position in prosodic structure. This provides an interesting context to test the behavior of all the three categories related to the simplified tokens. The effects of prosodic domain should manifest themselves differently depending on whether the process is controlled and planned by the speaker, or it is mechanical and uncontrolled. It is predicted that the prosodic boundaries will affect the duration of the clusters differently due to their physical and physiological versus intended nature. The duration of both reduced and overlapped tokens should be affected across the board.

There were some methodological challenges in this study. For example, it wasn't easy to record or process enough tokens to allow subject and context to be considered simultaneously as factors. This study took a reasonable approach for dealing with this. Long recording with the headset was another challenge. It is highly recommended to limit the duration of articulatory experiment to 30-40 minutes. Synchronizing the ultrasound frames with the acoustic data is quite crucial. There were some subjects who could not produce very abrupt and fast clicks. It is better to either train the subjects how to click “fast” before the start of the experiment or use another method for synchronization. Clicking fast is very important to make sure that the synchronization process is very accurate and exact. Since capturing the palate image is very significant, subjects should be asked to keep “good amount” of water in their mouth at least for 5-7 seconds to make sure that a high quality image of the palate is captured.

To conclude, this research showed that the simplification of consonant clusters in Persian is a gradient process. The optional nature of this phenomenon was linked to two mechanisms which could be active simultaneously in individuals: a mechanical process

which is not controlled by the speaker and language-specific patterns run by phonological rules or cognitive system. The main findings of the study were possible through collecting articulatory data from different individuals. This shows that all the valuable information about human's production will be omitted if we don't apply the methods, such as ultrasound technology, which allow us to gain this insight. The multiple layers of the simplified tokens could not be discovered without having access to articulatory data. It is suggested that data from individuals' production will be a valuable source of information which should not be missed in phonological studies.

## References

- Assadi, S. (2007) Sound deletion in colloquial Persian. Retrieved from <http://www.icphs2007.de/conference/Papers/1159/1159.pdf>.
- Anttila, A. (2007). Variation and optionality. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 519-536). Cambridge: Cambridge University Press.
- Avery, P. & Rice, K. (1988). Underspecification and the coronal node. In P. Avery (Ed.). *Toronto Working Papers in Linguistics*, 9, pp. 101-119.
- Babel, M. & Johnson, K. (2010). Accessing psycho-acoustic perception with speech sounds. *Laboratory Phonology*, 1(1), 179-205.
- Baker, A. (2005). *Palatoglossatron 1.0*. University of Arizona, Tucson, Arizona. <http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf>.
- Baker, A., Mielke, J., & Archangeli, D. (2008). More velar than /g/: Consonant coarticulation as a cause of diphthongization. *Proceedings of the 26nd West Coast Conference on Formal Linguistics*, pp. 60-68.
- Barry, M. (1985). A palatographic study of connected speech processes. *Cambridge papers in Phonetics and Experimental Linguistics*, 4, 1-16.
- Barry, M. (1992). Palatalisation, assimilation, and gestural weakening. *Speech Communication*, 11, 393-400.
- Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. In Tarone, E., S.M., Gass & A.D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 157-181). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85 (4), 785-821.
- Beddor, P. S, McGowan, K.B, Boland, J, Coetzee, A, & Brasher, A. (2013). The perceptual time course of coarticulation. *Journal of the Acoustical Society of America*, 133, 2350-2366.
- Blevins, J. (2004). *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.
- Blustein, S. (1981). An acoustic invariance in speech. In J. perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 178-193). Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Boersma, P. & Weenink, D. (2008). *Praat: doing phonetics by computer*. (Version 5.0.35) [Computer program]. <http://www.praat.org/>
- Boersma, P. & Weenink, D. (2012). *Praat: doing phonetics by computer*. (Version 5.3.19) [Computer program]. <http://www.praat.org/>
- Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303–328.
- Bombien, L. (2011). *Segmental and prosodic aspects in the production of consonant clusters: On the goodness of clusters*. Unpublished doctoral dissertation, LMU München: Fakultät für Sprach- und Literaturwissenschaften, Germany.
- Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.
- Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45, 140-155.
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 151-206.
- Browman, C. P. & Goldstein, L. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18, 299-320.
- Browman, C. P., & Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, (pp. 341-376). Cambridge: Cambridge University Press.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.
- Browman, C. P., & Goldstein, L. (2001). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parléé*, 5, 25-34.
- Byrd, D. (1992). Perception of assimilation in consonant clusters: a gestural model. *Phonetica*, 49, 1-24.
- Byrd, D. (1994). *Articulatory timing in English consonant sequences*, PhD dissertation, University of California, Los Angeles.
- Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209-224.

