Laryngeal Contrast and Phonetic Voicing:

A Cross-dialectal Laboratory Phonology Approach for Arabic

Latifa Alkuwaiz

Doctor of Philosophy

University of York

Language and Linguistic Science

June 2022

Abstract

The thesis investigates the relationship between VOT and the phonological representation of laryngeal contrasts in Arabic, in a series of three studies, based on (i) plain and emphatic coronal plosives in eight Arabic dialects in data from the IVAr corpus; (ii) experimental data investigating the effect of speech rate on VOT in the full range of plosives in the Najdi and Hijazi dialects of Saudi, and the Tunisian dialect; and (iii) a parallel speech rate experiment on VOT in plain and emphatic coronal plosives in Modern Standard Arabic (MSA) produced by the same speakers (Najdi / Hijazi / Tunisian). Recorded data of real/nonsense words yielding emphatic/plain minimal pairs from 88 speakers in total across eight Arabic dialects (Moroccan, Tunisian, Egyptian, Syrian, Jordanian, Iraqi, Kuwaiti, Omani) were used from the IVAr corpus. The findings reveal a continuum of variation (rather than a clear-cut dichotomy) between dialects with three VOT categories (plain voiceless ~ emphatic voiceless ~ voiced) versus two VOT categories (voiceless ~ voiced). This finding adds to the classification of dialects as having either two or three VOT categories based on the literature on VOT and laryngeal contrasts (Bellem, 2007, 2014). Data for the speech rate studies were from 64 speakers (18 Najdi / 18 Hijazi / 28 Tunisian) in both registers (dialect speech and MSA). Evidence from speech rate effects in dialect and MSA speech indicate a complex set of relationships between the number of VOT categories and the number of active phonological features, resulting in over-specification at some stages of the putative sound change from one end of the continuum to the other. The implications for Arabic phonological representations, and for the modelling of a theory-driven typological analysis of laryngeal features in Arabic dialects and of phonetic uniformity in the context of sound change, are discussed.

Contents

List of Tables

List of Figures

Abbreviations

AI	Artificial Intelligence
[CP]	Constricted Pharynx
ERP	Event-Related Potential methods
F1	First formant frequency
F2	Second formant frequency
fMRI	functional Magnetic Resonance Imaging
FOD	Final Obstruent Devoicing
FUL	Featurally Underspecified Lexicon
L1	First Language
L2	Second Language
ms	millisecond
MSA	Modern Standard Arabic
PSM	Parallel Structures Model
ROI	Region of Interest
RT	Reaction Time
[RTR]	Retracted Tongue Root
RVA	regressive voicing assimilation
sahi	Saudi Hijazi dialect
sanj	Saudi Najdi dialect
SD	Standard Deviation
[sg]	[spread glottis]
SPE	The Sound Pattern of English
tuns	Tunisian dialect
VOT	voice onset time

Acknowledgements

It has been difficult to face the reality that this journey is almost over and start writing this piece of acknowledgment. Many times, during this journey, it seemed like it was never-ending, yet here we are. I am genuinely grateful to many people who helped me bring this thesis to a successful end.

My first and deepest gratitude goes to my supervisor, Prof. Sam Hellmuth. Her genuine intellectual generosity is a true blessing, a generosity that triggers a natural "pay it forward" response. I am forever indebted to her for her guidance, support, and confidence in me. For all the great conversations, walks (with Fern), laughter, and the ability to sustain me through this journey, I hold immeasurable gratitude.

Completion of this thesis would not have been possible without the help and support of my family. My mother, Hussa Al Kuwaiz, and my father, Mohammed Al Kuwaiz, without their unconditional love, help, support, and prayers, I doubt I could have made it. My love for you two is incomparable. My siblings, Waleed, Ghada, Rana, Abdulaziz, Abdullah, and Sara, and my nieces and nephews, thank you for enduring my absence. Bear with me, I'm almost there.

I cannot begin to express my thanks to my friends, Raya, Adwaa, Shatha, Rana, Manal, Mariam and Kholood for their constant support and intuitive listening to my rambling about my thesis for the past years. I bet, by now, they can defend my thesis with me.

I am also grateful for everyone who participated in making this possible. Thanks to the external evaluator for spending hours helping me cross-evaluate the MSA speech for all the participants. Thanks to Mrs. Suha for writing the MSA script customized to my research needs. Many thanks to the fieldwork assistants in Tunisia (Nourhen Nemri), Riyadh (Hatoon and Eman), and Jeddah (Abdulrahman, Lamia, and Rawaa), and to all the participants in this study.

My gratitude is extended to Cosmo Kramer, Sheldon Cooper, Michael Scott, Dwight Schrute, George Costanza, Rick and Morty for making me laugh and keeping me sane during this tough journey.

This work was supported by Imam Abdulrahman bin Faisal University and the Saudi Arabian Cultural Bureau, to them I owe gratitude for their continuous help and generous provisions throughout my scholarship.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter I

Introduction

1. Introduction

1.1 Aim of the thesis

This thesis aims to investigate segmental typology in spoken Arabic, particularly in the realm of laryngeal contrasts and phonetic voicing and their interaction with emphasis. The thesis is framed within the Laryngeal Realism theory, an approach that maps a phonetic exponent, Voice Onset Time (VOT), to one or more phonological features modelling the laryngeal contrast (Honeybone, 2005; Iverson & Salmons, 1995, 2006).

Previous research on Arabic plosives found differences in the mapping of VOT as a cue to the plain versus emphatic contrast across dialects (Bellem, 2007): some dialects display a two-way VOT distinction (voiced ~ voiceless), and other dialects display a three-way VOT distinction (voiced ~ voiceless long lag ~ voiceless short lag). Firstly, therefore, this thesis serves to investigate this as yet little researched aspect of Arabic typology, that is, the intersection of the two phonological dimensions: emphasis and voicing. Through a corpus-based experimental study, this thesis aims first to clarify the nature of this classification comprehensively by examining VOT and F2 in following vowels in four coronal plosives in eight Arabic dialects: /t/ and /d/ and their emphatic counterparts /ṭ/ and /ḍ/. This is, to the best of my knowledge, the first cross-dialectal examination based on directly parallel data, which also provides a detailed VOT description in some dialects for the first time (e.g., Tunisian, Omani). Although the data are all from a single point in time, the results indicate a continuum of variation, rather than a clear split, consistent with interpretation of the variation in terms of a sound change from a three-way to a two-way voicing distinction (Bellem, 2014).

Building on this cross-dialectal foundation, the remainder of the thesis looks to understand the phonological specifications that might underpin that variation. Speaking rate typically displays asymmetric effects on VOT cross-linguistically: as speech rate decreases, values of long lag and voicing lead both increase, but short lag VOT is not affected. Phonologists have thus proposed that in two-way laryngeal contrast languages (e.g., English: [spread glottis] vs. [Ø]) only one side of the phonological contrast is active and affected by speech rate (Summerfield, 1975). However, recent results from Swedish (Beckman, Helgason, McMurray, & Ringen, 2011), which is a two-way voicing contrast language, found rate effects on both the voiced and voiceless plosives, which was interpreted as 'over-specification' of both [voice] and [sg] as active features in that language. The variation in the mapping of surface VOT values to underlying laryngeal contrasts in Arabic dialects, in a manner that might indicate sound change in progress, provides rich ground to explore the status of laryngeal features as active or not; if any over-specification pattern is found we might expect to see it in a system potentially undergoing change. Secondly, therefore, in two additional experimental studies, this thesis also examines variation in phonetic voicing and laryngeal contrasts in data from speakers of three Arabic varieties – Najdi, Hijazi, and Tunisian – producing both vernacular and Modern Standard Arabic (MSA) speech. The theoretical goal of this thesis is thus to identify the nature of underlying laryngeal contrast underpinning the spectrum of cross-dialectal variation. The investigation reveals a complex set of mappings between the number of voicing categories and the number of active laryngeal features. The investigation indicates that, in some Arabic dialects, there might be instances of over-specification, specifically in the dialects that are in the transitional phase and potentially currently undergoing sound change.

1.2 Outline summary of the thesis

Following this introductory chapter, chapter 2 provides a brief background on VOT as a phonetic parameter, both cross-linguistically and across Arabic dialects. The chapter also provides an overview of key theories of laryngeal phonology in general then focuses on Laryngeal Realism as the theoretical framework adopted in this thesis. Some details of prior studies that employed speaking rate as a diagnostic to test for active laryngeal features are also discussed. An overview on the linguistic context of the thesis – Arabic dialects – is also presented in this chapter, in addition to a review of Arabic diglossia and prior work that investigated and compared the two registers of Arabic (MSA and dialect speech). Finally, chapter 2 closes by restating the purpose of the thesis, identifying research gaps, and setting out the research questions arising from the review of literature.

Chapter 3 reports the results of systematic acoustic analysis of corpus-based production data aimed at examining the degree of variation in realization of the voicing contrast in Arabic dialects. Examining the coronal plosives (plain /t/ ~ /d/ and emphatic /ṭ/ ~ /ḍ/) in eight Arabic dialects – Moroccan, Tunisian, Egyptian, Syrian, Jordanian, Iraqi, Kuwaiti, and Omani – through multiple linear mixed effects models in R results in a continuum of variation across dialects rather than a clear-cut dichotomy. Some dialects show a three-way VOT distinction and other dialects show a two-way VOT distinction, while two dialects show hybrid results that might indicate 'transition' from a three-way to a two-way VOT distinction. This dialectal variation is mainly noticeable in the degree of voicing lag observed in VOT values for /t/ and /ṭ/, in that two-way dialects display overlapping values. Despite this overlap, the coarticulatory cues of emphasis in the following vowel are unaffected; F2 lowering after /ṭ/ is found to be consistent and not affected by dialect or gender. I conclude the chapter by discussing the results and the potential interpretation of them as evidence of sound change based on observed gender differences in the production of VOT, in two dialects in the middle of the continuum.

Chapter 4 reports the results of a study to establish the phonological representations of the voicing contrasts in three Arabic dialects, as reflected in VOT of their plosives using speech-rate manipulation as a diagnostic tool. Having established the spread of dialectal variation in chapter 3, chapter 4 examines the voicing contrast in three dialects expected to display a different number of laryngeal categories – Saudi (Najdi): three-way, Saudi (Hijazi): possibly three-way, and Tunisian: two-way – and compares the dialects through an initial description of the laryngeal contrast within the Laryngeal Realism framework. Multiple acoustic cues in the full set of Arabic plosives (voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, (g)/), with a focus on VOT, and at two speaking rates (fast ~ slow), are examined in a quantitative analysis using linear mixed effects models in R. The results indicate that the three investigated dialects have the same number of active (underlying) phonological features despite displaying different numbers of (surface) phonetic VOT categories. One dialect (Tunisian) appears to be over-specified with two active phonological features [voice] and [spread glottis]. In light of the results, the chapter provides a general discussion and conclusion.

Chapter 5 examines MSA speech produced by the same speakers as examined in chapter 4, aiming to test for the number of active underlying phonological feature(s) in their MSA production. The chapter starts with a brief overview on the linguistic context of MSA, diglossia, and dialectal variation in realization of MSA. Multiple cues to the emphatic contrast in MSA /d/ ~ /ḍ/ and /t/ ~ /ṭ/, at two speaking rates (fast ~ slow), are examined in data from speakers of the three dialects. The results suggest different patterns in the MSA production compared to the counterpart dialect speech production, in some dialects. Two of the three investigated dialects (Najdi and Hijazi) show the same number of active (underlying) phonological features in MSA and dialect speech, however, they display a different number of (surface) phonetic VOT categories. The remaining dialect (Tunisian) shows a different number of underlying active phonological features in MSA compared to its dialect speech production and compared to the other dialects, yet it shows a similar number of (surface) phonetic VOT categories as its dialect speech production. In the MSA speech production, one dialect (Hijazi) appears to be over-specified with two active phonological features [voice] and [spread glottis]. At the end of chapter 5, an interim discussion and conclusion is presented.

Chapter 6, which is the main discussion chapter of the thesis, interprets the combined results of three experiments of chapters 3 – 5 from four different phonetic and phonological perspectives. After summarizing the results, the first section proposes a theory-driven typological analysis of laryngeal features in Arabic dialects. The second section argues for interpretation of the observed dialectal variation as potential sound change in progress, based on the mismatches found between the surface categories and underlying representation of the laryngeal contrast in the two registers – dialect speech and MSA – and based also on the observed gender differences in production of plosives, in only those dialects that are suspected to be undergoing change. The third section discusses potential reasons for why this putative sound change process might be further advanced in the MSA register, as indicated by the discrepancies between the two registers in Tunisian and Hijazi productions. In addition, the section sketches a model of predictability in a system undergoing change following the principle of Phonetic Uniformity (Chodroff, Golden, & Wilson, 2019). Finally, the fourth section revisits the phonological representations of Arabic plosives within the Laryngeal Realism approach. The section restates phonological representations of emphasis in Arabic, and specifically the representations proposed by Youssef (2013) adapting Morén's (2003) Parallel Structures Model (PSM) of feature geometry. The section offers a proposal of a representational analysis of the observed over-specification patterns (which might indicate sound change in progress) in the PSM framework.

The thesis concludes with chapter 7, in which the contribution of the thesis and the limitations of the research are summarized, followed by possible future research directions.

In a broad sense, in the terms used by Youssef (2013), this thesis can be considered as "a study of synchronic microvariation" (p. 19), aiming to investigate the phonetic and phonological variation of closely related dialects of Arabic in their current, present-day usage. One advantage of studying multiple Arabic dialects is that the shared origin serves as a natural control factor, which facilitates detection of micro-parameters of variation (Youssef, 2013). In this way, the test case of Arabic dialects facilitates investigation and documentation of related phonetic and phonological phenomena in a sound system that is not stable, that is, in which some dialects might be undergoing change.

Chapter II

Background

2 Background

2.1 Introduction

The degree to which phonetic realization should influence our understanding of phonological representations has been a point of prolonged debates. Despite some views of phonology as autonomous from phonetics (e.g., substance-free phonology (Hale & Reiss, 2000)), many agree that there exists some link between the two (e.g., (Clements, 1985; Flemming, 2005)). It is also debated to what extent this link is direct. Most phonological theories and phonological typology approaches are concerned with this link, that is, how different languages encode similar phonetic cues (Hyman, 2014; Kiparsky, 2018; Youssef, 2021).

The link between phonological feature representation and phonetic realization pertains to all kinds of features, and particularly, to the laryngeal features ([voice], [spread glottis], [constricted glottis]). The contrasts among laryngeal features provide excellent testing grounds for this link. As stated by Cyran (2017) "at the heart of laryngeal phonology lies the nature of the relationship between phonology and phonetics" (p.477). At the surface, the question whether a plosive is voiceless or voiced is rather straightforward. However, defining laryngeal categories and the appropriate phonetic cues that encode them has not been straightforward in phonological theories (Wojtkowiak & Schwartz, 2018). The work of Lisker & Abramson (1964) drew attention to Voice Onset Time (VOT) as a phonetic component that is linked to the laryngeal system. The resulting synergy between the physical and cognitive aspects of speech sounds is what motivates this current research that follows an interdisciplinary Laboratory Phonology approach.

In the following sections of this chapter, I will start in section § 2.2 by reviewing Lisker & Abramson's (1964) observations about the relationship between phonetic realization and laryngeal contrast through the phonetic parameter VOT. Then, in section § 2.3, I will review key theories of laryngeal phonology, adopting Laryngeal Realism as the theoretical framework of this thesis. Section § 2.4 is a survey of prior works that employed speech-rate as a diagnostic tool to test for active laryngeal features of a studied language. In section § 2.5, I shall review previous works on the Arabic voicing contrast across its dialects, along with section § 2.6 explaining observed variation in the Arabic dialects' voicing contrast. Section § 2.7 provides a review on prior work that investigated and compared the two registers of Arabic MSA and dialect speech. Finally, in section § 2.8 the chapter concludes with restating the purpose of the thesis and identifying research gaps in light of the review of literature provided in this chapter.

2.2 Voicing contrast: across languages

A temporal variable, Voice Onset Time (VOT), is one key phonetic exponent of the laryngeal contrast in plosives in the languages of the world. VOT refers to the time difference between the release of the plosive closure, which is treated as a reference point (i.e., 0 ms in time), and the onset of vocal fold vibration (in the following vowel). Lisker & Abramson (1964) examined the VOT of utterance-initial plosives in eleven languages. They established a benchmark for VOT values in three categories: prevoiced, having negative VOT values due to the vibration of vocal folds during the closure interval before the burst of the plosive (voicing lead); unaspirated, having positive values of VOT caused by a short delay of vocal folds vibration after the release of the plosive (short lag); or aspirated, with longer delay after the release of the plosive and before the vocal folds vibration, resulting in greater positive VOT (long lag) (Abramson & Whalen, 2017; Lisker & Abramson, 1964).

Lisker & Abramson (1964) found that there are two types of languages having a two-way laryngeal contrast. One type has a laryngeal contrast between unaspirated (short lag) for plain voiced (lenis) plosives and aspirated (long lag) for voiceless (fortis) plosives. These languages are referred to in the literature as *aspirating languages*. This type includes most Germanic languages, e.g., standard varieties of English, German, Icelandic, Danish, and Norwegian (except Dutch (Honeybone, 2005), and Durham English (Harris, 1994)). The other type has a laryngeal contrast between prevoiced (voicing lead) for voiced (lenis) plosives and unaspirated (short lag) for plain voiceless (fortis) plosives, and these languages are termed *true voice languages*. The *true voice languages* type includes the Romance and Slavic languages; examples are French, Russian, and Spanish (Beckman, Jessen, & Ringen, 2013; Iverson & Salmons, 1995; Jansen, 2004; Lisker & Abramson, 1964; Ringen & Kulikov, 2012). Lisker & Abramson (1964) also examined languages which have a three-way laryngeal contrast, including Eastern Armenian and Thai, in addition to two languages with a four-way laryngeal contrast like Hindi and Marathi. Figure 2-1 below illustrates the VOT-based typology.

aspirating languages
Germanic

true voice languages
Romance and Slavic

	No. of categories	[d]	[t^{o}]*	[t^{h}]	[$d^{ɦ}$]
a. French	2	✓	✓		
b. English	2		✓	✓	
c. Thai	3	✓	✓	✓	
d. Hindi	4	✓	✓	✓	✓

Figure 2-1 Two-way '*true voice*' and '*aspirating*' languages, three-way and four-way languages in a VOT-based typology. *The non-IPA superscript diacritic '^{o}' used here is adopted from Honeybone (2005, p. 332) to represent the 'neutral', 'voiceless', or the underlyingly unspecified segments.

2.3 Laryngeal Phonology

In phonological theory, distinctive features have been central to the study of phonological patterns, within and between languages. The traditional generative phonology of SPE (Chomsky & Halle, 1968) viewed distinctive features as articulatory features having binary values. For laryngeal contrasts, assigning the features [+voice] and [-voice] became problematic for identifying the different laryngeal contrasts which distinguish aspirated vs. unaspirated phonemes in the two different two-way laryngeal contrast language types. Many post-SPE views argued against the binarity of distinctive features, on the basis that things in the world are not defined in terms of what they are not; we describe a thing by its positive attributes, and its negative attributes are considered unnecessary extra information (Beckman et al., 2013; Bellem, 2007; Honeybone, 2005; Iverson & Salmons, 1995; Jansen, 2004).

The laryngeal features have received great attention in phonological theory. Even though they are widely studied, and this entailed a number of proposed features and representations, they remain a central point of disagreement (Clements & Hallé, 2010).

The typology initiated by Lisker & Abramson (1964), that groups languages in terms of their voicing contrast in word-initial plosives through VOT, led scholars to fall into two main camps. The first are those that advocate use of the feature [voice] to represent the laryngeal contrast in both *true voice languages* and *aspirating languages* (Keating, 1984; Kingston & Diehl, 1994; Wetzels & Mascaró, 2001). The second camp are those that argue against using [voice] in *aspirating languages* (Beckman et al., 2013; Harris & Lindsey, 1995; Honeybone, 2005; Iverson & Salmons, 1995; Jessen & Ringen, 2002). The second group of phonologists and phoneticians argue that the feature distinguishing the contrast in *aspirating languages* is [spread glottis], not [voice][1], so that [voice] is used to distinguish the voicing contrast in *true voice languages*, only.

This second view is called Laryngeal Realism; it is a view that maps VOT to phonological specifications transparently and became the mainstream approach to modelling laryngeal contrasts. In the next sections below, we discuss the popularity of the laryngeal realism approach, then the critiques of this approach; finally, we review some other proposed approaches to laryngeal phonology.

Table 2-1 shows different phonological representations proposed for laryngeal specifications that are broadly comparable to the Laryngeal Realism. The table is adapted from Honeybone (2005, p. 325 (6)).

Table 2-1 Different phonological representations proposed for laryngeal specifications that are broadly comparable to the Laryngeal Realism. Adapted from Honeybone (2005).

Lombardi (1994)	**Harris (1994)**	**Iverson & Salmons (1995)**	**Honeybone (2005)**
[aspiration]	|H|	[spread glottis]	|spread|
[voice]	|L|	[voice]	|voice|
[glottalization]	|ʔ|	[constricted glottis]	|constricted|

2.3.1 Laryngeal Realism

The term Laryngeal Realism was coined by Patrick Honeybone (2005), but the idea itself was not new. The idea was proposed by Iverson and Salmons (1995) under the name of 'multiple feature hypothesis', based on the ground-breaking work of Kim (1970) on aspiration. Laryngeal Realism also follows in the footsteps of Element

[1] In Element Theory these translate to element |H| for aspiration and element |L| for voicing, see (Backley, 2011; Harris, 1990; Harris & Lindsey, 1995) and Table 2-1 for a summary of other representations.

Theory (Backley, 2011; Harris, 1990; Harris & Lindsey, 1995) and Dependency Phonology (Anderson & Ewen, 1987) in adopting privativity of features in phonological representation.

Laryngeal Realism is a proposal that classifies glottal phonation types structurally into basic phonological feature representations (Iverson & Salmons, 2006). Specifically, Laryngeal Realism claims that all two-way voicing contrast systems have a single privative laryngeal feature. In both systems, *true voice languages* and *aspirating languages*[2], the unaspirated (short lag VOT) series are said to be unmarked[3], while the fully voiced (VOT voicing lead) or the aspirated (long lag VOT) are marked and indicate a presence of a laryngeal feature. In *true voice languages* the proposed feature is [voice] (or the element |L|) and in *aspirating languages* the feature is [spread glottis] (or the element |H|) (Beckman et al., 2013; Harris, 1990; Helgason & Ringen, 2008; Honeybone, 2005; Iverson & Salmons, 1995, 2006). The representations Laryngeal Realism proposes are privative and realist. For instance, the representation of a two-way *true voice language* like French, is the [voice] feature for /b, d, ɡ/ that contrasts with [Ø][4] for /p, t, k/, where [Ø] refers to the non-existence of a laryngeal specification and [voice] is the active phonological feature in that language. As for *aspirating languages* like English, the feature [spread glottis] contrasts with [Ø]; the [spread glottis] feature represents voiceless aspiration in /p, t, k/, and /b, d, ɡ/ are unspecified with a laryngeal feature [Ø]. (Beckman et al., 2011; Beckman et al., 2013; Mester & Itô, 1989).

The Laryngeal Realism approach gained popularity for its elegant simplicity. Firstly, the approach captures phonological representation of the laryngeal contrast through empirical examination of phonetic cues. Secondly, it encodes other observed generalizations about phonological phenomena and language types and their ‘marked’ laryngeal feature. For example, Honeybone (2005, p. 329) and Iverson & Salmons (2006, p. 3) list the characteristics of an *aspirating language*:

i) ‘voiceless’ plosives are aspirated in most environments.
ii) ‘voiced’ plosives show inconclusive evidence of passive voicing.

[2] *Aspirating languages* and *true voice languages* are referred to as ‘type A’ and ‘type B’ languages respectively in Honeybone’s (2005) analysis.
[3] It is necessary to clarify what is meant by ‘marked’ and ‘unmarked’ in this thesis. As noted by Schwartz & Arndt (2018) ‘marked’ in this thesis is used to denote ‘bearing phonological specification’, and ‘unmarked’ denotes ‘phonologically unspecified’.
[4] In this thesis, segments unspecified for any laryngeal feature are represented with a null sign between brackets [Ø].

iii) assimilation in clusters is prominent for ‘voiceless’ plosives (devoicing) e.g., English *p**l**an* [pl̥an].

Whereas the characteristics of *true voice languages* are:

i) ‘voiceless’ plosives are unaspirated.
ii) ‘voiced’ plosives are fully voiced.
iii) assimilation in clusters is more common for ‘voiced’ plosives e.g., Dutch *po**t**dicht* [pɔddɪxt] ‘tight’.

However, this distinction of assigning the marked feature to the language type is not always straightforward. In *aspirating languages* like English, instances of both VOT lead and short lag VOT can occur on voiced stops (Docherty, 1992). Patterns of VOT have mostly been examined in initial position, but they are more complex in medial position. Some aspirating languages show passive voicing in word-medial voiced stops which means they are phonetically voiceless, and these passive voicing patterns are different from active voicing (Al-Tamimi & Khattab, 2018; Beckman et al., 2013). Passive voicing is different from active voicing in that the amplitude in passive voicing gradually decreases, whereas active voicing does not show this amplitude drop. For example, passive voicing in word-medial voiced stops in German, which is an *aspirating language*, are found to have decreasing voicing amplitude, while in Russian, which is a *true voice language*, word-medial voiced stops do not show this drop in amplitude (Ringen & Kulikov, 2012). Furthermore, Jansen (2004) indicates that plain voiceless (short lag) stops in *true voice languages* and plain voiced (short lag) stops in *aspirating languages* are not the same, as the short lag stops in *true voice languages* never show passive voicing. Rising issues of passive voicing along with other challenges to the Laryngeal Realism view are discussed in the next section below.

2.3.2 Challenges to the Laryngeal Realism approach

Despite its popularity, many assumptions of Laryngeal Realism are strongly criticized. For instance, Vaux & Samuels (2005) argue against the assumption in Laryngeal Realism that aspirated plosives are marked and unaspirated plosives are unmarked based on empirical evidence involving neutralization, ease of articulation, fast speech, and early acquisition in child phonology, among other evidence all in favour of the plain voiceless plosives which indicate markedness. For example, in terms of ease of articulation, Vaux and Samuels (2005) argue that producing aspirated plosives entails less articulatory precision than the plain unaspirated (short lag) plosives. They claim that unaspirated plosives "require specific laryngeal instructions" (p. 401) and speakers are not obligated to limit their production into a smaller temporal window (between 0 ms and roughly 35 ms), i.e., the production of the aspirated (long lag) plosives is considered easier since it has a large temporal margin. In regard to early acquisition and production of unaspirated (short lag) plosives by children, Vaux and Samuels (2005, p. 404) list a number of studies that indicate variability in plosive production by children, in fact in a number of acquisition cases aspirated or voiced (voicing lead) plosives are acquired before the unaspirated (short lag) plosives.

Another criticism of Laryngeal Realism pertains to the privativity of laryngeal features, and that word-internal plosive devoicing can occur in a language that does not show word-final devoicing. Wetzels and Mascaró (2001) claim that plain voiceless plosives in *true voice languages* can appear to be active phonologically and may spread to neighbouring segments. They provide empirical evidence from different languages, e.g., Parisian French, Bosnian/ Croatian/ Serbian, and Romanian. They argue against the claim of privative [voice] feature and attribute the binary [-voice] feature to account for these patterns of assimilation.

Criticism of Laryngeal Realism also comes from the case of Swedish investigated by Beckman et. al. (2011). Swedish is a language with a two-way contrast system which displays both voicing lead and long lag VOT patterns. This makes it 'over-specified' with two phonological features [voice] and [spread glottis] thus exhibiting maximal dispersion along the continuum of VOT. This over-specification view contradicts the notion of privativity and economy of representations (Clements, 2003).

Finally, many recent views, as we will briefly review below, question the phonological status of voicing altogether. These approaches propose an alternative ontology of 'voicing' in phonology, claiming that 'voice' is a simply salient carrying signal, thus languages do not need to associate a feature [voice], even in *true voice languages* (Cyran, 2011, 2017; Schwartz, 2013; Schwartz & Arndt, 2018; Wojtkowiak & Schwartz, 2018).

2.3.3 Other approaches to the Laryngeal Phonology

Several recent proposals followed, that try to adhere to the privative approach and yet offer a way to resolve the issue of over-specification and passive voicing. These approaches criticize the over-reliance on phonetic detail to encode phonological representations and claim that this over-reliance does not necessarily entail better understanding of the phonological phenomena. Two approaches, Laryngeal Relativism (Cyran, 2011, 2017) and Onset Prominence (Schwartz, 2010, 2013), call into question the role of [voice] in *true voice languages*.

The main claim of Cyran's (2011) Laryngeal Relativism model is that the relation between phonetics and phonology is 'arbitrary' and relative. Cyran claims that representations in phonology cannot be directly read from phonetic details; in this view, the phonological representation of a contrast is arbitrary, and a voicing series does not need to be marked as long as sufficient phonetic distance between the two series is maintained. Cyran also argues that reliance on observable phonological phenomena, such as regressive voicing assimilation (RVA) and final obstruent devoicing (FOD), to define categorical distinctions can be contradictory and misleading (Cyran, 2017). Another claim of the Laryngeal Relativism model regarding the nature of voicing in phonology is that the feature [voice] is not needed even in *true voice languages*. Cyran argues that voicing is not necessarily phonological, and the presence of voicing lag does not entail [voice] in a two-way series. This approach, he claims, explains the Swedish over-specification riddle: the voicing lag in *aspirating languages* is merely a phonetic observation. Cyran (2017) adds that the Swedish two-way contrast "goes for maximal dispersion rather than for sufficient phonetic distance" (p. 502).

Another theoretical alternative to the Laryngeal Realism approach, based on the assumptions of Modulation Theory (Traunmüller, 1994)[5], is the Onset Prominence representational framework (Schwartz, 2010, 2013). In this framework, the representations are hierarchical. Onset Prominence agrees with Laryngeal Relativism in questioning the status of [voice]. While Laryngeal Relativism questions the necessity of the [voice] feature, Onset Prominence rejects the existence of phonological voicing properties. Schwartz claims that voicing is an element of the acoustic carrier signal; it is the background that bears the linguistic message, so it does not qualify to be a product of phonological specification. Since Schwartz's Onset Prominence is hierarchical, it proposes that the difference between plain voiceless plosives and aspirated plosives is at

[5] The main assumption, for our purposes, in Traunmüller's (1994) Modulation Theory is that phonological features are a reflection of salient modulations on an inherently voiced carrier signal. The theory claims that the [voice] feature does not play any role in the laryngeal contrast of any voicing system (Wojtkowiak & Schwartz, 2018).

the level where [sg] is assigned; plain voiced plosives and plosives with voicing lead are unspecified, as seen in Figure 2-2 below from Schwartz & Arndt (2018, p. 103).

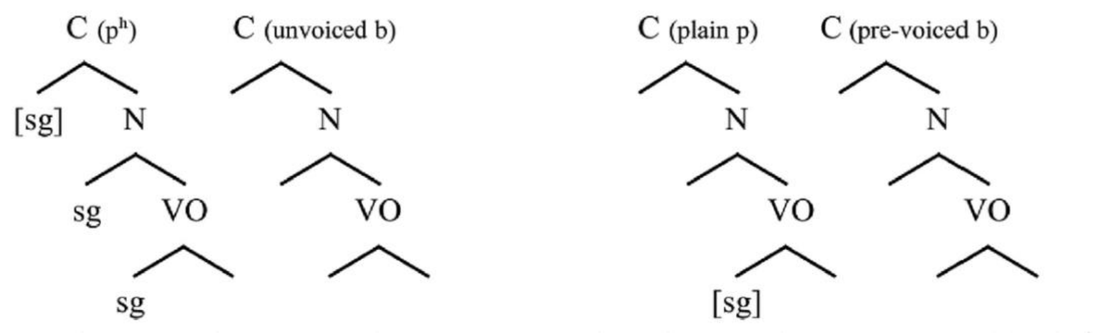


Figure 2-2 OP representations for *aspirating languages* (left pair of trees) and *true voice languages* from Schwartz & Arndt (2018, p. 103), where C is closure, N is noise, and VO is vocalic onset in the OP hierarchy.

2.3.4 Summary

In this section, we reviewed several approaches to laryngeal phonology, starting with the generative phonology approach using SPE distinctive features and leading up to recent models that followed the most popular approach to laryngeal phonology: Laryngeal Realism. This section also discussed the main reason behind widespread adoption of Laryngeal Realism, namely the proposed transparency between phonological representation and observed phonetic realizations of laryngeal contrasts. Critiques of and challenges to Laryngeal Realism are also reviewed. Lastly, this section listed other proposed approaches to laryngeal phonology that revisit the ontology of voicing in phonology, and as a result allow for voiced obstruents to be unmarked in *true voice languages*.

In this thesis, following Beckman et al. (2013), we assume privativity of laryngeal features; that is, laryngeal features are not binary, but are defined by either the presence or absence of the feature. This creates a distinction between *true voice languages* and *aspirating languages*. To give an example, in *true voice languages* like Russian, the [voice] feature contrasts with [Ø], where [voice] results in the active voicing in medial position as well. As for *aspirating languages* like German, the feature [spread glottis] contrasts with [Ø]; the [spread glottis] feature represents voiceless aspiration and unspecified plosives allow for medial position passive voicing (Beckman et al., 2011; Beckman et al., 2013; Mester & Itô, 1989). Since this thesis explores Arabic variation, we adopt the Laryngeal Realism approach in assigning different features for voicing [voice] and aspiration [spread glottis].

2.4 Diagnostics for the number of laryngeal contrasts

VOT is a temporal speech cue to voicing; it is described, as already mentioned, by the temporal relation between two acoustic and articulatory events. Another feature of speech is the highly variable speaking rate of humans. There has been considerable interest in the relationship between these two variables, and several studies have shown that speech rate affects a number of speech cues, including VOT, in an asymmetric fashion. Summerfield (1975) first reported the results of this relationship in English for VOT; as speech rate decreased, VOT values of voiceless plosives increased and those for voiced plosives remained relatively unchanged.

Further cross-linguistic production and perception studies followed on speaking rate effects on VOT. Initially, studies examining English supported the finding of Summerfield (1975) (Allen & Miller, 1999; Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller, Green, & Reeves, 1986; Summerfield, 1981). Later, further research has shown that this asymmetrical pattern is not only apparent in English; it has been found that it is also a property in other *aspirating languages* like Icelandic (Pind, 1995).

As for other types of languages, such as *true voice languages*, and languages having a three-way contrast, Kessinger and Blumstein (1997) provided a comprehensive cross-linguistic analysis. They examined speech rate effects on VOT production in three different types of languages: English, as a two-way *aspirating language*; French, as a two-way *true voice language*; and Thai, which has a three-way voicing contrast between long lag (aspiration), short lag, and voicing lead. In their experiment, participants produced words that are minimal or near-minimal pairs in three conditions: in isolation as a wordlist, at a slow speaking rate in a carrier phrase, and at a fast speaking rate in a carrier phrase. The findings showed asymmetric effects of speaking rate on the voicing contrast in all three languages. The English results supported previous findings with slow speech rate having minimal or no effect on short lag VOTs but an increase in the duration of (aspirated) long lag VOTs. Whereas in French, the duration of negative VOT in sounds with a voicing lead increased when speech rate decreased, while short lag VOTs were not affected. Finally, in Thai, the duration of negative voicing lead and long lag (aspirated) VOTs both increased in slow speech, while again, the short lag VOTs remained unaffected. Similarly, Magloire & Green (1999) found that Spanish and English speakers' short lag VOTs were very slightly affected by slow speech rate, yet negative voicing lead VOTs in Spanish and long lag (aspirated) VOTs in English were highly affected by slow speaking rate.

Theoretical proposals to account for the asymmetric effect of speech rate on voicing contrasts are limited. Studies which adopt privative phonological features as units to describe phonological representations suggest that the VOT of plosives specified with [voice] to describe voicing lead (Lombardi, 1995), and with [spread glottis] for aspiration (Iverson & Salmons, 1995), are the categories that are mainly affected by rate of speech (Beckman et al., 2011; Kessinger & Blumstein, 1997; Magloire & Green, 1999).

Crucially, it is suggested that the category affected by speech rate in a language reflects the feature used in the phonology of that language. This asymmetry can be used in a laboratory phonology approach as a tool to detect the presence of an active voicing feature in a language. In light of this, the findings of Kessinger & Blumstein (1997) can be interpreted regarding the active laryngeal feature in the phonology of the languages studied. As the speech slows, the duration of the phonetic exponent (VOT) increases, yet it primarily affects the active feature in the language. This is, for French the active feature is [voice] and for English it is [spread glottis]. At slow speech rates in Thai, a three-way language, there are effects on both features [voice] and [spread glottis].

Although there appears to be strong evidence of the effect of speaking rate at both ends of the voicing continuum, [voice] and [spread glottis], evidence for the lack of an effect of speech rate on short lag VOTs seems to vary, with short lag plosives said to be either slightly affected by speech rate or not at all affected. Cross-linguistically, the reason for them not to show an effect of speech rate is that they are proposed to be unspecified for any feature in the phonology (Beckman et al., 2011; Iverson & Salmons, 1995).

The principle of economical representations in phonology, following Clements (2003), assumes the use of the minimal number of features necessary to determine the contrasts of a language. If we assume economical representations, then two-way contrast languages only need one feature. Hence, *true voice languages* (having a two-way voicing contrast) are expected to employ [voice] as their active feature, and *aspirating* *languages* (also having a two-way voicing contrast) employ [spread glottis] as their active feature. Until recently, it was claimed that no two-way contrast language employs two active laryngeal features (Beckman et al., 2011; Iverson & Salmons, 1995).

However, evidence recently emerged that a two-way contrast language may in fact have two active features [voice] and [spread glottis]. Helgason & Ringen (2008) and Beckman, Helgason, McMurray & Ringen (2011) suggested that Central Standard Swedish has two active laryngeal features, like Thai, even though Swedish (unlike Thai) has a two-way voicing contrast. Swedish is thus said to be over-specified, as a two-way contrast language, in employing both features. The difference between the VOT of voiced and voiceless plosives in Swedish is around 120 ms, which is considered a very large difference compared to the difference between voiced and voiceless plosives in both *true voice* and *aspirating languages* i.e., roughly 60 ms.

To further investigate this, Beckman et al. (2011) used speaking rate as a diagnostic to determine the active phonological features that specify the laryngeal contrast in Swedish. In their study, the participants were eight Swedish speakers (4 females and 4 males). The speakers were asked to produce 18 target words of plosives in initial position (2 voicing x 3 place of articulation x 3 vowels). The target words were elicited in three conditions: a list of isolated words, a carrier phrase sentence with no instruction to speaking rate (slow), and a carrier phrase sentence with instruction to increase speaking rate without sacrificing accuracy (fast). They found that as speaking rate decreases, the VOT of both aspiration and voicing lag increase. The average values of VOT voicing lead in Swedish voiced plosives was -107.9 ms (slow) vs. -78.5 ms (fast), leaving a difference of 29.4 ms between the two conditions. While the average values for long lag VOT in Swedish voiceless plosives was 74.5 ms (slow) vs. 55.8 ms (fast), which makes the difference between the two conditions 18.7 ms. Beckman et al. (2011) compared the difference of their results to results of previous cross-linguistic data on speaking rate (e.g., Magloire & Green (1999) and Kessinger &Blumstein (1997)); they found close correspondence in the difference of values of the voicing lead and long lag plosives in two speaking rates in their study and the previous studies. Their findings support the claim that the Swedish phonological contrast is ‘over-specified’ with two active features [voice] and [spread glottis].

This pattern of over-specification has so far been found in Swedish (Beckman et al., 2011; Helgason & Ringen, 2008), Turkish (Petrova, Plapp, Ringen, & Szentgyörgyi, 2006), and Norwegian (Ringen & van Dommelen, 2013), and in one dialect of Arabic (which will be discussed in the next section). Of these, only Swedish and the Arabic dialect were investigated experimentally using speech rate as a variable to test for over-specification.

There are other additional diagnostics proposed by Beckman, Helgason, McMurry, and Ringen (2011), and Beckman et al. (2013) to test the active phonological feature through acoustic cues. In addition to the effects of speaking rate on VOT in word-initial positions, Beckman et al. (2013) suggest that the percentage of voicing in word-medial plosive closure serves as a second diagnostic for active features. They examine the amount of voicing in the closure duration of voiced plosives /b, d, g/ in *true voice languages* vs. *aspirating languages*. They found that in Russian, a *true voice language*, the percentage of voicing during closure (i.e., voicing during the closure is more than 90%) in /b, d, g/ is 97%, whereas in German, an *aspirating language*, 62% of the voiced plosives /b, d, g/ have more than 90% voiced closures. Beckman et al. (2013) propose that the continuation of voicing throughout the closure duration in Russian /b, d, g/ is due to the active and controlled voicing produced by speakers which suggests [voice] as the active phonological feature in Russian. In contrast, the inconsistent voicing during closure in German is considered passive voicing that is a consequence of voicing bleeding from the preceding vowel, which suggests that /b, d, g/ in German are not specified by a laryngeal feature.

The third diagnostic of an active feature is the extent to which voicing assimilation applies (Al-Gamdi, Al-Tamimi, & Khattab, 2022; Beckman et al., 2013; Iverson & Salmons, 1995). That is, the behaviour, in terms of RVA, of voiced and voiceless plosives across word boundary. In an across word boundary plosive-plosive cluster environment, the extent to which the preceding plosive is triggered in voicing or devoicing is used to account for the active phonological feature[6]. This thesis, however, will focus mainly on the use of speech rate – and briefly on percentage of voicing in voiced plosives – as diagnostics to investigate the active phonological features in Arabic dialects, and leaves assimilation processes across different laryngeal categories in Arabic dialects to future work.

[6] For instance, if [_C $_{\text{voiced}}$ C $_{\text{voiceless}}$ _] triggers devoicing to the previous voiced plosive, and [_C $_{\text{voiceless}}$ C $_{\text{voiced}}$ _] triggers voicing to the preceding voiceless plosive in systems, then both [voice] and [spread glottis] are active phonological features in that system.

2.5 Voicing contrast: across Arabic dialects

The literature on VOT in Arabic plosives is gradually growing, and a picture is emerging of possible inter-dialectal variation in the mapping of VOT to the number of laryngeal contrasts. The Modern Standard Arabic plosive inventory, as seen in Table 2-2, consists of bilabial /b/, alveolar /t, d/ and their emphatic counterparts /ṭ, ḍ/, velar /k/, and uvular /q/. Most spoken dialects use the same plosives, but the status of /q/ is variable, in some dialects of Arabic the voiced velar /g/ is also used. Emphatic consonants in Arabic are a set of coronals that involve a secondary articulation. Researchers have suggested that this secondary articulation in emphatics is either velarization, uvularization, or pharyngealization. More information about the Arabic emphatics will follow in § 2.6.2.

Table 2-2 Surface plosive inventory of Modern Standard Arabic (emphatics in bold).

	Bilabial	**Alveolar**	**Velar**	**Uvular**
voiceless		t **ṭ**	k	q
voiced	b	d **ḍ**		

Cross-dialectally, the VOT of Arabic plosives has been studied for different purposes. Gender effects on VOT were studied in different Arabic dialects, for instance, in a south-western variety of Saudi Arabic (Al Malwi, 2017), Jordanian (Abudalbuh, 2010; Khattab, Al-Tamimi, & Heselwood, 2006), Syrian (Almbark, 2008), and Colloquial Egyptian (Rifaat, 2003). Three studies, on Lebanese (Khattab, 2002), Saudi (Abha variety) (Al Malwi, 2017), and Jordanian (Al-Tamimi, Tarawneh, & Howell, 2021) focused on the acquisition of Arabic VOTs. Some studies were more general in scope and investigated formant frequencies and the length of the vowel following the plosive, as well as VOT (Mitleb, 2009; Rifaat, 2003), while others focused solely on differences in VOT between the Arabic plain and emphatic plosives (Heselwood, 1996; Khattab et al., 2006; Kriba, 2010).