- Byrd, D. (1996b). A phrase window frameworks for articulatory timing. *Phonology*, 13, 139-169.
- Byrd, D. & Tan, C. C. (1996). Saying consonant clusters quietly. *Journal of Phonetics*, 24, 263-282.
- Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225-252.
- Clements, G. N. (1990). The role of sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I*, (pp. 283-333). Cambridge: Cambridge University Press.
- Clements, G.N. & Hume, E. 1995. The internal organization of speech sounds. In J. Goldsmith (Ed.), *Handbook of phonology*, (pp. 245-306). Blackwell.
- Chitoran, I. (1998). Georgian harmonic clusters: phonetic cues to phonological representation. *Phonology*, 15, 121-141.
- Chitoran, I., Goldstein, L. & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp. 419-448). Mouton de Gruyter.
- Coetzee, A.W. (2004). *What it means to be a loser: non-optimal candidates in optimality theory*, Ph. D Dissertation, University of Massachusetts, Amherst.
- Côté, M-H. (2000). *Consonant cluster phonotactics: A perceptual approach*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Côté, M-H. (2004a). Syntagmatic distinctness in consonant deletion. *Phonology*, 21, 1-41.
- Côté, M-H. (2004b). Consonant cluster simplification in Québec French. *Probus*, 1, 151-201.
- Côté, M. H. & Kharlamov, V. (2008). The impact of experimental task on syllabification judgment: A case of Russian. Paper presented at CUNY Conference, City University of New York.
- Cunha, C. & Harrington, J. (2011). The perception of /pt/ and /kt/ in European and Brazilian Portuguese. *Proc. ICPHS*, Hong Kong.
- Darzi, A. (1991). Compensatory lengthening in modern colloquial Tehrani Farsi. *Studies in the Linguistic Science*, 21, 23-38.
- Davidson, L. (2003). *The atoms of phonological representation: Gestures, coordination and perceptual features in consonant cluster phonotactics*, Ph. D. dissertation, Johns Hopkins University.

- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Association of America*, 120(8), 407-415.
- Davidson, L. (2007). Coarticulation in contrastive Russian stop sequences. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 417-420.
- Davidson, L. & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. In *Proceedings of WCCFL 22*, pp. 165-178.
- Davis, S. (1991). Coronals and the phonotactics of nonadjacent consonants in English. In C. Parais & J. F. Prunet (Eds.), *Phonetics and phonology 2. The special status of coronals, internal and external evidence* (pp. 49-60). San Diego: Academic Press.
- De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32: 493–516.
- Derrick, D. & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 56.3, 307-319.
- Dorman, M. F., Studdert-Kennedy, M. & Raphael, L. J. (1977). Stop-consonant recognition: Release burst and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22(2), 109-122.
- Durand, J. (1990). *Generative and non-linear phonology*, London: Longman.
- Falahati, R. (2008). *Consonant cluster simplification in Persian*. Unpublished Manuscript. University of Ottawa.
- Falahati, R. (2011). *An experimental study of consonant deletion in Persian*. Unpublished Manuscript. University of Ottawa.
- Flemming, E. (1995). Auditory representations in phonology, Ph.D. dissertation, University of California, Los Angeles.
- Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology, *Phonology* 18, 7-44.
- Fougeron, C. & Keating, P. (1997). Articulatory strengthening at the edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728-3740.
- Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, 20 (2), 269-337.

- Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38, 148-158.
- Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology*, 16(1), 29-54.
- Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden and R. Temple (Eds.), *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation* (pp. 222-236). Cambridge: Cambridge University Press.
- Gick, B., F. Campbell, S. & Tamburri-Watt, L. (2006). Towards universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of phonetics*, 34(1), 49-72.
- Giles, S. B. and Moll, K. L. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, 31, 206-227.
- Goldsmith, J. A. (1976) *Autosegmental phonology*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.
- Gow, D. W. & McMurray, R. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole, J. I. Hualde (Eds.), *Laboratory Phonology 9* (173-199). Mouton de Gruyter.
- Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York.
- Halle, M. (1959). *The sound pattern of Russian: a linguistic and acoustical investigation*. The Hague: Mouton.
- Han, K. I. (2006). *Consonant cluster phenomena: A cross-linguistic comparison*. Boulder: University of Colorado at Boulder.
- Hansson, G. O. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass*, 2(5), 859-893.
- Hardcastle, W. J. (1994) Assimilation of alveolar stops and nasals in connected speech. In J. W. Lewis (Ed), *Studies in general and English phonetics in honor of Professor J. D. O'Connor*, (pp. 49-67). London: Routledge.
- Hardcastle, W. & Roach, P. (1979). An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollien & P. Hollien (Eds), *Current issues in the phonetic sciences*, (pp. 531-540). Amsterdam: John Benjamins.

- Hayes, B. P. (1999) Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In M. Darnell, E. A. Moravcsik, M. Noonan, F. J. Newmeyer & K. M. Wheatley (Eds.) *Functionalism and formalism in linguistics, Studies in Language Companion Series References 344 42*, (pp. 243-285). Amsterdam / Philadelphia, PA: John Benjamins.
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20, 253-306.
- Hayes, B. (1981). *A metrical theory of stress rules*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Hayes, B. & Stivers, T. (1996). *The phonetics of postnasal voicing*. Ms., Dept. of Linguistics, UCLA, Los Angeles.
- Holst, T. & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʃ] assimilation. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence, Papers in Laboratory Phonology IV* (pp. 315-333). Cambridge: Cambridge University Press.
- Hooper, J. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.
- Hura, S, Lindblom, B. & Diehl, R. (1992). On the role of perception on shaping phonological assimilation in rules. *Language & Speech*, 35(1-2), 59-72.
- Jakobson, R. Fant, C.G.M. & Halle, M. (1953). *Preliminaries to speech analysis*. Massachusetts: MIT Press.
- Jespersen, O. (1904). *Phonetische Grundfragen*. Leipzig and Berlin: Teubner.
- Jun, J. (2004). Place assimilation. In B. Hayes, R. B. Kirchner & D. Steriade (Eds.), *Phonetically based Phonology*, (pp. 68- 96). New York : Cambridge University Press.
- Kager, R. (1999). *Optimality Theory*. United Kingdom: Cambridge University Press
- Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Ph.D dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Kaisse, E. (1985). *Connected Speech: The Interaction of Syntax and Phonology*. San Diego: Academic Press.
- Kang, Y. (2000). *The phonetics and phonology of coronal markedness and unmarkedness*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Kavitskaya, D. (2001). Glottal stops and compensatory lengthening. Retrieved from <http://www.meertens.knaw.nl/glow2002/kavitskaya.pdf>.

- Kawasaki, H. (1982). *An acoustical basis for universal constraints on sound sequences*, Ph.D. dissertation, University of California, Berkeley.
- Keating, P. (2006). Phonetic encoding of prosodic structure. In J. Harrington & M. Tabain (Eds), *Speech production: Models, phonetic processes, and techniques*, (pp. 167-186) Psychology Press, New York and Hove, 2006
- Kent, R. D. & Read, C. (1992). *The acoustic analysis of speech*. San Diego: Singular Publishing Group
- Kenstowicz, M. (2005). *Generative phonology*. Ms. Department of Linguistics, MIT, Massachusetts.
- Kirchner, R. (2001). *An Effort-Based Approach to Consonant Lenition*. London: Routledge.
- Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanation. . In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling*, (pp.69-92). Netherlands: Kluwer Academic Publishers.
- Kochetov, A. (2006). *Syllable position effects and gestural organization: Articulatory evidence from Russian*. Simon Fraser University MS.
- Kochetov, A. (2008). Self-organization through misperception: Secondary articulation and vowel contrasts in language inventories. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in Phonology*, (pp. 193-216). New York: Mouton de Gruyter Press.
- Kochetov, A. & Pouplier, M. (2008). Phonetic availability and grammatical knowledge: an articulatory study of Korean place assimilation. *Phonology*, 25, 399-431.
- Krakow, R. A. (1989). *The Articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Doctoral dissertation, Yale University, New Haven, CT.
- Kühnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In H.C. Yehia, Demolin, D. Demolin, & R. Laboissière (Eds.), *Proceedings of 7th International Seminar on Speech Production*, (pp. 327-334). UFMG Belo Horizonte.
- Lazard, G. (2006). *A grammar of contemporary Persian*. Hermes Publisher, Tehran.
- Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 68, 1-13.
- Liberman, A., Cooper, F.S., Shankweiler, D. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review*, 74, 431-461.

- Lieberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249-336.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.
- Maddieson, I. (1992). The structure of segment sequences. In J. Ohala, T. Nearey, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings of 2nd International Conference on Spoken Language Processing* (pp. 1-4). Edmonton: University of Alberta.
- Mahootian, S. (1997). *Persian, Descriptive Grammars*, London / New York, NY: Routledge.
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Mattingly, I.G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 415-420). North Holland Publishing Company.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229–258.
- McCarthy, J. (1988). Feature geometry and dependency: a review, *Phonetica*, 45, 84-108.
- McCarthy, J. J. & Taub, A. (1992). Review of Paradis and Prunet 1991. *Phonology*, 9, 363-370.
- Mielke, J. (2003). The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica*, 60, 208-229.
- Mielke, J. (2008). Interplay between perceptual salience and contrast: /h/ perceptibility in Turkish, Arabic, English, and French. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in Phonology*, (pp. 173-192). New York: Mouton de Gruyter Press.
- Mielke, J. (2009). Ultrasound for studying articulatory variation. Presented at *NWAV 38*, University of Ottawa.
- Mielke, J. (2011). An articulatory study of rhotic vowels in Canadian French. *Proceedings of the Canadian Acoustical Association*.
- Mielke, J. & Archangeli, D. (2005). Ultrasound research in linguistics. LASSO.
- Mielke, J., Baker, A. & Archangeli, D. (2005). An ultrasound study of excrescent schwa before /r/ and /l/ in English. Presented at *Ultrafest 3*, Tucson.

- Mielke, J., Baker, A., Archangeli, D., & Racy, S. (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. In D. Siddiqi, and B. V. Tucker, (Eds.), *Coyote Papers* (pp. 97-108), 14, University of Arizona.
- Mielke, J., Baker, A. & Archangeli, D. (2010). Variability and homogeneity in American English /r/ allophony and /s/ retraction. In L. Goldstein, D.H. Whalen & C. Best (Eds.), *Laboratory Phonology 10*, (pp 699–730). New York: Mouton de Gruyter Press.
- Mielke, J., Olson, K. S., Baker, A. & Archangeli, D. (2011). Articulation of the Kagayanen interdental approximant: An ultrasound study. *Journal of Phonetics* 39 (3), 403-412.
- Molly, B. & Johnson, K. (2010). Accessing psycho-acoustic properties with speech sounds. *Laboratory Phonology 1(1)*: 179-205.
- Natel Khanlari, P. (1987). *Tarikh-e Zaban-e Farsi*. Tehran: Nashr-e No.
- Ohala, J. (1981a). The listener as a source of sound change. In C. S. Masek, R. A. Hendrik, M.F. Miller (Eds.), *CLS 17-2*, (pp. 178-203). Chicago: CLS.
- Ohala, J. (1981b). Speech timing as a tool in phonology. *Phonetica*, 38: 204-212.
- Ohala, J. (1990). Respiratory activity in speech. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling*, (pp.23-53). Netherlands: Kluwer Academic Publishers.
- Ohala, J. (1992). The descriptive role of segments: evidence from assimilation. In G. J. Docherty & D. R. Ladd (Eds.), *Gesture, segment, prosody, Papers in Laboratory Phonology II* (pp. 286-287). Cambridge: Cambridge University Press.
- Ohala, J. (1995). The perceptual basis of some sound patterns. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence, Papers in Laboratory Phonology IV* (pp. 87-92). Cambridge: Cambridge University Press.
- Ohala, J. & Kawasaki, H. (1985). Prosodic phonology and phonetics, *Phonology yearbook 1*, 113-127.
- Paradis, C. & Prunet, J. F. (1993). A note on velar nasals: the case of Uradhi. *Canadian Journal of Linguistics* 38: 425-439.
- Pisowicz, A. (1985). *Origin of the new and middle Persian. Phonological Systems*. Krakow.
- Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: constraint interaction in generative grammar*, Ms. Rutgers University, New Brunswick, NJ / University of Colorado, Boulder.