Single dialect studies on VOT have included work on Lebanese (Al-Tamimi & Khattab, 2018; Khattab, 2002; Yeni-Komshian, Caramazza, & Preston, 1977) where the results show minimal variation in the mean VOT values among the voiceless plosives /t, ṭ, k, q/ (short lag voiceless plosives), as opposed to the voiced plosives (voicing lead), suggesting that Lebanese might be a two-way voicing contrast dialect. Recent results from Al-Tamimi and Khattab (2018) also confirm the placement of Lebanese Arabic into the *true voice language* category displaying voicing lead in voiced plosives (mean -67.04 ms) vs. short lag VOT in voiceless ones (mean 8.70 ms). Similarly, Jesry (1996) reported parallel results for speakers of the Syrian dialect, indicating that it might also be a two-way voicing contrast dialect. Other dialects, Palestinian (Tamim, 2017) and Egyptian (Rifaat, 2003), reported similar results consistent with a two-way voicing contrast, at least in respect of VOT. So, there are a number of dialects in which the voiceless plosives are unaspirated, specifically /t/ and /ṭ/, and considered in the ‘short lag’ category (Bellem, 2014), as seen in Table 2-3 below.

Table 2-3 Summary of mean VOT values reported in studies on Levantine and Egyptian dialects.

Dialects		**Lebanese** (Al-Tamimi & Khattab, 2018)	**Lebanese** (Yeni-Komshian et al., 1977)	**Syrian** (Jesry, 1996)	**Palestinian** (Tamim, 2017)		**Egyptian** (Rifaat, 2003)
Method		word list and spontaneous speech	carrier sentence	carrier sentence	carrier sentence		word list
Position		both	initial	initial	initial	word-medial	initial
voiceless	/t/		25	25	25	17	
	/k/	9	28	32	41	28	30
	/ṭ/		23	24	22	18	
	/q/	--	30	29	--	--	--
voiced	/b/		-65	-70	-91	-64	--
	/d/	-67	-57	-67	-93	-55	
	/ḍ/		-60	-68	-94	-57	-77
	/g/	--	--	--	--	--	
contrast		2	2	2	2		2

In contrast, studies on more Eastern (Mashriqi) dialects, i.e., Iraqi and dialects spoken in the Arabian Peninsula, show a different pattern. VOT results from Iraqi dialects (Al-Ani, 1970; Bellem, 2007; Heselwood, 1996) show a pattern in the voiceless plosives in which /t/ and /k/ have noticeably longer VOT values (long lag) than /ṭ/ and /q/ that are unaspirated (short lag), as seen in Table 2-4. The Iraqi dialect results reported in both Al-Ani’s (1970) analysis and Bellem’s (2007) analysis of Alkalesi’s (2001)[7] production included the voiced plosives, and the VOT of the voiced plosives show an apparent voicing lead. Similar results indicating a three-way voicing contrast, based on VOT at least, is shown in Yemeni (Al-Nuzaili, 1993), and two varieties of Saudi: Abha and

[7] The VOT values reported for Alkalesi (2001) were taken from Bellem (2007), who analysed Baghdadi Arabic tokens that are taken from the CDs accompanying Alkalesi’s (2001) textbook about Modern Iraqi Arabic.

Najdi (Al Malwi, 2017; Bellem, 2007). The results show that both Yemeni and Saudi (Abha and Najdi) speakers have VOT values of /t/ and /k/ that are twice as long as /ṭ/ and /q/; consequently, these dialects display VOT values consistent with a three-way voicing contrast.

Table 2-4 Summary of mean VOT values reported in studies on the dialects of Iraq and the Arabian Peninsula[8].

Dialects		**Iraqi** (Al-Ani, 1970)	**Baghdadi** (Bellem, 2007)	**Baghdadi** (Heselwood, 1996)	**Yemini** (Al-Nuzaili, 1993)	**Saudi** (Abha) (Al Malwi, 2017)	**Saudi** (Najdi) (Bellem, 2007)
Method		isolated words	One male subject	in carrier phrase	isolated words one subject	picture naming task	from a database
Position		initial	initial	initial	initial	initial	both
voiceless	/t/	50	31	28	35	59	35
	/k/	70	39	--	46	59	44
	/ṭ/	25	11	15	10	15	16
	/q/	--	15	--	25	--	18
voiced	/b/	-85	-77	--	-70	-67	-63
	/d/	-90	-69	--	-76	-60	-66
	/ḍ/	-90	-99	--	-45	--	-84
	/g/	--	-55	--	-65	-68	--
contrast		3	3	3	3	3	3

It is important to note that we cannot assume that the dialect of a whole nation, or even a whole region, will show the same voicing patterns. In her analysis, Bellem (2014) reports several instances where, within the same country, the number of voicing contrasts differ between speakers from rural and urban regions. For instance, in the Syrian dialect, not all tested speakers had a two-way voicing contrast; some speakers from a north-eastern rural area produced a three-way contrast. According to Bellem, this was not surprising because the Levantine dialects vary along an urban – sedentary – Bedouin continuum[9]. It can therefore be misleading if one only mentions the nationality of the participants in a VOT investigation; thus, it is important to specify the speakers' background in detail rather than just mention that they are, for example, Jordanian or Syrian (Bellem, 2014).

Due to differences between studies in methodology and in the number of plosive contrasts included – and thus the lack of the comparative aspect – it is difficult to

[8] The reported results in the table for Alkalesi's (2001) VOT values of voiced plosives are averaged. Also, Al Malwi (2017) reported results in the table are the average of male and female VOT values reported in the study.
[9] This view will be discussed extensively below in § 2.6.1.

accurately determine the patterns of dialectal variation in voicing contrasts across Arabic dialects. Recently, more attention has been paid to Arabic cross-dialectal voicing categorization using comparative methods. Heselwood (1996) started this by investigating the voiceless emphatic and non-emphatic coronal plosives in two Arabic dialects: Baghdadi of Iraq and Cairene of Egypt. Heselwood documented a difference in the realization of the voiceless emphatic and non-emphatic among Baghdadi speakers where VOT values of /t/ and /ṭ/ are relatively far apart, while Cairene productions of the same two targets show closely related VOT values. He concluded that the variation was not an arbitrary difference, but rather a dialect feature. Similarly, in a comparative study of Cairene and San'aani phonology, Watson (2002) categorizes Cairene as having a two-way voicing contrast and San'aani with three-way voicing contrast. The most detailed study to date, Bellem (2007, 2014), classifies a range of Arabic dialects as either having a three-way voicing contrast or a two-way voicing contrast.

The acoustic analysis in Bellem (2007, 2014) was based on existing data from different sources[10] and comparisons from previous studies, hence the data are not directly parallel. Bellem (2014) focused on an aspect of emphasis that is not widely discussed, which is its voicing categories. Although the laryngeal system of Semitic languages is assumed diachronically to have involved a three-way contrast (Bellem, 2007), in a comparative analysis, Bellem was the first to articulate a typology of laryngeal contrast across Arabic dialects. Bellem classified dialects as dyadic or triadic. The dyadic dialects are those which display VOT values consistent with a two-way laryngeal contrast between voiceless and voiced plosives. As for the triadic dialects, those are the dialects which retain a three-way laryngeal contrast of voiced, voiceless plain, and voiceless emphatic plosives. The triadic dialects tend to be more conservative than the dyadic. This led Bellem to contextualize the typology, and further observe a correlation between this classification of the laryngeal contrast system, and what she refers to as, the 'dialect type'. Those dialects that have a triadic contrast prove all to be of a Bedouin or sedentary origin[11]; these, according to Bellem's observations, include San'aani, Baghdadi, some Ammani, rural-north-eastern Syrian, Saudi, Negev, and Fes and

[10] Bellem (2007) analysed recordings of one male Baghdadi speaker taken from CDs accompanying Alkalesi's (2001) textbook. In addition, Bellem (2007) analysed tokens from a database of seven male Saudi speakers from different regions in Saudi (mainly Najd). The database is part of a speech technology project at the Computer and Electronics Research Institute of King Abdulaziz City for Science and Technology in Saudi Arabia. See www.kacst.edu.sa.
[11] This will be discussed below in § 2.6.1.

Meknes Moroccan. Dyadic dialects are mostly urban; they include urban Syrian Damascene, Lebanese, and Cairene. Bellem notes that this variation is not idiolectal, but instead it is systematic dialectal variation.

None of these studies used speech rate, or any other diagnostic beyond surface VOT values, to determine the active laryngeal features involved in any of the Arabic dialects. Recently, Kulikov (2020) suggested that, like Swedish, Qatari is over-specified for both features [voice] and [spread glottis]. Kulikov examined the effects of speaking rate on five plosives in Qatari Arabic: voiced /b, d, g/ and voiceless /t, k/. He observed a significant effect of speaking rate on all examined plosives. Initially, Kulikov investigated only a subset of Qatari plosives, without including the emphatics in the (2020) study. In a later production and perception study, Kulikov (2021) examined three coronal plosives in Qatari Arabic /t/, /d/, and emphatic /ṭ/. The author concluded that in production the VOT values for the three plosives display a contrast between voiced /d/, voiceless unaspirated /ṭ/, and voiceless aspirated /t/. This pattern is consistent with a three-way voicing contrast, suggesting that the Qatari dialect falls in the set of so called ‘conservative’ dialects that retain a three-way contrast among plosives when including emphatics (Bellem, 2007; Watson, 2002). In this view, the Qatari laryngeal contrasts, like Thai, are not over-specified with two active features.

Considering this overview, we note the need to investigate the laryngeal system in more Arabic dialects with directly parallel data. In addition, there is a need to apply diagnostics, such as speech rate manipulation, to more Arabic dialects to test for the active laryngeal feature(s) in a context that varies among dialects. Lastly, we suggest that there is also a need to investigate what MSA, a different register, has to offer to our understanding of the laryngeal system of Arabic, and compare it to dialect speech. The purpose of this thesis is to address these needs.

2.6 Arabic Dialects

The Arabic Language may refer to the standard written form of the language or to the numerous existing regional spoken dialects. Yet, the use of the term in Arabic (اللغة العربية *allughatu al’arabiya*), to an Arab speaker, would imply the standard form of the language, which linguists often refer to as *Modern Standard Arabic* (MSA). MSA is the standardized, regulated variety that is used in formal written and spoken occasions, and it is the variety taught in schools. The mother tongue of an Arab speaker is not MSA; an

Arab child will speak the regional dialect that the child is surrounded by, from family and friends, while MSA is taught as part of the child's education (Watson, 2002).[12]

The regional dialects are used in daily spoken communication; they are also used to a limited extent in written forms, in informal online communication such as in blogs, chatrooms, forums, and texting. In addition, in recent years, a growing number of advertising billboards, and commercial signs in shops use the dialect form in their advertisements. This has increased the visibility of dialect speech as a written form in the landscape of many Arabic speaking countries, e.g., Kuwait (Akbar, Taqi, & Al-Gharabally, 2020), Tunisia (Ben Hamadi, 2019), Egypt (Plumlee, 2017). Arabic dialects, unlike MSA, are not governed by a prescriptive set of grammatical rules, though speakers of course have intuitions about grammatical and ungrammatical forms in their dialects. The various dialects are mostly mutually intelligible within the eastern-western divide. The extent of an individual's understanding of other dialects depends on the person's exposure to the other dialects' culture, media, and literature.

Classification of the regional dialects of Arabic can be complex or even in some cases arbitrary (Bellem, 2007; Versteegh, 2014). Typically, the classification depends partly on geography. Many linguists acknowledge broad geographical classification of Eastern vs. Western that is divided roughly at the Egyptian Delta[13], with the following main subgroups of Arabic dialects (Bellem, 2007; Holes, 2004; Versteegh, 2014; Watson 2002, 2011a):

- Eastern (Mashriqi)
    - **Arabian Peninsula**: Arabic spoken in the Gulf region; mainly in Saudi Arabia, Kuwait, Bahrain, Qatar, Emirates, and Oman. It also includes the Bedouin varieties and the Yemeni dialect.
    - **Mesopotamian**: Arabic spoken in Iraq.
    - **Levantine**: Arabic dialects spoken in the following countries: Lebanon, Palestine, Jordan, and Syria.
    - **Egyptian**: Arabic spoken in Egypt.
- Western (Maghrebi)

[12] More on MSA will follow in § 2.7 and § 5.1 of this thesis.
[13] Versteegh's (2014) distinction between the Eastern and Western dialects is characterized by the isogloss of the prefix *n*- as the first-person singular of the imperfect verb, e.g., Western (Maghrebi) /nəktəb/ vs. Eastern (Mashriqi) /ʔəktob/ 'I write' (p.178).

- **Maghrebi**: Arabic dialects spoken in North Africa, including Morocco, Algeria, Tunisia, and Libya.

Other dialects in the Arab world are often regarded as their own class e.g., Sudanese, and the Arabic of Mauritania[14]. Although they are all dialects of the same language, there are many differences in the phonetic-phonological, rhythmical, and lexical structures across these varieties (Watson, 2011a).

2.6.1 The typology of Bedouin vs. sedentary

Alongside the geographical Eastern ~ Western classification, consideration is also given to 'ecolinguistic' classification of Arabic. This classification ties the cultural, social, and geographical aspects to the linguistic features used by speech communities. Cadora (1992) introduced the notion of 'ecolingustics' referring to the linguistic correlates that are related to environmental differences in speech communities, based on three ecological structures in the Arab world: nomadic (hereafter Bedouin), sedentary rural, and sedentary urban. This ecological structural development in Arab communities is mainly due to contact, Bedouin migration and/or settlement in rural and urban places. Consequently, the changes experienced by these communities caused them to linguistically adapt. Cadora (1992) presents transitional systems of ecological structure and their linguistic adaptation (p.1):

Bedouin → Bedouin-Rural → Rural → Rural-Urban → Urban

Bedouin dialects, originating from the heart of the Arabian Peninsula, were regarded in the early time of the Islamic empire as the form of Arabic that most truly represents the classical Arabic of the Holy Quran and pre-Islamic poetry (Holes, 2004; Versteegh, 2014). There is no strict areal definition for Bedouin dialects, but they are often recognized by their tribal affiliation (Bellem, 2007; Ingham, 1994). Waves of migration, and settlement by some of the Bedouin tribes, created a dialect dichotomy between Bedouin vs. sedentary speech communities in many areas of the Arab world. The effect of Bedouin migration is seen in Iraq, along with tribally defined[15] speech communities in Kuwait, Bahrain, Qatar, and the Emirates. Similarly, the Syro-Mesopotamian dialects include Bedouin dialects of Jordan and Syria (Holes, 1990, 2006; Ingham, 1994).

[14] Although some systems classify the Sudanese dialect with the Egyptian dialect and Mauritanian Arabic with Maghrebi dialects (Versteegh, 2014; Watson, 2011a).
[15] Tribes from the North-east (Najd) of the Peninsula e.g., the tribes of Shammar and Aniza (Ingham, 1994).

Although the linguistic features of the Bedouin dialects nowadays are clearly of what is termed the 'New Arabic' type, i.e., exhibiting changes from the Classical Arabic in that they do not have grammatical case endings, nevertheless they are still considered more linguistically conservative than sedentary dialects. As stated by Holes (2006):

> It is not that the dialects of central Arabia have not changed, they have; but until very recently the changes have been mainly the result of internal evolutionary processes to which all languages are subject, rather than a consequence of invasions, immigration, or most disruptive of all, the mass learning of the language by conquered foreigners (p. 27).

Sedentary dialects, in contrast, developed through contact, mainly via Bedouin migration and settlement outside the Peninsula before and after the Islamic conquests in the seventh century. Speakers of Arabic already living in these areas before the Bedouin arrived were primarily villagers who either farmed, fished, or wove, and had been in the area for a long period of time. Their sedentary rural dialects shared some features that separate the sedentary dialects from the dialect of the Bedouin incomers (Holes, 2006; Versteegh, 2014). Sedentary urban dialects, however, were spoken mainly by Christians and Jews outside the Arabian Peninsula, and these dialects were highly changing and innovative. In terms of prestige, these sedentary dialects were not as prestigious as Bedouin dialects in the early period after of the Islamic conquests. Yet, urban dialects nowadays in large urban centres have attracted the focus of civilization and gained power (Versteegh, 2014). Holes (2004) adds to the Bedouin ~ sedentary distinction a third socioeconomical group which he refers to as 'ruralite'. This group, mainly in the Levant, is different from both the 'Bedouin' type and the 'urban/city' type in that this third speech community consists of long-established farmers in villages.

This classification of Arabic dialects is examined in the literature about Arabic and can be termed as ecolinguistic (Cadora, 1992), lifestyle distinction (Youssef, 2021), dialect type (Bellem, 2014), or dialect group (Palva, 1991). The classification is reflected in typological similarities among Arabic dialects, particularly in terms of laryngeal contrasts as we have seen (Bellem, 2007; Watson, 2011a). A further characteristic of relevance to this thesis is the degree of emphasis and the laryngeal production of emphatics in these two groups of dialects. Bedouin dialects tend to have greater cues to emphasis, including a greater VOT distinction between the voiceless emphatic and non-emphatic than sedentary rural and urban dialects. This underpins the description of

Bedouin dialects as ‘conservative’ (Bellem, 2007; Watson, 2002). That is, the Bedouin dialects are still preserving the Classical Arabic ‘three-way’ laryngeal contrast, in contrast to ‘innovative’ dialects where laryngeal features of emphatics seem to be merging with non-emphatic counterparts. The classification of ‘conservative’ and ‘innovative’ (or ‘progressive’ (Watson, 2002)) comes from the development of the obstruent system in the innovative sedentary urban dialects, in that they merge the laryngeal distinction of emphatic and non-emphatic voiceless plosives (e.g., Cairene and Lebanese). In contrast, in conservative dialects with a three-way laryngeal contrast, the voiceless emphatic plosive is in fact not a true counterpart of the voiceless non-emphatic one because the contrast is not minimal (Bellem, 2007, p. 132).

The central status of emphatics in the categorization of Arabic dialects invites us to review the emphatics of Arabic in more detail. The following section § 2.6.2 reviews the Arabic emphatics and debates on the number of emphatic consonants, as well as the acoustic and auditory cues to emphasis.

2.6.2 The Arabic emphatics

The Arabic ‘emphatics’ have attracted the attention of linguists for a long time. There have been prolonged debates on how many emphatic consonants there are in Arabic, the realization of their secondary articulation, and whether emphasis is a segmental or suprasegmental property. However, there is a general agreement that there are, at least, emphatic counterparts to four plain coronal obstruents in MSA /t, d, ð, s/ which exhibit a simultaneous secondary articulation somewhere between the post-velar and pharyngeal part of the tract (Khattab et al., 2006, p. 141). In this thesis, in common with other works on Arabic (e.g., (Owens, 2013)), emphatic plosives are not represented in a strict IPA manner. Instead, a dot underneath the represented sound is used to distinguish emphatics from non-emphatic counterparts in order to avoid an a priori assumption about how emphatics are realized phonetically. See e.g., (Israel, Proctor, Goldstein, Iskarous, & Narayanan, 2012; McCarthy, 1994; Watson, 1999).

Emphasis is an actively employed phonological feature in major modern Arabic dialects but can be realised differently in different dialects. That is, emphasis may not share the same articulatory, acoustic, and perceptual correlates in all Arabic dialects. Dialects also differ in terms of how many, and which, emphatics are present in their consonant inventory. Conservative (Bedouin) varieties tend to have three emphatic coronals in their inventory /ḍ̱, ṣ, ṭ/, whereas progressive (urban) dialects may have some or all the following emphatic coronals /ḍ, ṣ, ṭ, ẓ/. Some Arabic dialects also produce emphatic counterparts of other coronal and non-coronal consonants /b, m, n, l, r/ (Jakobson 1957; Lehn, 1963; Youssef, 2014). Appendix (A.1) in this thesis provides a full list of the cognate consonant reflexes in Arabic dialects.

There are competing views on the secondary simultaneous articulation involved in emphasis and on where in the vocal tract the articulation is triggered. Instrumental advances in recent years have helped in determining the articulatory characteristics of emphasis. The majority of scholars have reported pharyngealization as the articulatory configuration involved in emphatics across Arabic dialects (Al-Ani & El-Dalee, 1983; Al-Ani, 1970; Al-Tamimi, Alzoubi, & Tarawnah, 2009; Ali & Daniloff, 1972; Davis, 1995; Ghazeli, 1977; Hassan, 2005; Herzallah, 1991; Laufer & Baer, 1988; Lehn, 1963; Wahba, 1996; Younes, 1982). Some studies focus on the backwards movement of the tongue dorsum towards the upper pharynx, since the upper pharyngeal wall is not active when articulating the emphatics (Al-Ani & El-Dalee, 1983; Ali & Daniloff, 1972). Lehn (1963) notes that when articulating the emphatics both tongue retraction and muscular tension are involved. Other researchers report an associated movement of the lips in the production of emphatics in some dialects (Hetzron, 2013; Jakobson 1957). For instance, in the Egyptian dialect of Arabic, Lehn (1963) reports a degree of lip protrusion and rounding.

However, a number of studies proposed that, either along with or instead of pharyngealization, velarization takes place (Hetzron, 2013; Norlin, 1987), or uvularization (McCarthy, 1994; Zawaydeh, 1998). There is not necessarily one secondary articulatory feature involved in the production of emphatics in comparison to the non-emphatic counterparts, so the general use of either ‘pharyngealization’, ‘velarization’, or ‘uvularization’ as one term describing the phenomenon of emphasis has the potential to be misleading. Furthermore, trying to define the articulation of emphatics is troublesome, hence the disagreement of researchers on whether it is generally pharyngeal, upper pharyngeal ‘uvularization’, or lower pharyngeal ‘pharyngealization’. As a result, many scholars prefer to use the term ‘pharyngealization’ due to the general role of the pharynx in articulating the emphatics (Bellem, 2007, p. 45).

There have also been debates over whether emphasis is a property that is consonantal, vocalic, or suprasegmental. Within this, the phonological feature representing emphasis has been debated, to be either a feature of the consonant or the vowel. A typical generative phonology view is of emphasis as a consonantal feature of primary coronal emphatics, and that this feature tends to spread (Bellem, 2007; Youssef, 2014). The initial description of emphasis within the framework of distinctive phonology features was an acoustic feature [flat], by Jakobson (1957), referring to the observed flattening of the acoustic spectrum, generally speaking. However, movement towards more articulatory distinctive features argued against [flat] as also being a feature for labialization and retroflexion (Bellem, 2007). Many articulatory features have been proposed for the Arabic emphatics including [CP] i.e., constricted pharynx (Hoberman, 1987), [pharyngeal] (Herzallah, 1991; McCarthy, 1994), and [RTR] i.e., retracted tongue root (Davis, 1995; Goad, 1991; Shahin, 1996; Zawaydeh, 1998). Accounts of emphasis as prosodic rather than segmental include Lehn’s (1963) analysis of emphasis as a phonologic component that takes the whole syllable as its domain.

Relatively few acoustic studies have examined VOT values as potential cues to the emphatic ~ non-emphatic contrast. Even fewer studies have incorporated this difference into phonological representation of emphatics. Generally, as discussed earlier, VOT values are greater in voiceless non-emphatic coronals and shorter in emphatic counterparts in some Arabic dialects like Iraqi, Jordanian, Saudi, Qatari, and Yemeni (Al Malwi, 2017; Al-Ani, 1970; Al-Nuzaili, 1993; Bellem, 2007; Bukshaisha, 1985; Heselwood, 1996; Khattab et al., 2006). Yet, in other dialects, like Egyptian, Lebanese, and Syrian, there is no consistent difference between emphatic and non-emphatic coronals in terms of their VOT values (Heselwood, 1996; Rifaat, 2003; Yeni-Komshian et al., 1977).

Characterization of emphasis as a feature that spreads arises from acoustic and auditory cues beyond the emphatic segment itself. In terms of formant frequencies of the following vowel, it is generally observed that when producing the emphatics, there is a significant lowering of the second formant as a result of an enlarged oral cavity, and some degree of raising of the first formant as a result of a reduced pharyngeal cavity caused by retracted tongue root (Bellem, 2007, p. 46; Watson, 2002, p. 269). The majority of acoustic studies reported these results (significant F2 lowering and slight F1 raising) in different dialects of Arabic, including Jordanian (Abudalbuh, 2010; Al-Masri

& Jongman, 2004; Ghazeli, 1977; Khattab et al., 2006), Palestinian (Card, 1983), and Alexandrian Egyptian (Wahba, 1996) among other dialects (Al-Ani, 1970; Bukshaisha, 1985; Heselwood, 1996; Laufer & Baer, 1988).

While F2 lowering in vowels following emphatic consonants is a robust indicator of emphasis, the lowering effect is not the same for all three phonemic vowel qualities in Arabic. When analysing the acoustic effects of emphasis on /a/, /u/, and /i/ in Jordanian Arabic, Jongman et al. (2011) investigate lowering of F2 after emphatics in all three vowels, at vowel onset, midpoint, and offset. The interaction of vowel quality with emphasis was significant for the various measurement points, in that the effect of F2 lowering in emphasis was stronger and more enduring through the vowel in /a/, than in /i/ and /u/. Jongman et al. (2011) also reported the effect of emphasis on F1, which was more prominent in /a/, than in /i/ and /u/. In addition, results of the perception task in Jongman et al. (2011), with cross-spliced data, show that perception of emphasis heavily relies on the following vowel more than the target itself.

As much as Arabic emphatics are discussed in the literature, there is still, as we mentioned above, little discussion on one aspect of emphasis, i.e., its laryngeal categories. The current thesis will discuss this, across Arabic dialects. Modern Standard Arabic is particularly interesting in this matter because it is said to retain all four coronal emphatics /ḍ, ṭ, ð̣, ṣ/, and so may shed further light.

2.7 Dialectal influence on MSA

Given the diglossic situation of Arabic[16], it is intriguing to ask whether dialectal variation in patterns of VOT values is reflected in the speakers' production of MSA. Generally, it is widely assumed that phonological and phonetic properties in speakers' dialect are detectable in their MSA (Gibson, 2002). Benkirane (1998) suggested that in spoken MSA, speakers will apply the stress assignment rules of their colloquial dialect. Yet, evidence for this claim is in fact lacking in the literature. Particularly on the prosodic level, the native dialect is assumed to influence the production of MSA, hence audible variation when spoken by different mother tongue dialects. ElZarka & Hellmuth (2009) tested this generalization with respect to intonational properties, and they found

[16] More on this in chapter 5, § 5.1 of this thesis.

that Egyptian dialect and MSA production are similar in some intonational properties, but not others.

Elgibali (1993) examines and compares the relative stability of Badawi's (1973) proposed five levels of Arabic diglossia (Classical Arabic, MSA, educated colloquial, literate colloquial, and illiterate colloquial) in two Arabic dialects: Kuwaiti and Cairene. He analyzes certain salient variables, including phonological variables (the use of /q/ and /θ/), in which the standard variety differs from the dialectal ones based on recorded data of read speech and spontaneous speech in natural conversations. Elgibali calculated the ratio of occurrence of the different realizations of the MSA /q/ and /θ/ in both Cairene (where /q/ can be realized as [ʔ] and /θ/ as [s] or [t]), and Kuwaiti (where /q/ can be realized as [g]). His results indicate that, phonologically, there are substantial differences cross-dialectally. Cairene use of [s] instead of standard /θ/ in MSA production is higher than the use of [ʔ] as /q/; suggesting that /θ/ use is the best indicator for the degree of formality. In contrast, Kuwaiti MSA use of /q/ is a better diagnostic for the degree of formality.

In their review, Khamis-Dakwar & Froud (2019) discussed the role of diglossia in child language development in Arabic speaking communities. They cite Amayreh's (2003) examination of the trajectory of MSA consonant development in Jordanian Arabic children, at the point when exposure to MSA intensifies as they start learning MSA officially in school. Amayreh examined consonants of both the dialect and MSA in a picture naming task, and the results showed that, despite intensified exposure to MSA in schools, some children replaced MSA consonants with their familiar dialect ones even though they were asked to use MSA in this task. The results of both Elgibali (1993) and Amayreh (2003) suggest that there is dialectal influence on the production of MSA, and that the degree of influence is not the same across dialects.

The only study I have found on VOT in MSA is AlDahri's (2013) investigation of highly trained Quranic reciters that reported the unexpected result of a two-way *aspirating* contrast, with no voicing lead in any of the plosives. This is a pattern that is not present in any dialect reviewed above. We should also note that many of the previously mentioned studies on VOT were framed at the time as studies of 'Arabic'; the carrier sentences in most of them, as well as the target words, were in Standard Arabic, and the stimulus was presented in Arabic orthography. Nevertheless, speakers may, to some degree, be aware of some of the pronunciation differences between MSA and their native dialect, but not others. For example, if a speaker aims to speak the 'Standard' Arabic, factors like VOT might be unconscious and thus uncontrolled (Bellem, 2014). That might predict that the underlying laryngeal system for Arabic is the same in both MSA and dialect speech. This prediction underpins the need to investigate the laryngeal categories in MSA production and compare it to the dialectal production.

2.8 Purpose of the thesis

The aim of this thesis is to clarify the typology of dialectal variation in VOT through parallel investigation of different Arabic dialects, which are measured in a consistent manner, using the same plosives, and measuring different variables consistently. A major gap in the literature is a comprehensive, consistent, and systematic cross-dialectal examination of this aspect of the Arabic language.

There is no directly parallel evidence of the merger of VOT values in the voiceless emphatics and non-emphatic plosives in the relevant dialects of Arabic. In addition, there is no parallel evidence of which phonological features are active in the laryngeal systems of Arabic dialects. These gaps are the first motivation of this attempt at a comprehensive cross-dialectal overview. The originality of this thesis lies in its multidimensional approach through which, in three experimental studies, it explores and compares Arabic variation across different dialects, across different speaking rates, and across the two registers (dialect and MSA).

The corpus-based study in chapter 3 looks at VOT in emphatics, and their plain counterparts, to provide directly parallel evidence regarding the degree of variation across dialects. This will help us determine the typology of dialect varieties in terms of their laryngeal categories; that is, which dialects display a two-way voicing contrast and which display a three-way contrast.

The first experimental study in chapter 4 tests for the active underlying phonological feature(s) in three dialects expected to display different numbers of laryngeal categories, by manipulating speaking rate. The study aims to explore whether two-way voicing dialects are over-specified with two active phonological features.

The second experimental study in chapter 5 examines MSA speech produced by the same speakers investigated in chapter 4, to test the active underlying phonological feature(s) in MSA production. This will give insight into the laryngeal phonology of MSA and whether it differs by speaker origin.

To summarize, this thesis aims to answer the following research questions:

- **1a)** By examining VOT in emphatic and non-emphatic coronal plosives in eight dialects, what is the number of laryngeal contrasts in each dialect?
- **1b)** Does the degree of F2 lowering in the following vowel co-vary with the presence or absence of VOT as a cue to the emphatic ~ non-emphatic contrast?
- **2a)** How does speaking rate affect VOT in different voicing categories, and which laryngeal features are active in each dialect?
- **2b)** Are two-way voicing category dialects over-specified with two active phonological features?
- **3a)** Are the observed VOT patterns in dialect speech mirrored in MSA produced by the same speakers?
- **3b)** Do speakers of a dialect display the same active feature(s) in the two registers of Arabic?

In a Laboratory Phonology approach, then, this thesis aims to explore the laryngeal phonology of Arabic taking into account its different dialects and its diglossic nature. The first step to accomplish this is to explore and typologize the number of laryngeal categories in Arabic dialects. The following chapter (Chapter 3) will examine VOT and other cues to laryngeal and emphatic contrast in the coronal set of plosives, both plain /t, d/ and emphatic /ṭ, ḍ/, in eight Arabic dialects.

Chapter III

Cross-dialectal Corpus-based Study

3 Cross-dialectal corpus-based study

This chapter aims to classify eight dialects of Arabic into Bellem's (2014) laryngeal contrast typology of Arabic dialects in either having a two-way or three-way voicing contrast. VOT values in four plosive consonants (plain /t/ and /d/ and their emphatic counterparts, /ṭ/ and /ḍ/) in the eight dialects of the Intonational Variation in Arabic (IVAr) corpus (Hellmuth & Almbark, 2019) elicited and measured in a consistent manner confirms this observed inter-dialectal variation and typology in terms of the number of laryngeal contrasts. Instead of a clear-cut dichotomy, the results show a continuum of variation across the different synchronic grammars: some dialects display a three-way voicing distinction and other dialects display a two-way voicing distinction, while there are dialects in the middle. This might suggest a potential sound change; some synchronic grammars retain the diachronic Proto-Semitic three-way voicing distinction, but other dialects now show a two-way voicing distinction, and others are in the process of change from a three-way to a two-way distinction. In this chapter, I will begin to explore the hypothesis that: synchronic variation is, or might be, an indication of sound change in progress, based on synchronic corpus data.

Examination of formant frequencies in the following vowels is used to determine whether there is any sort of compensation for the relative presence/absence of VOT as a laryngeal cue to the emphatic contrast and in the strength of formant differences following plain/emphatic plosives. Our hypothesis is that there will be no compensation or trade-off, based on the observations of perception studies that concluded the reliance of listeners on vocalic cues of the following vowels to determine the presence or absence of an emphatic consonant (Hayes-Harb & Durham, 2016; Jongman et al., 2011).

Our interest is focused not only on how speakers realize the emphatic/plain opposition, but also on how speakers keep these oppositions phonetically distinct in speech. These observations of differences across dialects might be useful for practical purposes such as accent/dialect detection for forensic or artificial intelligence (AI) applications.

In this chapter, I first discuss the methods in § 3.1, then the process of data analysis in § 3.2. In § 3.3 I present the results, then in § 3.4 I discuss these results in light of the literature review in chapter 2. Finally, § 3.5 provides a brief summary of this chapter and introduces the following chapter.

3.1 Methods

In this study, for the purpose of comprehensiveness, cohesiveness, and direct comparison, data from eight representative dialects from the Intonational Variation in Arabic (IVAr) corpus are studied to answer the following research questions: **1a)** By examining VOT in emphatic and non-emphatic coronal plosives in eight dialects, what is the number of laryngeal contrasts in each dialect? **1b)** Does the degree of F2 lowering in the following vowel co-vary with the presence or absence of VOT as a cue to the emphatic ~ non-emphatic contrast?

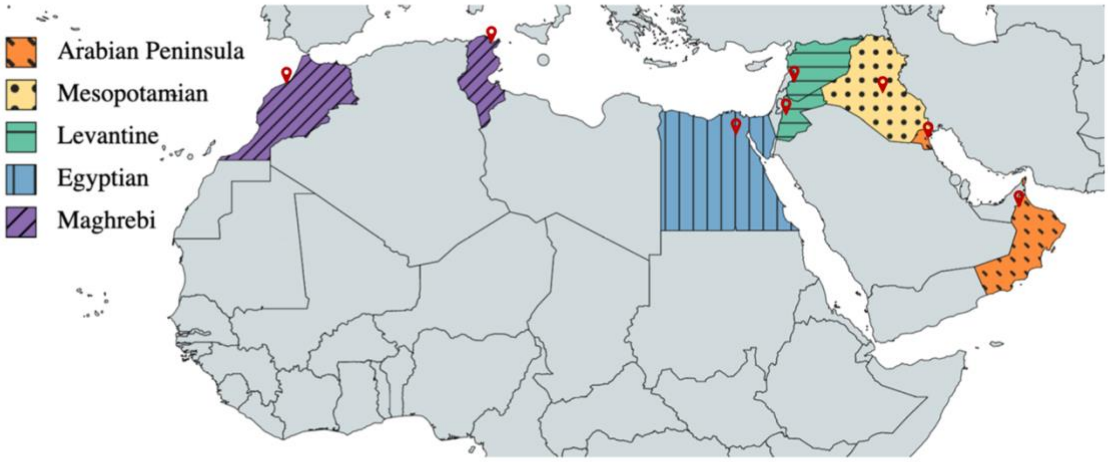
IVAr is a project initiated by the Department of Language and Linguistic Science at the University of York that provides an open access corpus of Arabic speech (Hellmuth & Almbark, 2019). The corpus includes the production of sentences, stories, and conversations in eight colloquial Arabic dialects. At least one dialect from each regional dialect group is included. The data analysed in this chapter are supplementary word list data collected at the same time as main IVAr corpus with the same participants, but the supplementary data are not published in the IVAr corpus.

3.1.1 Participants

The IVAr corpus provides data collected from 88 speakers in total, typically with 12 speakers (6 female / 6 male) per dialect, except for Iraqi (6 female / 4 male), Syrian (3 female /3 male), and Egyptian (5 female / 7 male). The speakers represent eight regionally defined varieties of Arabic as listed in Table 3-1 and Map 3-1. All speakers were aged 18 years or over and provided their consent for use and distribution of their speech data. Recording sessions took place on location in the Middle East and North Africa, in the town or city of residence of speakers. Participants were, in all cases, born and raised in the stated city/region, and in the majority of cases were also resident in that city; all speakers from Damascus and Baghdad were resident (and recorded) in Amman, Jordan (Hellmuth & Almbark, 2019).

Table 3-1 Abbreviations of the dialects selected in the study and the number of participants in each dialect.

code	dialect	F	M
moca	Moroccan (Casablanca)	6	6
tuns	Tunisian (Tunis)	6	6
egca	Egyptian (Cairo)	5	7
joka	Jordanian (Karak)	6	6
syda	Syrian (Damascus)	3	3
irba	Iraqi (Muslim Baghdadi)	6	4
kwur	Kuwaiti (Urban)	6	6
omba	Omani (Buraimi)	6	6



Map 3-1 Location of the Arabic dialects in the IVAr corpus[17].

3.1.2 Materials

The experiment investigates four contrastive plosives that appear in the Arabic language, two of which are voiced /d/ and /ḍ/[18], and two are voiceless /t/ and /ṭ/. The choice of plosives was restricted to those that have a counterpart across the voicing contrast in all dialects investigated (i.e., /b/ and /k/ were not included). These plosives are investigated in word-initial and word-medial positions, followed by three long vowels /a:, i:, u:/ and three short vowels /a, i, u/. The total number of targets is 45 as demonstrated in Table 3-2 below. Stimuli for each dialect are minimal pairs, where

[17] Map customized and zoomed in from https://mapchart.net/

[18] It is important to note that the alveolar emphatic plosive /ḍ/ has varying surface reflexes in the modern dialects. The emphatic /ḍ/ is realized as the plosive /ḍ/, the fricative /ẓ/, or the interdental /ð̣/ in different dialects (Youssef, 2021, p. 9). In this thesis, stimuli are elicited using the Arabic orthographic form of /ḍ/ (i.e., ض) for completeness, regardless of how /ḍ/ is produced in the target dialects. Please see § 3.2.1 on how I treated /ḍ/ when produced as the interdental /ð̣/ in some dialects.

possible, across the emphatic contrast, so the target word list contains a mix of real and nonsense words. Also, see Appendix (B.1) and Appendix (B.2) for full list of target words and English gloss.

Table 3-2 The total count of target words per target consonant and position.

target	**position**		**total**
	word-initial	word-medial	
/t/	9	1	10
/ṭ/	9	2	11
/d/	7	5	12
/ḍ/	6	6	12
total	31	14	45

3.1.3 Procedure

The recording sessions were run by a paid local fieldwork assistant whose first language was the dialect in question. Recordings were made using a Marantz PMD661 solid state data recorder directly to digital format (.wav) at 44.1kHz 16 bit, using Shure SM10A-CN head-worn dynamic cardioid microphones.

Target words are presented to the subjects in Arabic orthography as shown in Appendix (B.1) and Appendix (B.2) within a carrier phrase that translates to "write __ twice" in all cases. The carrier phrase varied in lexical choice among dialects to foster dialectal production, using the informal orthographic norms of each dialect in order to direct participants towards producing the phrases in the colloquial register and divert them from producing the formal register (Modern Standard Arabic) as seen in Appendix (B.3) (Siemund et al., 2002). The target word list was presented to participants on printed paper sheets in pseudo-random order, among a large number of distractor utterances elicited for other purposes in the corpus.

3.2 Data Analysis

The scripted speech data recordings were segmented, using Praat (Boersma & Weenink, 2016), into individual tokens for further analysis. The total number of tokens for all dialects were potentially 3960 (88 speakers x 45 target items). The dataset included 692 disfluent tokens; these tokens were excluded from the analysis leaving 3268 tokens.

3.2.1 VOT

Using Prosody Lab Aligner (Gorman, Howell, & Wagner, 2011), an orthographic transcription was force-aligned to the audio in Praat textgrids. The resulting levels of segmentation in the textgrids – two interval tiers, one for phones and one for words – were used to identify word duration and assist in labelling VOT in target consonants and formants in following vowels. In some instances, the word/phone alignment for the target word and/or phone were inaccurate and had to be manually corrected. The word-level segmentation was used for measuring the total duration of the target word in milliseconds. To measure VOT, a point tier was added using a Praat script where two boundaries were inserted in the tier, one labelled (B) for the plosive burst and the other labelled (V) for the onset of periodic striations of vocal folds vibration, as seen in Figure 3-1 below. The position of these two boundaries in the point tier was manually adjusted in each textgrid to the place of the burst or vocal fold vibration onset, based on the visual inspection of the waveform and spectrogram displays in Praat. Another Praat script was then used to extract word duration from the word-level tier and VOT duration from the added point tier.

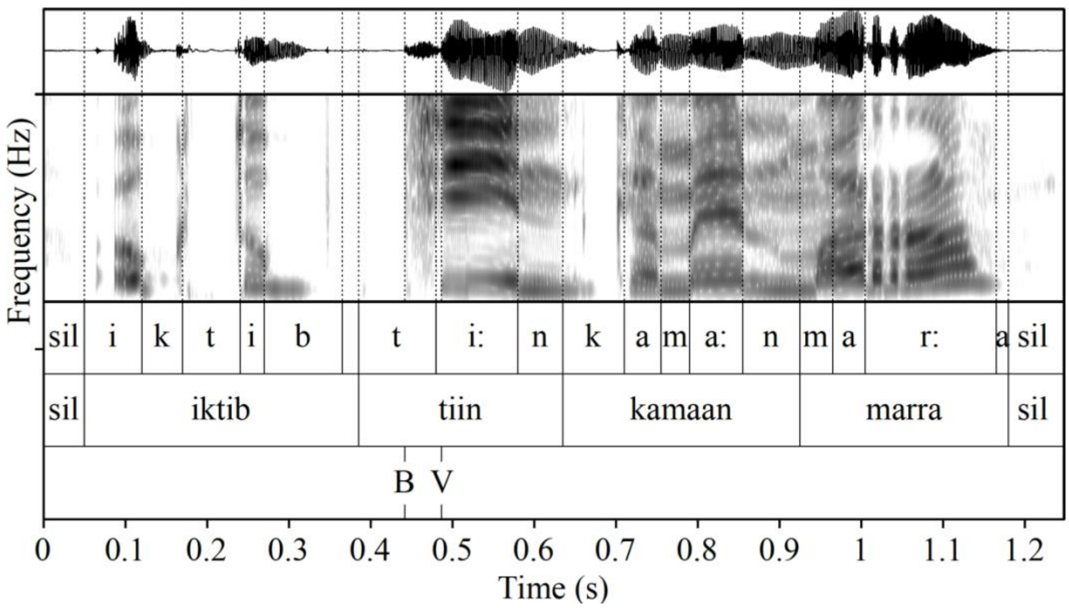


Figure 3-1 Segmentation of the word /ti:n/ produced by an Egyptian female speaker, with tiers (1) phones, (2) words, and (3) VOT points.

The extracted measurements were further analysed using R (R Core Team, 2018) to produce a visualization of the variation in VOT values across dialects in both voiceless and voiced plosives. VOT plots were produced using *ggplot2* (Wickham, 2016). The results were explored in a series of Linear Mixed-Effects Models (LMM) using the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) with *VOT* as the dependent variable, and the following factors as predictors: *dialect* (the eight dialects listed above in Table 3-1); *target* (/t/, /ṭ/, /d/, /ḍ/); *position* (initial, medial); *gender* (male, female); *voicing* (voiced, unvoiced); and *type* (plain, emphatic). The categorical factors were sum-coded to centre our categorical predictors around the mean. *Speaker* and *item* were included as random intercepts in all models. To determine the best fit model, the likelihood ratio comparison test was used.[19]

It is a particular challenge for this design that the chosen dialects, as we mentioned earlier, treat the /ḍ/ and /ð̣/ merger differently. In all the dialects, target words were elicited with /ḍ/ in Arabic orthography (i.e., ض) as shown in Appendix (B.1) and Appendix (B.2). It was noticed during the segmentation process that some speakers in some dialects (e.g., Kuwaiti, Omani, Iraqi, Jordanian, and Tunisian) produced the interdental fricative /ð̣/, and these tokens were discarded. Other instances were produced as an interdental fricative yet there appeared an obvious burst. These instances were kept. Figure 3-2 is an example of the segmentation procedure followed in each instance of an interdental fricative with a burst.

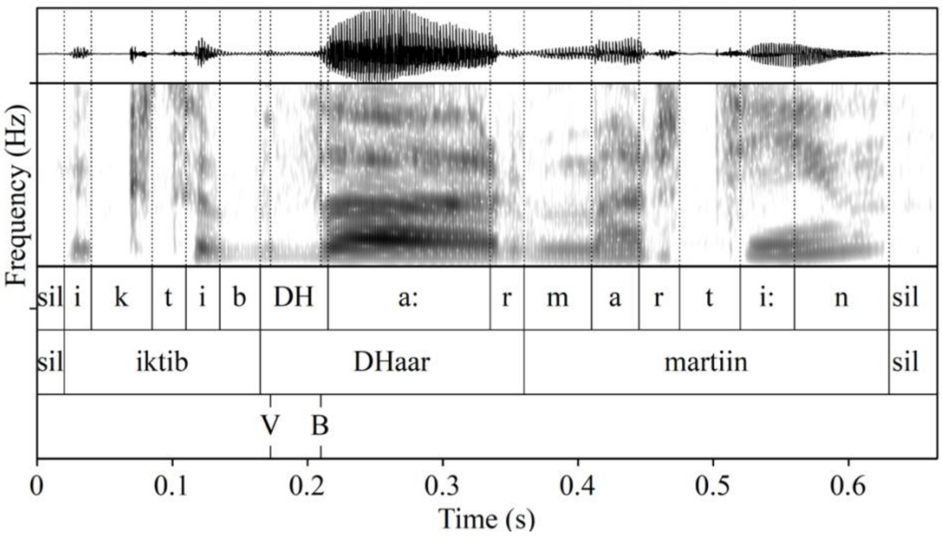


Figure 3-2 Segmentation of the word /ḍa:r/ produced as a voiced interdental fricative with a burst by a Tunisian female speaker, with tiers (1) phones, (2) words, and (3) VOT points.

[19] The resulted best fit models for the voiced and the voiceless subsets:
vot ~ dialect * target * gender + position + (1|item) + (1|subject), data= voiced
vot ~ dialect * target * gender + position + (1|item) + (1|subject), data= unvoiced

3.2.2 Other cues to emphasis

In addition to VOT, F1 and F2 values in the vowel following the target plosive were also obtained. I followed the single-point measurement approach in using a Praat script to extract formant resonance frequencies at the midpoint of the vowel to represent the central tendency of that vowel. This simple measure of vowel quality is used (F1/F2 at the midpoint of the vowel) to provide a first indication whether dialectal variation in VOT of plain/emphatic plosives is matched by variation in F1/F2 on the following vowel. It is worth noting that F1 and F2 are not the only acoustic features of vowels, and by only plotting formant values at the midpoint of the vowels we may not fully specify the information that is perceptually salient to listeners. To allow for direct visual comparison and detection of outliers, the formant frequencies in each dialect and target were box plotted in R (R Core Team, 2018), then plotted separately in a standard F1xF2 plane using *ggplot2* (Wickham, 2016). Outliers were then removed manually by looking at extreme values on the boxplots.

For demographic considerations, that is, to remove the influence of gender physiological differences, formant measurements of vowels are normalized (Foulkes, Scobbie, & Watt, 2010). Using NORM (Thomas & Kendall, 2007), the dataset was normalized following the Lobanov (1971) z-scores method of normalization given that the data were vowel-extrinsic but formant-intrinsic. That is, the normalization algorithm applied is to more than one vowel but a single formant value in each vowel. To calculate the distance between vowels following plain versus emphatic consonants in F1 x F2 space, we considered using Euclidian Distance (the distance between these two points, e.g., plain /a/ and emphatic /a/) represented by both coordinates of F1 and F2, but there were some missing values which prevented use of this approach.

The F1/F2 midpoint data were explored in a series of linear mixed-effects models (LMM) using *lme4* (Bates et al., 2015) in *R* (R Core Team, 2018), with each acoustic measure in turn as dependent variable (F1/F2), *dialect* and *target* (/t/, /ṭ/, /d/, /ḍ/) and their interaction as fixed factors, and a random intercept for *item* and *speaker*.[20]

[20] normf2 ~ dialect * target * gender + position + vowel + (1|item) + (1|subject), data= unvoiced

3.3 Results

3.3.1 Overview

The boxplots in Figure 3-3 below illustrate values of VOT in the investigated plosives (/t ~ ṭ/, /d ~ ḍ/) in all eight dialects. All dialects in this figure are positioned according to the manually calculated difference in the mean between /t/ and /ṭ/ in each dialect. There is a clear split between voiced and voiceless plosives in all dialects, with the phonologically voiced plosives /d/ and /ḍ/ showing clear voicing lead during closure, while the phonologically voiceless plosives /t/ and /ṭ/ show voicing lag. In /d/ and /ḍ/, all dialects display voicing lead, in this simple measure of VOT, with similar values regardless of the plain/emphatic contrast. In contrast, in the voiceless subset, there is variation between dialects in the degree of voicing lag between /t/ and /ṭ/.

Looking closer at the main locus of variation between dialects, which is the voiceless subset /t/ and /ṭ/, we notice that the range of values for /t/ and /ṭ/ is nearly coextensive in some dialects, such as in the Egyptian, Syrian, Tunisian, and Omani dialects; this suggests that these dialects have a two-way distinction in VOT only (voiced/voiceless). Whereas other dialects have a discrete range of VOT values for /t/ and /ṭ/, as in the case of Moroccan, Kuwaiti, Jordanian, and Iraqi, suggesting that these dialects have a three-way VOT distinction (voiced/ long lag voiceless/ short lag voiceless). This grouping is in accordance with the dichotomy, proposed by Bellem (2007, 2014), between ‘progressive’ and ‘conservative’ dialects.

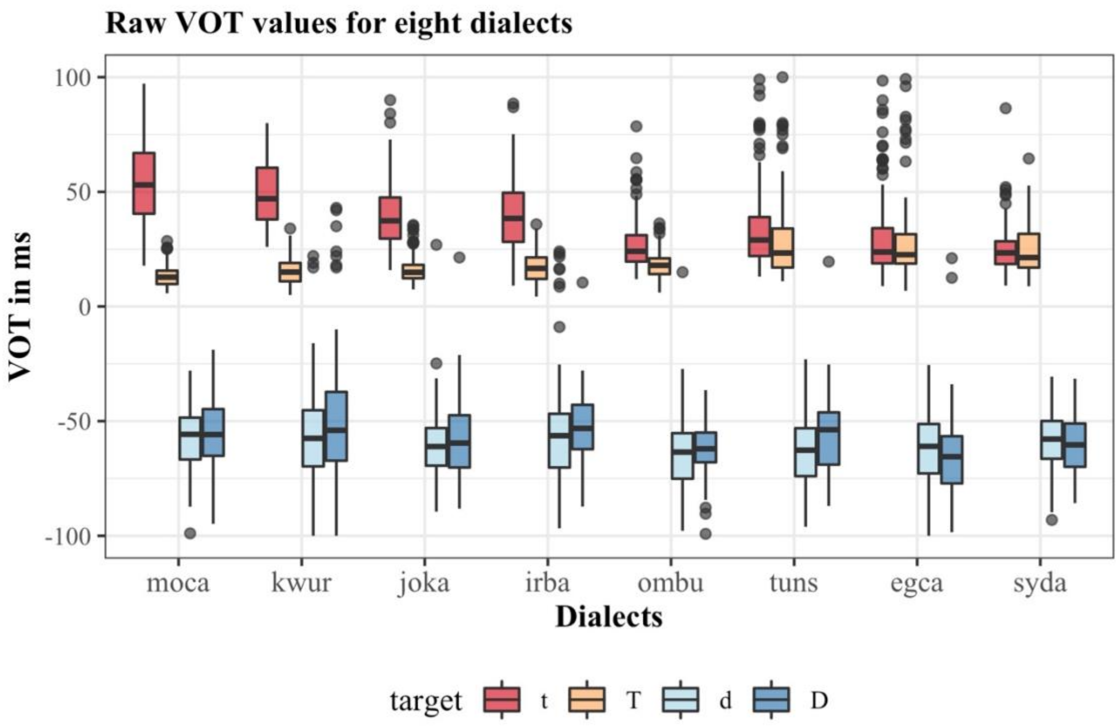


Figure 3-3 Median and interquartile range for raw VOT values in four plosives, from left /t/, /ṭ/, /d/, then /ḍ/ by dialect (listed in Table 3-1).

However, the picture here is more subtle: there isn't a clear-cut distinction between dialects in the voiceless subset, instead, it is more of a continuum of variation. On closer inspection, Figure 3-3 shows a clear overlap of /t/ and /ṭ/ in Egyptian and Syrian, whereas in Omani and Tunisian the overlap is partial. The four leftmost dialects show no overlap in VOT values in /t/ and /ṭ/, but greater variation in values of long lag /t/. Compared to the other Gulf dialects (Kuwaiti) and Iraqi, /t/ and /ṭ/ in Omani look like they might have merged: they have lost the long lag aspiration in plain /t/ and there is little variation in VOT values in emphatic /ṭ/. The overlapping values for /t/ and /ṭ/ might cause us to consider Omani as a two-way dialect. However, the tight range of emphatic VOT values in Omani is similar to the tight values of VOT in /ṭ/ in the other three-way dialects (left in the plot) and different to the spread of values for /ṭ/ in the remaining two-way dialects (right in the plot). Tunisian, however, behaves more like a two-way dialect, compared to Omani, in terms of the spread of values for the emphatic /ṭ/. Jordanian, on the other hand, looks conservative, it shows a clear distinction between /t/ and /ṭ/. We can also see that the dialects that have a clear contrast between /t/ and /ṭ/ (e.g., Moroccan, Kuwaiti, Jordanian, Iraqi) have a much more compact distribution for /ṭ/ whereas values for /t/ are spread.

This continuum hints at a sound change in progress in some dialects. This is further supported by a gender difference in the dialects that are ‘in transition’, as shown in Figure 3-4 below. From the Figure below, we can see that male speakers show more merged values i.e., with distributions that are more overlapped for /t/ and /ṭ/ in both Omani and Tunisian.

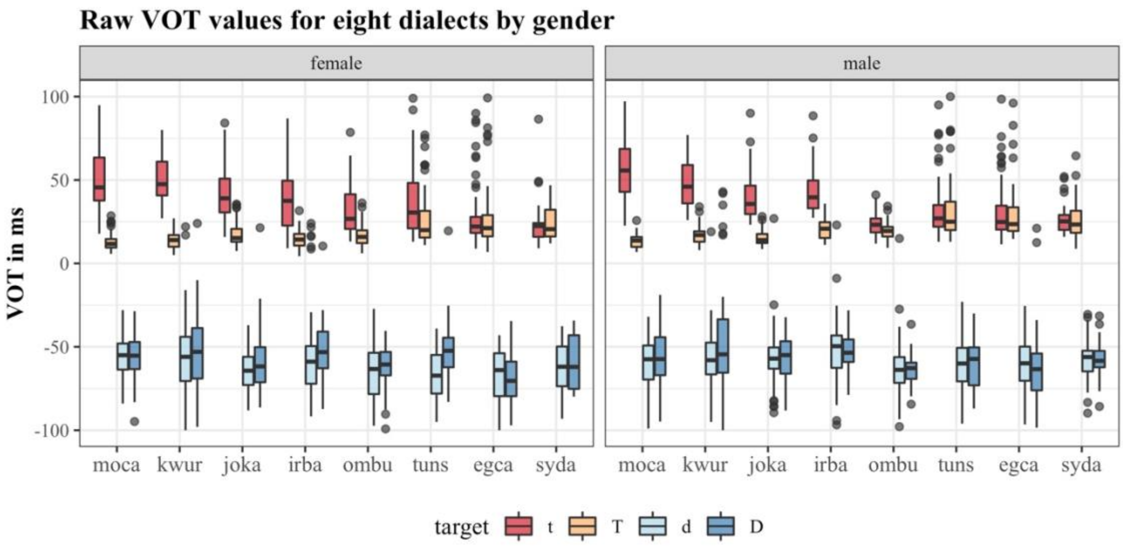


Figure 3-4 Median and interquartile range for raw VOT values in four plosives from left /t/, /ṭ/, /d/, then /ḍ/ by dialect and gender.

The overview of the raw results hints at a continuum of change in the realization of /t/ and /ṭ/, but all dialects show a clear distinction between voiced and voiceless plosives. The degree of variation in voicing lead in /d/ and /ḍ/ is minimal across dialects, while /t/ and /ṭ/ show clear differences in the degree of values across dialects. In the coming sections, we therefore examine in more depth how VOT plays a role in the plain vs. emphatic contrast across dialects and gender in voiced and voiceless plosives separately. Section § 3.3.2 examines the voiced contrast /d/ and /ḍ/, and § 3.3.3 the voiceless contrast /t/ and /ṭ/; we discuss in § 3.3.4 the quality of following vowel as cues to emphasis.

3.3.2 The /d/ ~ /ḍ/ contrast

As mentioned earlier, all dialects display voicing during the closure in both examined voiced consonants /d/ and /ḍ/, and the VOT values are overlapping across the plain vs. emphatic contrast. A closer look at Figure 3-5 below, shows a slight effect of gender in some dialects. The overlap of values /d/ and /ḍ/ is less for females (i.e., shorter voicing lead for /ḍ/) particularly in Tunisian female speakers.

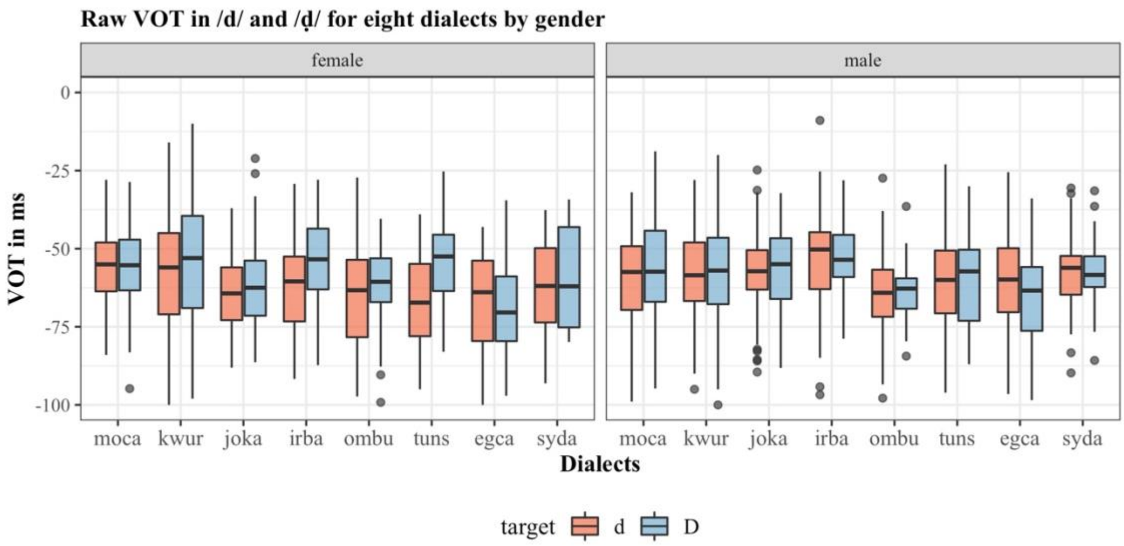


Figure 3-5 Median and interquartile range for raw VOT values in two voiced plosives /d/ and /ḍ/ by dialect and gender.

We explored the voiced subset of the data with sum-coded categorical factors in a linear mixed-effects model with *VOT* as the dependent variable, and *dialect*, *target* (/d/ and /ḍ/), and *gender*, and the interaction among them, as fixed factors, plus *position* as an additional fixed factor, and finally random intercepts for *item* and *speaker*. The coding of the categorical predictors with sum-coding was as follows: *gender* (female = 1, male = -1), *position* (initial = 1, medial = -1), *target* (/d/ = 1, /ḍ/ = -1), and *dialect* (seven predictors of "each dialect" ~ "syda", in which the first listed dialect was set to 1, "syda" to -1, and all unlisted dialects to 0). Figure 3-6 shows 95% confidence intervals (CI) around the predicted marginal mean VOT values by dialect, target, and position.

As we expected, there is a main effect of *position* (β = -3.23, t = -2.99, p = .007), in that voicing lead, in both voiced plosives, is somewhat longer in word-initial position than it is in word-medial position across all dialects. There are no significant three-way interactions among *dialect*, *target*, and *gender*. The main interest here is to investigate whether there is inter-dialectal variation in the use of VOT to differentiate the plain /d/ and emphatic /ḍ/. Figure 3-6 shows an overlap in the 95% CI around the predicted marginal mean values of VOT for /d/ and /ḍ/, which indicates that there is no use of the laryngeal cues to distinguish /d/ and /ḍ/ in any of the dialects.

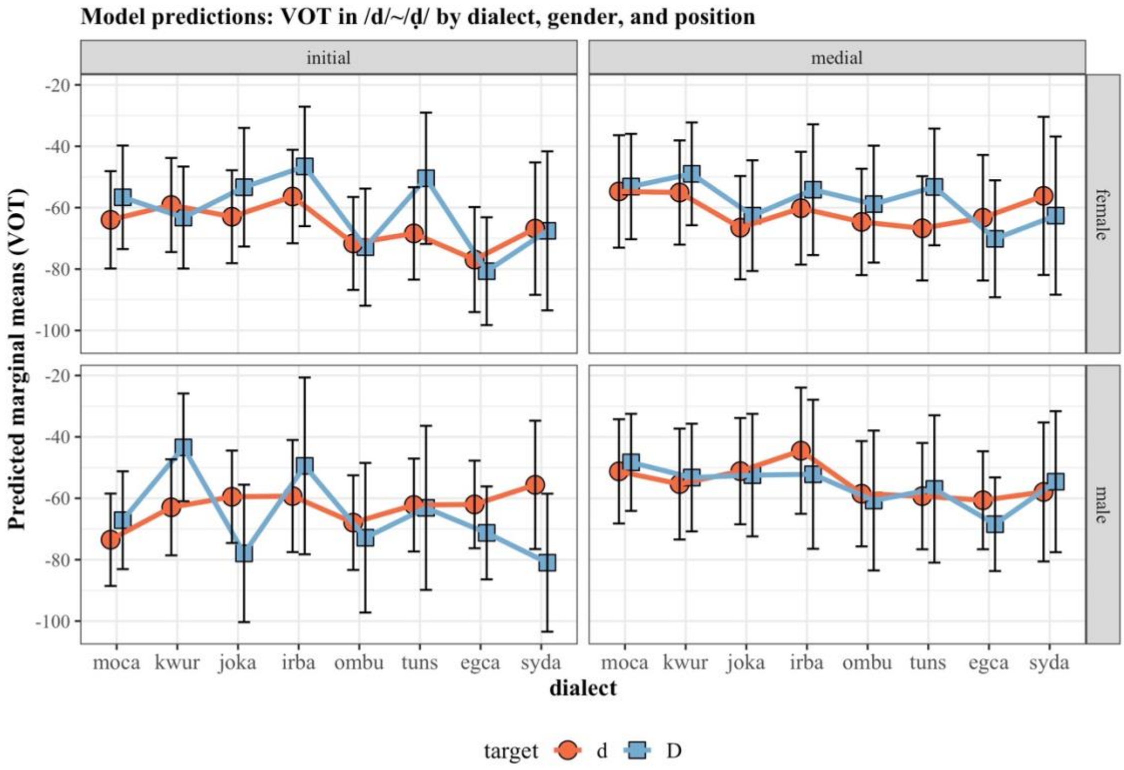


Figure 3-6 Predicted marginal mean (and 95% CI) for VOT in /d/ ~ /ḍ/ by dialect, gender, and position.

These results confirm our predictions of no variation between the voiced plosives /d/ and /ḍ/ across Arabic dialects in use of VOT to differentiate the plain and emphatic contrast in voiced plosives. That is, VOT is not a phonetic exponent of the /d/ vs. /ḍ/ contrast. The full model result summary is provided in Appendix (B.4). In the next section, we explore in greater depth the main observed locus of variation across dialects, which is the voiceless plain and emphatic contrast /t/ and /ṭ/.

3.3.3 The /t/ ~ /ṭ/ contrast

All dialects show voicing lag in both /t/ and /ṭ/, with varying degrees of difference between the plain and emphatic segments in each dialect. A closer look at Figure 3-7 in the voiceless subset of the raw data, clearly shows a continuum of variation rather than a dichotomy. There also appears to be a potential effect of gender in some dialects, with rather less overlap and more spread values in /t/ and /ṭ/. Specifically, we see longer VOT values in /t/ produced by female speakers, and less overlap between the /t/ and /ṭ/ contrast also by female speakers, in some dialects, and especially in the ‘in-transition’ dialects in the middle of the plot (Tunisian and Omani).

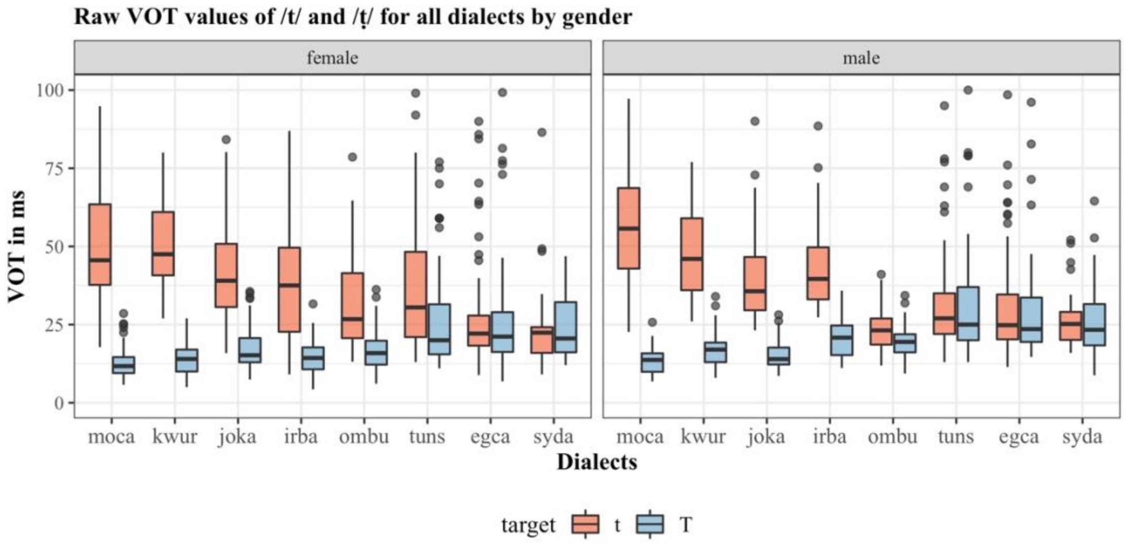


Figure 3-7 Median and interquartile range of raw VOT values in /t/ ~ /ṭ/ by dialect and gender.

The voiceless subset of the data was explored in a series of linear mixed-effect models yielding a best fit model that predicted *VOT* as the dependent variable with *dialect*, *target* (/t/ and /ṭ/), and *gender*, and the interaction among the three, as fixed factors, plus *position* as an additional fixed factor, with random intercepts for *item* and *speaker*. The categorical factors in this subset were sum-coded as well. The coding of the categorical predictors with sum-coding were the same as the voiced subset, except for *target* (/t/ = 1, /ṭ/ = -1). The model results, illustrated in Figure 3-8, show the 95% confidence interval around predicted marginal mean values of voiceless VOT by *dialect*, *target*, and *gender*.

As in the voiced subset, there was an overall main effect of *position* (β = -4.67, *t* = -3.06, *p* = .003); in the voiceless subset, voicing lag in both of the voiceless plosives is somewhat shorter word-initially than word-medially, across all dialects. The model shows a significant three-way interaction among *gender*, *dialect*, and *target* for only Omani and Tunisian, reflected in shorter voicing lag in /t/ by both male speakers of Omani (β = 1.38, *t* = 1.97, *p* = .05), and male speakers of Tunisian (β = 1.95, *t* = 2.69, *p* = .007) resulting in greater overlap of values for /t/ and /ṭ/ produced by males in both dialects.

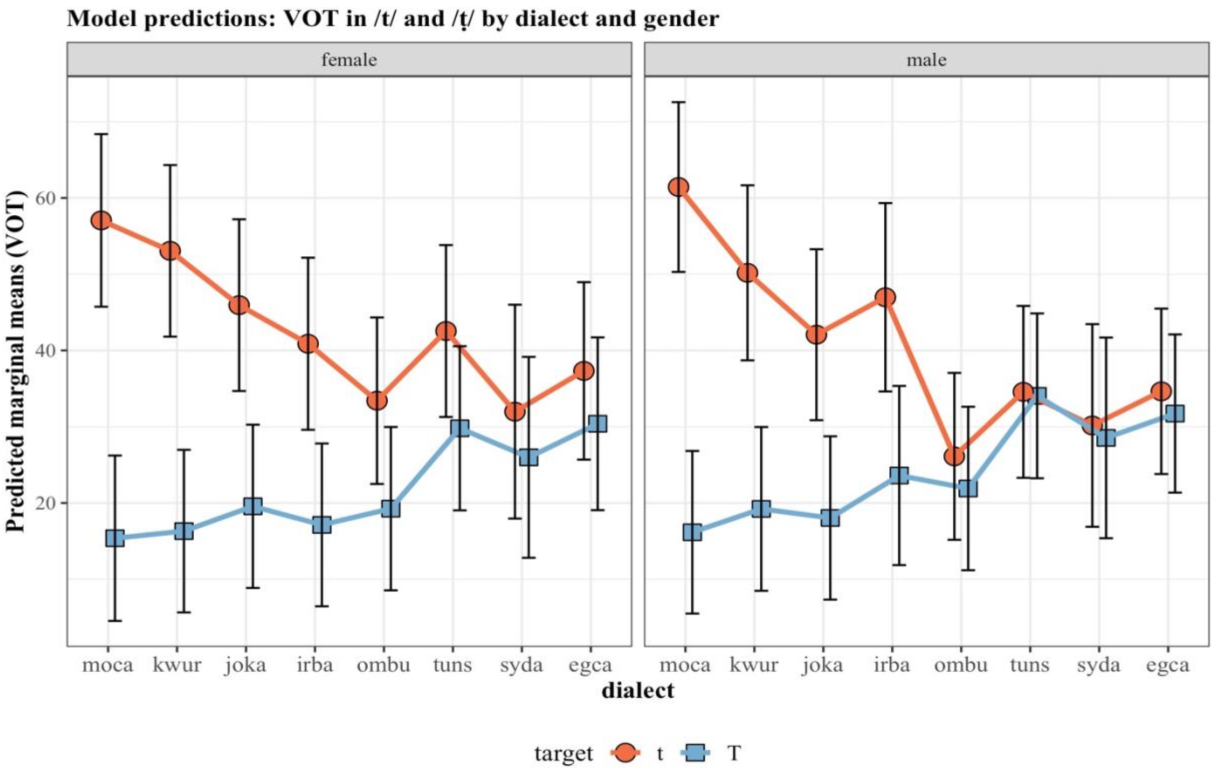


Figure 3-8 Predicted marginal mean (and 95% CI) for VOT in /t/ ~ /ṭ/ by dialect and gender.

In this model, the estimate of dialect interacting with target is decreasing, which means that target gets an additional boost of difference between /t/ and /ṭ/ with respect to dialect effect as shown in Table 3-3 below. The full model result summary is provided in Appendix (B.5).

Table 3-3 Model results for the two-way interaction of dialect*target in the voiceless subset[21].

Interactions of Interest	*Estimate*	*Std. Error*	*t. value*	*p. value*
dialect moca : target t	12.325	0.701	17.588	< 001*
dialect kwur : target t	7.512	0.719	10.448	< 001*
dialect joka : target t	3.187	0.702	4.538	< 001*
dialect irba : target t	2.365	0.771	3.068	.002*
dialect ombu : target t	-4.824	0.710	-6.792	< 001*
dialect tuns : target t	-6.097	0.734	-8.312	< 001*
dialect egca : target t	-6.955	0.718	-9.687	< 001*
dialect syda : target t	-7.511	1.044	-7.197	< 001*

The model results support our initial observations from the visualization of the data, i.e., the obvious overlap in the distribution of VOT values for /t/ and /ṭ/ in Egyptian, Syrian and Tunisian, which maps to increasing estimate values in these four dialects at the

[21] To estimate the held-out factor "syda", we rotated the levels of the model.

bottom of Table 3-3 above. Visually, the data suggested that Omani and Tunisian might be marginal dialects, in transition towards merging the VOT values for plain and emphatic voiceless plosives; however, the model (excluding the interaction with gender) indicates that the merger might be already complete in Tunisian, as the estimate is greater when comparing VOT values for /t/ and /ṭ/ to those of the reference level. The remaining dialects – Moroccan, Kuwaiti, Jordanian, and Iraqi – showed little or no visual overlap in the distribution of VOT values for /t/ and /ṭ/, which is confirmed by a significant difference between /t/ and /ṭ/ in VOT values for each of these four individual dialects.

In summary, the locus of dialectal variation in VOT is in the contrast in the voiceless subset between /t/ and /ṭ/. The results for this subset suggest a continuum of change rather than a clear-cut dichotomy; indeed, four dialects on the left end of this continuum in Figure 3-3 above (Moroccan, Kuwaiti, Jordanian, Iraqi) show an overall three-way VOT distinction (voiced ~ voiceless plain ~ voiceless emphatic). In contrast, the two 'in-transition' dialects at the middle of the continuum (Omani and Tunisian) display gender effects where the female speakers appear to retain a somewhat three-way contrast (voiced ~ voiceless plain ~ voiceless emphatic), while male speakers in those two dialects display a two-way contrast (voiced ~ voiceless) with overlapping values of /t/ and /ṭ/. However, dialects on the right side of the continuum (Egyptian, Syrian) show completely overlapping values for /t/ and /ṭ/ which equates to a two-way laryngeal contrast (voiced ~ voiceless).

In the next section, we investigate whether this continuum of change is mirrored in the F2 lowering of the vowel after the emphatic in a similar pattern. In other words, are there any differences among dialects in the degree of F2 lowering? The choice to focus on F2 is based on previous studies, in which the differences observed on F1 were not as consistent as the lowering differences obtained in F2, which led many scholars to consider F2 as the main acoustic cue of emphasis in Arabic (McCarthy, 1994; Watson, 2002).

3.3.4 F2 lowering effects in /t/ ~ /ṭ/

First, we lay out the observed patterns in descriptive terms. Since the F2 lowering of the vowels following the emphatic consonant is not the same for the three vowels in Arabic, we chose to visually compare F2 values of the vowel that shows the most effect of lowering; the back long vowel /a:/ to detect the variation after the voiceless plosives as shown in Figure 3-9 and Figure 3-10 below. Raw results in Figure 3-9 and Figure 3-10 suggest that there is an effect of F2 lowering in short and long /a/ following the emphatic /ṭ/. Visual comparison of the Lobanov-normalized F2 values after the plain and emphatic voiceless plosives show only somewhat random variation across dialects. It seems from Figure 3-9, F2 values of the long vowel /a:/, that the F2 difference between /t/ and /ṭ/ is greater in the Tunisian and Egyptian dialects compared to the other dialects.

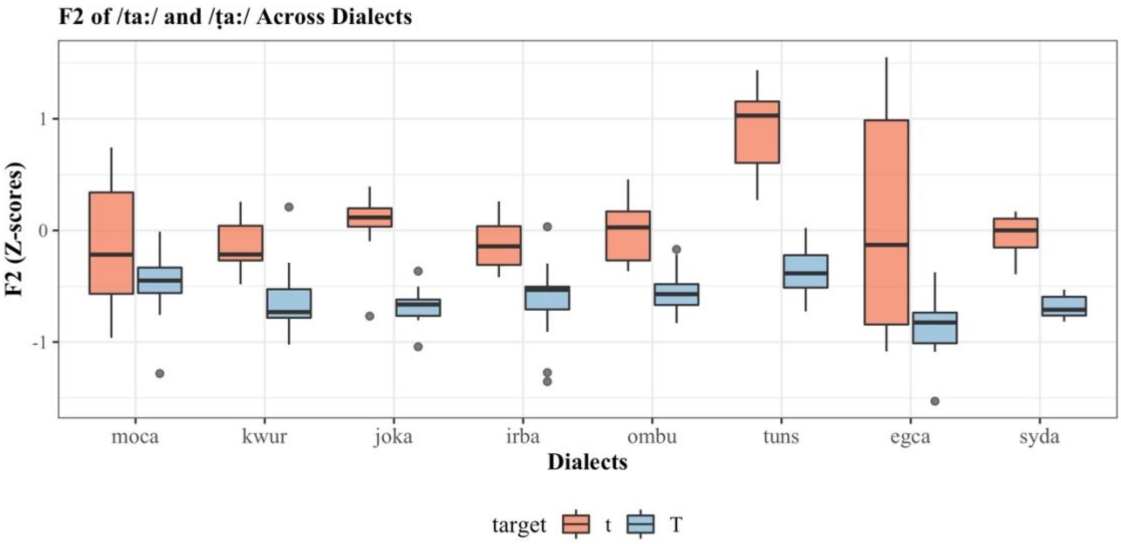


Figure 3-9 Median and interquartile range for Lobanov-normalized F2 in the long vowel /a:/ following plain /t/ and emphatic /ṭ/ by dialect.

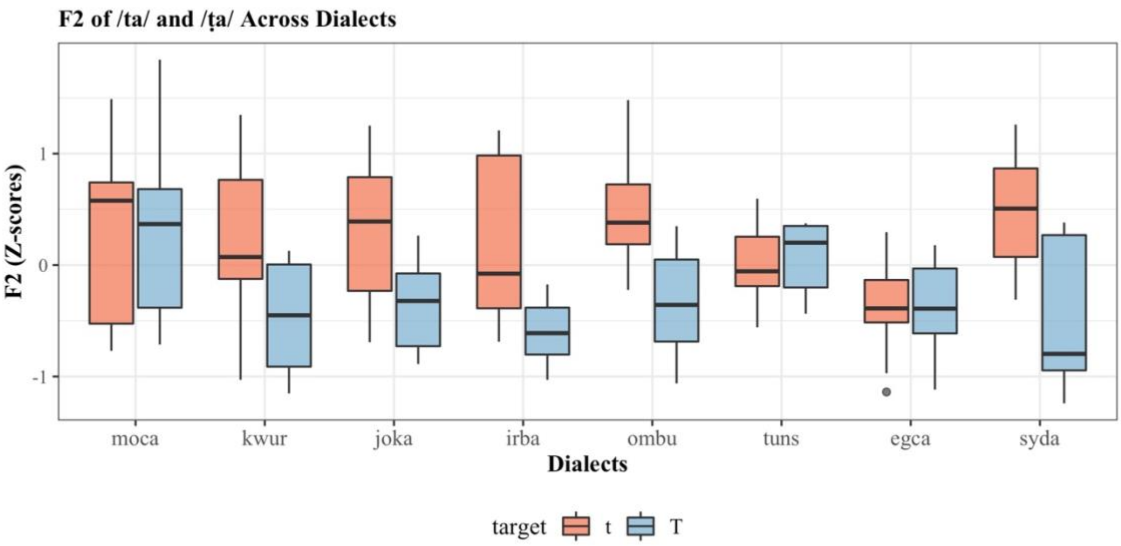


Figure 3-10 Median and interquartile range for Lobanov-normalized F2 in the short vowel /a/ following plain /t/ and emphatic /ṭ/ by dialect.

To assess these differences statistically, the voiceless subset of the data was also explored in a series of linear mixed-effects models (LMM) using *lme4* (Bates et al., 2015) in *R* (R Core Team, 2018). The model predicted *normf2* as the dependent variable, which refers to the z-score normalized F2 of the vowels following concerned targets, with *dialect*, *target* (/t/ and /ṭ/), *gender*, as well as the interaction among the three as fixed factors, plus *position*, and *vowel* as additional fixed factors, with random intercepts for *item* and *speaker*. The categorical factors in this model are sum-coded in the same codes we mentioned earlier, in addition to *following vowel* (five predictors of /iː/ ~ /a/, /uː/ ~ /a/, /aː/ ~ /a/, /i/ ~ /a/, and /u/ ~ /a/ in which the first listed vowel was set to 1, /a/ to -1, and all unlisted vowels to 0). The model results, illustrated in Figure 3-11, show 95% confidence intervals around the predicted marginal mean values of the normalized F2 by dialect, target, and gender.

There was no main effect of *position* ($\beta = 7.10$, $t = 0.43$, $p = .67$) or *gender* ($\beta = -3.28$, $t = -0.23$, $p = .82$). There was, however, a main effect of *target*: as expected, /t/ was positively related to /ṭ/ ($\beta = 2.77$, $t = 3.86$, $p = .002$). This means that, for each decrease in F2 values after /ṭ/ by one standard deviation, F2 values of vowels following /t/ increased by 2.8. There was only one significant three-way interactions among *gender*, *dialect*, and *target* in the Iraqi dialect (*irba* x */t/* x *female*: $\beta = -7.41$, $t = -2.62$, $p = .009$). This means that the gender difference between F2 values of /t/ and /ṭ/ was greatest in Iraqi, in which female speakers have closer F2 values between /t/ and /ṭ/ than male speakers. The full model result summary is provided in Appendix (B.6).

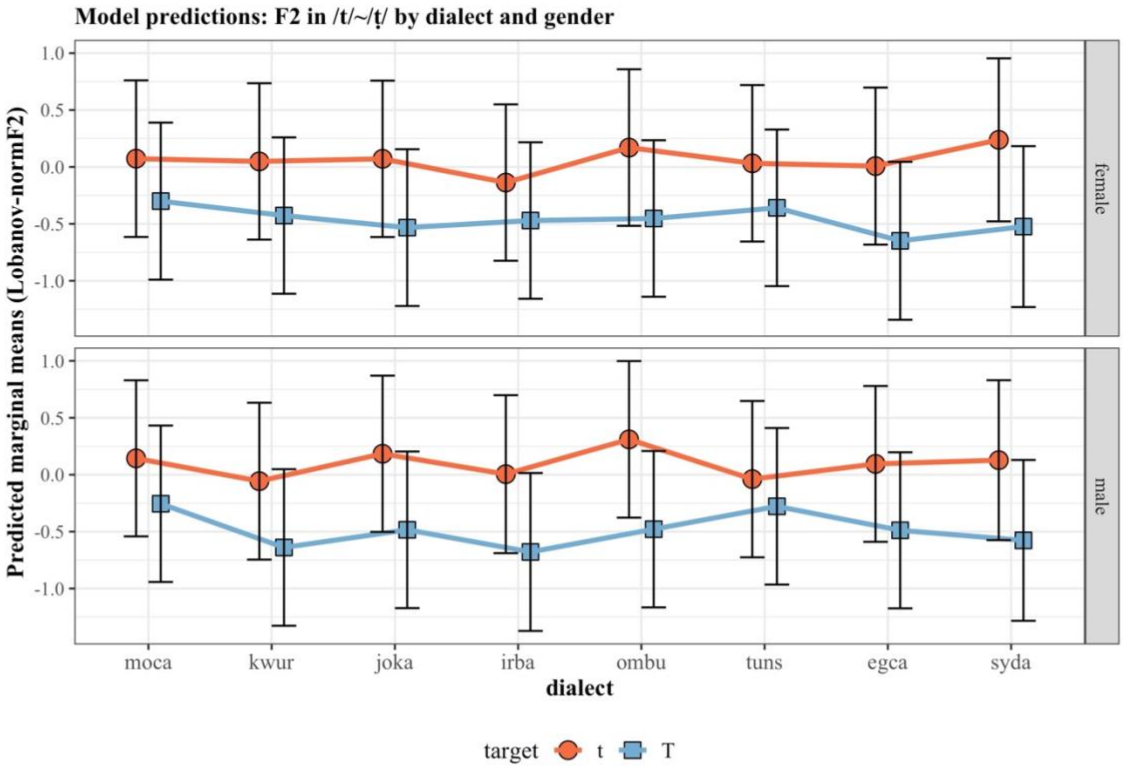


Figure 3-11 Predicted marginal mean (and 95% CI) for Lobanov-normalized F2 in vowels following plain /t/ and emphatic /ṭ/ by dialect and gender.

The main interest here is to explore whether vowels following the voiceless emphatic plosive display similar F2 lowering regardless of the dialect or gender effects observed in the variation of VOT values across the same contrast. In Figure 3-11, we can see that the 95% CI around the predicted mean for the normalized F2 values of vowels following /t/ and /ṭ/ that all dialects use F2 lowering to a similar extent, confirming the primacy of the vocalic cue to distinguish the plain and emphatic contrast cross-dialectally.

In summary, the results for normalized F2 show that all dialects have a similar lowering effect in the vowels following emphatic /ṭ/ regardless of their laryngeal cues to the emphatic contrast (three-way vs. two-way VOT distinction).

3.4 Discussion

3.4.1 VOT

The literature reports differences in the number of VOT categories across Arabic dialects and this motivated us to identify the scope of variation in the number of laryngeal contrasts in VOT patterns across Arabic dialects. Our main interest is to confirm this variation and to present a coherent laryngeal typology of Arabic dialects in respect of VOT.

Our results support the claim of cross-dialectal variation in VOT. This variation stems from the fact that the voiceless emphatic and plain plosives are not realized uniformly across the dialects. In the results section, we confirmed that the VOT values of /d/ and /ḍ/ are coextensive across all the dialects. In addition, at first glance, the VOT values of /t/ and /ṭ/ are consistent with the pattern noticed by Bellem (2007) namely that dialects are grouped into two different categories: the dialects in one category retain an overall three-way VOT distinction among voiced and voiceless plosives, hence the separation of values between /t/ and /ṭ/ (conservative dialects: Moroccan, Kuwaiti, Iraqi, Jordanian); the other category of dialects merges the VOT values of voiceless plain and emphatic plosives resulting in an overall two-way VOT distinction (progressive dialects: Syrian, Egyptian, Tunisian, Omani).

In our analysis, we were mainly interested in the voiceless subset of the data because it is the locus of this variation. Generally, these results are similar to the degree of variation reported by Heselwood (1996). The SD of Moroccan, Kuwaiti Jordanian, Iraqi and Omani /t/ and /ṭ/ are distinct as Heselwood found for Baghdadi, but the SD of Tunisian, Egyptian, and Syrian /t/ compared to /ṭ/ are closer together as Heselwood found for Cairene /t/ and /ṭ/, as seen in Table 3-4 below. In our data, however, Omani and Tunisian look different to Egyptian and Syrian; it seems that Omani and Tunisian are still in transition towards losing the VOT distinction between /t/ and /ṭ/, and the transition is less advanced in Omani than in Tunisian.

Table 3-4 The standard deviation (SD) for VOT values of /t/ and /ṭ/ in eight dialects from the present study compared to Heselwood's (1996) reported SD for /t/ and /ṭ/ in Baghdadi and Cairene.

dialect	**SD**		**dialect**	**SD**		**Heselwood (1996)**	**SD**	
	/t/	**/ṭ/**		**/t/**	**/ṭ/**		**/t/**	**/ṭ/**
Moroccan	19	5	Omani	12	6			
Kuwaiti	14	6	Tunisian	22	18	Baghdadi	12	5
Jordanian	16	6	Egyptian	21	18	Cairene	14	10
Iraqi	17	7	Syrian	14	12			

It is striking to see Omani in transition towards merging the two voiceless consonants, since it is a dialect of the Gulf and is expected to be of the conservative dialect type. As we mentioned earlier, however, the data collected representing the Omani dialect is from Buraimi, which is an area in the north-west of Oman that shares a border with the United Arab Emirates. According to Holes (1989), the Buraimi dialect of Oman acts as an exceptional case compared to the rest of the Omani dialects. Holes states that there is a heterogeneous bundle of features shared by all Omani dialects in all areas except for Buraimi. This might be attributed to the transitioning of this dialect into a progressive dialect. Consequently, as mentioned in § 2.5, it is important to note that we cannot assume that the dialect of a whole nation, in this case Omani, will show the same voicing patterns.

Further exploration, especially of gender differences in the data, make it clear that the variation cannot be described as a classification into two groups per se; it is, rather, a continuum of change. It is notable that a significant gender effect is not present in all dialects, but only in those that are 'in-transition'. Most studies on the phonetic categories of voicing in dialects of Arabic as a function of gender have stated that there is no effect of gender. One case is a study on colloquial Egyptian Arabic (Rifaat, 2003), which is a dialect included in our current study, and similar results are reported; Rifaat found no apparent gender differences in VOT values for male and female participants. Similarly, in a study of Syrian VOT, Almbark (2008) also reported no significant effect of gender on VOT in the plain and emphatic contrast, and these results are in accordance with our results here for the Syrian dialect.