- Proctor, M. I. (2009). *Gestural Characterization of a Phonological Class: The Liquids*. Ph.D. dissertation, Yale University: New Haven.
- “R” Development Core Team (2012). “R” (*Version 2.15.1*): software for Statistical Analyses. <http://www.r-project.org/>
- Ramirez, C. J. (2006). Acoustic and Perceptual Characterization of the Epenthetic Vowel between the Clusters Formed by Consonant + Liquid in Spanish. *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 48-61). MA: Cascadilla Proceedings Project.
- Rohany Rahbar, E. (2012). *Aspects of Persian Phonology and Morpho-phonology*. Ph.D. dissertation, University of Toronto: Toronto.
- Sadeghi, V & Bijankhan, M. (2007) Compensatory lengthening in Persian. Paper presented in ICPHS Germany. Retrieved from <http://www.icphs2007.de/conference/Papers/1159/1159.pd>
- Samareh, Y. (2002). *Ava Shenasiye Zabane Farsi*. Tehran: Makaze Nashre Daneshgahi.
- Scobbie, J. M., Stuart-Smith, J. & Lawson, E. (2008). Looking variation and change in the mouth: developing the sociolinguistic potential of ultrasound tongue imaging. Research Report for ESRC Project RES-000-22-2032
- Shademan, S (2003). Glottal-deletion and compensatory lengthening in Farsi. *UCLA Working Papers in Phonetics*, 104, 61-81.
- Slud, E., Stone, M., Smith, P., and Goldstein, M. (2002). Principal components representation of the two-dimensional coronal tongue surface, *Phonetica* 59, 108-133.
- Sole', M. J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35, 29-43.
- Sole', M. J. (1995). Spatio-temporal patterns of velo-pharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38 (1), 1-23.
- Sole', M. J. (2007). Controlled and mechanical properties in speech: A review of the literature. In M. J. Sole', P. S. Beddor, and M. Ohala (Eds.), *Experimental Approaches to Phonology*, (pp. 302-322). New York: Oxford University Press.
- Sole', M. J. & Estebas, E. (2000). Phonetic and phonological phenomena: V.O.T.: A cross-language comparison. Proceedings of the 18th AEDEAN Conference. Vigo, 437-444.
- Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291-311.

- Stevens, K. N. & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In E. Miller (Ed.), *Perspectives on the study of speech*, (pp. 1-38). Erlbaum: Hillsdale.
- Steriade, D. (1997). Phonetics in phonology: case of laryngeal neutralization. Retrieved from <http://www.linguistics.ucla.edu/people/steriade/papers/phoneticsinphonology.pdf>.
- Steriade, D. (2009). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In K. Hanson and S. Inkelas (Eds.), *The nature of the word: Essays in Honor of Paul Kiparsky*, (pp. 151-179). Cambridge, Mass.: MIT Press.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19, 455-501.
- Slud, E., Stone, M., Smith, P., & Goldstein, M. (2002). Principal components representation of the two dimensional coronal tongue surface. *Phonetica*, 59, 108–133.
- Tuller, B. & Kelso, J.A.S. (1990). Phase transitions in speech production and their perceptual consequences. In M. Jeannerod (Ed.), *Attention and Performance XIII* (429-452), Hillsdale, NJ: Erlbaum.
- Wang, Q. E. (1995). *Are syllables units of speech motor organization? A kinematic analysis of labial and velar gestures in Cantonese*. Ph.D. dissertation, Department of Linguistics, University of Connecticut.
- Warner, N. & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America*, 130 (3), 1606 - 1617.
- Wright, R. A. (1996). *Consonant clusters and cue preservation in Tsou*, Ph.D. dissertation, University of California, Los Angeles.
- Zsiga, E. C. (1992). Acoustic evidence for gestural overlap in consonant sequences. *Haskins Laboratories Status Report on Speech Research*, 112, 43-62
- Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, 22, 121-140.
- Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV: Phonology and phonetic evidence* (pp. 282–302). New York: Cambridge University Press.
- Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28 (1), 69–102.

**Appendix 1 The topics/questions used in the conversations with participants  
(in Persian)**

1. در مجموع فکر می کنی نفت برای ایران خوب بوده یا بد بوده؟
2. این که ایران کشور نفت خیزی است چه تاثیری روی اقتصادش داشته؟
3. به نظر تو در آینده نفت ارزون میشه؟
4. یک اصطلاحی هست که توش مفت و کوفت استفاده شده آیا میدونی اون چیه؟
5. اصطلاح مفت گرونه یعنی چی؟
6. اگر جنسی مفت باشه به نظرت الزامن کیفیتش هم پایینه؟
7. یک اصطلاحی هست که توش مفت و ارزون استفاده شده آیا این اصطلاح رو می شناسی؟
8. آیا دوست داری سخت کوش/سخت کار/سخت گیر باشی؟
9. به نظرت یک آدم سخت کوش/سخت کار/سخت گیر چه آدمیه؟
10. زمانی که آخرین دفعه می خواستی تو ایران از خانواده ات خداحافظی کنی خیلی سخت بود؟
11. کدام یک از بچه ها تو دانشگاه رشته سخت افزار می خونه؟
12. چند دسته و رده از حیوانات و جانوران را نام ببر؟
13. آیا خرچنگ سخت پوسته؟