In contrast, a previous study on Jordanian Arabic did find gender effects on the /t/ and /ṭ/ contrast. Khattab et al (2006) found that female speakers from Irbid, Jordan, had significantly longer values of /ṭ/ compared to those of male speakers; that is, the female speakers showed greater tendency to merge the plain ~ emphatic contrast than male speakers. On the other hand, our results show no significant gender effects in the Jordanian data (*joka* x */t/* x *female*: $\beta = -0.52$, $t = -0.74$, $p = .46$). However, most participants in Khattab et al.'s (2006) study were from Irbid, in the north of Amman the capital of Jordan, whereas the participants in this study are from Karak, in the south of Amman. Abudalbuh's (2010) results for Jordanian /t/ and /ṭ/ were similar to our results; VOT values of /ṭ/ were significantly shorter than /t/ regardless of gender, and their data were recorded in one of the public universities in Jordan located on the outskirts of the city of Mafraq, north of Amman. It is worth noting that the rural dialect of Karak, unlike urban Amman, is described to use the [g] variant of /q/ (Albdairat, 2021), which is a feature generally ascribed to Bedouin ‘conservative’ dialects (Watson, 2011a).

To summarize, our results confirm variation in VOT patterns across Arabic dialects. The locus of variation is in the voiceless plain and emphatic contrast and their degree of VOT lag. The results indeed show that some dialects have three VOT categories (voiced ~ voiceless plain ~ voiceless emphatic) and others have eliminated laryngeal cues as distinction to the plain and emphatic contrast, resulting in two VOT categories overall (voiced ~ voiceless). It is not a clear-cut dichotomy but instead it appears to be a continuum of change. Two of the dialects in the middle of this continuum display gender differences, as female speakers have a strong tendency to maintain greater VOT distinction between /t/ and /ṭ/, and male speakers are leading the change in elimination of VOT as a cue to the plain vs. emphatic contrast.

3.4.2 F2 lowering

As mentioned earlier, previous literature on the Arabic emphatics indicates that F2 lowering in emphatic vowels is a robust indicator of emphasis, though the lowering effect is not uniform for all three vowels in Arabic. F2 lowering after emphatics is known to be stronger and more enduring through the vowel in /a/, than in /i/ and /u/. The VOT results which confirmed variation in VOT across Arabic dialects led us to explore whether the loss of the laryngeal cue to the emphasis contrast, i.e., merger of the plain and emphatic plosive VOT values in some dialects, is compensated for by additional F2 lowering in the emphatic vowel.

From our analysis above, we concluded that the degree of F2 lowering is consistent across all dialects, and in all the vowels, and the status of dialects as having two vs. three VOT categories did not affect this acoustically. Given that the lowering effect is greater in /a:/ and /a/, we saw in Figure 3-9 that F2 lowering is found across all dialects regardless of whether it is a two-way or three-way dialect. Figure 3-9 also indicates that the F2 difference between /t/ and /ṭ/ is greater in the Tunisian and Egyptian dialects compared to the other dialects, which is due to the stronger fronting and closing of /a:/ after plain consonants in those dialects (Fathi, 2013; Gibson, 2008). We can also notice from the figure that the properties of the post emphatic /a:/ (blue) is similar across the dialects, whereas the differences here are in the F2 values of the plain /a:/ (red) which indicates that the realization of plain /a:/ might vary across dialects (Almbark & Hellmuth, 2016).

These observations underline the importance of the adjacent vowel as a cue to emphasis. Previous findings from perception studies with L2 learners of Arabic highlight listeners' reliance on following vowels to identify emphatic environments. Zaba (2007) used a cross-language identification task to determine the role of adjacent vowels to accurately perceive emphasis. She found that native English speakers identified Arabic /a:/ after the plain consonant /t/ as English /æ/, and they identified Arabic /a:/ after the emphatic consonant /ṭ/ as English /ɑ/, but both Arabic /i:/ and /u:/ were identified as mostly English /i/, and English /u/ or /ʌ/ after plain and emphatic contexts. So, native English listeners were not able to accurately identify emphatics after the high front vowel /i/ but they were able to accurately identify the emphatics after the high back vowel /a/.

Similar results were found by Hayes-Harb & Durham (2016) in a study which investigated perception of plain vs. emphatic voiced plosive /d/ and /ḍ/ through cross-language vowel identification and perceptual discrimination tasks. They found that native English listeners relied on adjacent vowels more than the consonants themselves to identify Arabic emphatics, and that listeners were more accurate when the following vowel was /a:/, than for /u/ and /i/. In a similar perception study done by native Arabic speakers, Jongman et al. (2011) concluded that what comes after the consonant contributes significantly more to the perception of emphasis than the consonant itself.

In our results, the F2 lowering tendency was not affected by gender. Al-Masri & Jongman (2004), however, found that a stronger emphasis effect for female speakers than male speakers in their Jordanian data, i.e., the extent of F2 lowering in emphatic vowels is significantly greater for females than males. Opposite effects of gender on the degree of emphasis, with stronger production of emphasis by males than females, were found in the majority of the literature on emphasis (Kahn, 1975; Khattab et al., 2006; Royal, 1985; Wahba, 1996).

These findings along with our current results further indicate that emphasis in Arabic is realized mostly in the vowel. The fact that some dialects can tolerate a change in VOT patterns, which blurs the laryngeal distinction between plain and emphatic /t/ and /ṭ/, in addition to the lack of distinction in VOT values of /d/ and /ḍ/ in any case, across all the dialects, indicates the primacy of the vocalic cue to the plain vs. emphatic contrast.

3.5 Summary

In this chapter, we used a parallel dataset (from IVAr) to confirm the typology of Arabic dialectal laryngeal contrasts proposed by Bellem (2007), and also the primary status of F2 lowering in adjacent vowels as a cue to emphasis. The phonetic differences highlighted here have implications for phonological representations of the Arabic voicing contrast continuum. From the conclusion drawn above, and to pursue the 'over-specification' analysis of the active laryngeal features discussed in the literature review in chapter 2, as the next step we decided to collect variable speech-rate data from three different dialects: **Tunisian**, a dialect in transition towards being progressive as shown from our results; **Najdi**, a variety of Saudi dialect from the central region which is known to be conservative and to display a three-way VOT category (Al Malwi, 2017; Bellem, 2007); and **Hijazi**, another variety of the Saudi dialect from the western region which has never been studied for its laryngeal categories, and which is suspected to be a progressive dialect.

Chapter IV

Speech-rate effects on VOT in dialect speech

4 Speech-rate effects on VOT in dialect speech

After confirming the typology of variation in the number of laryngeal contrasts in chapter 3, we are interested to know what the active underlying phonological features are in three dialect types that are expected to display different laryngeal categories (three-way, in transit, and two-way). Most importantly, we are interested to know whether or not the two-way VOT category dialect is over-specified with two active phonological features.

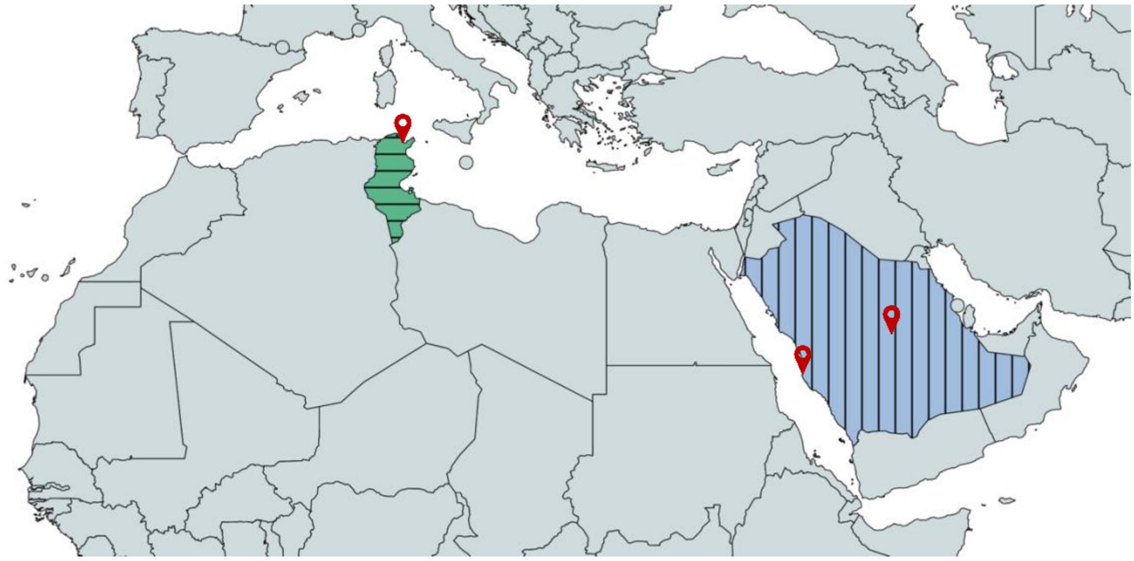
Three Arabic dialects are chosen, with at least one dialect from each laryngeal category pattern. **Tunisian**, which we concluded based on the results from the IVAr corpus to be a dialect merging its plain and emphatic voiceless plosives and behaving almost like a two-way laryngeal contrast dialect, the **Najdi** variety of the Saudi dialect, which is predicted to be a conservative dialect preserving a three-way laryngeal contrast, and the **Hijazi** variety of the Saudi dialect, which is predicted to be a progressive dialect having a two-way laryngeal contrast, but possibly in transition.

In the upcoming chapters (4 and 5), I will treat Tunisian as a two-way contrast dialect by comparison to Najdi and Hijazi. The three dialects are treated like a zoomed-in continuum, with Najdi at one end (three-way), Hijazi in the middle, and Tunisian at the other end (two-way). In a perfect world, it would have been ideal to have a true two-way contrast dialect like urban Egyptian or Syrian, but due to difficulties in reaching these areas at the time of the study, the Tunisian dialect was chosen instead.

In this chapter, § 4.1 starts with a brief overview on the linguistic context of the chosen dialects, then in § 4.2 and § 4.3 I present the methods, measurements and segmentation used in this chapter. In sections § 4.4 – 4.6 I present detailed results on the speeded data of the plain vs. emphatic coronal plosives, then detailed results on the whole set of plosives present in the discussed dialects. In light of the results, in § 4.7 I provide a general discussion and conclusion.

4.1 Background on chosen dialects

This section serves as a general background to the choice of dialects of Arabic examined in chapters 4 and 5 of this thesis, in the context of the overall dialectal variation in Arabic. From the conclusion drawn in the previous chapter (Chapter 3), and to look for potential ‘over-specification’ of active laryngeal features as discussed in the background of the previous literature chapter (Chapter 2), I decided to collect variable speech-rate data from three different dialects: firstly, the **Najdi** dialect, which is a variety of Saudi Arabic from the central region and known from previous literature to have three VOT categories (voicing lead ~ short lag ~ long lag) (Al Malwi, 2017; Bellem, 2007); secondly, **Hijazi** dialect, which is another variety of Saudi Arabic, from the western region, which has never been studied for its laryngeal contrast, and which is suspected to be in transition towards having two VOT categories (voicing lead ~ short-ish lag); lastly, **Tunisian** dialect, which has two VOT categories (voicing lead ~ short-ish lag) as shown from our results in chapter 3. The three dialect data samples represent the urban variety spoken in major cities and their surrounding areas (see Map 4-1). In the following subsections we present each dialect and its linguistic characteristics and landscape. First, § 4.1.1 reviews the Najdi dialect of Saudi. Then, § 4.1.2 examines the Hijazi dialect of Saudi. Lastly, Tunisian dialect is discussed in § 4.1.3.



Map 4-1 Location of the chosen dialects in the Arab world– Vertical lines = Saudi dialects, Horizontal lines = Tunisian dialect [22].

4.1.1 Najdi

Najdi dialect is associated geographically with the region of Najd, the Central region in Saudi Arabia. This variety belongs to the Northern-Central Arabian group of dialects spoken in the Arabian Peninsula. Ingham (1994, p.4-5) lists dialects that can be classified in a general sense as Najdi Arabic. These sub-dialects are mutually intelligible, although they vary slightly, phonologically and morphologically (Al-Essa,

[22] Map zoomed in from https://mapchart.net/

2009; Ingham, 1994; Johnstone, 1967; Prochazka, 1988). The urban variety of Najdi originates from the capital of Saudi Arabia, Riyadh, and is the dialect of interest in this section. One well recognized phonological feature of Najdi is the affrication of /k/ and /g/ to [ts] and [dz] in certain environments (see examples (1c) below) (Ingham, 1994; Versteegh, 2014). The following are more of the common consonantal inventory features in Najdi Arabic:

(1) Najdi:

(a) MSA /q/ is realized in Najdi cognate words as either [q] or [g] (lexically determined variation)

[q] ***q****aːnuːn* 'law'

[g] (MSA) ***q****aːl* ≈ (Najdi) ***g****aːl* 'he said'

(b) /ḍ/ and /ð̣/ are merged into a single phoneme [ð̣]

[ð̣] (MSA) *biḍ****ḍ****abṭ* ≈ (Najdi) *bið̣****ð̣****abṭ* 'exactly'

(c) affrication in the environment of front vowels (Al-Essa, 2009) (sociolinguistically determined variation)

[k] ~ [ts] (e.g., *ummi****k*** ~ *ummi****ts***) 'your mother'

[g] ~ [dz] (e.g., *ṭiriː****g*** ~ *ṭiriː****dz***) 'road'

4.1.2 Hijazi

Geographically, Hijazi dialect is associated with the Western region of the Arabian Peninsula, within the borders of Saudi Arabia. It is mainly spoken in the following cities: Jeddah, Taif, and the holy cities of Makkah and Madinah. In this thesis, Hijazi dialect is represented by the urban dialect spoken in Jeddah, a city located on the coast of the Red Sea. Jeddah is considered to be a cosmopolitan city with a heterogeneous population because of its strategic location on the routes of trade and as the gateway to the holy cities of Makkah and Madinah. The migration of many Muslim pilgrims from around the world to Jeddah and other main cities in Hijaz helped in shaping the linguistic identity of this dialect.

The Hijazi dialect is different from other dialects in the Arabian Peninsula in many features, both phonologically and morphologically. It is generally considered less conservative than the other Bedouin varieties that retain many Classical Arabic features. However, in contrast to other varieties in Saudi, Hijazi retains the emphatic /ḍ/ (Al-Essa, 2009; Watson, 2002). The following examples are some of the consonantal inventory features of Hijazi Arabic:

(2) Hijazi:

(a) merger of the interdental fricative /ð/ with [d, z]

[d] (Najdi) *ki**ð**a* ≈ (Hijazi) *ki**d**a* 'such'

[z] (MSA) *astaʔ**ð**an* ≈ (Hijazi) *astaʔ**z**an* 'took permission'

(b) merger of the interdental fricative /θ/ with [t, s]

[t] (MSA) ***θ**ala:**θ**a* ≈ (Hijazi) ***t**ala:**t**a* 'three'

[s] (MSA) *ma**θ**alan* ≈ (Hijazi) *ma**s**alan* 'for example'

(c) lexical variation [ḍ] and [ẓ]

[ḍ] (MSA) ***ḍ**afi:ra* = (Hijazi) ***ḍ**afi:ra* 'hair braid'

[ẓ] (MSA) *biḍ**ḍ**abṭ* ≈ (Hijazi) *biẓ**ẓ**abṭ* 'exactly'

4.1.3 Tunisian

Tunisian Arabic, one of the Maghrebi Arabic dialects referred to in § 2.6, is the dialect spoken within the borders of Tunisia, North Africa. The Tunisian dialect in this thesis is represented by the urban dialect spoken in the capital Tunis and surrounding areas. It is a de facto prestige variety that other varieties of Tunisia are shifting towards. The Tunisian dialect, like other Maghrebi dialects, is subject to influence from Berber (Amazigh), French, and the gradual spread of English. Thus, it is generally intelligible to speakers of other Maghrebi dialects but not as readily intelligible to Arabic speakers from other parts of the Middle East (Gibson, 2002, 2008; Sayahi, 2011).

The most salient phonological feature in the urban Tunisian dialect is the use of the uvular plosive /q/. The voiced phoneme [ɡ] occurs but is not frequent in the urban variety and instead more commonly used in rural varieties (Gibson, 2002). In Tunisian Arabic, short vowels are not frequently used, especially at the end of an open syllable. In a non-emphatic environment, the Tunisian /a:/ is strongly fronted and closed to become [eː] (Gibson, 2008; Jabeur, 1987). The following are some of the Tunisian consonantal and vocalic features:

(3) Tunisian (urban variety of Tunis):

(a) Tunisian /q/ observed in cognate words containing MSA /q/

[q] ***q**uːl* 'say'

(b) MSA /a:/ fronting in non-emphatic environment

[e:] (MSA) *b**aː**b* ≈ (Tunisian) *b**eː**b* 'door'

(c) /ḍ/ and /ð̣/ merged into a single phoneme [ð̣]

[ð̣] (MSA) ***ḍ**aːbṭ* ≈ (Tunisian) ***ð̣**aːbṭ* 'policeman'

4.2 Methods

4.2.1 Participants

There were 64 participants: 28 educated native speakers of Tunisian Arabic (14 female, 14 male), 18 educated native speakers of the Najdi variety of Saudi (8 female, 10 male), and 18 educated native speakers of the Hijazi variety of Saudi (10 female, 8 male). Their ages range from 17 – 45. None of the participants reported a history of hearing or speech impairment. The participants were not aware of the purpose of the experiment, nor were they paid to participate in this study. They were told that the focus of interest was their local dialect. To ensure dialect homogeneity, the speakers were all born or raised within a thirty-mile radius of the city where the recordings happened (Tunis, Riyadh, Jeddah). Participants were asked to speak in their native dialects as if communicating with family and friends.

Most participants are bilingual in French or English, besides Arabic (and some are multilingual). Recruitment of participants was through friends and acquaintances. Digital recordings took place in quiet rooms or classrooms (for Tunisian dialect, it was mostly in the Faculty of Arts and Humanities in the University of Manouba, Tunisia). The recordings were made directly to wav format at 44.1KHz 16bit, using a Marantz PMD660 and head-mounted Shure SM10 microphones.

To minimize the influence of the observer's paradox (Labov, 1972), a local research fieldwork assistant was recruited to give instructions to the participants in their native dialect during fieldwork in Tunisia and Jeddah. Fieldwork recordings in Riyadh were run solely by the author who is a native speaker of the Najdi dialect from Saudi Arabia.

4.2.2 Materials

To address the research questions about speech rate effects on VOT – **2a)** How does speaking rate affect VOT in different voicing categories, and which laryngeal features are active in each dialect? **2b)** Are two-way voicing category dialects over-specified with two active phonological features? – and based on the results of the first experiment study and our predictions, the previously mentioned three dialects were chosen. This experiment investigates eight target plosives, six of which are plain /b, d, t, k, ɡ, q/ and two are emphatic /ḍ, ṭ/[23]. These plosives are investigated in initial and word-medial positions, followed by three long vowels /a:, i:, u:/ and three short vowels /a, i, u/. Stimuli for each dialect are real and nonsense words (N = 64 as seen in Table 4-1), and they were presented to the participants in Arabic orthography as shown in Appendix (C.1) and Appendix (C.2). The participants read the target words embedded in a carrier phrase /ɡu:l _____ marti:n/ which translates to “Say ____ twice”. There were no typographic differences in carrier phrase presentation by dialect. Tunisian, however, were expected to produce /qu:l/ rather than /ɡu:l/. Note that ‘qu:l’ ends with a sonorant vowel-like segment, rather than a plosive to avoid misreading the beginning of the targeted segment.

Table 4-1 The total count of target words per target consonant and position.

target	**position**		**total**
	word-initial	word-medial	
/t/	7	1	8
/ṭ/	7	2	9
/k/	6	-	6
/q/	6	-	6
/b/	6	-	6
/d/	6	5	11
/ḍ/	6	6	12
/ɡ/	6	-	6
total	50	14	64

[23] Bearing in mind that alveolar emphatic plosive /ḍ/ has varying surface reflexes in the modern dialects, as I mentioned in § 4.1.1 and § 4.1.3, the treatment of /ḍ/ in this chapter is identical to the previous chapter. If participants produced the interdental fricative /ð̣/, tokens were discarded. More in § 3.2.1.

4.2.3 Procedure

The target words listed in Appendix (C.1) and Appendix (C.2) were presented to participants on printed paper sheets. The target words were elicited at two different speeds. In the first time reading the script, there was no instruction regarding speaking rate (coded as “Slow”). Then, the participants read the same list of embedded target words, but they were asked to increase their speaking rate – as fast as possible – without having to sacrifice speech accuracy (coded as “Fast”). The target words embedded in the list were pseudo-randomly ordered to disguise the nature of the research experiment. The participants were asked to read a stimulus word and phrase again if they made a mistake or hesitated in production. Each recording session lasted approximately 35 minutes.

In a pilot study, two different stimulus presentation scenarios were trialled, with one participant. One was a screen-based method, following Allen & Miller’s (1999) methodology, and the other was a script-based method, as described above and as adopted by many scholars such as Beckman, Helgason, McMurray, & Ringen (2011) and Kulikov (2020). In the screen-based method, test words were presented to the participant visually on a laptop screen. To elicit normal speech rate, each word was visible on the screen for 1500 ms, then the next word was displayed afterwards while the previous word started to fade away from the screen. For the faster speech rate, the procedure was the same except that the duration of the words being displayed was only 750 ms, to stimulate a faster rate of speech. The script-based method showed better and more obvious separation of speech rates with a difference of 201 ms in the word duration of the scripted fast vs normal compared to a 104 ms difference between the screen-based fast vs normal speech rates. The script-based method was therefore used in the main study, and degree of rate differences is reported in the results § 4.5.1.

4.3 Measurements and Segmentation

Acoustic measurement and segmentation were undertaken by using Praat version 6.0.16. Firstly, Prosody Lab Aligner (Gorman et al., 2011) was used to force-align an orthographic transcription to the audio in Praat textgrids, yielding two levels of segmentation– two interval tiers, one for phones and one for words, as shown in Figure 4-1. The tiers were used to identify word duration and to assist in labelling VOT in target consonants and formants in following vowels.

The total number of potential tokens for the Najdi dialect was 2304 (18 speakers x 64 target items x 2 speech-rates), but 269 tokens were disfluent, leaving 2035 tokens in the Najdi subset. As for the Hijazi dialect, the total number of potential tokens was 2304 (18 speakers x 64 target items x 2 speech-rates), but 274 tokens were disfluent, resulting in 2030 tokens in the Hijazi subset. Lastly, the total number of potential tokens for the Tunisian dialect was 3584 (28 speakers x 64 target items x 2 speech-rates). This subset of the data included 633 disfluent tokens (mostly because all tokens of [g] were discarded because it was realized in Tunisian as [q]); these tokens were excluded from the analysis leaving 2951 tokens. This results in a total of 7016 tokens used in the dataset across all three dialects.

The boundaries of segments are determined visually and labelled manually by inspecting waveforms and broadband spectrograms in Praat. VOT is measured for all plosives in milliseconds (ms). Negative VOT values are considered as the interval from the voice onset to the onset of the plosive burst. Both spectrograms and waveforms are used also to visually identify the onset of periodic striations as an indication of vocal fold vibration. As for positive VOT values, they are measured as the interval from the plosive burst to the onset of voicing. Overall word duration was also measured to indicate speech rate.

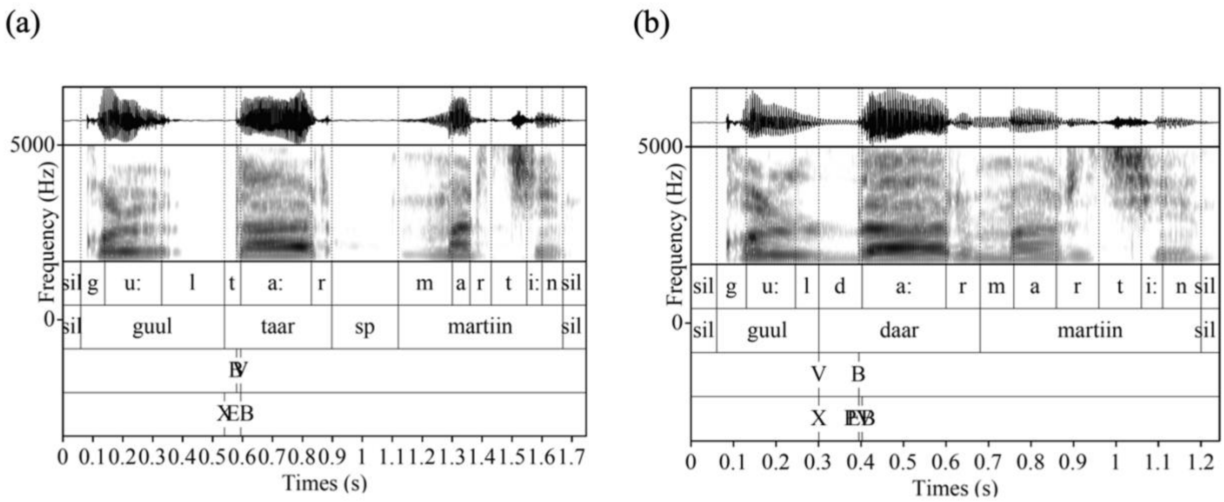


Figure 4-1 Segmentation of the words (a) /ta:r/ and (b) /da:r/ produced by a Tunisian male speaker, with four tiers: (1) phones, (2) words, (3) VOT point (B = burst; V = onset of voicing), and (4) closure duration + %voicing.

A number of acoustic correlates were manually labelled then automatically measured using a Praat script. A total of 14 acoustic temporal and non-temporal variables were extracted. The following sections describe the criteria and description of the acoustic variables chosen to further analyse the nature of voicing in the three dialects.

4.3.1 Word Duration (ms)

The word-level segmentation was used for measuring the total duration of the target word in milliseconds. Target word duration was measured as a proxy for speaking rate.

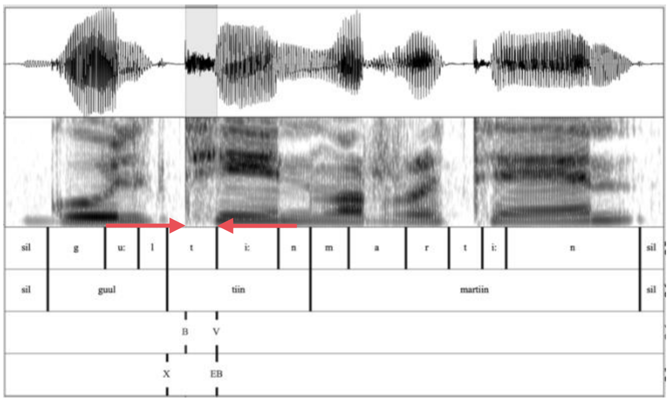
4.3.2 VOT (ms)

To measure VOT, a point tier was added using a Praat script with two boundaries inserted in the tier: one labelled (B) for the plosive burst and the other labelled (V) for the onset of periodic striations of vocal folds vibration. These two boundaries in the point tier were then placed and adjusted manually to the position of the burst or the vocal fold vibration onset, based on the waveform and spectrogram displays in Praat.

Certain phonemes differ according to the target dialect. If participants produced a non-dialect pronunciation, the token was discarded. For instance, in certain dialects where [q] is not in the dialect's inventory, replacing velar [g] with uvular [q] as pronounced in MSA, was not acceptable. And vice versa, if the target word was intended to be produced in velar [g] but the dialect does not have this realization for [q] then the token was discarded as well.

It was noticed during the segmentation process that in the context of /i:/ as a following vowel, the VOT production of the voiceless /t/ and /ṭ/ in the Tunisian dialect was perturbed by palatalization. According to Youssef (2015), the context of a front high vowel /i:/ is reported to trigger a palatalization effect for Cairene Arabic, and a similar pattern was found in our Tunisian data. The following figures are examples of the segmentation process followed in this study for a minimal pair /ti:n ~ ṭi:n/ in two dialects, Najdi (Figure 4-2) and Tunisian (Figure 4-3). There is no palatalization effect in the context of the front high vowel for e.g., Najdi speakers, as seen in in Figure 4-2. There is, however, a clear palatalization effect is in the context of the front high vowel for the Tunisian speakers, as seen in Figure 4-3.

(a)



(b)

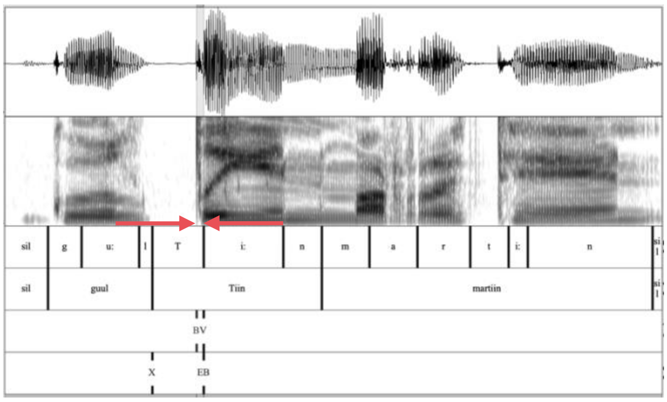
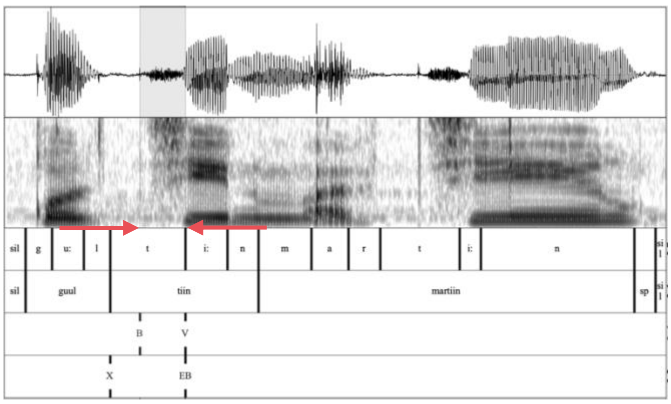


Figure 4-2 Segmentation of (a) /ti:n/ (VOT = 65 ms) and (b) /ṭi:n/ (VOT = 15 ms) by a Najdi female speaker with tiers: (1) phones, (2) words, (3) VOT points, and (4) closure duration + %voicing. VOT values in segmented lines between red arrows.

(a)



(b)

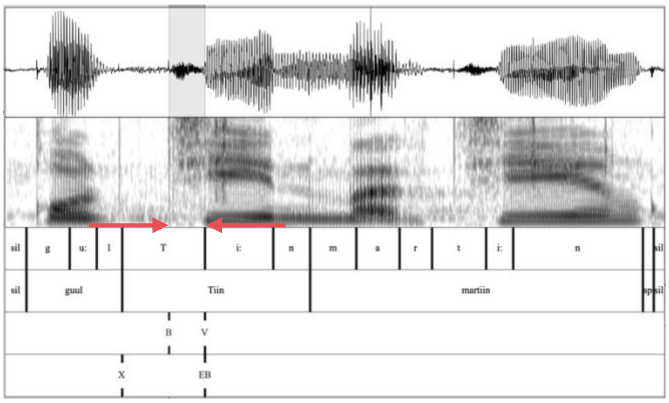


Figure 4-3 Segmentation of (a) /ti:n/ (VOT = 86 ms) and (b) /ṭi:n/ (VOT = 66 ms) by a Tunisian female speaker with levels (1) phones, (2) words, (3) VOT points, and (4) closure duration + %voicing. VOT values in segmented lines between red arrows.

4.3.3 Closure Duration (ms)

Closure duration was measured in (ms). The same script mentioned above was used to insert a second point tier (tier 4 in the figures above) with two boundaries ‘X’ the beginning of the target word, and ‘EB’ i.e., end of burst or the release of the plosive which were then manually adjusted to the correct position. Closure duration was measured on the distance from ‘X’ to ‘EB’. In voiced plosives, closure duration may overlap with voicing lead VOT values if fully voiced, but in voiceless plosives it is the closure duration before the burst. In addition to VOT, measuring closure duration can show greater consistency, in the duration of glottal gestures across places of articulation, than VOT (Abramson & Whalen, 2017).

4.3.4 Burst Duration (ms)

Burst duration is measured in (ms). It is labelled from ‘B’, the start of the burst, on tier (3) to ‘EB’, the end of the burst on tier (4). Burst duration will overlap with voicing lag durations in voiceless plosives, but in voiced plosives burst duration might be a cue to the plain and emphatic contrast, since VOT was shown not to be an indicator of the voiced emphatic contrast.

4.3.5 Percentage of voicing (%)

During manual labelling, when the closure of a voiced plosive was fully voiced, a ‘PV’ label on tier (4), that stands for %voicing, was inserted at the same position of the burst ‘B’ on tier (3). If the closure was observed, auditorily and impressionistically by reference to the spectrogram, to be not fully voiced, the 'PV' label was manually adjusted to where the voicing amplitude ends. The Praat script calculated the distance between 'PV' on tier (4) from 'V' on tier (3) and divides it by the duration of the closure to give us the percentage of voicing during the closure.

For a voiced plosive to qualify as 'prevoiced', the voicing lag threshold of 50% is followed (Abramson & Whalen, 2017). For voiced plosives, this percentage might also be an indicator of the plain and emphatic contrast and the degree of emphasis. The usefulness of this criterion has been used to test voicing in the singleton vs. geminate contrast of Lebanese Arabic (Al-Tamimi & Khattab, 2018).

4.3.6 Fundamental frequency, F1, and F2 in following vowels (Hz)

The fundamental frequency f0 of the vowel following the target consonant was estimated at multiple points and extracted in semitones. F1 and F2 values in the vowel following the target plosive were also obtained in Hz. Using a Praat script, we followed a dynamic multiple-point measurement approach, extracting formant resonance frequencies at the 25% point, midpoint, and 75% point of the vowel (Almurashi, Al-Tamimi, & Khattab, 2020). Formant frequencies were estimated using the default Burg algorithm in Praat. The formants were measured with a maximum frequency of 5000 Hz for male and 5500 Hz for female speakers. These reflexes of vowel quality (f0/F1/F2 at the multiple points of the vowel) are used to provide a first indication whether dialectal variation in VOT of plain/emphatic plosives is matched by variation in f0/F1/F2 in the following vowel. The fundamental frequency tends to be higher in voiceless contexts compared to voiced ones (Al-Tamimi & Khattab, 2018), and here, we test it as a potential cue to the plain and emphatic contrast.

Due to demographic considerations, that is, to remove the influence of gender physiological differences, formant measurements of vowels are normalized (Foulkes et al., 2010). Using NORM (Thomas & Kendall, 2007), the F1/F2 values were normalized following the Lobanov (1971) z-score method of normalization given that the data were vowel-extrinsic but formant-intrinsic.

4.4 Results 1: Overview

In the next section, § 4.4.1, we firstly lay out the nature of the emphatic contrast in Najdi, Hijazi, and Tunisian in comparison with the findings for the IVAr data explored in the previous chapter. Then in the following section, § 4.5, we firstly assess the effectiveness of the speech-rate manipulation by examining word duration as a function of the two different speech-rates in § 4.5.1. In the later sections, § 4.5.2 – 4.5.5, we address speech rate effects on the emphatic contrast (/d/ ~ /ḍ/, /t/ ~ /ṭ/) in the chosen dialects. Finally, in § 4.6, we investigate the effect of speech rate manipulation on the full set of voiced and voiceless plosives in Arabic (/b, d, ḍ, (g)/ ~ /t, ṭ, k, q/), in the three dialects.

4.4.1 Where do these dialects fall in the continuum?

In this section, I report the results for all three dialects (Najdi, Hijazi, and Tunisian) by comparing the behaviour of four plosives (emphatic contrasts) /t/ ~ /ṭ/ and /d/ ~ /ḍ/ in the normal (slow) speech-rate only, to see where they might fall in the continuum of variation revealed in the previous chapter. This initial check serves also to determine a baseline of the behaviour of these plosives in the one previously unexplored dialect (Hijazi).

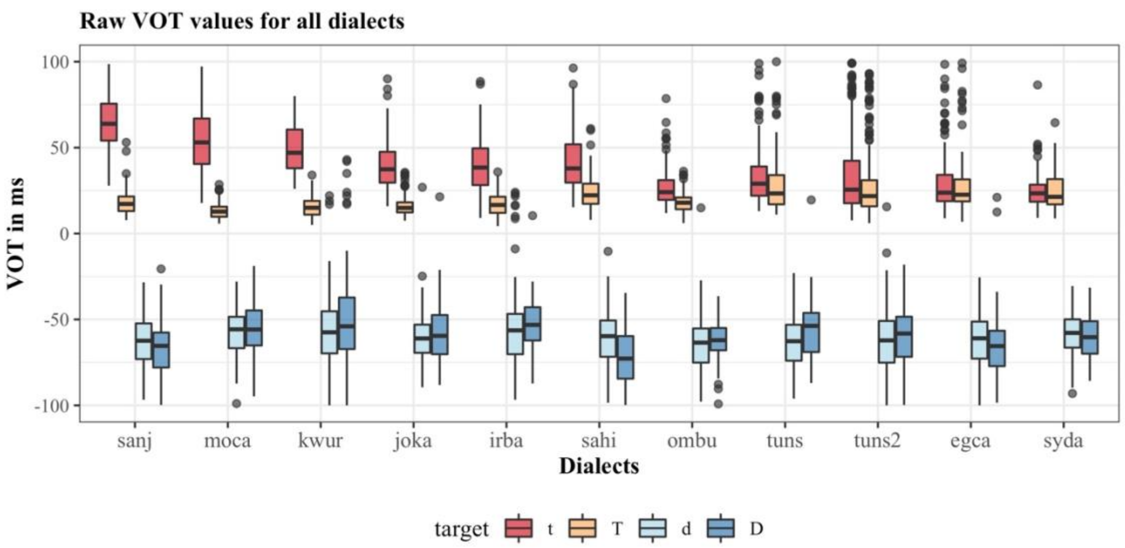


Figure 4-4 Median and interquartile range for raw VOT of /t/, /ṭ/, /d/ and /ḍ/ in the IVAr dialects (Chapter 3) plus Najdi (sanj), Hijazi (sahi), and Tunisian (tuns2) in Experiment 1.

Figure 4-4 shows raw VOT values for the four coronal plosives (/t ~ ṭ/ and /d ~ ḍ/) in the Najdi (= sanj), Hijazi (= sahi), and Tunisian (= tuns2) dialects extracted to compare their values to the IVAr dialects from chapter 3. All dialects are positioned in this figure according to the manually calculated difference in the mean between /t/ and /ṭ/ in each dialect. As we can see, the Hijazi dialect falls right in the middle, and looks roughly intermediate between Iraqi and Omani. Visually, the Hijazi values of /t/ and /ṭ/ are not merged, but they are closer to being merged than those of Iraqi. Hijazi looks similar to the three-way contrast dialects, however, in that it has more spread values for /t/ compared to the five dialects on the right of the continuum. Najdi is located leftmost, with more distinct values between /t/ and /ṭ/ than in Moroccan. The new Tunisian data, on the other hand, reflect closely the measurements for the earlier IVAr Tunisian dataset (collected in 2014) discussed in chapter 3.

First, using linear mixed-effects models, we ran a model on the subset of data containing the three new dialects, for four plosives at one speech-rate (slow). The model's dependent variable was *VOT,* and the fixed factors included *dialect*, *target*, *gender*, as well as the interaction between them, plus *position* as a further fixed factor. All the categorical factors were sum-coded to centre our categorical predictors around the mean. The coding of the categorical predictors with sum-coding was as follows: *gender* (female = 1, male = -1), *position* (initial = 1, medial = -1), and *target* (three predictors of /t/ ~ /ḍ/, /ṭ/ ~ /ḍ/, and /d/ ~ /ḍ/ in which the first listed segment was set to 1, /ḍ/ to -1, and all unlisted segments to 0). Random intercepts were included for *speaker* and *item* as well as a random slope of *target* by *speaker*.

4.4.2 Overview

For all three investigated dialects, the values of the four plosives (/t ~ ṭ/ and /d ~ ḍ/) are illustrated in Figure 4-5 below. The dialects display clear separation between the voiced and the voiceless consonants with voicing lead in the voiced set /d/ and /ḍ/, and voicing lag in the voiceless set of sounds /t/ and /ṭ/.

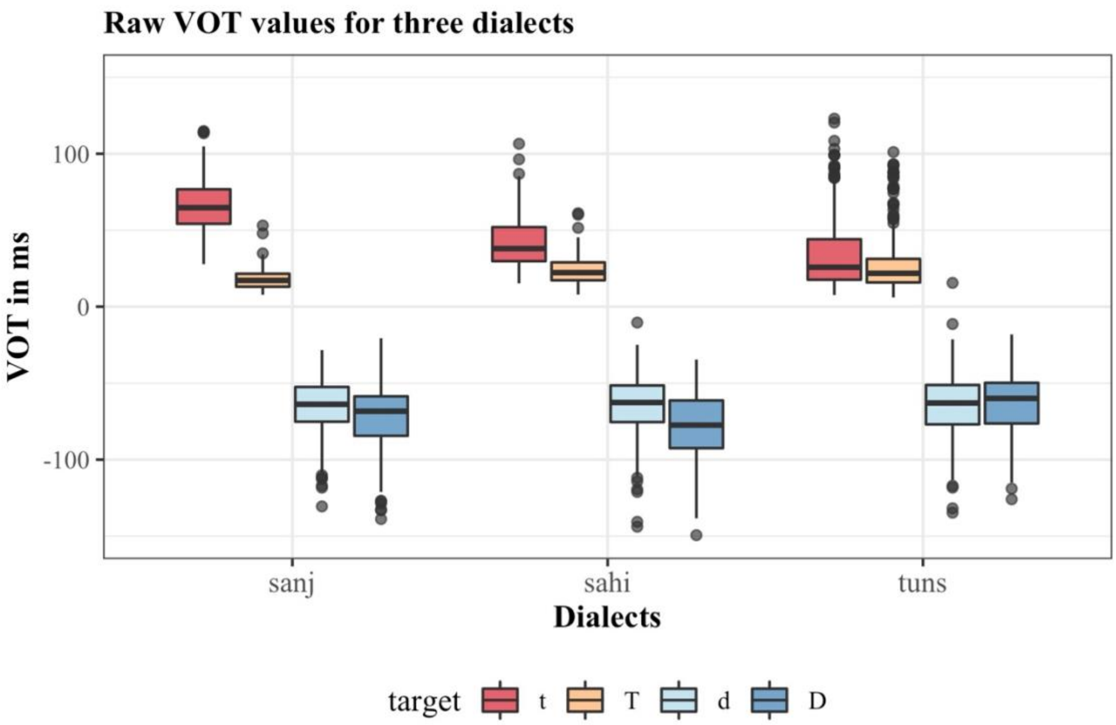


Figure 4-5 Median and interquartile range for raw VOT of Najdi, Hijazi, and Tunisian in /t/, /ṭ/, /d/ and /ḍ/ by dialect in slow speech-rate only.

The variation here lies in the extent of voicing lag in the voiceless subset, i.e., the degree of VOT difference between /t/ and /ṭ/. From Figure 4-5 above, we can see that Najdi dialect (N = 677) (on the left of the figure) shows a three-way distinction in VOT (voiced, voiceless emphatic, voiceless plain), and Tunisian (N = 950) (on the right of the figure) has a two-way VOT distinction (voiced, voiceless). Meanwhile, Hijazi (N = 691) (in the middle) has values that are intermediate: not quite merged but not as discrete as in the Najdi variety. This visualization is consistent with our finding in chapter 3 that Arabic voiceless emphatics are likely undergoing a process of sound change in some dialects, evident by different VOT behaviour across Arabic dialects/varieties.

The partial overlap of values for /t/ and /ṭ/ in Hijazi is accompanied by greater spread of VOT values for /ṭ/, while the long lag VOT values of /t/ are more uniform visually. In Figure 4-6, it is not clear whether gender differences in the dialects are present or not; it looks like there may be a difference in the Najdi dialect, as female speakers' /t/ and /ṭ/ are further apart in this variety. Tunisian female speakers have overall more dispersed values of /t/ and /ṭ/ than male speakers. Hijazi, on the other hand, shows no obvious influence of gender on VOT.

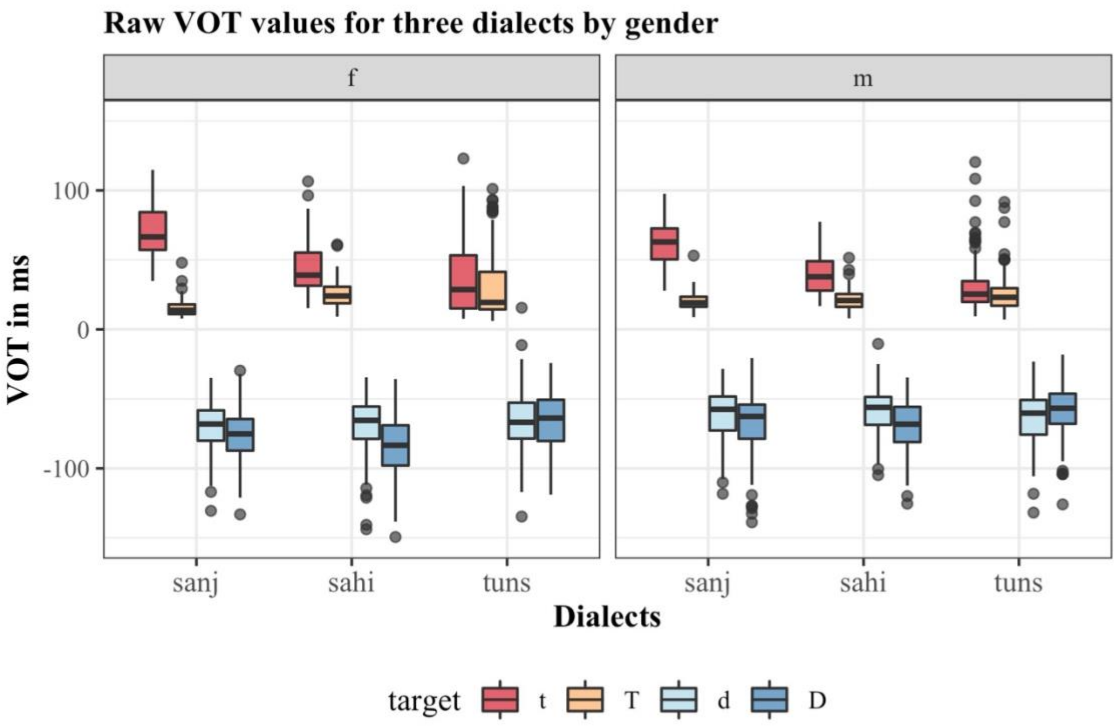


Figure 4-6 Median and interquartile range for raw VOT of Najdi, Hijazi, and Tunisian in /t/, /ṭ/, /d/ and /ḍ/ by dialect and gender in slow speech-rate only.

In this initial overview, we have confirmed that all the dialects have at least a two-way distinction between voicing lead in /d/ ~ /ḍ/ and different degrees of the voicing lag in /t/ ~ /ṭ/. In the next subsection, we examine this subset of the data statistically — first in the voiced subset then the voiceless — before drawing conclusions on the role of VOT as a phonemic cue to the plain and emphatic contrast in these dialects.

4.4.3 VOT in /d/ ~ /ḍ/

All three dialects display voicing during the closure for both voiced consonants /d/ and /ḍ/. A linear mixed-effects model [24] on the voiced subset revealed a main effect of *target* (β = 3.54, t = 2.46, p = .021), *gender* (β = -4.29, t = -3.31, p = .002), and *position* (β = -5.88, t = -4.37, p < .001). The effect of target is reflected in overall longer VOT values of /ḍ/ compared to /d/, while in gender and position VOT values are shorter for male speakers compared to female speakers, and shorter word-medially than word-initially. The model revealed a significant two-way interaction between dialect and target in only two of the three dialects. Hijazi speakers produced a greater difference in

[24] Model used: vot ~ group * target * gender + position + (1 | item) + (1 + target | speaker), data = voiced (slow /d ~ ḍ/).

the VOT values between /ḍ/ and /d/ than the other dialects (*sahi* x */d/:* β = 3.70, *t* = 3.33, *p* = .002) while Tunisian speakers VOT were the opposite, with the least difference between /d/ and /ḍ/ compared to the other dialects (*tuns* x */d/:* β = -4.54, *t* = -4.30, *p* <.001). The model results are illustrated in Figure 4-7 below, and the model summary is provided in Appendix (C.3).

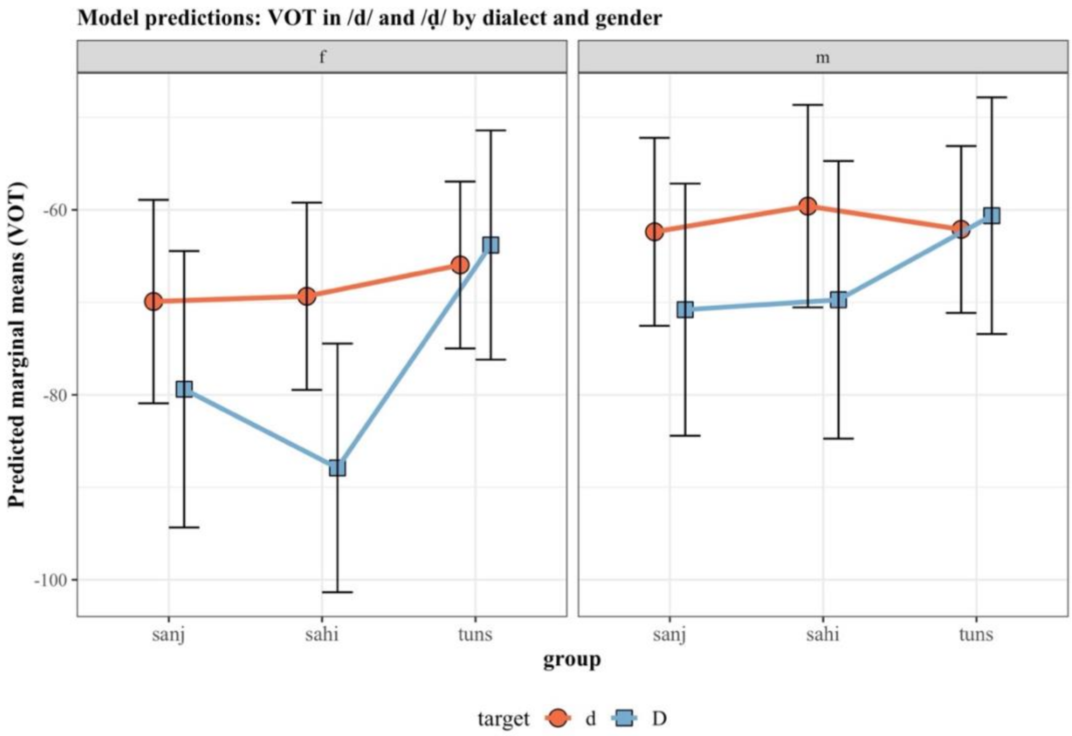


Figure 4-7 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian VOT in /d ~ ḍ/ by dialect and gender.

The main interest in this section is to investigate whether dialects vary from each other in the extent to which VOT is used to distinguish /d/and /ḍ/. In Figure 4-7 above, the 95% confidence interval around predicted mean VOT shows varying yet overlapping distributions in all three dialects, suggesting that this laryngeal cue is not used to differentiate the contrast between /d/ and /ḍ/ in any dialect.

4.4.4 VOT in /t/ ~ /ṭ/

The main locus of variation is in the voiceless emphatic contrast, as VOT values of the plain and emphatic /t/ and /ṭ/ vary across dialects. All three dialects display voicing lag for both voiceless consonants /t/ and /ṭ/. The linear mixed-effects model[25] on the

[25] Model used: vot ~ group * target * gender * position + (1 | item) + (1 | speaker), data = voiceless (slow /t/ ~ /ṭ/).

voiceless subset also revealed a main effect of *target* (β = 14.95, *t* = 4.47, *p* < .001) with longer VOT values in /t/ compared to /ṭ/ across the three dialects, and of *gender* (β = 3.37, *t* = 3.70, *p* < .001) with longer VOT values by female speakers, but there was no main effect of *position* (β = -6.34, *t* = -1.90, *p* = .08). The model results are shown in Figure 4-8 below, and the model summary is provided in Appendix (C.4).

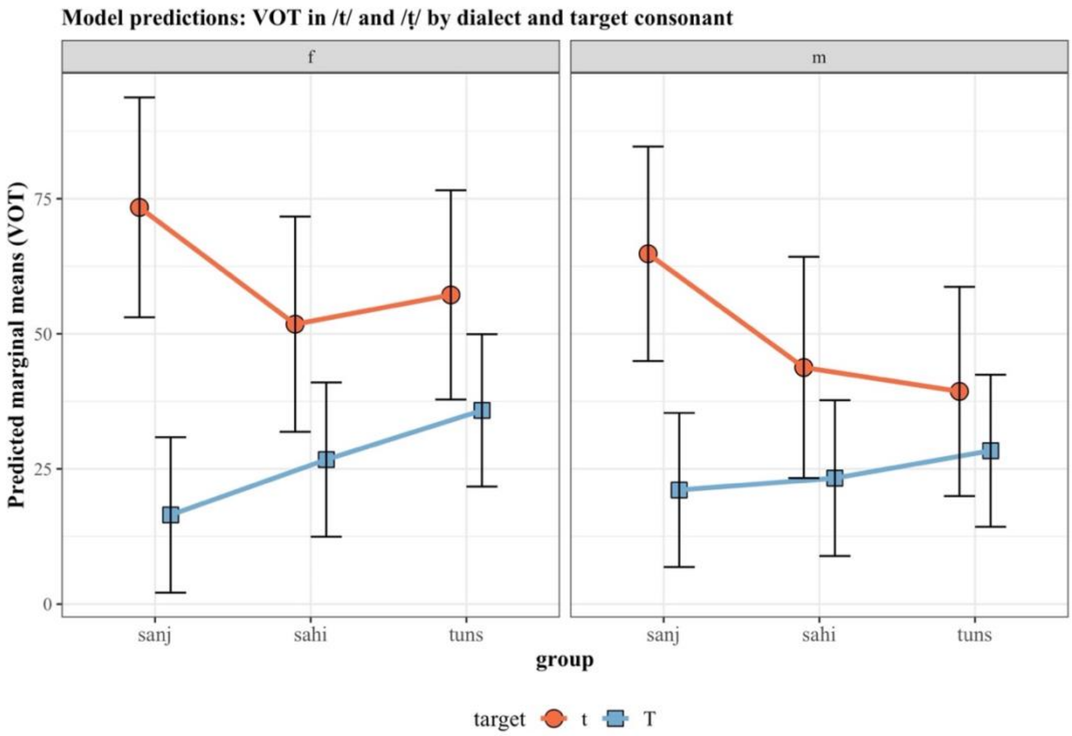


Figure 4-8 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian VOT in /t ~ ṭ/ by dialect and gender.

The model also showed a significant two-way interaction between dialect and target across all three dialects, albeit to varying degrees. Najdi shows the most distinct predicted mean VOT values in /t/ and /ṭ/ regardless of *gender* and *position* (*sanj* x */t/:* β = 10.19, *t* = 12.09, *p* < .001) followed by Hijazi with overlapping values of VOT (*sahi* x */t/:* β = -3.42, *t* = -3.96, *p* < .001), then Tunisian with complete overlap between /t/ and /ṭ/ (*tuns* x */t/:* β = -6.77, *t* =-8.83, *p* < .001). It was only Tunisian that recorded a two-way interaction between *dialect***gender* reporting more overlap between the plain and emphatic contrast as produced by male speakers (*tuns* x *female:* β = 2.97, *t* = 2.49, *p* = .015).

As we saw in the IVAr data, the conservative dialect (Najdi) shows a clear split between the short lag values of VOT < 25 ms in /ṭ/ and the long lag values of VOT > 28 ms and higher in /t/. Whereas progressive dialects (Tunisian and somewhat Hijazi) display more merged VOT values of /t/ and /ṭ/ centring around 32 ms. Gender differences are only apparent in Tunisian, with female speakers having greater difference between /t/ and /ṭ/ than male speakers, which exactly reflects the Tunisian results in the IVAr data.

4.4.5 F2 lowering in /t/ ~ /ṭ/

In this section, F2 values in the vowel following the voiceless plain and emphatic plosives are explored in six vowels (long and short /i, a, u/) and in word-initial and word-medial positions. Since VOT is not a robust phonetic cue to the plain vs. emphatic contrast in the voiceless plosives in Hijazi and Tunisian, we are interested in looking also at the most prominent phonetic cue agreed to distinguish this contrast, namely F2 lowering. The goal is to know whether these two dialects maintain a cue to emphasis through systematic F2 lowering after /ṭ/ regardless of the loss of the voicing contrast in /t/ and /ṭ/, as was the case for the dialects in the IVAr data (see § 3.3.4).

In this dataset, we obtained dynamic formant measurements at three points of the vowel. Before running linear mixed-effect models, these three points of the vowel (25-50-75) were plotted to visualize and assess the point that showed the greatest degree of F2 lowering. The onset point of the vowel (point 25) had the greatest degree of lowering as expected.

In the voiceless subset of the data at one speech rate (slow), we explored the degree of F2 lowering at 25% into the following vowels using a linear mixed-effect model. The dependent variable in the model was F2 in the following vowel (*f2.25)* and the fixed factors included *dialect*, *target*, *following vowel* and *gender* as well as the interaction between *dialect***target*[26]. The categorical factors were sum-coded in this subset, and the model included a random intercept for *item*. The model showed a significant main effect of *target* ($\beta = 0.36$, $t = 4.22$, $p < .004$) indicating higher F2 values in vowels following /t/ compared to the overall average of values across all dialects. The model did not show a significant main effect of *gender* ($\beta = -0.03$, $t = -1.84$, $p = .066$) though female speakers showed lower F2 values in general. Figure 4-9 shows the predicted marginal mean (and 95% CI) for F2 values in the vowel following the voiceless plain and emphatic plosives in all three dialects by gender. The full model summary is provided in Appendix (C.5).

[26] Model used: f225 ~ group * target + gender + folv + (1 | item), data = unvoiced (slow /t ~ ṭ/).

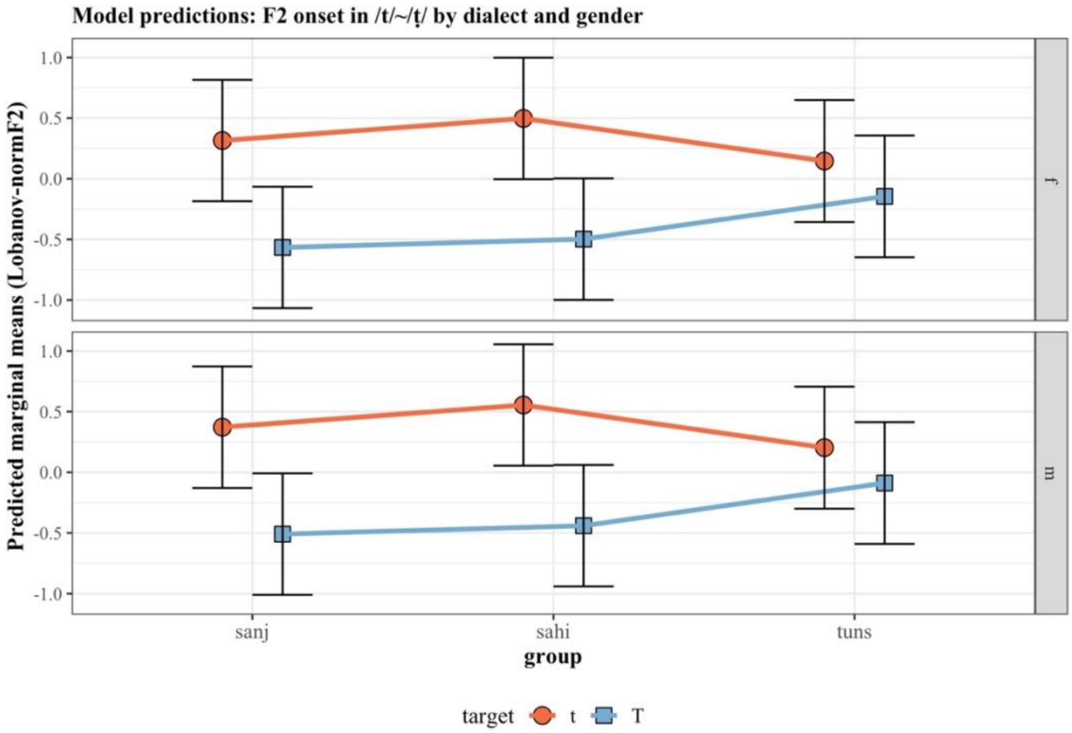


Figure 4-9 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian normalized F2 onset values of vowels following /t ~ ṭ/ by gender.

The model also showed a significant interaction between dialect and target across all three dialects. The interaction in the Tunisian dialects was negative, such that F2 values after /t/ were lower than the overall average (*tuns x /t/:* $\beta = -0.22$, $t = -10.04$, $p < .001$) resulting in reduced difference between the plain and emphatic vowels F2 values.

4.4.6 Results 1: Summary

In summary, Hijazi and Tunisian show a similar VOT distribution for /t/ and /ṭ/, while Najdi shows distinct VOT distribution in the same contrast. Considering the results for both the voiced and voiceless plosives, we can conclude that Najdi is a three-way voicing contrast dialect, as expected, with voicing lead /d ~ ḍ/, long lag /t/, and short lag /ṭ/. The Tunisian results indicate that it is a two-way voicing contrast dialect with /d ~ ḍ/ and /t ~ ṭ/. Hijazi is in the middle, with not completely merged values between /t/ and /ṭ/. Compared to the IVAr data, Hijazi and Tunisian might still be considered hybrid dialects, though. The cues to emphasis were also explored by looking at the extent of F2 lowering in vowels following the voiceless plosives /t/ and /ṭ/; the results showed an effect of lowering in all three dialects, though the lowering was the least in the Tunisian dialect.

4.5 Results 2: Speech-rate effects on the emphatic contrast

4.5.1 Speech-rate manipulation

In reviewing the literature, we mentioned in § 2.4 that previous studies showed a relationship between speaking rate and VOT, in which speaking rate influences VOT asymmetrically. This asymmetry lies in an effect of speaking rate only on some of the plosives of the investigated languages. Since the founding work of Summerfield (1975), speaking rate effects on VOT have been used as diagnostic of the active phonological features in a given language (Allen & Miller, 1999; Beckman et al., 2011; Kessinger & Blumstein, 1997; Kulikov, 2020; Magloire & Green, 1999; Midtlyng, 2011; Miller et al., 1986; Pind, 1995; Summerfield, 1975, 1981).

The main interest of this chapter is to know whether speech-rate manipulation has an effect on the VOT values of both voiced and voiceless plosives in the examined dialects of Arabic. If, based on the findings in the previous section, the Tunisian dialect is to be described as having a two-way voicing contrast, then only one feature is expected to be active, in this case [voice]; that is, only voiced plosives would be affected by an increased speech-rate. If, on the other hand, both voiced and voiceless prove to be affected by speech-rate manipulation, then this might be a matter of ‘over-specification’ of laryngeal features as in the case of Swedish (Beckman et al., 2011). As for Hijazi, since it is possibly a hybrid dialect between having a two-way and a three-way voicing contrast, then it is unknown whether it has two active features or not. For Najdi dialect, we have established in the previous section that it can be described as having a three-way voicing contrast, so in this case, both the voiced plosives and the aspirated long lag plosives would be affected by an increase in speech-rate; which means, two features are expected to be active [sg] and [voice].

Before attempting to investigate these questions, it is necessary first to validate the effectiveness of the speech-rate manipulation. Therefore, *word duration* was examined as a function of the *speech-rate* variable (slow ~ fast).

4.5.1.1 Word duration

It is crucial to ensure that the speech-rate manipulation has achieved the intended effect. Word duration in milliseconds was used as a proxy for speech-rate in the repeated measures. Figure 4-10 shows raw values of word duration at two speeds by dialect.

Visually we can see that the speech-rate manipulation was successfully achieved; there is little to no overlap between the *fast* and *slow* 'normal' speech-rates in all dialects.

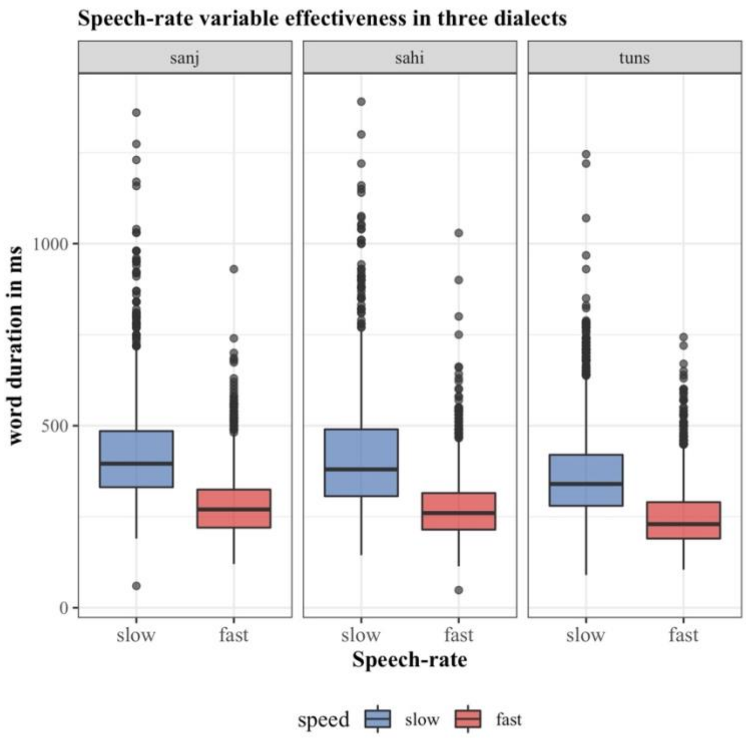


Figure 4-10 Effect of speech-rate on word duration by dialect (Najdi – Hijazi – Tunisian).

To provide further statistical support, a linear mixed-effects model[27] was run on the whole dataset. *Word duration* was the dependent variable and *speed* (fast, slow) interacting with *dialects* were the fixed factors, the model included a random intercept for *item*. The categorical factors, *speed* and *dialect*, were sum-coded (*speed*: slow = 1, fast = -1) and (*dialect*: two predictors of *sanj* ~ *tuns* and *sahi* ~ *tuns* in which *sanj* was set to 1, *tuns* to -1, and *sahi* to 0). As expected, the model revealed a significant main effect of speech-rate ($\beta = 66.27$, $t = 67.41$, $p < .001$), indicating that word duration in slow speech was significantly longer than it was in fast speech in all dialects. The full model summary is provided in Appendix (C.6).

4.5.2 VOT ~ Speech-rate (overview)

Since the speech-rate manipulation attained the desired results in all chosen dialects, it is safe now to examine the question in hand: does speech-rate affect VOT across the targets /d ~ ḍ/ and /t ~ ṭ/? In this section, § 4.5.2, we firstly assess the overall effect of

[27] Model used: worddur ~ speed * dialect + (1 | item), data = all the dataset

speech-rate on the emphatic contrast (/d ~ ḍ/ and /t ~ ṭ/) in all three dialects. Then, § 4.5.3 is a closer look on speech rate effects on (/d ~ ḍ/ and /t ~ ṭ/) in Najdi, followed by Hijazi in § 4.5.4 then finally Tunisian in § 4.5.5.

As a first step, we explored raw values of VOT as shown in Figure 4-11 below. The boxplots display VOT values for voiced /d ~ ḍ/ and voiceless /t ~ ṭ/ in the Najdi, Hijazi, and Tunisian dialects at two speech-rates. Overall, there is a clear separation between the phonologically voiced and voiceless plosives, with voicing lead in the voiced plosives and voicing lag for the voiceless ones, at both speech-rates. Speech-rate displays some impact on the voiced subset in all dialects, meaning /d ~ ḍ/ at the slow 'normal' speech-rate have longer values of voicing lead than /d ~ ḍ/ at the speeded speech-rate. In contrast, in the voiceless category, there seems to be varying impact of speech-rate on the plain voiceless contrast. There is some influence on /t/ (longer voicing lag at slower rates) but this influence of speech-rate is not the same in all dialects; it is evident in Najdi, but it is more marginal in Hijazi and Tunisian, and not as strong as the impact on the voiced subset. The two voiceless plosives /t/ and /ṭ/ display minimal to no impact of speech-rate in Hijazi and Tunisian, as there is an obvious overlap of VOT values. The emphatic /ṭ/ displays the least (and almost no) impact of speech-rate in all three dialects. VOT values for /ṭ/ in Najdi show a compact distribution at both speech rates, compared to a more spread distribution of values for /ṭ/ in Hijazi and Tunisian at both speech rates.

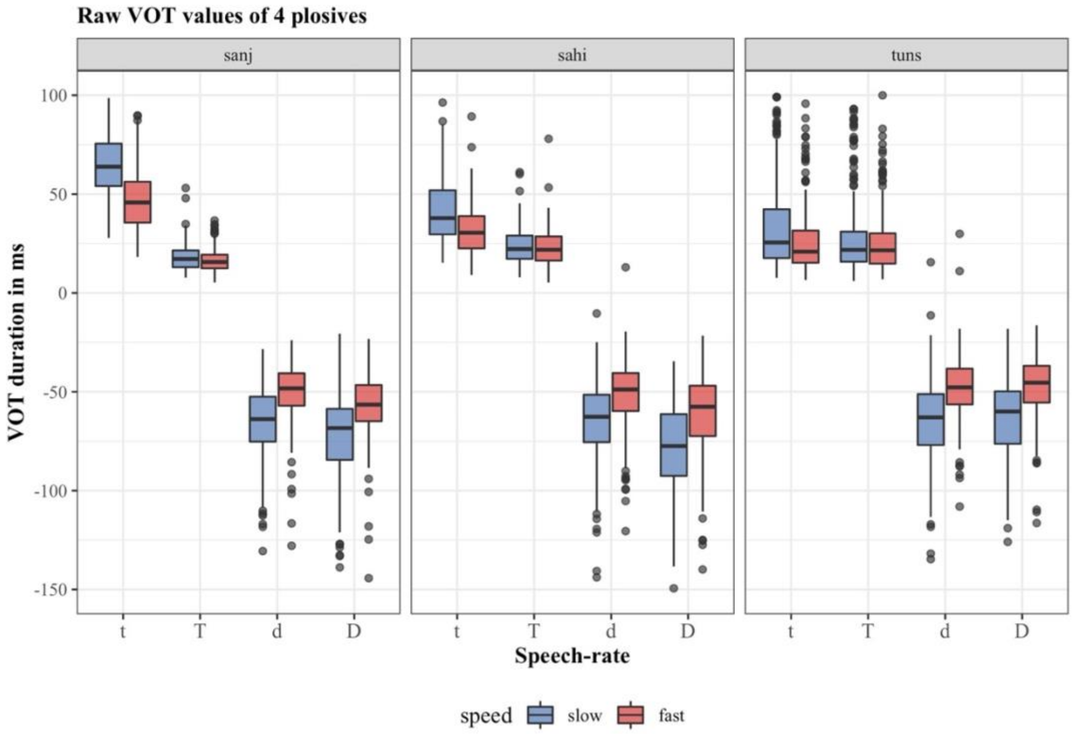


Figure 4-11 Median and interquartile range of raw VOT in Najdi, Hijazi, and Tunisian /t ~ ṭ/ and /d ~ ḍ/, by speech rate.

In the following sections, we will investigate the effect of speaking rate on the VOT of the emphatic contrasts /d ~ ḍ/ and /t ~ ṭ/ in each dialect separately.

4.5.3 Najdi Arabic

Earlier, in § 4.4.1, we confirmed that the Najdi variety of Saudi has a three-way voicing contrast. The Najdi dialect showed discrete values of /t/ and /ṭ/, and when compared to the other dialects in the IVAr dataset, it showed the greatest separation of VOT values in this voiceless emphatic contrast.

In the speeded dataset for Najdi dialect, we can see in a closer look in Figure 4-12 that the asymmetric effect of speech rate on VOT in the voiceless set of plosives is clear, in addition to the clear effect of speech rate on the voiced subset. No obvious effect of gender is present in this visualization. VOT values for /ṭ/ appear to be in a compact range at both speech rates.

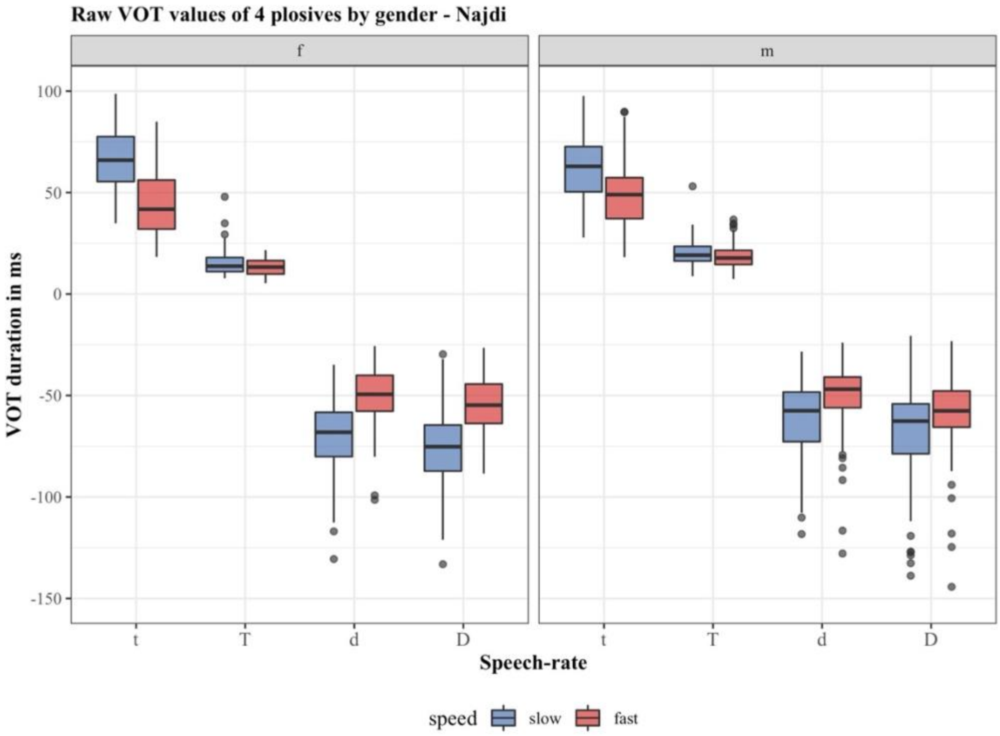


Figure 4-12 Median and interquartile range of raw VOT in Najdi /t ~ ṭ/ and /d ~ ḍ/, by speech rate and gender.

Based on the visualization above, we can initially claim that there are two active features in Najdi, [sg] and [voice], as speech rate has a clear asymmetric effect on VOT in the voiceless emphatic contrast.

To investigate this effect statistically, several linear mixed-effects models were run on the Najdi subset of the data. The best fit model included *VOT* as dependent variable with *speed*, *target*, *position*, and *gender* as fixed factors. The model included the interaction between the following factors: *speed*target*. The model also included random intercepts for *speaker* and *item* as well as a random slope by *target*[28]. As in previous models, categorical factors were sum-coded as follows: *voicing* (voiceless = 1, voiced = -1), *gender* (female = 1, male = -1), *position* (initial = 1, medial = -1), *target* (three predictors of /t/ ~ /ḍ/, /ṭ/ ~ /ḍ/, and /d/ ~ /ḍ/ in which the first listed segment was set to 1, /ḍ/ to -1, and all unlisted segments to 0). The full model result summary is provided in Appendix (C.7).

[28] Model used: vot ~ speed * target + position + gender + (1 | item) + (1 + target | speaker), data = Najdi /t, ṭ, d, ḍ/.

The model showed no significant main effect of *gender* (β = -1.26, *t* = -1.21, *p* = .246), female speakers showed slightly shorter VOT values in general. As expected, there was a two-way interaction between *speed* and *target* in all investigated plosives as shown in Table 4-2[29] and Figure 4-13 below. For the voiced plosives, speech rate showed an inverse relationship with VOT: as speech rate decreases voicing lead lengthens by about 6.5 ms in /d/ and 7.4 ms in /ḍ/. In the voiceless plosives, voicing lag in slow /t/ is estimated to be 11 ms longer than in fast /t/, while in /ṭ/ it is estimated to be 2.8 ms longer in slow speech. Although there was significant two-way interaction between *speed* and *target* in all plosives, the size of the estimate was smaller in the target /ṭ/.

Table 4-2 Model results for the two-way interaction of speed*target in the Najdi subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	11.019	0.839	13.133	< 001*
slow : ṭ	2.803	0.799	3.507	< 001*
slow : d	-6.459	0.775	-8.338	< 001*
slow : ḍ	-7.364	0.781	-9.429	< 001*

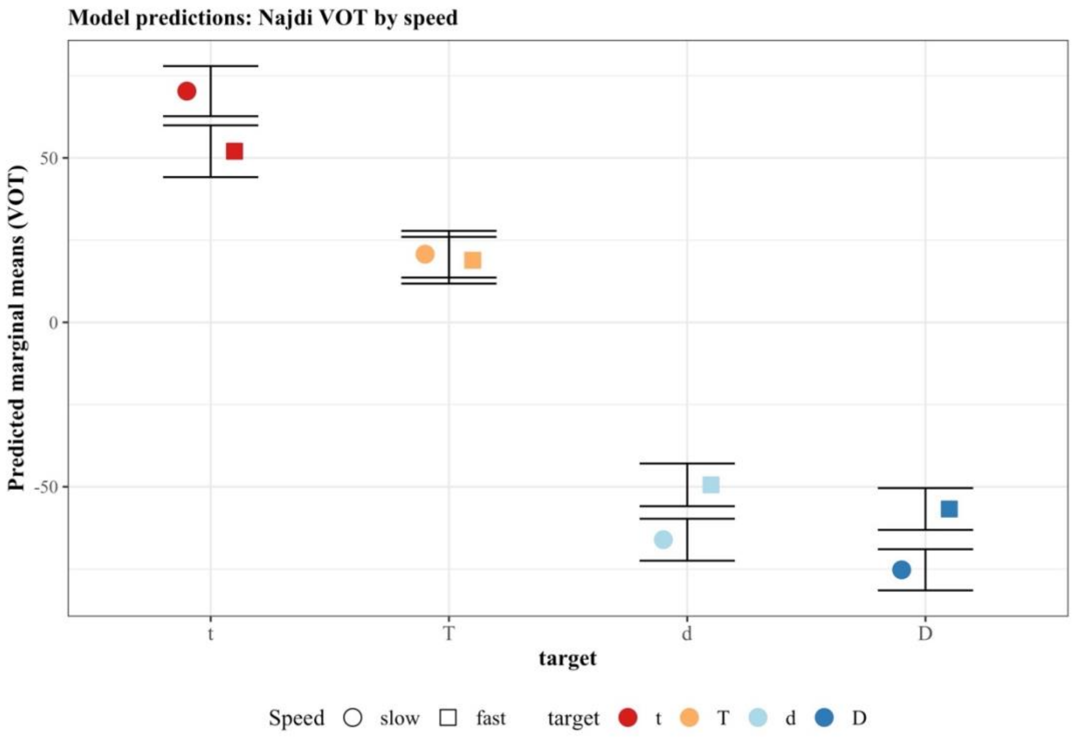


Figure 4-13 Predicted marginal mean (and 95% CI) for Najdi VOT in /t ~ ṭ/ and /d ~ ḍ/ by speech rate.

[29] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

This means that the effect of speech rate on VOT was significant in all four plosives, but this significance varied. The effect of speech rate on the VOT of the emphatic /ṭ/ was less compared to the other set of plosives in the Najdi dataset. In Figure 4-13, the overlapping CIs are only apparent for /ṭ/. Moreover, Figure 4-14 below, shows a clear 'flat' distribution for /ṭ/, unlike /t/, /d/ and /ḍ/.

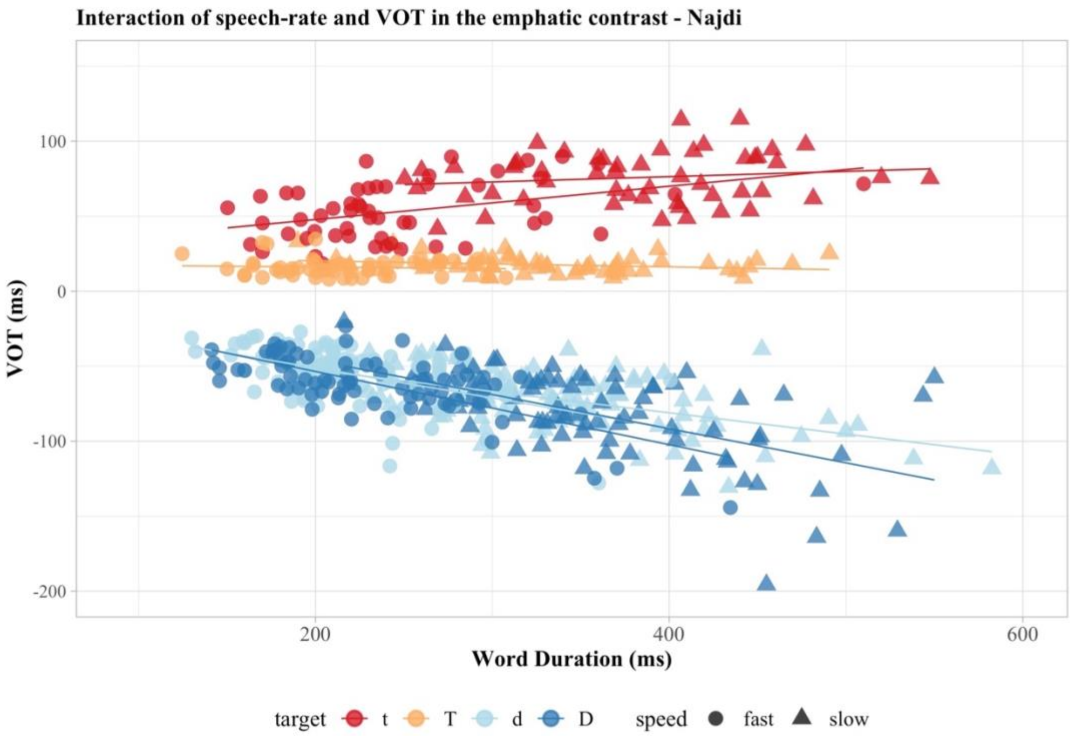


Figure 4-14 Scatter plot of word duration and VOT for /t ~ ṭ/ and /d ~ ḍ/ in both speech-rate conditions – Najdi data.

4.5.4 Hijazi Arabic

For the Hijazi dialect, we first established a baseline of the nature of its voicing contrast. Earlier in section § 4.4.1, we concluded that Hijazi falls in the middle of the continuum – to the left of Tunisian and Omani – and thus likely to be in transition to become a dialect that merges VOT across voiceless plain ~ emphatic contrast. In this section, we explore the speech rate variable and its effect on VOT in the plain and emphatic Hijazi dialect plosives.

In Figure 4-15 below, we can see a clear effect of speech rate on voicing lead in the voiced subset, similar to that seen in the Najdi results. As for the voiceless subset, it seems that the only obvious effect of speech rate on VOT is on the plain plosive /t/, i.e., longer voicing lag for /t/ in slow speech, and especially for female speakers. The presence or absence of an effect of speech rate on the emphatic /ṭ/ is not clear from this raw visualization.

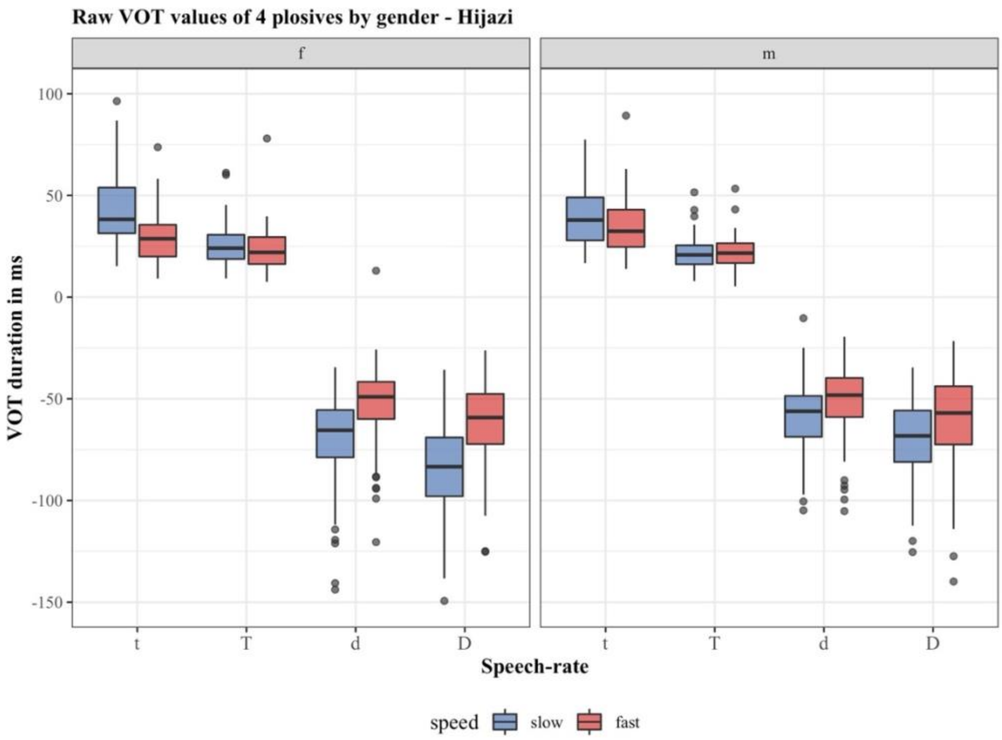


Figure 4-15 Median and interquartile range of raw VOT in Hijazi /t ~ ṭ/ and /d ~ ḍ/, by speech rate and gender.

Similar to the Najdi subset, we can initially claim that there is an active feature [voice] in Hijazi, but the asymmetric effect of speech rate on the VOT of the voiceless /t ~ ṭ/ is not so clear cut. Following the same model structure used in the Najdi data above, the best fit model for the Hijazi subset included *VOT* as dependent variable with *speed*, *target*, *position*, and *gender* as fixed factors. The model included the interaction between the following factors: *speed*target*. The model also included random intercepts for *speaker* and *item* as well as random slopes by *target*[30].

This model showed no main effect of *gender* (β = 0.77, t = 0.61, p = .55), but a significant interaction between speed and target in all four plosives as shown in Table 4-3[31] and Figure 4-16. In the voiced set of plosives, voicing lead in /d/ is estimated to decrease by -4.1 ms in faster speech rate and by -6.8 ms in /ḍ/ in faster speech. In the voiceless plosives, speech rate showed a positive relationship with VOT: as speech rate

[30] Model used: vot ~ speed * target + position + gender + (1 | item) + (1 + target | speaker), data = Hijazi /t, ṭ, d, ḍ/.
[31] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

decreases voicing lag lengthens by about 7.2 ms for /t/, while in /ṭ/ it is estimated to be only 3.7 ms longer in slow speech compared to fast speech. Although the two-way interaction between *speed* and *target* was highly significant in all plosives, /ṭ/ had the smallest estimate size compared to /t/, /d/, and /ḍ/. The full model results summary is provided in Appendix (C.8).