**چند تا سوال ریاضی ساده دارم لطفن جواب بده.**

14. اگر ما روزی یک بار به ایران زنگ بزنیم در مجموع در یک هفته چندبار به ایران زنگ زده ایم؟
15. اگر ما روزی یک کیلومتر پیاده روی کنیم در طول یک هفته چند کیلومتر پیاده روی کرده ایم؟
16. برای تو در جدول ضرب شماره هفت آسونه یا هشت؟
17. وقت کشی کردن یعنی چی؟
18. یک اصطلاحی هست که معادل ارزون از دست دادنه. می دونی اون چیه؟
19. از نظر تو چه کارهایی وقت گیره؟

20. وقت می کنی در شبانه روز حداقل شش ساعت بخوابی؟

21. به نظرت آیا وقت آسون به دست میاد؟

22. آخرین اخبار درباره ایران چی می گفت؟

23. اگر تو توی فرهنگستان کار بکنی واژه جایگاه سوخت گیری را انتخاب می کنی یا پمپ بنزین را؟ چرا؟

24. به نظرت در آینده سوخت ارزون می شه؟

**چند تا بهت جمله میدم تو اون ها رو کامل کن.**

25. غذا داشت می سوخت که.....

26. صدف غذا پخت که.....

27. او داشت از تب می سوخت که.....

28. ایران دویست تراکتور به افغانستان فروخت که.....

29. تو قصد داری چند ساله درست رو تموم کنی؟

30. بعضی از شرکت های تولید کامپیوتر طوری اونارو می سازن که بیشتر از سه یا چهار سال کار نکنن. به نظرت

اونها از قصد این کارو انجام میدن؟

31. به نظرت آدم بدبخت کیه؟

32. به نظرت آدم بدبخت از اول بدبخت به دنیا اومده؟

33. تعریف تو از دوست چیه؟

34. میگن قدیم ها توی جاده ها آدم ها رو لخت می کردن. تو این جمله لخت کردن یعنی چی؟

35. آیا فکر می کنی لخت کردن هنوز هم اتفاق می افته؟

36. اصطلاح سنگ روی یخ کردن یعنی چی؟

**فکر کنیم که مثلن محسن یک کت و شلوار جدید دوخته.**

37. محسن کت و شلوار جدید رو دوخت که با کی بره بیرون؟

38. محسن کت و شلوار جدید رو دوخت با کی بره بیرون؟

39. محسن کت و شلوار جدید رو دوخت برای چی؟

40. کت و شلوازی که محسن دوخت از جنس چی بود؟

**Appendix 1 The topics/questions used in the conversations with participants (translated into English)**

1. In general, do you think oil has been good for Iran or bad?
2. How has the situation of Iran as a country with a lot of oil reserves affected its economy?
3. Do you think that oil will be cheaper in future?
4. There is an expression in which both “cheap” and “damn” have been used. Do you know this expression?
5. What does the expression “moft geruneh” mean?
6. If something is cheap, does it necessarily mean that its quality is low?
7. There is an expression in which both “cheap” and “gratis” have been used. Do you know this expression?
8. Do you want to be hard worker/ very strict?
9. In your opinion, who is a hard working / strict person?
10. Was it too hard for you the last time you wanted to say goodbye to your family in Iran?
11. Who is doing software engineering at the University of Ottawa?
12. Please mention some types or categories of animals?
13. Is crab a type of crustacean?

I have a few simple math questions. Please answer them.

14. If we call Iran once a day, how many times have we called Iran in a week?
15. If we walk one kilometer a day, how many kilometers have we walked in a week?
16. Which of the following numbers is simpler for you in multiplication table: seven or eight?
17. What does “killing the time” mean?
18. There is an expression which means “to lose easily”. Do you know what that is?
19. In your opinion, what type of works are very time consuming?
20. Can you sleep at least six hours a day?
21. Do you think it is easy to find time?
22. What was the latest news about Iran?
23. If you are a member of Persian Academy, will you choose “gas station” or “filling station”? Why?
24. Do you think that gas will be cheaper in future?

Please complete the following sentences.

25. Food was burning that . . . .
26. Sadaf made food to . . . .

27. He had a bad temperature that . . . .
28. Iran sold 200 tractors to Afghanistan to . . . .
29. When do you plan to finish your studies?
30. There are some computer manufacturing companies that their parts won't last more than 3-4 years. Do you think that they purposefully produce their parts in low quality?
31. In your opinion, who is a poor guy?
32. Do you think that a poor person is born as a miserable person?
33. What is your definition of friend?
34. It has been said that in old days people were robbed on the roads. What does robbing mean?
35. Do you think that robbing still happens?
36. What does this expression mean: "sang-e ruj-e jax kardan"?