Table 4-3 Model results for the two-way interaction of speed*target in the Hijazi subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	7.2173	0.7864	9.177	< 001*
slow : ṭ	3.6952	0.7522	4.912	< 001*
slow : d	-4.1204	0.7075	-5.824	< 001*
slow : ḍ	-6.7921	0.7033	-9.657	< 001*

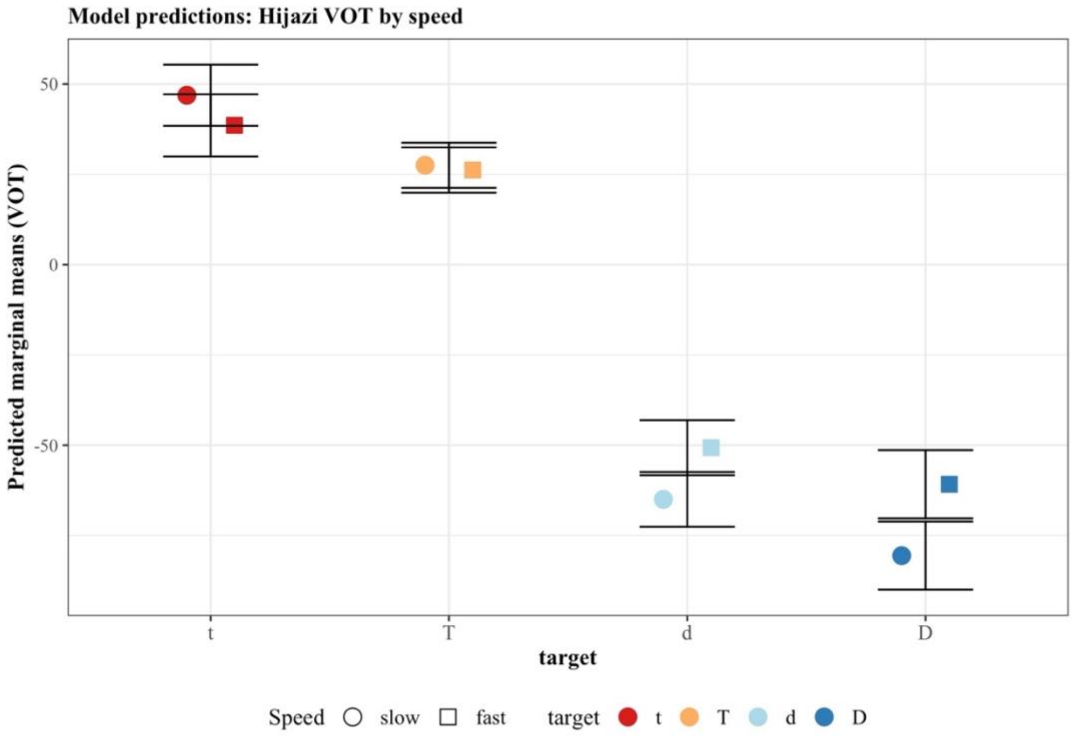


Figure 4-16 Predicted marginal mean (and 95% CI) for Hijazi VOT in /t ~ ṭ/ and /d ~ ḍ/ by speed.

Again, the effect of speech rate on VOT was significant in all four plosives of the Hijazi dialect, but the size of estimate varied. The effect of speech rate on the VOT of the emphatic /ṭ/ was smaller compared to the other set of plosives in this dataset. Figure 4-16 shows that the overlapping CIs are apparent for /ṭ/ and somewhat for /t/.

Furthermore, Figure 4-17 below, shows a flatter distribution for /ṭ/, also for /t/, unlike /d/ and /ḍ/, which could be an indication of transition.

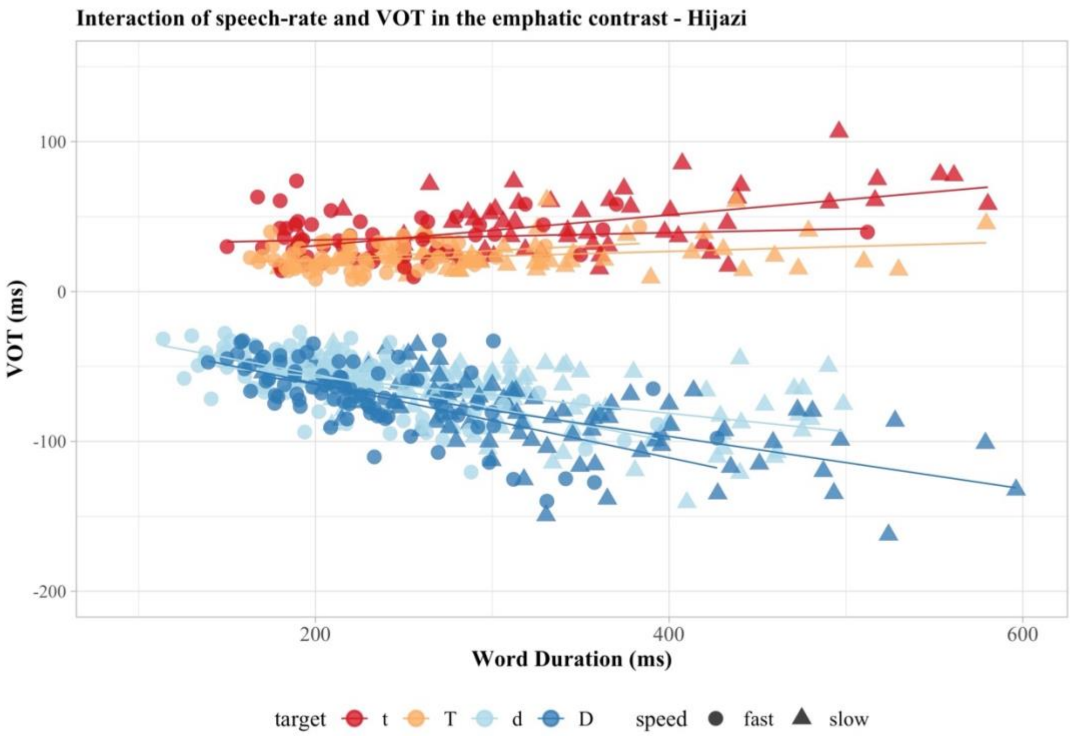


Figure 4-17 Scatter plot of word duration and VOT for /t ~ ṭ/ and /d ~ ḍ/ in both speech-rate conditions – Hijazi data.

4.5.5 Tunisian Arabic

A closer look at the Tunisian plain and emphatic contrasts in Figure 4-18 below, tells us that this dialect shows no laryngeal cues to the plain and emphatic contrast, regardless of gender (although female speakers have more spread values for /t/ than male speakers at both speech rates). This differs from our previous IVAr dataset results where there were some gender differences, although the overall results showed partial overlap of VOT and indicated a likely change in progress towards losing VOT as a cue to the emphatic contrast.

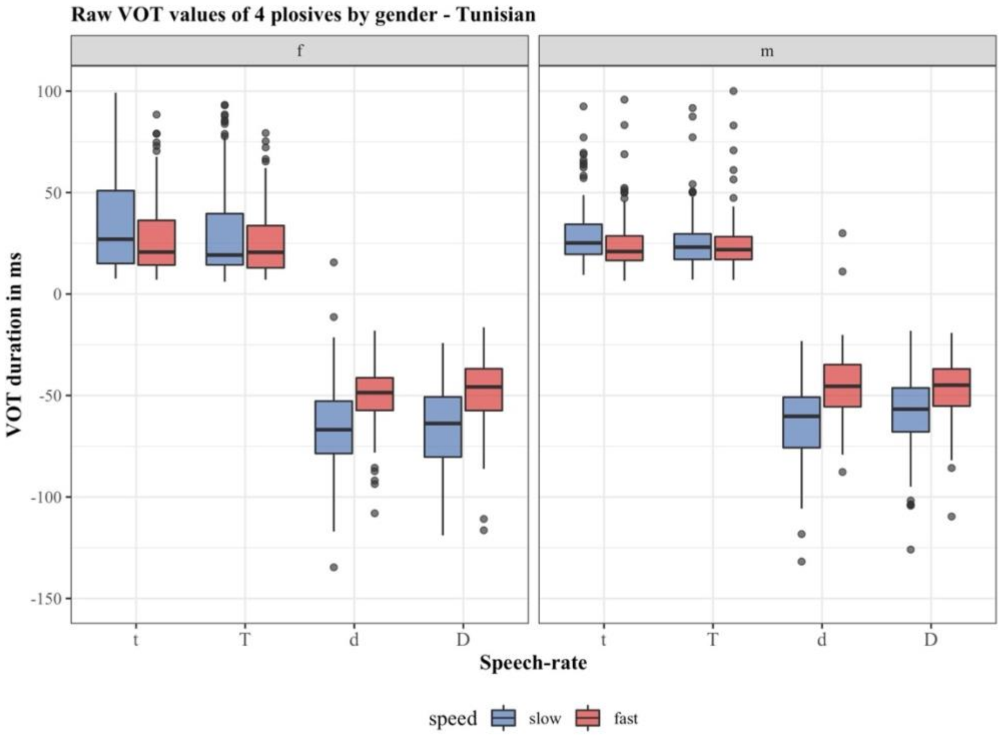


Figure 4-18 Median and interquartile range of raw VOT in Tunisian /t ~ ṭ/ and /d ~ ḍ/, by speech rate and gender.

From the raw visualizations of the plot above we can initially claim that in Tunisian there is an active feature [voice], but in voiceless /t ~ ṭ/ there appears to be no obvious effect of speech-rate on this contrast, visually.

To investigate this statistically, with the categorical factors sum-coded as well, we ran several linear mixed-effect models for the Tunisian subset. The resulting model had *VOT* as dependent variable, and *speed*, *target*, *position*, *gender*, and *following vowel* as fixed factors. The model included the interaction between *speed*target*. The model also included random intercepts for *speaker* and *item*[32].

This model showed no main effect of *gender* (β = 0.11, *t* = 0.12, *p* = .91), but, most importantly, there was a significant interaction between *speed* and *target* in all four plosives as shown in Table 4-4[33] and Figure 4-19 below. For the voiced plosives, speech rate showed an inverse relationship with VOT: as speech rate decreases voicing lead lengthens by about 5.9 ms in /d/ and 5.3 ms in /ḍ/. In voiceless plosives, slow voicing lag in /t/ is estimated to be 6.8 ms longer than in fast /t/, while in /ṭ/ it is estimated to be

[32] Model used: vot ~ speed * target + position + gender + folv + (1 | item) + (1 | speaker), data = Tunisian /t, ṭ, d, ḍ/.
[33] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

4.4 ms longer in slow speech. There is a significant two-way interaction between *speed* and *target* across the board, but the size of the estimate was smallest in the target /ṭ/, as in other dialects. The full model summary is provided in Appendix (C.9).

Table 4-4 Model results for the two-way interaction of speed*target in the Tunisian subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	6.8042	0.6516	10.442	< 001*
slow : ṭ	4.4081	0.6182	7.131	< 001*
slow : d	-5.8628	0.5790	-10.125	< 001*
slow : ḍ	-5.3495	0.6512	-8.215	< 001*

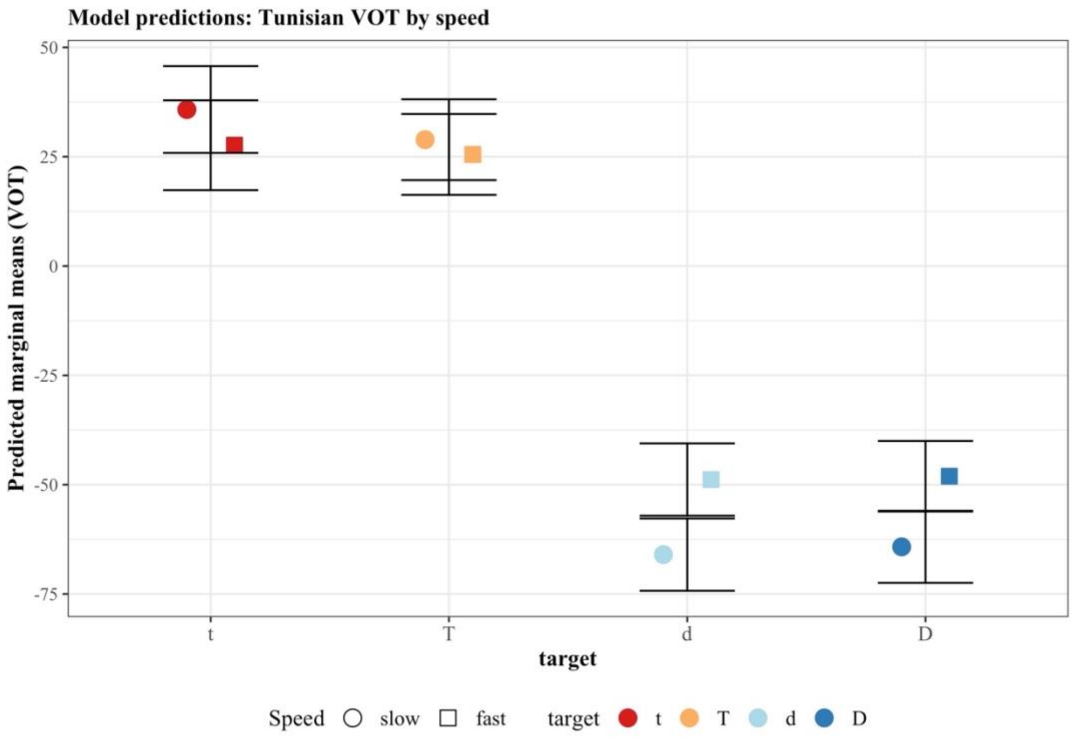


Figure 4-19 Predicted marginal mean (and 95% CI) for Tunisian VOT in /t ~ ṭ/ and /d ~ ḍ/ by speed.

Similar to Najdi and Hijazi, the effect of speech rate on VOT was significant in all four plosives, but this effect size varied. The effect of speech rate on the emphatic /ṭ/ VOT was less compared to the other set of plosives in the Tunisian dataset, and it was the most (4.4 ms) in comparison with the Najdi (2.8 ms) and Hijazi (3.7 ms) speech rate effect on emphatic /ṭ/. Figure 4-19 shows some separation of the CIs in Tunisian /ṭ/ and more overlap for /t/. Figure 4-20 below shows dispersed distribution for /ṭ/, overlapping with /t/.

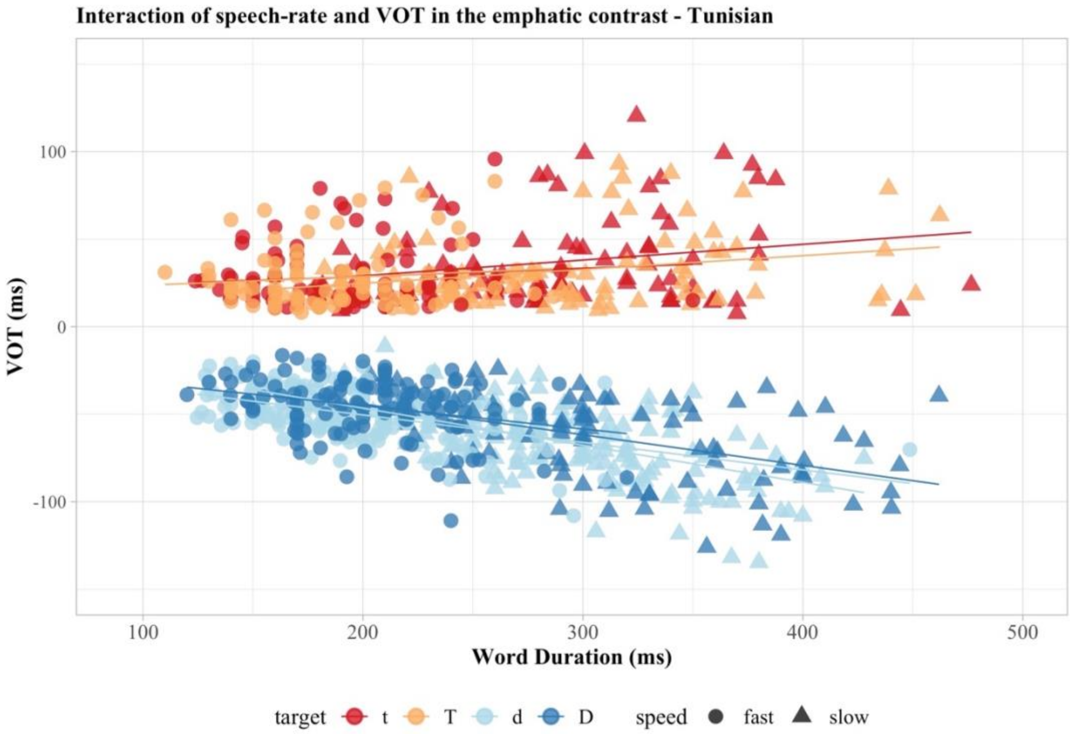


Figure 4-20 Scatter plot of word duration and VOT for /t ~ ṭ/ and /d ~ ḍ/ in both speech-rate conditions – Tunisian data.

4.5.6 Results 2: Summary

Overall, therefore, voiced vs. voiceless plosives in the Tunisian and both Saudi dialects showed no overlap. That is, voiced plosives display voicing lead with a maximum of -10 ms and voiceless plosives show clear voicing lag with the lowest value of 5 ms across all the dialects in both speaking rates.

The results also show that there is indeed an asymmetric effect of speaking rate on VOT in Arabic dialects. This effect of speaking rate is more consistent in the voiced plosives than it is in the voiceless in all the investigated dialects. The voiced emphatic contrast /d/ ~ /ḍ/ in Najdi, Hijazi, and Tunisian exhibit a clear effect of speech rate in which voicing lead in /d/ and /ḍ/ is longer in slower speech than in faster speech. This can be interpreted as confirmation that the phonological feature [voice] is active in all three dialects.

As for the voiceless plosives, the extent to which the emphatic /ṭ/ is affected by speech rate is different in the three dialects. The least affected dialect is Najdi and the most affected is Tunisian, while Hijazi was in the middle. In contrast, plain /t/ showed a significant effect of speech rate across all three dialects; voicing lag in /t/ was longer in slower speech, which again, can be interpreted as confirmation that in all three dialects there is an active phonological feature of [sg].

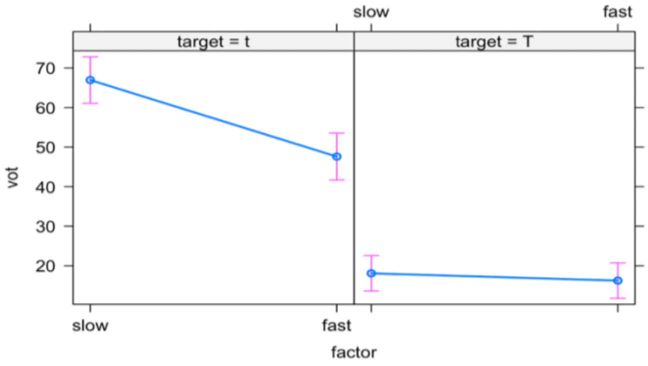


Figure 4-21 Marginal mean (and 95% CI) for Najdi VOT in /t ~ ṭ/ by speed. Effect plot from the model reported in § 4.5.3.

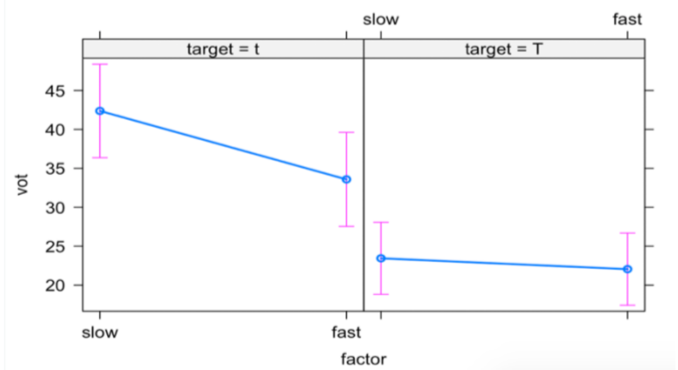


Figure 4-22 Marginal mean (and 95% CI) for Hijazi VOT in /t ~ ṭ/ by speed. Effect plot from the model reported in § 4.5.4.

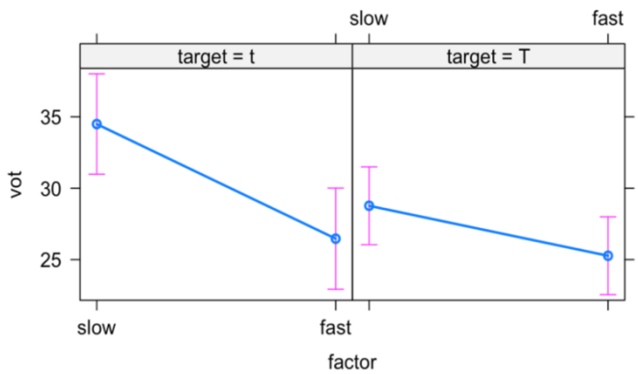


Figure 4-23 Marginal mean (and 95% CI) for Tunisian VOT in /t ~ ṭ/ by speed. Effect plot from the model reported in § 4.5.5.

Najdi voiceless results, as seen in Figure 4-21, showed little effect of speech rate on emphatic /ṭ/ but clear effect on plain /t/. Being a three-way contrast dialect (voiced, voiceless emphatic, voiceless plain), it is safe to say that Najdi dialect has two active features, [sg] and [voice], and the voiceless emphatic /ṭ/ is not specified with a laryngeal feature. Najdi dialect is behaving as we expected and is not over-specified.

For Hijazi, the interpretation depends on whether we consider /ṭ/ as affected by speaking rate or not. In Figure 4-22, the effect of speech rate on /ṭ/ is a lot less that it is in Tunisian /ṭ/ as shown in Figure 4-23 and more like Najdi emphatic /ṭ/ in Figure 4-21 above. As for the plain /t/, compared to the Tunisian data, Hijazi dialect displayed clearer effect of speech rate on VOT values of /t/. In Hijazi, both [sg] and [voice] are thus active phonological features.

Previously, we suspected that the Tunisian dialect has a two-way voicing contrast (voicing lead ~ short lag), which was predicted to entail one active feature, which would be [voice]. However, the Tunisian dialect, based on these four plosives, appears to be over-specified. At least one voiceless plosive in the voiceless emphatic contrast /t/ ~ /ṭ/ is clearly affected by speech-rate (in this case, /t/), so both features [sg] and [voice] are active in the Tunisian dialect. In order to confirm and understand this over-specification analysis, further investigation of the full set of plosives in both speaking rates is needed, which will be explored in the following section, § 4.6.

In summary, the results demonstrated that both voiced and voiceless plosives in the emphatic contrast are affected by speech rate. This effect is consistent for the voiced plain and emphatic plosives /d/ ~ /ḍ/ in all three dialects. However, plain voiceless /t/ and emphatic voiceless plosive /ṭ/ had varying degrees of speech rate effect on VOT; Tunisian dialect showed the most effect on emphatic /ṭ/ and the least effect on plain /t/. In contrast, Najdi dialect showed the least effect on emphatic /ṭ/ and the most effect on plain /t/. The Hijazi dialect showed intermediate effect on both /ṭ/ and /t/, however, shorter lag in /t/ than Najdi. Both [sg] and [voice] appear to be active phonological features in all investigated dialects.

4.6 Results 3: Speech-rate effects on the laryngeal contrast

4.6.1 VOT ~ Speech-rate (overview)

In the previous section, we found that speech rate does indeed have an asymmetric effect on VOT in the subset of plosives with a plain ~ emphatic contrast. In all three investigated dialects, faster speech affects the voiced plosives /d/ ~ /ḍ/ resulting in shorter voicing lead. Meanwhile, the voiceless plosives /t/ ~ /ṭ/ showed different effects of speech rate across the chosen dialects. These patterns may be better explained when considered in the context of the full set of plosives in each dialect (voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/).

Firstly, raw VOT values at both speech rates in all three dialects were explored as shown in Figure 4-24. As we mentioned earlier, there is an overall clear separation between the phonologically voiced and voiceless plosives in all dialects, with voicing lead in the voiced plosives and voicing lag in the voiceless ones, regardless of speaking rate. In the Tunisian dialect, VOT values for the voiceless plosives are clustered around one mean, with most data points below 50 ms. This observation lends support to the hypothesis that Tunisian is a two-way contrast dialect.

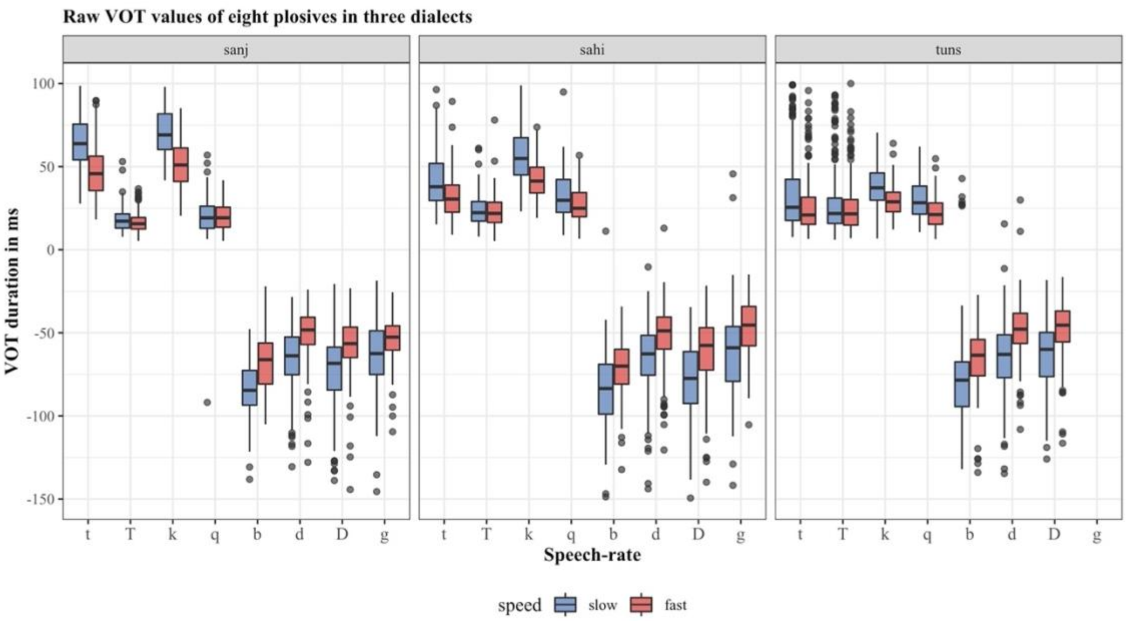


Figure 4-24 Median and interquartile range for raw VOT in Najdi, Hijazi, and Tunisian voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/, by speech rate.

In all three dialects, there is a clear impact of speech rate on the voiced subset: /b, d, ḍ, g/ have longer voicing lead in slow speech rate (blue) than in fast speech rate (red). On the other hand, speech rate has varying degrees of impact on the voiceless subset, which makes it harder to interpret. From the visualization in Figure 4-25, mirroring the previous section, speaking rate has a clear impact on /t/ in Najdi and Hijazi, while the impact is less in Tunisian. The emphatic counterpart /ṭ/ shows no clear speech rate effect in all three dialects. The uvular /q/ seems to be affected by speech rate in Tunisian and Hijazi but not in Najdi [34]. Speech rate has a clear impact on the velar /k/ across all three dialects (voicing lag is longer at slower speech rate), yet values of both fast and slow /k/ are higher in Najdi and Hijazi (aspirated) than they are in Tunisian.

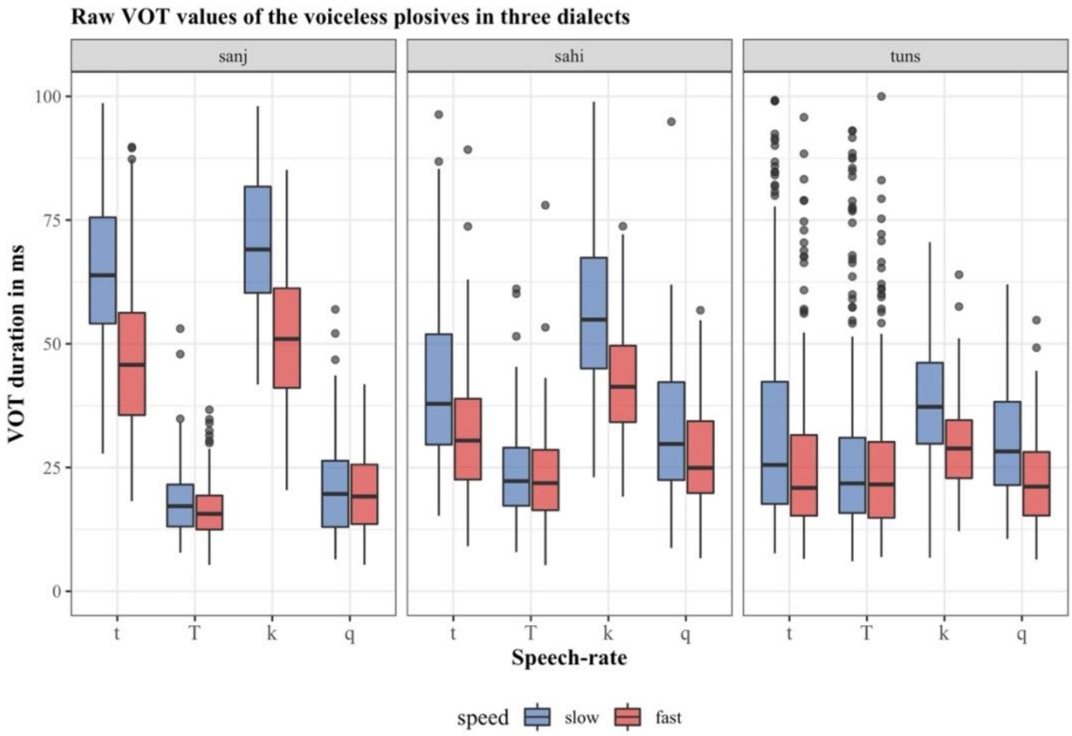


Figure 4-25 Median and interquartile range for raw VOT in Najdi, Hijazi, and Tunisian voiceless /t, ṭ, k, q/, by speech rate.

To summarize, /t/ and /k/ are visually affected by speaking rate in all three dialects. Both Tunisian and Hijazi /q/ are slightly affected by speech rate, but not Najdi /q/. The emphatic /ṭ/, like the Najdi /q/, shows no impact of speech rate on all three dialects. The overall picture is complex and required further statistical investigation to confirm these observations.

In the following sections, we explore in more detail the effect of speaking rate on VOT across the laryngeal contrast (voiced ~ voiceless) in each dialect separately.

4.6.2 Percentage of voicing

For completeness, we calculated the percentage of voicing, during the closure duration, in all voiced plosives to check for variation at both speech rates. All voiced plosives in

[34] This could be because /q/ is a native phoneme in Tunisian, while in Najdi and Hijazi /q/ is marginal (mainly for MSA loanwords) and /g/ is the main cognate of MSA /q/ in both Najdi and Hijazi.

all three dialects at both speaking rates showed voicing during the closure. Figure 4-26 is a visualization of proportion of tokens produced with different degrees of voicing during the closure, by speech rate and by dialect. Using the *dplyr* package (Hadley Wickham, François, Henry, & Müller, 2019), we grouped, combined and summarized the data separately by number of observations and mutated a new column to report the proportions. The figure shows that among all voiced plosives in all three dialects, 76.5% and above of the observations for each plosive show 100% voicing during the closure. That is, the blue part of the bars in the figure represents the proportion of fully voiced closure duration (100% voicing) tokens, for each plosive. We can see that neither speech rate nor target influences the percentage of voicing during the closure in voiced plosives. We therefore conclude that there is no evidence in this dataset that % voicing during closure plays a role as an exponent of the active phonological feature at different speaking rates, and do not discuss this potential cue further.

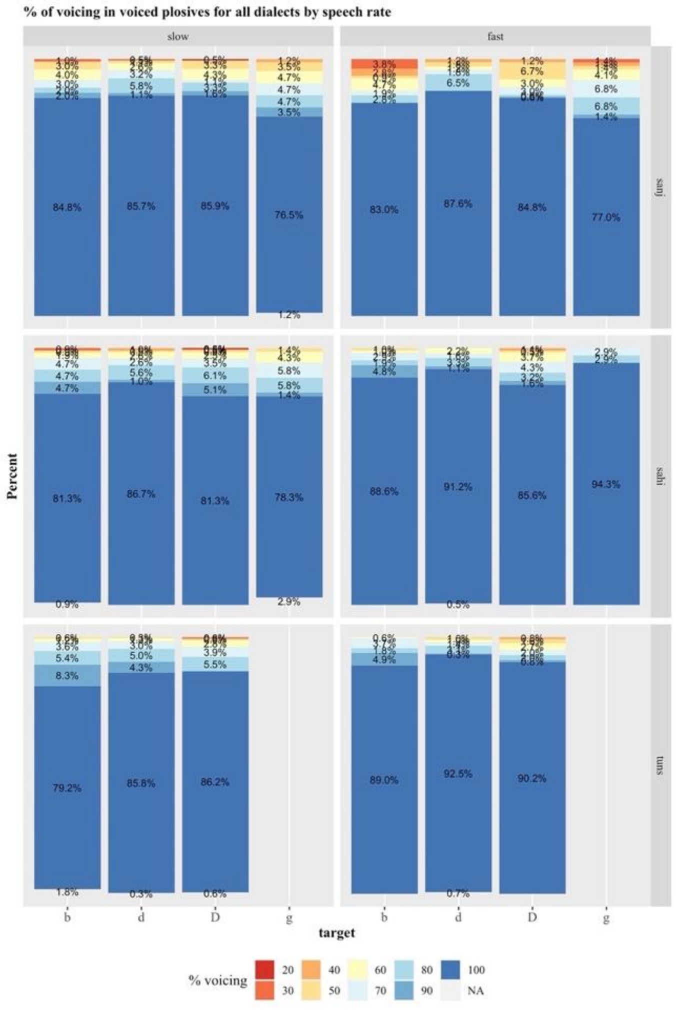


Figure 4-26 Proportion of tokens produced with different degrees of voicing during the closure in voiced plosives, by dialect and speech rate.

4.6.3 Najdi Arabic laryngeal contrast

4.6.3.1 Najdi VOT

Figure 4-27 visualizes of the Najdi raw VOT data, and the effect of speech rate on Najdi plosives is very neat and as expected. It is clear that the influence of speaking rate affects all the voiced plosives [voice], yielding longer voicing lead in slower speech, while in the voiceless plosives speech rate influences only the long lag /t/ and /k/ [sg], but not the short lag /q/ and emphatic /ṭ/ [Ø]. The spread of VOT values for /ṭ/ and /q/ in both speech rate conditions is visually overlapping.

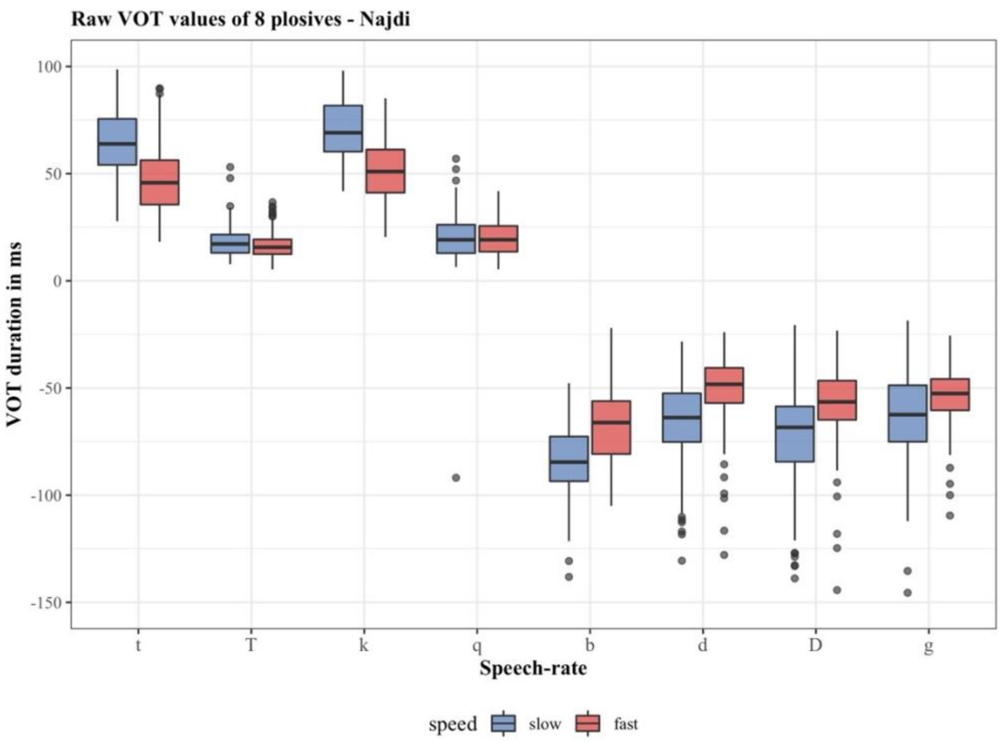


Figure 4-27 Median and interquartile range for raw VOT in Najdi voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/, by speech rate.

To explore the data statistically, the Najdi data was sum-coded as previously mentioned then a linear mixed-effect model was fitted that considered *VOT* as the dependent variable and *speed*, *target*, *position*, and *gender* as fixed factors. The interaction between *speed*target* was included, as well as random intercepts for *speaker* and *item*[35].

[35] Model used: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker), data = Najdi

The model showed no main effect of *gender* (β = -1.39, *t* = -1.38, *p* = .19). In Table 4-5[36] below, we can see that the interaction between *speed* and *target* was significant in all the target plosives except for voiceless uvular /q/ (*sanj slow* x */q/:* β = 0.95, *t* = 0.77, *p* = .45). The interaction in the voiced plosives was significant across all targets but was the least in the voiced velar /g/ (*sanj slow* x */g/:* β = -3.36, *t* = -2.83, *p* = .0047). The voiceless emphatic /ṭ/ shows the smallest effect (after /q/) among the voiceless plosives (*sanj slow* x */ṭ/:* β = 2.31, *t* = 2.64, *p* = .0083). Figure 4-28 shows the predicted marginal mean and 95% confidence interval from the model. The full model summary is provided in Appendix (C.10).

Table 4-5 Model results for the two-way interaction of speed*target in the Najdi subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	10.5200	0.9269	11.350	< .001*
slow : ṭ	2.3074	0.8738	2.641	0.008*
slow : k	12.1787	1.0359	11.756	< .001*
slow : q	0.9483	1.2536	0.756	0.45
slow : b	-7.7663	1.0586	-7.336	< .001*
slow : d	-6.9650	0.8417	-8.274	< .001*
slow : ḍ	-7.8608	0.8496	-9.252	< .001*
slow : g	-3.3622	1.1883	-2.830	0.005*

[36] To estimate the held-out factor /g/, we rotated the levels of the model.

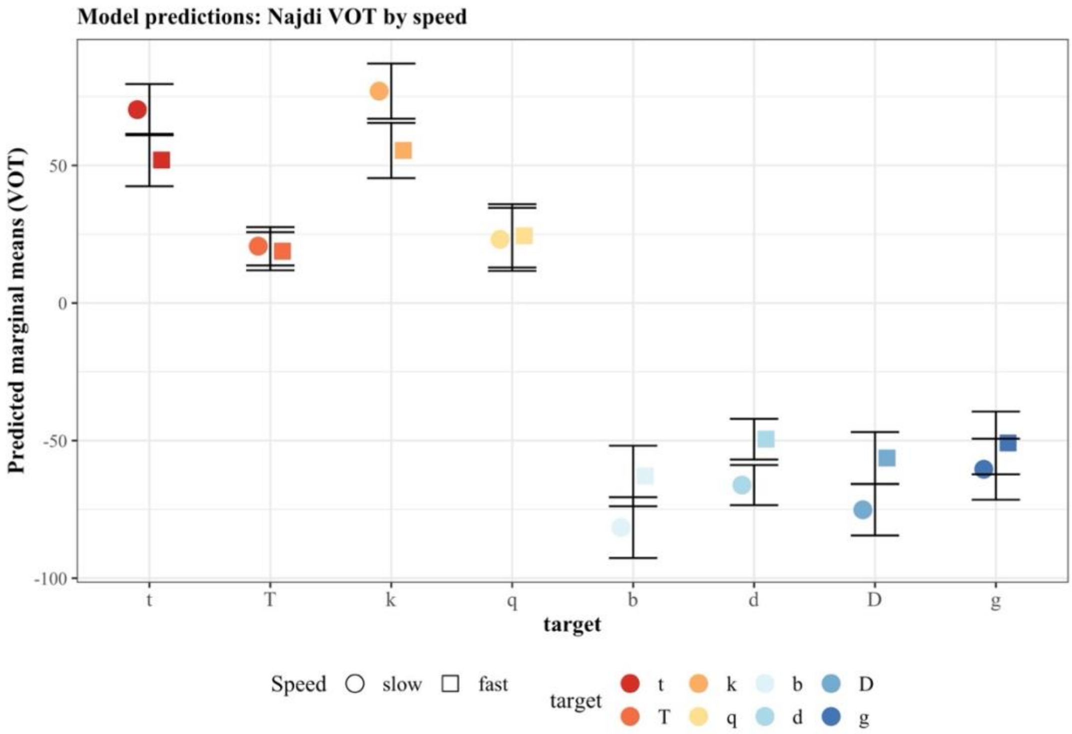


Figure 4-28 Predicted marginal mean (and 95% CI) for Najdi VOT in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ by speech rate.

To summarize, the Najdi data aligned with our predictions, in that speech rate had a significant effect on voiced plosives /b, d, ḍ, g/ with longer voicing lead in slower speech indicating an active feature [voice]. Speech rate also had a significant effect on the long lag voiceless plosives in Najdi: as speech rate decreases, voicing lag in /t/ and /k/ increase, indicating an active feature [sg]. VOT values in the voiceless emphatic /ṭ/ and the uvular /q/ are overlapping in both speech rate conditions ‘fast’ and ‘slow’, indicating no active feature is present.

4.6.3.2 Najdi Closure duration

According to Al-Tamimi and Khattab (2018), closure duration, in their data, proved to be the most important cue to distinguish voicing in the context of gemination. In the current study, we investigated closure duration to see to what extent speaking rate has an effect on this cue, in all the plosives, focusing on the voicing contrast. In Figure 4-29 below, we can see that there is no clear visual difference in the closure duration of the two voicing series (voiced ~ voiceless). However, the influence of speaking rate on individual plosives varies. The emphatic /ṭ/ seems to show the most influence of speech rate on the closure duration, followed by the uvular /q/, in which closure duration in slower speech of producing these plosives results in higher values in milliseconds.

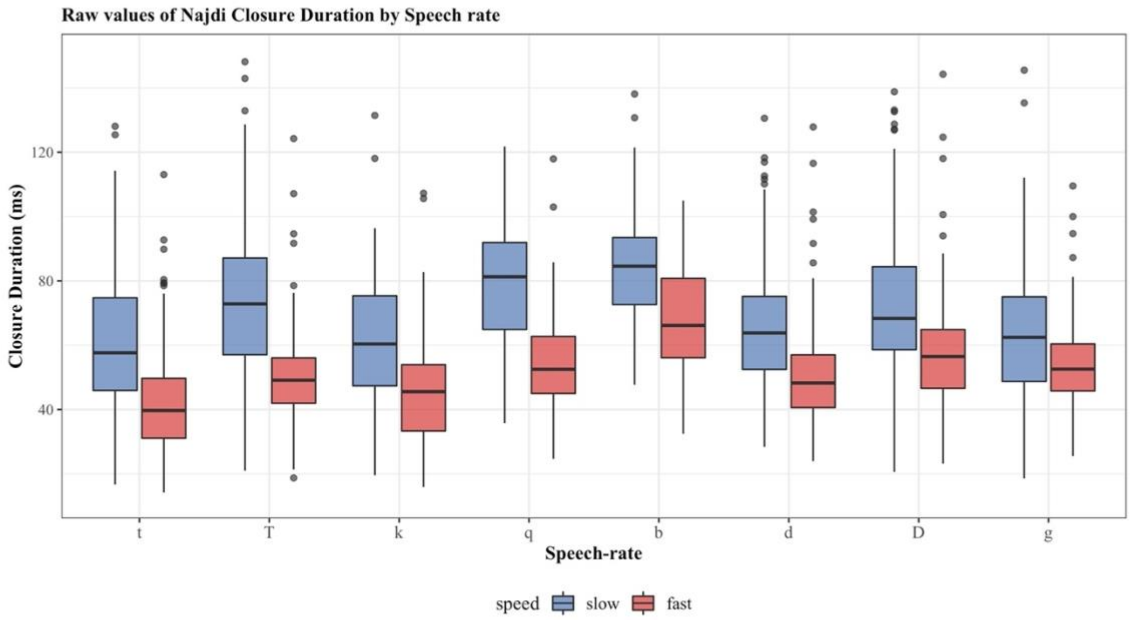


Figure 4-29 Median and interquartile range for Closure duration values of 8 plosives both voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/, by speech rate - Najdi dialect.

We also investigated this variable statistically using linear mixed-effect models. The categorical variables in the Najdi subset were sum-coded, and the best fit model had *Closure Duration* as the dependent variable and the fixed factors were *speed*, *target*, *position*, and *gender*. The model considered the interaction between *speed*target*, and random intercepts for *speaker* and *item*[37].

The model resulted in a significant main effect of *speed* (β = 9.23, *t* = 21.53, *p* < .001): closure duration at slower speaking rate showed higher values than the average. The model also revealed a significant effect of *position* (β = 4.38, *t* = 3.95, *p* < .001) in which closure duration in initial position shows higher values, as expected. There was no significant effect of *gender* (β = 1.47, *t* = 1.01, *p* = .329). In Figure 4-30, we can see the predicted marginal mean and 95% confidence interval by *speed* and *target*: for two plosives there was a significant effect of speaking rate on the closure duration, /ṭ/ (β = 2.70, *t* = 2.76, *p* = .006) and /q/ (β = 2.86, *t* = 2.04, *p* = .041), which show the most difference between fast and slow speech. In contrast, /g/ (β = -4.23, *t* = -3.18, *p* < .001)

[37] Model used: closdur ~ target * speed + gender + position + (1 | item) + (1 | speaker), data = Najdi

showed the least difference. The summary of model results is provided in Appendix (C.11).

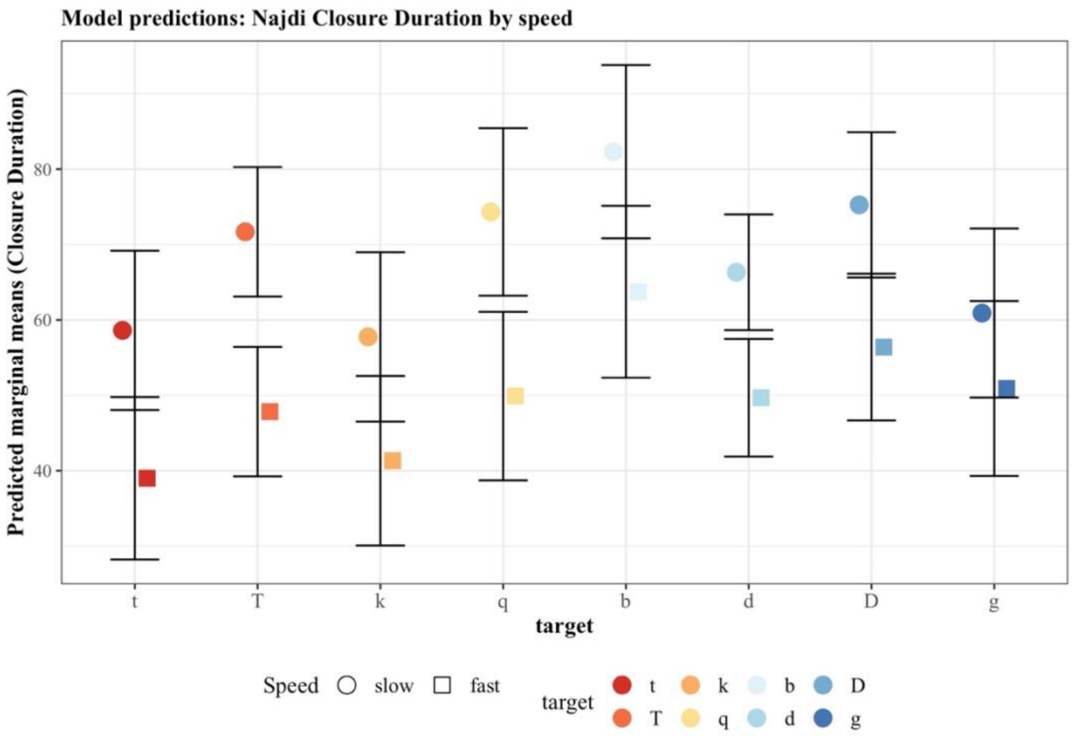


Figure 4-30 Predicted marginal mean (and 95% CI) for Najdi Closure Duration in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ by speech rate.

So, to sum up, both *position* and *speed* had a significant main effect on the closure duration of Najdi plosives, but not *gender*. There was a significant interaction of speed and target for two plosives, the emphatic /ṭ/ and the uvular /q/, which showed the greatest effect of speaking rate. Whereas it was only /g/ that showed the least effect of speaking rate on the closure duration of this plosive. Among the voiceless plosives, it was the plosives that are not assigned an active feature, /ṭ/ and /q/, that were mostly affected by speaking rate.

4.6.3.3 Najdi Fundamental frequency

Another known perceptual correlate of voicing contrast in initial position is f0 onset (Dmitrieva, Llanos, Shultz, & Francis, 2015). For both *aspirating* and *voice languages*, f0 onset is reported to be lower in voiced contexts and higher in voiceless contexts. We examined f0 at the 25% point of the vowel in Najdi dialect.

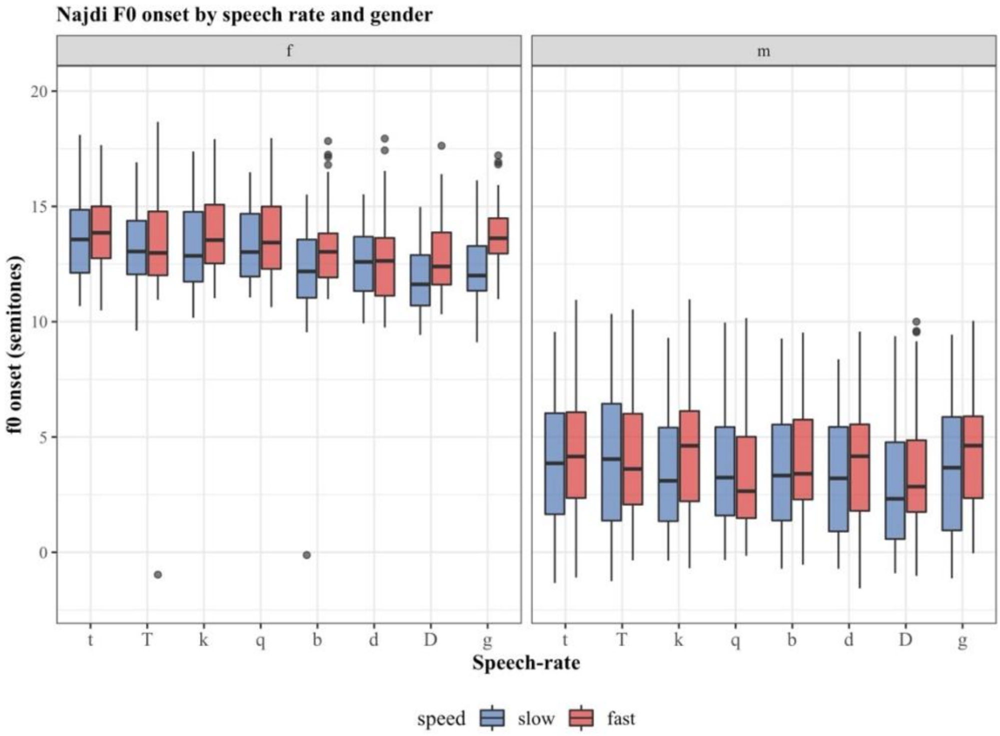


Figure 4-31 Median and interquartile range of f0 at 25% point in vowels following Najdi plosives, by gender and speech rate.

In Figure 4-31 above, there is a clear effect of gender as expected: female speakers have higher f0 values than male speakers. The influence of voicing and speaking rate is not clear in this raw visualization, so we needed to explore the data statistically. We ran a number of linear mixed-effect models on the Najdi subset of the data, the best fit model had *f0.25* as the dependent variable and the fixed factors were *speed*, *target*, *position*, *gender* and *following vowel*. The model included the interaction between *speed*target*, and random intercepts for *speaker* and *item*[38].

Categorical factors were sum-coded in the following manner: *speed* (slow = 1, fast = -1), *gender* (female = 1, male = -1), *target* (seven predictors of /t, ṭ, k, q, b, d, ḍ/ ~ /g/ in which the first listed segment /t/ was set to 1, /g/ to -1, and all unlisted segments to 0), and *following vowel* (five predictors of /a, a:, i, i:, u / ~ /u:/ in which the first listed vowel /a/ was set to 1, /u:/ to -1, and all unlisted vowels to 0).

There was a significant main effect of *speed* ($\beta = -0.24$, $t = -8.71$, $p < .001$) whereby f0 onset in slower speech is lower than the average. *Gender* also showed a significant main

[38] Model used: f025 ~ speed * target + gender + following vowel + (1 | item) + (1 | speaker), data = Najdi

effect (β = 4.63, t = 8.32, p < .001) where, as expected, female speakers have higher f0 values than male speakers on average.

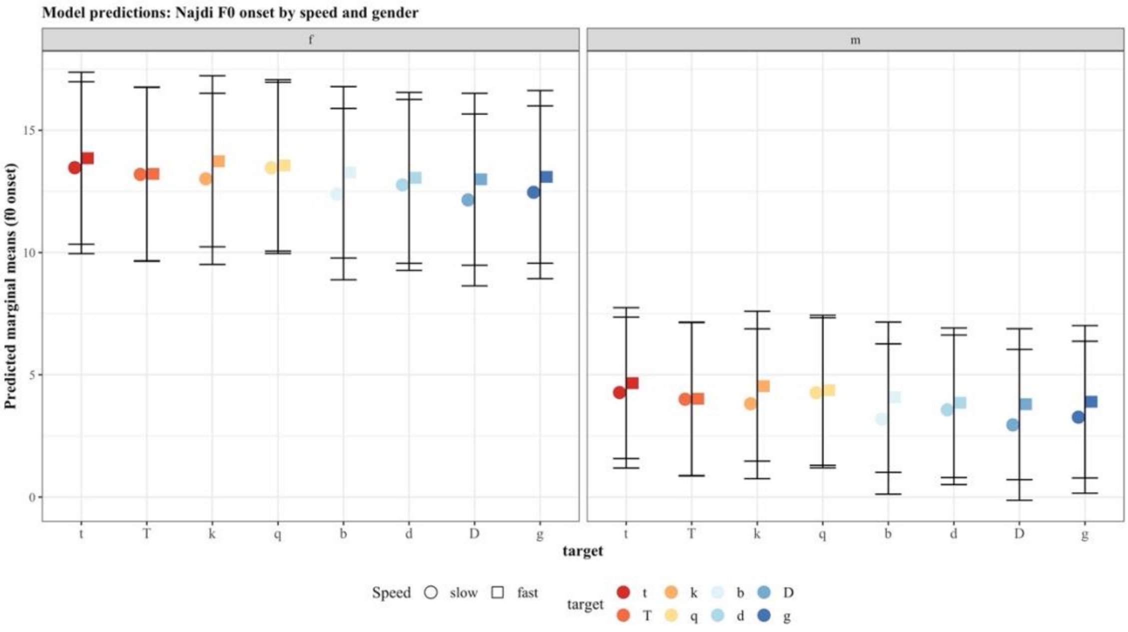


Figure 4-32 Predicted marginal mean (and 95% CI) for Najdi f0 at 25% point in vowels following voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ plosives by speech rate and gender.

The model also revealed a two-way interaction between *speed* and *target* for four of the plosives only. From Figure 4-32 above, we can see the effect of speaking rate as we mentioned earlier such that slower speech rate resulted in lower f0 values, but the degree of lowering varied among the plosives: /b/ (β = -0.20, t = -2.76, p = .006) and /ḍ/ (β = -0.18, t = -2.32, p = .020) showed lower f0 values in slower speech than in faster speech in both genders; in contrast, the emphatic /ṭ/ (β = 0.23, t = 350, p = .0005) and uvular /q/ (β = 0.19, t = 2.27, p = .023) varied from the average in showing only small effect of speech rate and f0 lowering difference. From the figure we can also see that voiceless plosives generally have higher values of f0 than voiced plosives, except for the voiceless emphatic /ṭ/. The full model summary is provided in Appendix (C.12).

In summary, f0 at 25% point of the vowel has lower values in slower speech than in faster speech, and for male speakers than female speakers. Only two plosives /b/ and /ḍ/ showed significantly above average effect of speech rate on f0 (lowering). The two plosives /ṭ/ and /q/ showed a significantly below average effect of speech rate on f0. Among the voiceless plosives, it was only the plosives that are not assigned with an active feature, /ṭ/ and /q/, that are not affected by speaking rate.

4.6.4 Hijazi Arabic laryngeal contrast

4.6.4.1 Hijazi VOT

Turning to the Hijazi full set of plosives, we can see from the visualization of the raw data below in Figure 4-33 that speech rate seems to be clearly affecting the voiced set of plosives, as we saw earlier in the Najdi data. Likewise, the Hijazi data display differences in the influence of speech rate on the voiceless subset. Both /t/ and /k/ show a clear impact of speaking rate with higher values of voicing lead in slower rate. The emphatic /ṭ/, however, displays little or no impact of speaking rate. Uvular /q/ also seems to be influenced, with longer voicing lag at slow speech, yet to a lesser degree than in /t/ and /k/. Based on this visualization below, we can initially claim that Hijazi data has both [sg] and [voice] as active features and that /ṭ/ is unspecified with a laryngeal feature as it is clearly not impacted by speech rate, though this needs to be explored statistically.

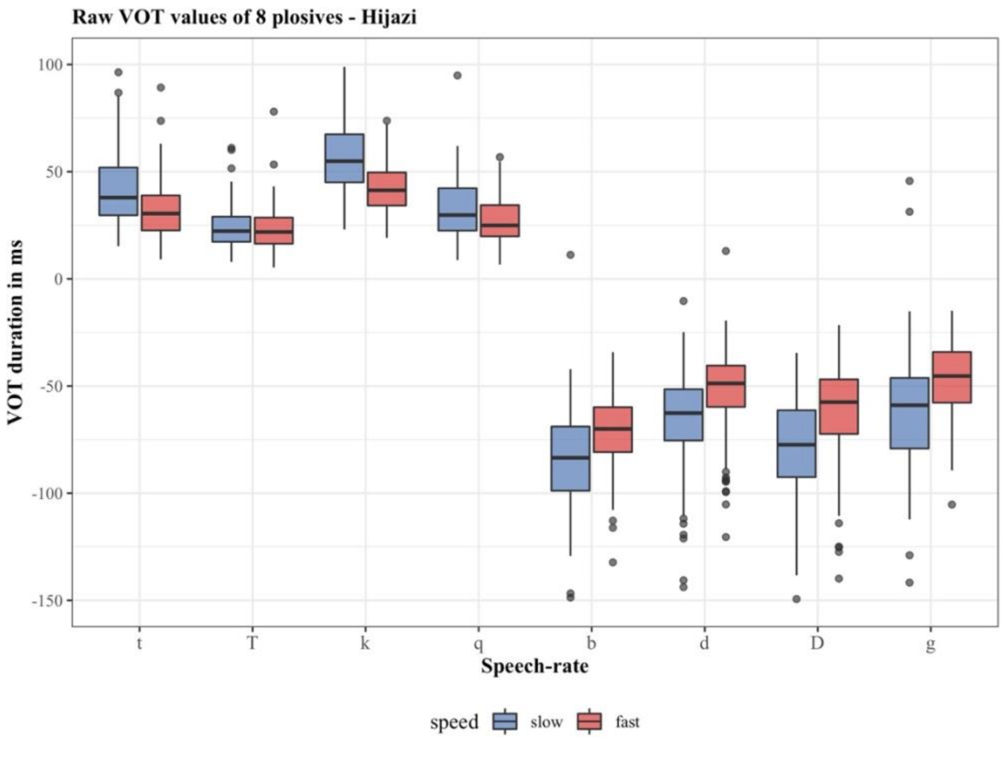


Figure 4-33 Median and interquartile range for raw VOT in Hijazi voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/, by speech rate.

As in all the previous subsets, the Hijazi data was sum-coded then a linear mixed-effects model was fitted that mirrored the previous model of the Najdi data, with *VOT* as the dependent variable and *speed*, *target*, *position*, and *gender* as fixed factors. In this model, we included the interaction between *speed*target*, and we also included random intercepts for *speaker* and *item*[39].

Table 4-6 Model results for the two-way interaction of speed*target in the Hijazi subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	6.1496	0.9742	6.313	< .001*
slow : ṭ	2.4702	0.9250	2.670	< .001*
slow : k	9.4159	1.1037	8.531	< .001*
slow : q	4.8224	1.3972	3.452	< .001*
slow : b	-5.1711	1.0864	-4.760	< .001*
slow : d	-5.2390	0.8591	-6.098	< .001*
slow : ḍ	-8.0128	0.8521	-9.404	< .001*
slow : g	-4.4352	1.3110	-3.383	< .001*

The model showed no main effect of *gender* (β = -1.18, t = 0.80, p = .437), like in Najdi. There was a significant interaction between *speed* and *target* for all the plosives as summarized in Table 4-6[40]. This pattern is clear in the visualization of the model predictions in Figure 4-34, which shows the predicted marginal mean and 95% confidence interval from the model. The speech rate effect on the voiced plosives is negative, in which voicing lead in slower speech is longer than the average by around 5.2 ms in /b/ and /d/, 8 ms in /ḍ/, while /g/ was the least with 4.4 ms less than the average. In the voiceless plosives, the amount of effect on the plosives varied. Again, the emphatic /ṭ/ showed the least effect. In the slow condition voicing lag in /ṭ/ is estimated to be only 2.5 ms longer than the average, and the two speech rates are greatly overlapping. Hijazi /k/ showed the greatest effect of speech rate among the voiceless subset: in slower speech, voicing lag /k/ is 9.4 ms longer than the average. In contrast, /q/ showed a smaller effect of speech rate with voicing lag of 4.8 ms longer in slow speech than the average. The full summary of the model results is provided in Appendix (C.13).

[39] Model used: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker), data = Hijazi
[40] To estimate the held-out factor /g/, we rotated the levels of the model.

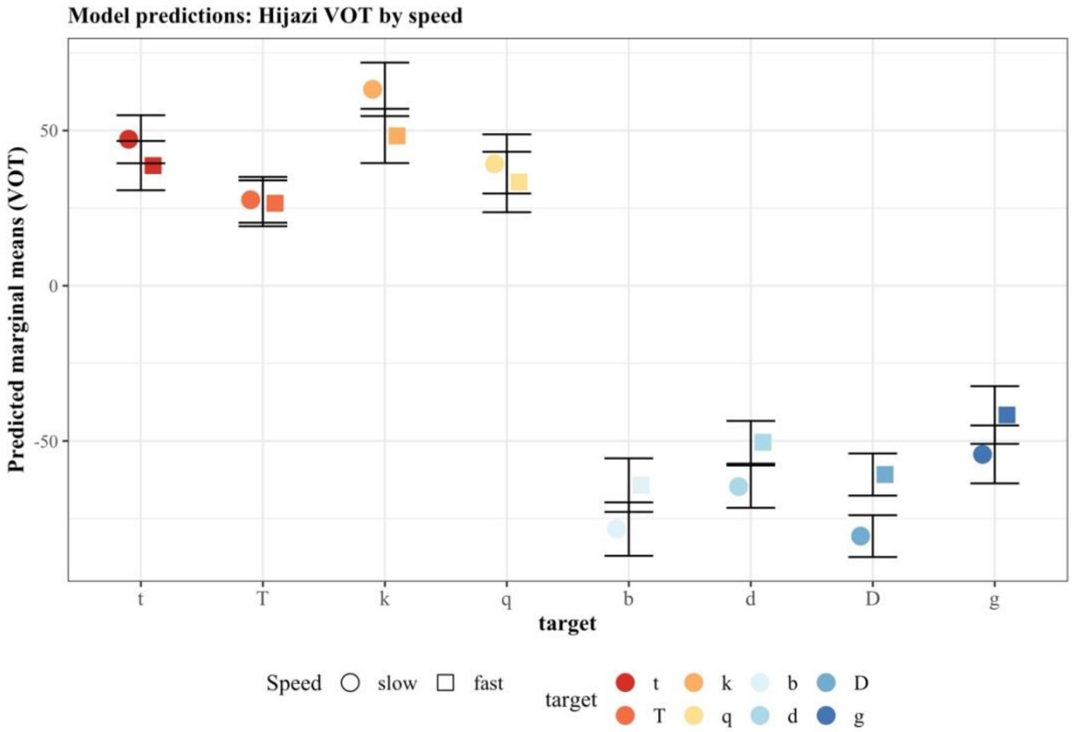


Figure 4-34 Predicted marginal mean (and 95% CI) for Hijazi VOT in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ by speech rate.

In summary, speaking rate had a significant effect on VOT values in all Hijazi plosives. This effect varied, however, especially in the voiceless plosives. Voiced plosives in Hijazi manifested a clear effect of speaking rate on their VOT indicating an active feature [voice]. In the voiceless plosives, emphatic /ṭ/ in Hijazi showed overlapping values of VOT in the two speech rate conditions indicating absence of an active feature; the voiceless emphatic showed the least effect of speech rate. The uvular /q/ followed the emphatic /ṭ/ in the degree of rate effect, in that voicing lag values in fast and slow /q/ in Hijazi are overlapping, but to a lesser degree than the emphatic /ṭ/. Hijazi /t/ and /k/ were significantly affected by speaking rate, with longer voicing lag in slower speech indicating an active feature [sg]. The velar /k/ in Hijazi was most affected by speaking rate.

4.6.4.2 Hijazi Closure duration

We also investigated the closure duration of plosives in the Hijazi subset of the data, Figure 4-35 below for raw closure durations shows a clear effect of speech rate on the closure duration of all plosives, though there was no clear difference visually between the voiced and the voiceless. Similar to the Najdi subset, the Hijazi emphatic /ṭ/ and uvular /q/ seem to show the most influence of speaking rate on closure duration. Unlike Najdi, in Hijazi the closure duration values slightly overlap in the two speaking rates.

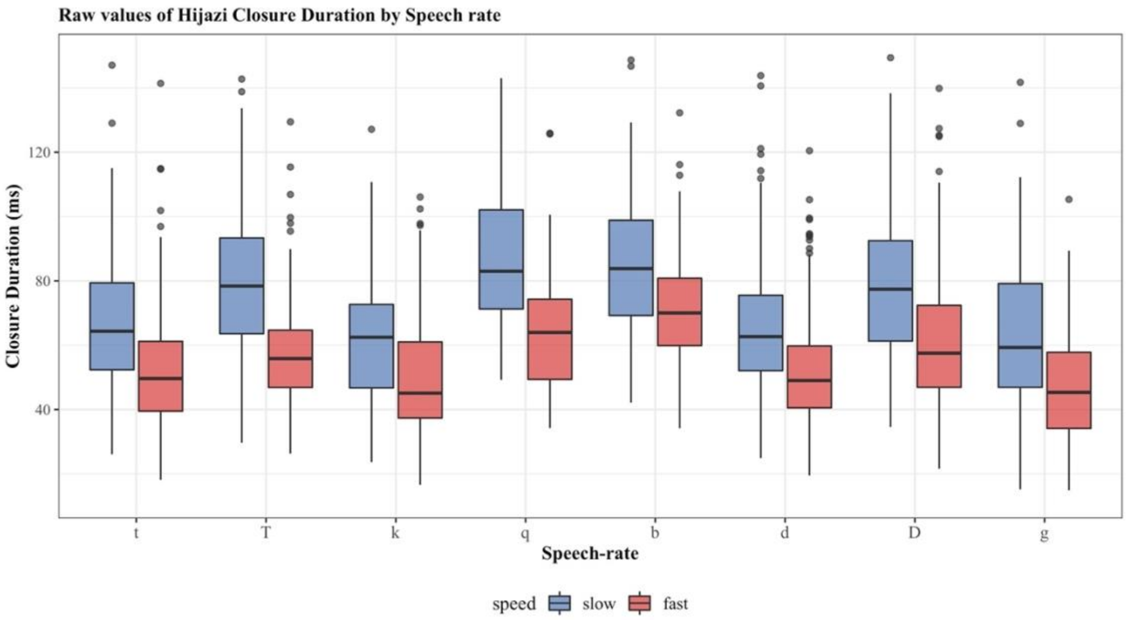


Figure 4-35 Median and interquartile range for Closure duration values of 8 plosives both voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/, by speech rate - Hijazi dialect.

The Hijazi subset was explored statistically using the same linear mixed-effect model of the Najdi subset with sum-coded categorical variables. The best fit model included *Closure Duration* as the dependent variable and were *speed*, *target*, *position*, and *gender* as the fixed factors. The model considered the interaction between *speed*target*, and random intercepts for *speaker* and *item*[41].

The model shows a significant main effect of *speed* (β = 8.61, t = 18.52, p < .001) with higher values of closure duration at slower speaking rates. *Position* also had a significant effect on closure duration (β = 3.67, t = 3.35, p < .001), with higher closure duration in initial position. There was no significant effect of gender in the Hijazi subset (β = 2.43, t = 1.35, p = .19), as was the case in Najdi. The interaction between *speed* and *target* was significant in only one plosive, the emphatic /ṭ/ (β = 2.74, t = 2.60, p = .009), in which the closure duration difference between fast and slow was the greatest. The full model results summary is provided in Appendix (C.14).

[41] Model used: closdur ~ target * speed + gender + position + (1 | item) + (1 | speaker), data = Hijazi

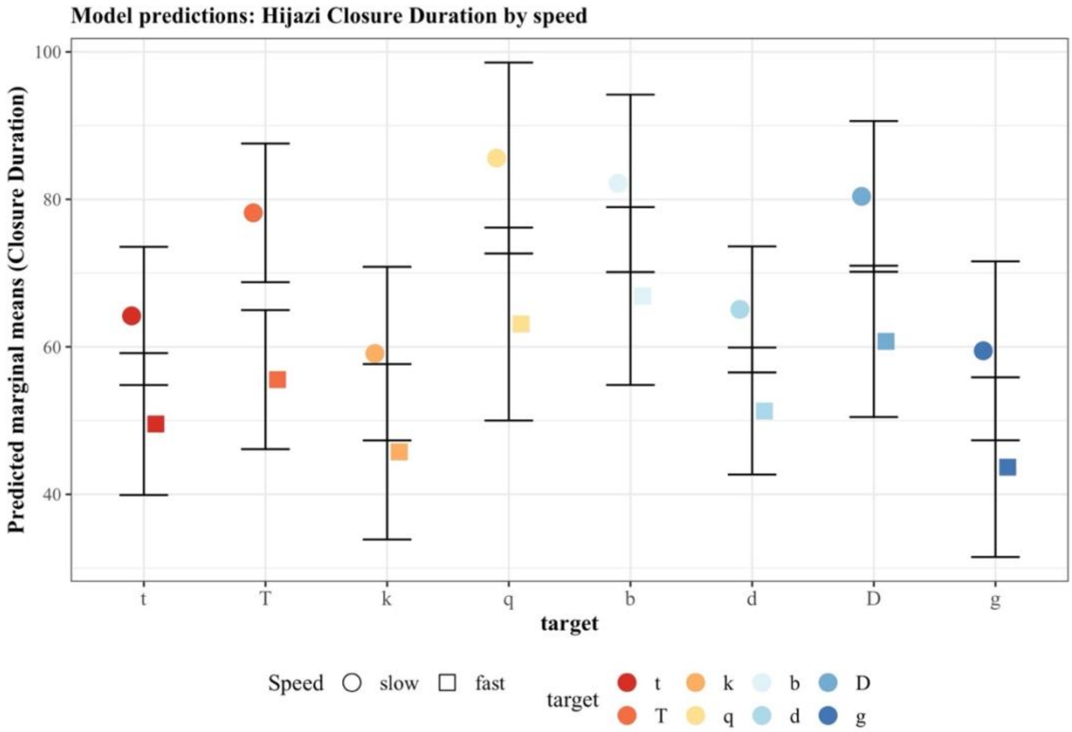


Figure 4-36 Predicted marginal mean (and 95% CI) for Hijazi Closure Duration in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ by speech rate.

We can see this effect clearly in Figure 4-36 above, where the emphatic /ṭ/ followed by the uvular /q/ display the most effect of speaking rate on the closure duration. The effect on /q/ ($\beta = 2.77$, $t = 1.74$, $p = .082$) was great (higher values) but not significant.

To summarize, closure duration in the Hijazi subset of the data was significantly affected by both *position* and *speed* but not *gender*. The emphatic /ṭ/ showed the most influence of speaking rate on its closure duration with the greatest difference between the two speaking rates. The uvular /q/ also had a similar effect of speech rate on the closure duration, though it was not significant.

4.6.4.3 Hijazi Fundamental frequency

Fundamental frequency at the 25% point of the vowel was also explored in the Hijazi subset of the data. Visualization of the raw data did not show clear influence of speaking rate nor voicing, the only clear influence was of gender. We explored the data statistically in several linear mixed-effect models after sum-coding the categorical variables, and the best fit model mirrored the Najdi model with *f0.25* as the dependent variable and *speed*, *target*, *position*, *gender* and *following vowel* as the fixed factors.

The model included the interaction between *speed*target,* and random intercepts for *speaker* and *item*[42].

The model showed a significant main effect of *speed* (β = -1.30, t = -3.76, p < .001) where f0 onset had higher values in faster speaking rate. As expected, *gender* (β = 4.31, t = 11.54, p < .001) showed a significant main effect with female speakers producing higher values of f0 than male speakers.

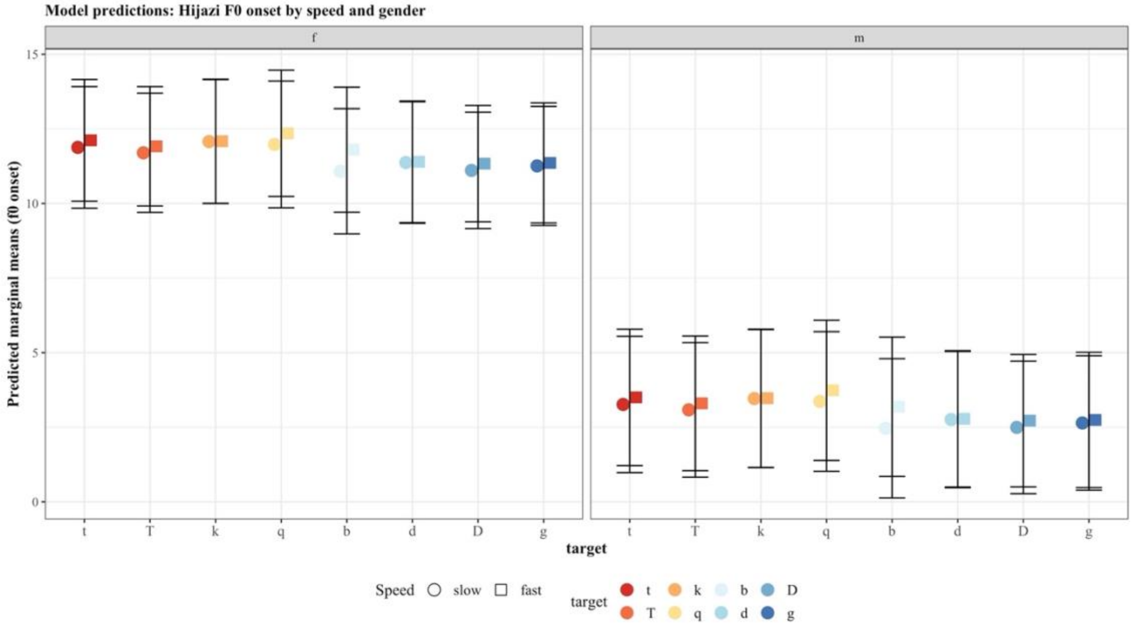


Figure 4-37 Predicted marginal mean (and 95% CI) for Hijazi f0 at 25% point in vowels following voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/ plosives by speech rate and gender.

There was a significant interaction between *speed* and *target* in only one plosive /b/ (β = -2.22, t = -2.56, p = .011) in which, in both genders, the f0 raising effect in vowels following /b/ in faster speech was larger than average. Figure 4-37 above also shows that, generally, voiceless plosives have higher values of f0 than voiced plosives. The full summary of the model results is provided in Appendix (C.15).

To summarize, voiced plosives, plosives at slower speaking rate, and plosives produced by male speakers had lower f0 values than their counterparts.

[42] Model used: f025 ~ speed * target + gender + following vowel + (1 | item) + (1 | speaker), data = Hijazi

4.6.5 Tunisian Arabic laryngeal contrast

4.6.5.1 Tunisian VOT

In this section, we explore, finally, the effect of speech rate on VOT in the full set of plosives in the Tunisian variety of Arabic[43]. The boxplots in Figure 4-38 below display VOT values of voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/ in the Tunisian dialect at two speech rates. Overall, the picture is more complex in this dialect: speech rate displays some influence on the voiced subset, meaning /b, d, ḍ/ at the slow speech rate have longer voicing lead than /b, d, ḍ/ in the speeded speech rate. However, in the voiceless subset, there seems to be varying influence of speaking rate across the range of plosives; there is some kind of impact on /k/ and /q/ (longer voicing lag at slow rates as well) but this impact does not appear to be as strong as the effect on the voiced subset. The impact on /t/ is small, with partial overlap across rates, whereas /ṭ/ displays little or no impact of speech rate, and the speech rate overlap is clearer.

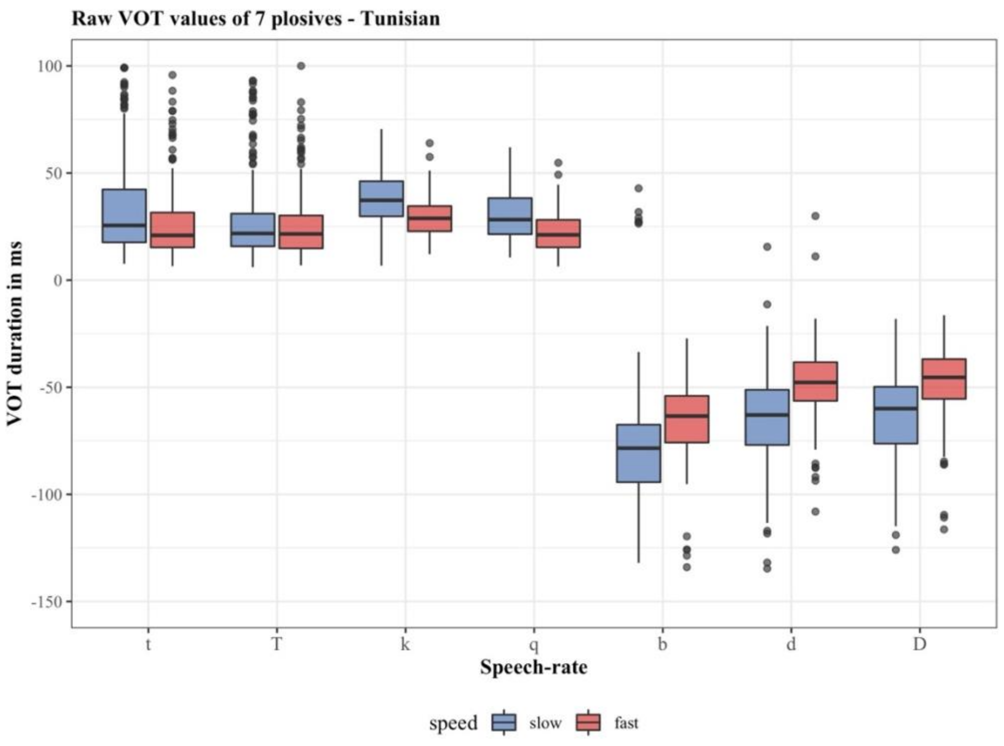


Figure 4-38 Median and interquartile range for raw VOT in Tunisian voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/, by speech rate.

[43] Which are seven plosives; voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/. Urban Tunisian does not include /g/ in its consonant inventory.

From the visualizations of the raw plots above we can initially state that, in Tunisian, there is an active feature [voice], but in the voiceless subset the picture is more complicated and needs to be explored statistically.

The Tunisian subset of the data was sum-coded then explored in a linear mixed-effects model where *VOT* was the dependent variable and *speed*, *target*, *gender*, and *position*, were fixed factors. The model counted for the interaction between *speed*target*. Random intercepts for *speaker* and *item* were also included in the model[44].

Table 4-7 Model results for the two-way interaction of speed*target in the Tunisian subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	5.1042	0.7525	7.039	< .001*
slow : ṭ	2.8405	0.6875	4.131	< .001*
slow : k	5.8605	0.8041	7.288	< .001*
slow : q	5.0937	0.8157	6.245	< .001*
slow : b	-4.7487	0.8064	-5.889	< .001*
slow : d	-7.4082	0.6342	-11.681	< .001*
slow : ḍ	-6.7421	0.7313	-9.219	< .001*

The model showed no main effect of *gender* ($\beta = -0.69$, $t = -0.78$, $p = .445$), like in Najdi and Hijazi. This model showed a significant interaction between *speed* and *target* in all seven plosives as shown in Table 4-7[45] above and Figure 4-39 below. In the voiced plosives, there was an inverse relationship between speaking rate and VOT in which slower speech rate results in longer voicing lead than the average by about 4.7 ms in /b/, 7.4 ms in /d/ and 6.7 ms in /ḍ/. Among the voiced plosives, /b/ shows the least difference between the two speaking rates, although speech rate still has significant effect on /b/. As for the voiceless subset, the effect of speech rate varied across the different voiceless targets. The effect was lowest for the emphatic /ṭ/ with voicing lag in slow /ṭ/ estimated to be only 2.8 ms longer than the average. The full summary of the model results is provided in Appendix (C.16).

[44] Model used: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker), data = Tunisian
[45] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

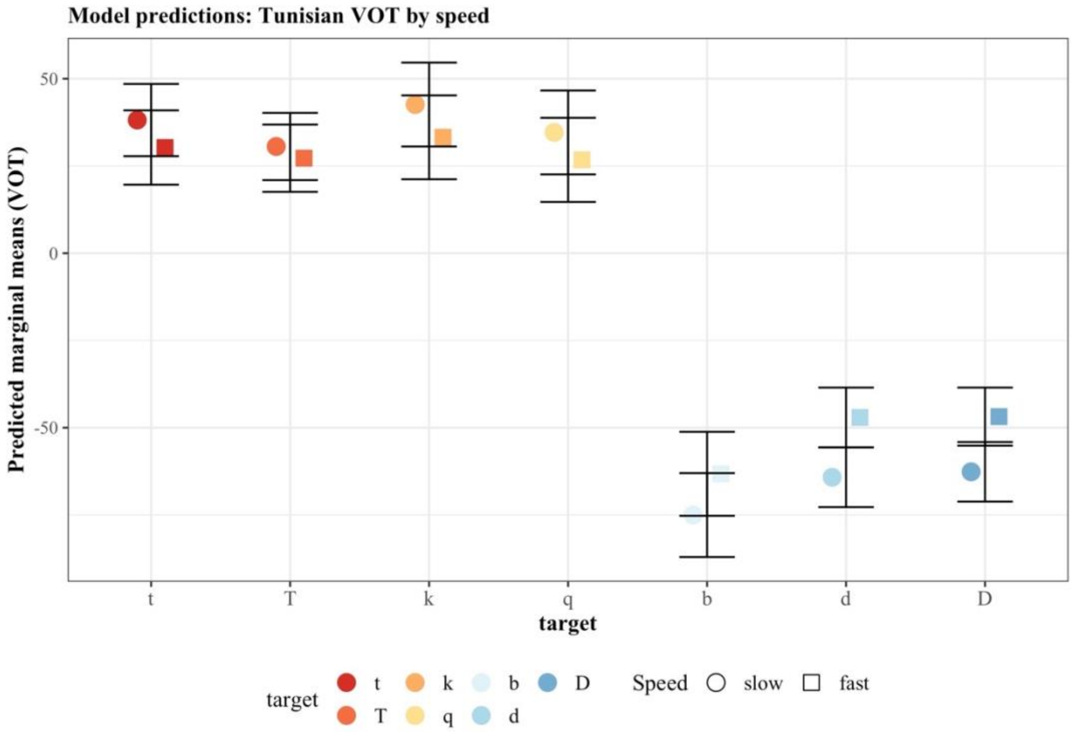


Figure 4-39 Predicted marginal mean (and 95% CI) for Tunisian VOT in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/ by speech rate.

To get a clearer view of the voiceless subset, we explored the four voiceless plosives in a separate linear mixed-effect model with sum-coded categorical variables. The model included *VOT* as the dependent variable and *speed*, *target*, *gender*, *position*, and *following vowel* as fixed factors. The model included the interaction between *speed*target*, and random intercepts for *speaker* and *item*[46]. The voiceless subset of the Tunisian data shows a significant main effect of *speed* (β = 3.52, *t* = 12.10, *p* < .001) which translates into overall longer voicing lag in slower speech. The model also revealed a significant two-way interaction for only two plosives, /ṭ/ and /k/ as seen in Table 4-8[47]. In slow speech, the Tunisian emphatic /ṭ/ is **less** affected by rate than the average, by 1.8 ms (*slow* x */ṭ/:* β = -1.79, *t* = -3.84, *p* < .001), while the Tunisian /k/ showed a significant effect of speech rate in that slower speech resulted in **longer** voicing lag than the average, by 1.2 ms (*slow* x */k/:* β = 1.17, *t* = 2.23, *p* =.026). The full summary of the model results is provided in Appendix (C.17).

[46] Model used: vot ~ speed * target + position + folv + (1 | item) + (1 | speaker), data = Tunisian voiceless subset
[47] To estimate the held-out factor /q/, we rotated the levels of the model.

Table 4-8 Model results for the two-way interaction of speed*target in the voiceless Tunisian subset.

Interactions of Interest	*Estimate*	*Std. Error*	*t.value*	*p.value*
slow : t	0.1586	0.4951	0.320	0.75
slow : ṭ	-1.7898	0.4658	-3.842	< .001*
slow : k	1.1724	0.5265	2.227	0.03*
slow : q	0.4587	0.5326	0.861	0.39

Considering the full set of plosives in Tunisia, this means that the effect of speech rate on VOT was significant in all plosives, but the size of the effect varied. The effect of speech rate on voicing lag in the emphatic /ṭ/ was the least compared to the other set of plosives in the Tunisian dataset. All the voiced plosives show a clear effect of speaking rate, indicating an active [voice] feature. In the voiceless subset, VOT values for /ṭ/ significantly overlap in the slow and fast conditions, hence /ṭ/ is not affected by speech rate indicating no active feature. Whereas, /t/, /q/, and /k/ are all affected by speaking rate, though to a lesser degree than the voiced subset. The velar /k/ showed the most effect of speech rate in the voiceless subset, with longer voicing lag for /k/ in slower speech indicating an active [sg] feature.

4.6.5.2 Tunisian Closure duration

Closure duration of the plosive was also investigated in the Tunisian subset of the data. In Figure 4-40, from the raw data, we can see clear influence of speech rate on closure duration in that it is longer in slower speech. However, unlike in Najdi and Hijazi, there appears to be some influence of closure duration on voicing: voiced plosives have slightly longer closure duration than voiceless plosives. Speech rate has a varied influence on the individual plosives, and /t/ seems to show the most overlap between the two speech rates.

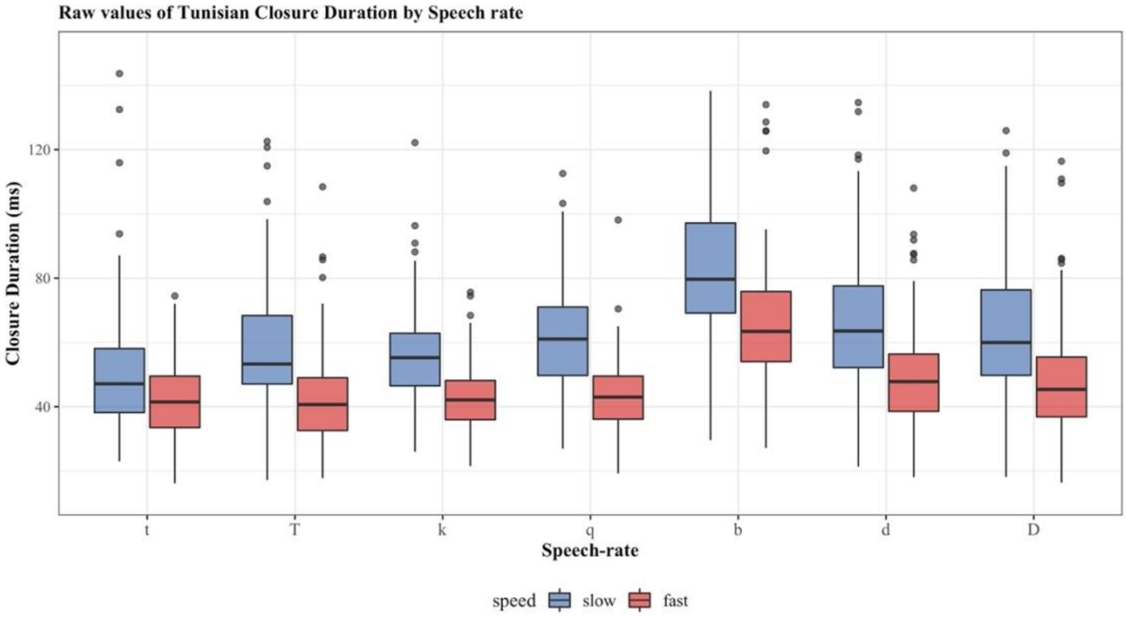


Figure 4-40 Median and interquartile range for Closure duration values of 7 plosives both voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/, by speech rate - Tunisian dialect.