Let's assume that Mohsen has made a new suit.

37. Mohsen made a new suit to go out with whom?
38. Mohsen made a new suit in order to go out with whom?
39. Why did Mohsen make a new suit?
40. What was Mohsen's new suit made of?

## Appendix 2 The short story used for data collection (in Persian)

لطفا متن زیر را داستان گونه بخوانید.

من قصد کردم هفت بار از روی تخت بپریم. احمد به من چشم دوخت که بپریم ولی وقتی نتونستم به من گفت که برو جلو بوق بزن و منو با این حرفش سنگ روی یخ کرد. یکی از لابلای جمعیت گفت اون بدبخت بی گناهِه..! همین دیشب مراسم عقد کنون داشت و قصد کرده با خانومش بره یزد. یکی دیگه از لای جمعیت گفت این به تو ربط پیدا نمیکنه...!

خیلی برام سخت بود. دلم می سوخت احمد منو به همین سادگی فروخت به هیچ. آبرویی رو ریخت که برام خیلی اهمیت داشت.

سخت بود که از احمد برای همیشه بپریم. در دلم هفت بار به اون فحش دادم. با خودم گفتم حتی حیوانات و حشرات هم با همدیگه اینطور برخورد نمیکنن. می خواستم یه چیز آبدار بهش بگم طوری که نه سیخ بسوزه نه کباب. بش گفتم منو تو با هم تفاوت داریم. بقول معروف کبوتر با کبوتر باز با باز... کند همجنس با همجنس پرواز. بعد از این دلم هری ریخت پایین. احمد به اونها چشم دوخت آیا میرن بیرون. در همین حال یه مرد با قیافه زمخت اومد داخل. آب توی بطریش ریخت اونجا و گفت اون آدمی رو که بدبخت کردید اینجاست.

## **Appendix 2 The short story used for data collection (translated into English)**

**Please read this in an informal and story-like manner.**

I intended to jump over the platform seven times. Ahmad was starring me to jump. When I could not succeed, he told me that I am too young for this job and this made me embarrassed before people. One person out of the crowd shouted this poor guy is innocent..! He just had engagement party last night and is planning to go to Yazd with his wife. Another person in the crowd shouted this is none of your business.

It was too hard for me to see how much Ahmad was indifferent to me. He ruined my reputation, the thing which was very important for me.

It was very difficult for me to break up with Ahmad for good. I cursed him seven times. I told him that even animals and insects do not treat each other like this. I wanted to tell him something which is neither too harsh nor that much gentle and soft. I told him we are very different from each other. There is a nice expression which says “birds of a feather flock together”.

After this, I suddenly got terrified. I starred them to see if they leave the room. Meanwhile, a guy with a very tough face arrived. He emptied his water bottle and said the guy whose life you made miserable is here..!

### Appendix 3 The list of unsimplified and simplified tokens extracted from seven subjects in the study.

(The orthography does not reflect the deletion of the coronal stops in the simplified tokens)

In my opinion oil	نفت به نظرم
Oil is an issue	نفت مقوله ای
Oil for	نفت برای
Oil has been	نفت بوده
Oil from	نفت از
Oil in the world	نفت در جهان
Oil is	نفت هست
Oil is not	نفت نیست
Oil should be	نفت باشه
Oil to	نفت به
Returned after	برگشت بعد
Oil got	نفت رفت
Oil had	نفت داشته
Oil us	نفت ما را
Oil will be cheap	نفت ارزون می شه
Oil in future	نفت در آینده
The price of oil should go down	قیمت نفت بیاد پایین
Oil will get	نفت بدست می آد
Oil will influence	نفت تاثیر میگذاره
Oil will finish	نفت تموم میشه
Because oil	نفت آخه
Oil is going to	نفت داره

Oil in general	نفت کلن
Oil will reduce	نفت کم می شه
Went up	رفت بالا
Oil has kept	نفت نگه داشته
A country rich in oil reserves	کشور نفت خیزیه
With oil	با نفت.
Oil goes	نفت می ره
Oil always	نفت همیشه
Is not that	نیست که
Oil expensive	نفت گرون
Rich in oil	نفت خیز که
Be cheap	مفت باشه
Be worthless	کوفت باشه
Is of little value	مفت گرونه
Has gone very cheap	مفت گذاشتن
Has become cheap	مفت شده
I am hard-worker	سخت کوشم
I am strict	سخت گیرم
He/she is strict	سخت گیره
He/she is not easy going	سخت پسنده
It was hard	سخت بود
Suddenly I got panic	هری دلمون ریخت
He/she will fix it	درست کنه
I will fix it	درست می کنم
It is not an expression	اصطلاح نیست.

To reach easy	مفت به دست
It is not pointless	الکی نیست.
Is that	هست که
It will become difficult	سخت میشه
To lose easy	مفت از دست دادن
To gain with no pain	مفت بهش برسه
I lose it	از دست بدم
It is not good	خوب نیست.
To lose	از دست دادن
Hard-working	سخت کوش
I am a hard-worker	سخت کارم
Twenty people	بیست نفر
I like	دوست دارم
Seven days	هفت روز
Was getting burnt	دآشت می سوخت
In my opinion, a friend	دوست به نظر من
To rip off	خفت کردن
Being hard	سخت گیری
Told you frankly	رک و راست بهت
To take it difficult	سخت بزنه
Time only	وقت فقط
I don't think about time	به وقت فکر نمی کنم
It is not like this that	اینطور نیست که
Be difficult	سخت باشه
I work hard	سخت کار می کنم

Anything in which there is hard	هر چی توش سخت باشه
Not strict	سخت گیر نه
I am strict	سخت می گیرم
It takes time	وقت می گیره
They will be poor	بدبخت می شن
Not punctual	وقت شناس نه
I am punctual	وقت شناس هستم
I should be more punctual	وقت شناس تر باشم
It burnt that	سوخت که
It was very difficult	خیلی سخت بود
Hard-ware	سخت افزار
Is crustacean	سخت پوسته
Crustaceans	سخت پوستان
Seven times	هفت بار
Seven kilometers	هفت کیلومتر
Eight is easier for me	هشت برام آسونتره
Seven is easy	هفت آسونه
Seven is easier than eight	هفت از هشت آسونتره
To kill the time	وقت کشی کردن
It is very time consuming	خیلی وقت گیره
You like it	دوست داری
I have time	وقت می کنم
It is not easy to find time	وقت آسون بدست نمی آد
I have time	وقت می کنم
It is easy to lose it	آسون از دست می ره

Gas station	جایگاه سوخت گیری
Not fuel	سوخت نه
Fuel won't be cheap	سوخت ارزون نمی شه
It was burning	داشت می سوخت
It was getting burnt	می سوخت که
Food was cooked to	غذا پخت که
Food was cooked and	غذا پخت و
Was from fever	داشت از تب
Two hundred tractors	دویست تراکتور
Sold to Afghanistan	به افغانستان فروخت
He/she sold to	فروخت که
Not to fix	درست نکردن
If poor	بد بخت اگر
A hopeless poor	بد بخت بی چاره
Poor from the beginning	بد بخت از اول
If time allows	اگه وقت بشه
Those two are	اون دو تاست دیگه
We intended	قصد کردم
It is not high	بالا نیست
Did it intentionally	از روی قصد این کار رو
Poor is a person	بد بخت آدمی
Poor is born	بد بخت به دنیا
He/she has been poor	بد بخت بوده
A friend is a person who	دوست کسی
Is that	هست که

To rob	لخت کردن
He/she sew that	دوخت که
He/she sew for	دوخت برای
Sew was made of	دوخت از جنس
Sew with	دوخت با

#### Appendix 4 The list of control (i.e., singleton) tokens from seven subjects in the study.

To wander around	علاف باشی
Loitering	علاف بازی
To waste	تلف کردن
I waste	تلف می کنم
If it wastes	تلف بشه
That was nonsense	مزخرف بود
Be nonsense	مزخرف باشه
Sadaf food	صدف غذا
He/she is going	طرف داره
He/she offered	تعارف کرد
You offer	تعارف می کنی
Famous expression bird	معروف کبوتر
Became famous	معروف شد
Was organized	ردیف بود
To be weak	ضعیف بودن
They are not weak	ضعیف نیستن
Got cold	یخ کرد
Skewer burns	سیخ بسوزه
To blow horn	بوق بزن
I encouraged	تشویق می کردم
I don't know exactly	دقیق نمی دونم
It happens	اتفاق می افته
It was thread	نخ بود

To be stubborn

شاخ بودن

He/she becomes stubborn

شاخ می شه