Using linear mixed-effects models, we explored this subset of the data statistically with sum-coded categorical variables. The best fit model was the same as the two previous dialects, with *Closure Duration* as the dependent variable and were *speed*, *target*, *position*, and *gender* as the fixed factors. The model considered the interaction between *speed*target*, and random intercepts for *speaker* and *item*[48].

Again, the results showed a significant main effect of *speed* (β = 8.45, *t* = -15.01, *p* < .001) in which closure duration had higher values in slower speech rates. Neither *position* (β = -0.29, *t* = -0.25, *p* = .807) nor *gender* (β = 2.36, *t* = 1.91, *p* = .067) had a significant main effect on closure duration in the Tunisian subset. There was a significant interaction between *speed* and *target* for three target plosives. From the voiced plosives, /b/ (β = 3.55, *t* = 2.38, *p* = .018) and /d/ (β = 2.43, *t* = 2.07, *p* = .038) showed a significant effect of speaking rate on the closure duration in which the closure duration difference between fast and slow was the greatest among the investigated plosives, as seen in Figure 4-41. As we noted in the raw data visualization, voiceless /t/ (β = -4.24, *t* = -3.20, *p* < .001) shows a significant interaction with speaking rate in that it showed the least difference of closure duration between slow and fast rates, compared

[48] Model used: closdur ~ target * speed + gender + position + (1 | item) + (1 | speaker), data = Tunisian

to average. The full summary of the model results is provided in Appendix (C.18).

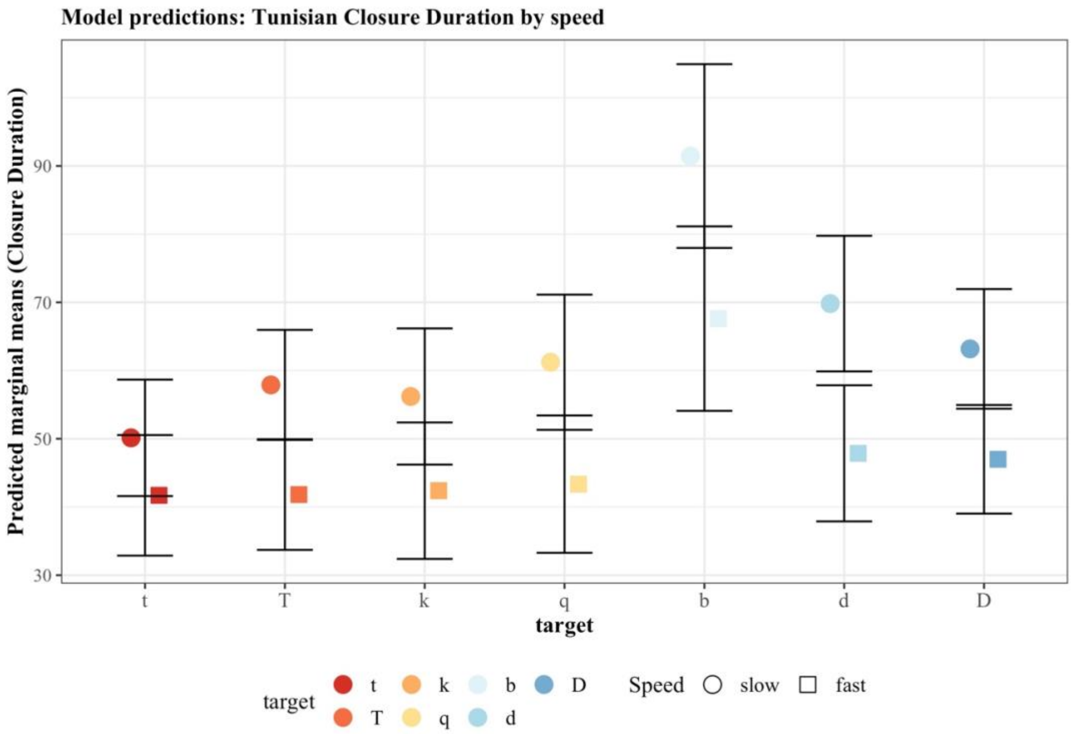


Figure 4-41 Predicted marginal mean (and 95% CI) for Tunisian Closure Duration in voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/ by speech rate.

To summarize, both *position* and *gender* had no significant effect on the closure duration of Tunisian plosives, however, *speed* did have a main effect on closure duration as expected. The interaction of speech rate and target plosive was significant for /d/ and /b/ (larger effect than average), and for /t/ (smaller than average). The closure duration of /t/ had the most overlap between the two speaking rates. The Tunisian results are different from the Najdi and Hijazi closure duration results, where in the latter two dialects the effect was mostly on the voiceless emphatic /ṭ/ and the uvular /q/.

4.6.5.3 Tunisian Fundamental frequency

We also explored f0 at the 25% point in the vowels following the plosives of the Tunisian subset. The raw data visualization showed clear influence of gender, but no obvious effect of speech rate or voicing. The Tunisian subset was investigated statistically with sum-coded categorical variables as well. The best fit model after comparing several linear mixed-effects models had *f0.25* as the dependent variable and *speed*, *target*, *position*, *gender* and *following vowel* as the fixed factors. The model included the interaction between *speed*target*, and random intercepts for *speaker* and *item*[49].

In the Tunisian subset, as for Najdi and Hijazi, the model showed a significant main effect of *speed* (β = -1.40, t = -5.66, p < .001) in that f0 is higher when speaking rate is faster. Unsurprisingly, *gender* (β = 5.06, t = 9.04, p < .0001) showed a significant main effect with higher values of f0 by female speakers.

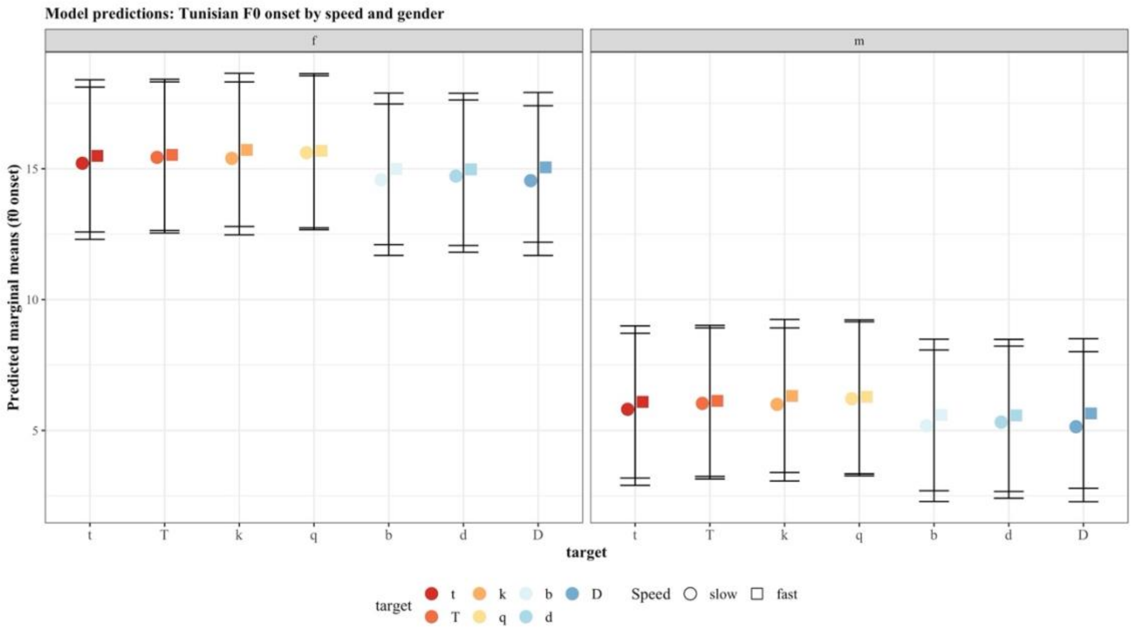


Figure 4-42 Predicted marginal mean (and 95% CI) for Tunisian f0 at 25% point in vowels following voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ/ plosives by speech rate and gender.

The Tunisian subset showed no significant interaction between *speed* and *target*, though the emphatic /ḍ/ (β = -1.08, *t* = -1.56, *p* = .115) showed the greatest f0 differential in values between the two speaking rates, and the emphatic /ṭ/ (β = 9.36, *t* = 1.56, *p* = .120) showed the least difference in f0 values between fast and slow speaking rates as seen in Figure 4-42. Again, none of the plosives in this dataset showed a significant effect of speech rate on f0 on the following vowel. The full summary of the model results is provided in Appendix (C.19).

[49] Model used: f025 ~ speed * target + gender + following vowel + (1 | item) + (1 | speaker), data = Tunisian

In summary, voiced plosives at slower speaking rate, and by male speakers, show lower f0 in the following vowel. The Tunisian dataset showed no significant interaction between speech rate and target in any of the plosives.

4.6.6 Results 3: Summary

In the results section for each dialect, we confirmed that VOT in voiced and voiceless plosives showed no overlap, even at a fast speech rate. Values of VOT in voiced plosives display voicing lead with a maximum of -10 ms and VOT values in voiceless plosives show clear voicing lag with the lowest value of 5 ms, across all the dialects at both speaking rates.

Our results also demonstrated that, in all three dialects, VOT values in voiced and voiceless plosives show an asymmetric effect as speaking rate changes. The rate effect is consistent and obvious in voiced plosives in all three dialects. Tunisian /d/ showed the greatest effect of speech rate among the dialect’s voiced plosives, while in Hijazi and Najdi it was the emphatic counterpart /ḍ/ that showed the largest effect of speaking rate. As a whole, the effect of speaking rate in the voiced plosives is clear and significant across the board. This can be interpreted as confirmation that [voice], as a phonological feature, is active in all investigated dialects.

The picture becomes more complex when looking at the voiceless plosives in these dialects. The extent to which the emphatic /ṭ/ and the uvular /q/ are affected by speech rate – by virtue of having short lag VOT in most Arabic dialects – is different in the dialects investigated. As we discussed above, Tunisian showed the largest effect of speech rate on the emphatic /ṭ/ and on the uvular /q/, compared to the other dialects. The Najdi emphatic /ṭ/ is the least affected and the Najdi uvular /q/ is not at all affected by changes in speech rate, while Hijazi emphatic /ṭ/ and uvular /q/ were in the middle. In contrast, in all the investigated dialects, /t/ and /k/ showed a significant effect of speaking rate: voicing lag in /t/ and /k/ were clearly longer in slower speech, which can be interpreted as confirmation that there is an active phonological feature of [sg] in all three dialects.

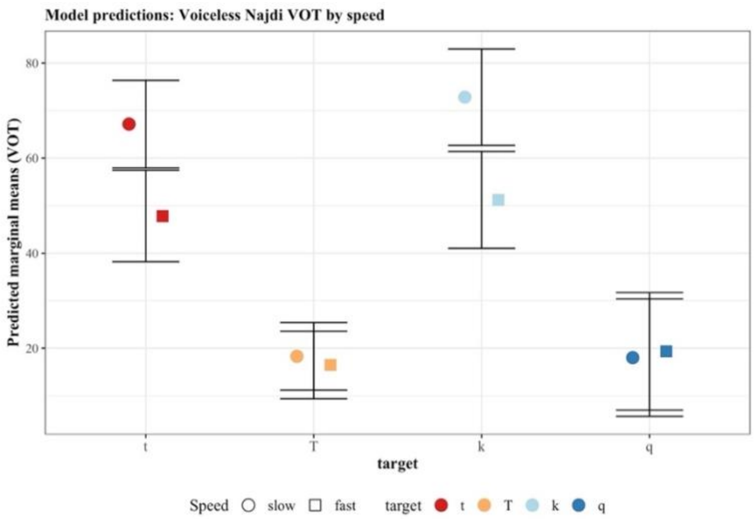


Figure 4-43 Predicted marginal mean (and 95% CI) for Najdi VOT in /t, ṭ, k, q/ by speech rate.

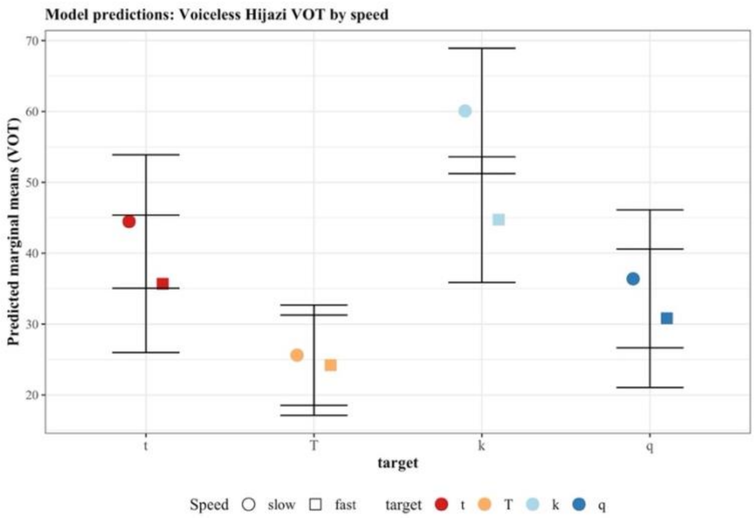


Figure 4-44 Predicted marginal mean (and 95% CI) for Hijazi VOT in /t, ṭ, k, q/ by speech rate.

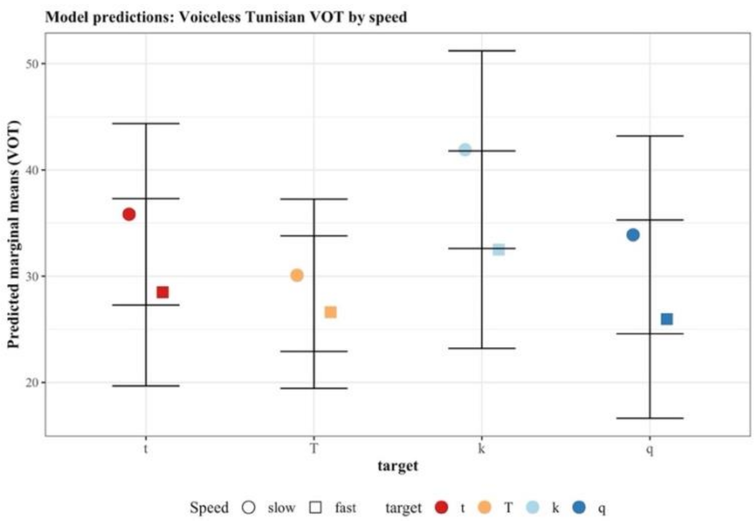


Figure 4-45 Predicted marginal mean (and 95% CI) for Tunis VOT in /t, ṭ, k, q/ by speech rate.

The Najdi dialect showed the least effect of speaking rate on the emphatic /ṭ/ and uvular /q/, and the most effect on /t/ and /k/. As a three-way contrast dialect (voiced, voiceless short lag, voiceless long lag), we can safely claim that this dialect has two active features, [sg] and [voice], and that the voiceless emphatic /ṭ/ and the uvular /q/ are not specified with a laryngeal feature. As concluded in § 4.5.6, the Najdi dialect is not over-specified; that is, it needs two active features to encode a robust three-way VOT contrast.

In the Hijazi dialect, likewise, both /t/ and /k/ are affected by speaking rate significantly. The emphatic /ṭ/ and the uvular /q/, however, are less affected by speech rate than the Tunisian dialect (compare Figure 4-44 and Figure 4-45). The impact of speech rate on the Hijazi voiceless plosives is closer to Najdi in Figure 4-43 than Tunisian. The Hijazi dialect can be considered as a hybrid dialect, in terms of its VOT categories, that has two active features. Therefore, Hijazi looks like an over-specified dialect.

In § 4.5.6, we concluded that the Tunisian dialect might be over-specified since it has only two VOT categories but shows two active phonological features, and we based this conclusion on the results of only four plosives (the plain and emphatic contrasts) in this dialect. This diagnosis of over-specification still stands after investigating the full set of plosives in this section. As we mentioned earlier, /t/, /k/, and /q/ are all clearly affected by speech rate indicating that [sg] is an active feature in this dialect, even though VOT values of /t/ and /k/ in Tunisian are shorter in normal speech compared to a three-way contrast dialect (e.g., Najdi).

The results so far are based on separate models (one per dialect). As a final step, to develop a full picture and compare all three dialects directly, we also explored the VOT values of the voiceless subset across all three dialects. The visualization of raw VOT across dialects at the start of this section in Figure 4-25 showed clear influence of speaking rate only on Najdi and Hijazi voiceless plain plosives; Tunisian voiceless plain plosives varied in the degree of speech-rate influence. The voiceless subset was investigated statistically with sum-coded categorical variables, *dialect* is sum-coded as follows (two predictors of *sanj* ~ *tuns* and *sahi* ~ *tuns* in which *sanj* was set to 1, *tuns* to -1, and *sahi* to 0). The best fit model after comparing several linear mixed-effects models was fitted that considered *VOT* as the dependent variable and *speed*, *target*, *dialect*, *following vowel* and *gender* as fixed factors. The interaction between *speed*target*dialect* was included, as well as random intercepts for *speaker* and *item*[50].

The voiceless subset of the dataset revealed a significant main effect of *speed* ($\beta = 4.27$, $t = 19.30$, $p < .001$) which translates into overall longer voicing lag in slower speech. The subset showed no main effect of *gender* ($\beta = 0.59$, $t = 0.87$, $p = .39$). The model also revealed significant three-way interaction in all plosives in all dialects except for /ṭ/, /k/ and /q/ in Hijazi: which are closer to the average (i.e., intermediate between Tunisian and Najdi in the degree of effect of speech rate). The model results are complex and difficult to interpret due to multiple interactions, but Figure 4-46 below illustrates the overall picture. The full model summary is provided in Appendix (C.20).

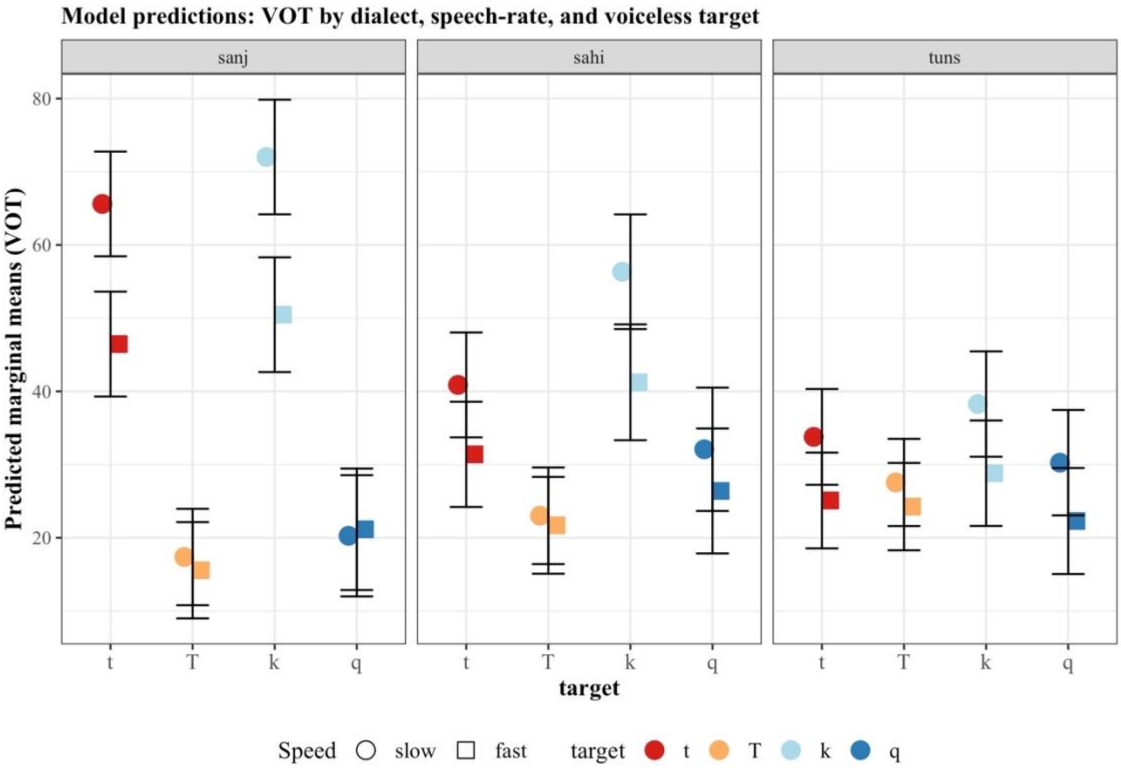


Figure 4-46 Predicted marginal mean (and 95% CI) for voiceless VOT values /t, ṭ, k, q/ in Najdi, Hijazi, and Tunisian by speech rate.

From Figure 4-46 we can see that the VOT values for voiceless plosives in Najdi has long lag and short lag, while in Hijazi the VOT values for voiceless plosives has intermediate lag[51] and short lag, finally Tunisian has intermediate lag for all its

[50] Model used: vot ~ speed * dialect * target + folv + gender + (1 | item) + (1 | speaker), data = unvoiced subset
[51] Values between the short lag category and the long lag category, more in § 4.7.

voiceless plosives. The rate effect on the long lag plosives is clear, while short lag plosives show no speech rate effect. The intermediate lag plosives also show residual rate effect. Overall, the results of this section showed that both voiced and voiceless plosives in Arabic can be affected by the rate of speech. In voiced plosives the effect of rate is consistent across dialects, but in voiceless plosives the effect of rate varies. Regardless of the number of observed VOT categories, the voiceless emphatic /ṭ/, with short lag VOT values, appears to be unspecified with a laryngeal feature, whereas both [sg] and [voice] are active features in all investigated dialects.

4.7 General discussion and conclusion

In the previous chapter (Chapter 3), we concluded that there are differences in the mapping of VOT to laryngeal contrasts across Arabic dialects, and we described a continuum of variation. Some dialects displayed three VOT categories of voiced (voicing lead), voiceless plain (long lag), and voiceless emphatic (short lag), while other dialects displayed two VOT categories of voiced (voicing lead) and voiceless (short lag). Other dialects (Tunisian and Omani) showed signs of a merger between the voiceless plain and emphatic values indicating potential sound change in progress.

This chapter was designed to investigate the nature of the laryngeal categories of Arabic by examining a number of acoustic cues to voicing, but mainly VOT, in three different dialects at two speaking rates: fast ~ slow. An asymmetric effect of speaking rate on VOT was used in previous studies as a diagnostic for active laryngeal features cross-linguistically. Assuming that a *true voice language* with two VOT categories (e.g., Spanish) has a [voice] feature (vs. [Ø]), and an *aspirating language* with two VOT categories (e.g., English) has a [spread glottis] feature (vs. [Ø]), then faster speaking rate affects the phonetic cue of only the 'active' phonological feature (Allen & Miller, 1999; Kessinger & Blumstein, 1997; Miller et al., 1986). Recent studies reported an 'over-specification' issue where a language appears to have two active features [voice] and [sg] despite having only two VOT categories, as in the case of Swedish (Beckman et al., 2011), where both *voiced* and *voiceless* plosives were affected by speech-rate manipulation.

By examining three Arabic dialects, the aim in this chapter was to investigate the effects of speaking rate on dialects expected to have three VOT categories (Najdi), and two VOT categories (Tunisian), and a dialect expected to be in the process of change

(Hijazi). We mainly want to answer the following research questions: **2a)** How does speaking rate affect VOT in different voicing categories, and which laryngeal features are active in each dialect? **2b)** Are two-way voicing category dialects over-specified with two active phonological features?

Before examining the speech-rate manipulation influence on the VOT categories of the dialects, we examined the realization of the emphatic contrast in the three dialects in § 4.4 and compared them to the IVAr data explored in chapter 3. The results of that section showed voicing lead in the VOT values of the voiced emphatic contrast /d/ and /ḍ/ in all three dialects. While the voiceless emphatic contrast showed somewhat similar distributions of VOT values in /t/ and /ṭ/ in both Hijazi and Tunisian (though with more merger of VOT values in Tunisian), Najdi showed distinct VOT distributions for /t/ and /ṭ/. Based on these results, Tunisian can be considered as having two VOT categories and Najdi as having three VOT categories. In contrast, Hijazi can be regarded as transitioning towards merging its voiceless plain and emphatic contrast /t/ and /ṭ/, but is still at a stage where it should be considered a hybrid dialect.

The results of speech-rate manipulation on /t, ṭ, d, ḍ/ in § 4.5 revealed that Najdi, as expected, can be argued to have both [voice] and [sg] as active phonological features, by virtue of impact on VOT values in response to speaking rate changes in only two of the three categories: *voiced* and *voiceless plain* plosives. The short lag *voiceless emphatic* plosive /ṭ/ was unaffected by changes in the speaking rate. The results also showed that Hijazi *voiced* and *voiceless plain* plosives, similar to Najdi, are affected by speech-rate, but not the *voiceless emphatic,* which can also be interpreted as evidence of having both phonological features [voice] and [sg] active in this dialect. Note that Najdi and Hijazi are expected to need two active features if they both have three VOT categories, and this was observed. The Tunisian results, however, showed an effect of speaking rate on all plosives investigated, both *voiced* and *voiceless*, though the *voiceless emphatic* was the least affected category. This suggests that Tunisian dialect, like Swedish, has two VOT categories but is over-specified with two active phonological features: [voice] and [sg].

The significant effect of speech rate on /t/, /ṭ/, /d/, and /ḍ/ in § 4.5 hints at over-specification of [voice] and [sg] in the Tunisian dialect. Table 4-9 compares the mean VOT values of languages with different numbers of categories of voicing from Kessinger & Blumstein’s (1997) study and the present study. While we must interpret the data with caution, due to possible variability in speakers' productions and in their interpretation of the instructions to speak slow or fast, across the two studies, we notice a close correspondence.

Table 4-9 Mean VOTs of different categories of voicing reported as a function of speaking rate in the two studies. English, French, and Thai from Kessinger & Blumstein (1997); Najdi, Hijazi, and Tunisian from this study.

Language	Plosive Type	Slow (ms)	Fast (ms)	Difference (ms)	Ratio (fast/slow)
English	Long lag	108	79	29	0.73
	Short lag	18	17	1	0.95
Thai	Long lag	80	53	27	0.66
	Short lag	13	12	1	0.92
	Voicing lead	-69	-42	27	0.61
French	Short lag	33	30	3	0.91
	Voicing lead	-110	-83	27	0.75
Najdi	/t/	67	47	20	0.70
	/ṭ/	18	16	2	0.89
	/d/	-66	-50	16	0.76
	/ḍ/	-75	-57	18	0.76
Hijazi	/t/	42	32	10	0.76
	/ṭ/	24	23	1	0.96
	/d/	-66	-52	14	0.79
	/ḍ/	-81	-61	20	0.75
Tunisian	/t/	35	26	9	0.74
	/ṭ/	28	25	3	0.96
	/d/	-65	-48	17	0.74
	/ḍ/	-64	-47	17	0.73

On average, VOT in short lag plosives (i.e., emphatic /ṭ/ in our study) vary across speech rates in the range of 1 and 3 ms, and that is not considered as 'movement' by Kessinger & Blumstein (1997). The mean VOT difference in the long lag category (i.e., plain /t/ in our study) for English, Thai, and Najdi shows a similar speech rate differential of between 20 and 29 ms. In contrast, Hijazi and Tunisian plain /t/ stands on the threshold of difference with an average of 10 – 9 ms.

Due to methodological differences between studies, comparing raw differences may be misleading, or may not give us a full picture. The raw mean VOT values of /t/ in Hijazi and Tunisian are intermediate 'mid lag' values, so the difference between fast and slow will be smaller than for long lag categories. In the same Table 4-9[52], a Ratio column is added to examine the relative increase at slow vs. fast rates. Ratio is proposed as a simple, yet more careful, measure to normalize the magnitude of the VOT difference between fast and slow across studies and categories. By using the ratio instead of the raw difference, we notice that a ratio around (0.9) could be a useful indicator of minimal or no difference. In this view, the ratio of fast to slow in Hijazi and Tunisian /t/ are (0.76) and (0.74) respectively, which would count as movement, regardless of their intermediate mid lag VOT values. The emphatic voiceless /ṭ/ in our study is behaving the same way in the three Arabic dialects, and also as in the short lag categories unaffected by rate in the three languages, English, French, and Thai from Kessinger & Blumstein (1997). We suggest that this is what an unspecified category looks like under the effect of speech-rate, (that is, it shows a Ratio of fast to slow of around 0.9 ms or higher).

Before discussing this apparent over-specification further, we review also the full range of plosives beyond the coronals. In § 4.6, aiming to understand the laryngeal contrast systems of the dialects as expressed through acoustic cues, we also included in the analysis bilabial /b/ and velar /g/ in the voiced set of plosives, and velar /k/ and uvular /q/ in the voiceless set. The voiced plosives /b, d, ḍ, g/ were all similarly affected by changes in the speaking rate, indicating an active [voice] feature present and active in all three dialects. In contrast, the voiceless plosives varied in the degree to which they show an effect of speaking rate. In all dialects, the voiceless emphatic /ṭ/ appears to be unaffected by changes in speaking rate indicating that it is unspecified with a laryngeal feature. The uvular /q/, however, is unaffected by changes in speaking rate only in Najdi dialect but is affected in Hijazi and Tunisian. The plain voiceless /t/ and /k/ have longer voicing lag than the emphatic /ṭ/ in all three dialects and were all affected significantly by speaking rate indicating that [sg] is an active phonological feature in all investigated dialects. It is worth noting that voicing lag in /t/ and /k/ in Tunisian fast speech are not shortened by the same amount at fast speech rates as the same plosives in Najdi and Hijazi, but they show similar ratios of (slow : fast) rate. For example, mean difference of slow and fast values of /k/: Najdi = 21 ms, ratio (0.71); Hijazi = 15 ms, ratio (0.74), Tunisian = 10 ms, ratio (0.75).

[52] To view the number of tokens in each target, see Appendix (C.21).

We now reflect on what the results tell us about the phonological laryngeal specifications in these dialects. It seems clear that Najdi has three VOT categories, and employs [voice] and [sg] as laryngeal phonological features. Hijazi, however, is a hybrid dialect that is transitioning towards having a two VOT categories, and still employs [voice] and [sg] as laryngeal phonological features. As for Tunisian, the dialect seems to have two VOT categories and is over-specified with two phonological features [voice] and [sg].

We notice that the VOT values of the different voiceless plosives in Tunisian /t, ṭ, k, q/ are all clustering around similar mean values, above 20 ms and below 40 ms, as we see in Figure 4-46. These values are not short lag (between 5-10 ms), which does not require any laryngeal activity or articulatory force. Yet, they are not >50 ms which could readily be classified as long lag aspirated. This atypical short lag VOT in Tunisian is still, however, sensitive to speaking rate. Discussion in the literature of an atypical VOT category, that is sometimes called 'intermediate lag' (or medium lag), in languages like Japanese (Riney, Takagi, Ota, & Uchida, 2007) Hebrew (Raphael et al., 1995) and Puerto Rican Spanish (Raphael, Tobin, & Most, 1983), might be relevant in this case. Here, in the Tunisian dialect /t, ṭ, k, q/ and the Hijazi /t, q/, we also observe a category which is intermediate when compared to the short lag versus long lag dichotomy of voiceless plosives used generally to characterize languages. In Table 4-9, we can see that the speech-rate mean difference in the 'short lag' voiceless emphatic /ṭ/ of Tunisian is the same as the 'short lag' difference for French voiceless from Kessinger and Blumstein's (1997) study (i.e., 3 ms). Nevertheless, Kessinger and Blumstein (1997) concluded that this phonetic category of short lag was unaffected by speaking rate. In addition, the ratio of fast to slow, as previously mentioned, places the Tunisian emphatic /ṭ/ in the 'no movement' category above (0.96), similar to the short lag French in the aforementioned study which was (0.91).

To account for this intermediate lag VOT, as opposed to the long lag seen in Najdi, it is tempting to attribute it to the fact that Tunisian speakers are typically bilingual in French, which is a de facto lingua franca in Tunis. French is a *true voice language* but has been reported to have similar 'intermediate' lag values for voiceless plosives compared to English short lag plosives (Caramazza & Yeni-Komshian, 1974; Kehoe & Kannathasan, 2021; Nearey & Rochet, 1994). The sustained community bilingualism of Tunisian speakers in French as an L2 might have affected their production of Arabic long lag voiceless plosives (such those seen in Najdi) in the same way that the anglophone environment has affected the Canadian French speakers in Caramazza et al.'s (1973) study of bilingual speakers of French and English: their VOT in French voiceless plosives were slightly longer than that of the French monolinguals. Caramazza and Yeni-Komshian (1974) also compared monolingual French speakers from Canada with monolingual French speakers from France and showed that the voicing lag of Canadian French speakers was longer than that of French speakers from France. They attribute this longer VOT in Canadian French voiceless plosives to the language being in contact with Canadian English, such that if the voiceless plosives in Canadian French are pulled toward the VOT of the English voiceless long lag plosives. The effect of bilingualism is greatly discussed in the literature of VOT, and the general consensus is that bilingual speakers have the ability to adapt their production of plosives based on the systems of each language (see (Khattab, 2002) and (Flege, 1995)). Although Arabic-French bilingualism might be a plausible explanation for Tunisian intermediate lag, bilingualism does not explain the potential intermediate lag of /t/ in Hijazi. The bilingual Hijazi speakers in our data reported English as their second language. Table 4-10 is a summary of the phonetic and phonological implications of the laryngeal contrast of the three investigated dialects.

Table 4-10 Results Summary: comparison of three dialects (dialect speech) in terms of their phonetic number of VOT categories and the possible phonological representations. The asterisks * indicate potential phonological 'over-specification'.

Task	**Dialect**	
Dialect	Phonetics	Phonology
Najdi	three VOT categories (voicing lead – short lag – long lag)	[voice] [Ø] [sg]
Hijazi	three VOT categories (voicing lead – short lag – intermediate lag)	[voice] [Ø] [sg]
Tunisian	two VOT categories (voicing lead – intermediate lag)	[voice] [Ø] [sg]*

To conclude, the three investigated dialects appear to have the same number of active phonological features despite displaying different surface patterns of VOT categories. Both voiced and voiceless plosives in Arabic dialects are affected by speech rate. The effect is consistent in the voiced plosives across the dialects, but not consistent in the voiceless plosives. The pattern discovered in Tunisian dialects in this chapter resembles the pattern in Swedish, which has two VOT categories but two active phonological features [voice] and [sg] (Beckman et al., 2011). However, unlike Swedish, the Tunisian voiceless emphatic plosive /ṭ/ is not affected by changes in speaking rate, suggesting that its phonological representation differs from /t/. This suggests that Arabic dialects can have surface patterns of two or three VOT categories, but in all cases the voiceless emphatic /ṭ/ is unspecified with a laryngeal feature, and both [sg] and [voice] are active features in all investigated dialects.

The next chapter explores one more avenue to further examine these patterns in another register – Modern Standard Arabic (MSA). The main goal of the next chapter is to see whether the phonological and phonetic properties in MSA parallel those in a speaker's dialect. A speech-rate variable is included in the MSA dataset also, to shed light on the phonological representations of the laryngeal contrast in this different register. Note that we are not claiming that MSA is the 'underlying form' of Arabic across dialects. Instead, we are curious to see whether Arabic speakers maintain their dialectal laryngeal contrast in the more formal register of MSA.

Chapter V

Dialect influence on MSA

5 Dialect influence on MSA

The data collected for this chapter are parallel data in MSA from the same speakers who participated in chapter 4, at two speech rates, aimed to investigate whether the phonological and phonetic properties in MSA are parallel to those in a speaker's dialect. In this chapter, we aim to answer the following questions: **3a)** Are the observed VOT patterns in dialect speech mirrored in MSA produced by the same speakers? **3b)** Do speakers of a dialect display the same active feature(s) in the two registers of Arabic?

In this chapter, § 5.1 starts with a brief overview on the linguistic context of MSA, diglossia, and dialectal variation in MSA. In § 5.2 and § 5.3 the methods, measurements, and segmentation of this study are presented. In § 5.4 and § 5.5 I present detailed results on the data of the plain vs. emphatic coronal plosives in MSA, then detailed results on the speeded data of the same plosives /t, ṭ, d, ḍ/ in three dialects (Najdi, Hijazi, and Tunisian). In light of the results, in § 5.6 I provide a general discussion and conclusion.

5.1 Modern Standard Arabic: A sketch

5.1.1 The diglossic situation

Arabic is a classic example of the sociolinguistic phenomenon of diglossia. Diglossia, as Ferguson (1959) defines it, is two or more varieties of the same language used by some speakers under different conditions. The co-existence of Modern Standard Arabic (henceforth MSA), alongside vernacular varieties of Arabic is one example of interest in this section. In diglossia, the two varieties of the same language are not quite interchangeable, as they are used in socially distinct contexts. In Arabic, MSA is the High variety used mostly in religious sermons, formal settings, educational contexts, and TV broadcasts. MSA is also the form that is used formally in writing, and it is no one's mother tongue but is taught in educational settings at an early age. On the other hand, the vernacular is considered in the literature of diglossia to be the Low variety. It is used in everyday life situations among family and friends, and is not used in formal writing. A written form of vernacular is used among family and friends for texting or online chatting. The vernacular register is the form of language that a child first acquires. The two varieties are linguistically related, however, in the phonology, morphology, and syntax, as well as a large shared set of lexical items (Amer, Adaileh, & Rakhieh, 2011; Ferguson, 1959; Saiegh-Haddad, 2012).

The High and Low prestige categorization is not discrete, but there are instead gradient speech levels, along a continuum corresponding to a variety of social situations in speech. Ferguson's (1959) model thus allows for 'intermediate forms of the language' that form this diglossic continuum (p.332). One of the earliest explorations of the continuum was Badawi's (1973) five levels of Egyptian spoken Arabic[53]. The 'levels' range from pure MSA '*fuṣħa:'* to illiterate colloquial '*ʕaamiyyat al-ʔummiyyiin'*, having '*ʕammiyyat al-muθaqqafiin*' as the middle level which translates to "the variety spoken by the cultured / well-educated intellectuals". This mid-level variety is described as having the standard and the vernacular combined with almost equal distribution in usage. Native Arabic speakers do not treat MSA as a coherent and self-contained system, but rather as part of a larger system that takes in the whole communicative continuum (Parkinson, 1996).

5.1.2 Dialect variation in MSA

The time of the emergence of Arabic diglossia has been much debated. Some scholars believe that diglossia in Arabic emerged in the 7th century A. D., with the Islamic conquests (Blau, 1977). Other scholars argue that Arabic diglossia is traced to a period predating the Islamic era, as the language of pre-Islamic poetry is found to be different from the colloquial (Altoma, 1969; Rabin, 1955). Rabin (1955) stresses that Medieval Muslim authors generally agreed that the language of pre-Islamic poetry is identical to the language spoken by the Bedouins of central and eastern Arabia, and that the language of Qur'an is the language of the prophet, which is the dialect of the Quraish Tribe in the west of Arabia. According to Rabin (1955), philologists acknowledge that considerable differences existed between the different dialects at that time. The recent advances of epigraphy have also made it possible to confirm that pre-Islamic Arabic was diverse and not at all linguistically homogenous (Al-Jallad, 2018; Harrell, 1960; Van Putten, 2017a, 2017b).

[53] (1) *fuṣħa al-tura:θ* 'classical Arabic' (2) *fuṣħa al-ʕ*aṣr 'Modern Standard Arabic' (3) *ʕammiyyat al-muθaqqafiin* 'educated colloquial (4) *ʕammiyyat al-mutanawriin* 'literate colloquial' (5) *ʕaamiyyat al-ʔummiyyiin* 'illiterate colloquial'.

In the modern era, the standardization of Arabic has generated a set of norms that early grammarians called /fuṣħa:/, which is a term used to refer to both Classical Arabic (or Qur'anic Arabic) and its modern descendant MSA. The limitation of using Classical Arabic or Qur'anic Arabic in recitation of the Qur'an has resulted in MSA being at one end of the general speech continuum (Al-Qenaie, 2011).

However, not all scholars agree on the existence of a single current standardized form of MSA. Harrell (1960) proposed the necessity of discovering regional and social varieties of MSA. In the diglossic context, Bentahila (1991) addresses the issue of standardization and concludes that MSA does not really achieve the highest degree of standardization. Parkinson (1993) believes that MSA is not the same for all Arabic speakers, and that, if this variation were studied, major differences in the average knowledge of MSA would be found.

From my personal observation, among many levels of variation, there does appear to be lexical variation in MSA across Arabic speaking countries. In Tunisia, for instance, a tenancy agreement or a contract is called "عقد كراء" /ʕaqd kira:ʔ/, whereas in the Middle East it is normally called "عقد إيجار" /ʕaqd ʔi:dʒa:r/. Even though it is widely assumed that MSA is the same across the Arab world, on closer reflection, Arab individuals notice lexical differences in the use of MSA in neighbouring countries, especially when comparing the written form of newspapers (Ibrahim, 2009).

5.1.3 Dialect influence on MSA

Earlier, in § 2.7, I discussed how it is generally assumed that in spoken production, MSA phonetic and phonological properties are influenced by those of the everyday dialect. Benkirane (1998) and ElZarka & Hellmuth (2009) note suprasegmental properties of MSA production that are, for the most part, affected by dialect speech, for instance, stress assignment and intonational properties. I also reviewed Elgibali's (1993) comparison of six sociolinguistic variables in data from speakers of Arabic from Kuwait and Cairo in both dialect and MSA. Elgibali concluded that MSA is different cross-dialectally, and that Kuwaiti MSA display more standard phonological features than Cairene MSA.

Turning to segmental features, Harrell (1960) compared MSA production in Egyptian Radio broadcasts with Classical Arabic, and noted that, by the use of the long vowels

/o:/ and /e:/ and the disappearance of short vowels word-finally, the Egyptian Radio production of MSA is a lot closer to colloquial Egyptian than it is to Classical Arabic.

Al-Fahid (2000) also discusses dialect influence on read speech MSA produced by Egyptian readers. In radio and TV broadcasts, it is evident that Egyptian readers tend to transfer some phonetic properties of their dialect when they read a text in MSA. For instance, Egyptian readers tend to replace the dental fricative /ð/ with their dialect realization of the sound as [z]. It is common to detect these phonetic properties of the Egyptian dialect in a formal radio channel or TV broadcast, such that you might hear the broadcaster say [hazihi] instead of MSA /ha:ðihi/ 'this (f./sg.)'.

Currently, there are no directly parallel data in the two registers (dialect and MSA) concerning voicing contrasts. It is not known, therefore, whether the patterns of VOT in MSA produced by speakers of different dialects of Arabic displays the same variation in dialectal speech that we have seen in the previous chapters.

5.2 Methods

5.2.1 Participants

The same participants as in the previous study in chapter 4 from Tunis, Jeddah, and Riyadh were asked to read a target short passage in Modern Standard Arabic. The total number of participants is 64: Tunisian speakers were 28 (14 female, 14 male), Saudi speakers of Hijazi were 18 (10 female, 8 male), and Saudi speakers of Najdi were 18 as well (8 female, 10 male). All participants were educated native speakers of their mother tongue dialect, and their age ranged 17-45 years old. All participants were asked to read the MSA passage as if they are reading in front of an audience or on a TV broadcast.

5.2.1.1 Self-reported evaluation of confidence

Before commencing the MSA experiment, the participants were asked to fill out a 'language background' questionnaire. When asked in the questionnaire about their confidence level in using MSA, most participants in all three dialects were rather confident. In the language background questionnaire, as seen in Appendix (D.1), a scale from 1-10 for confidence level in producing MSA was used, where (1) is not confident and (10) is very confident. Figure 5-1 shows the range of self-reported confidence scale, by dialect.

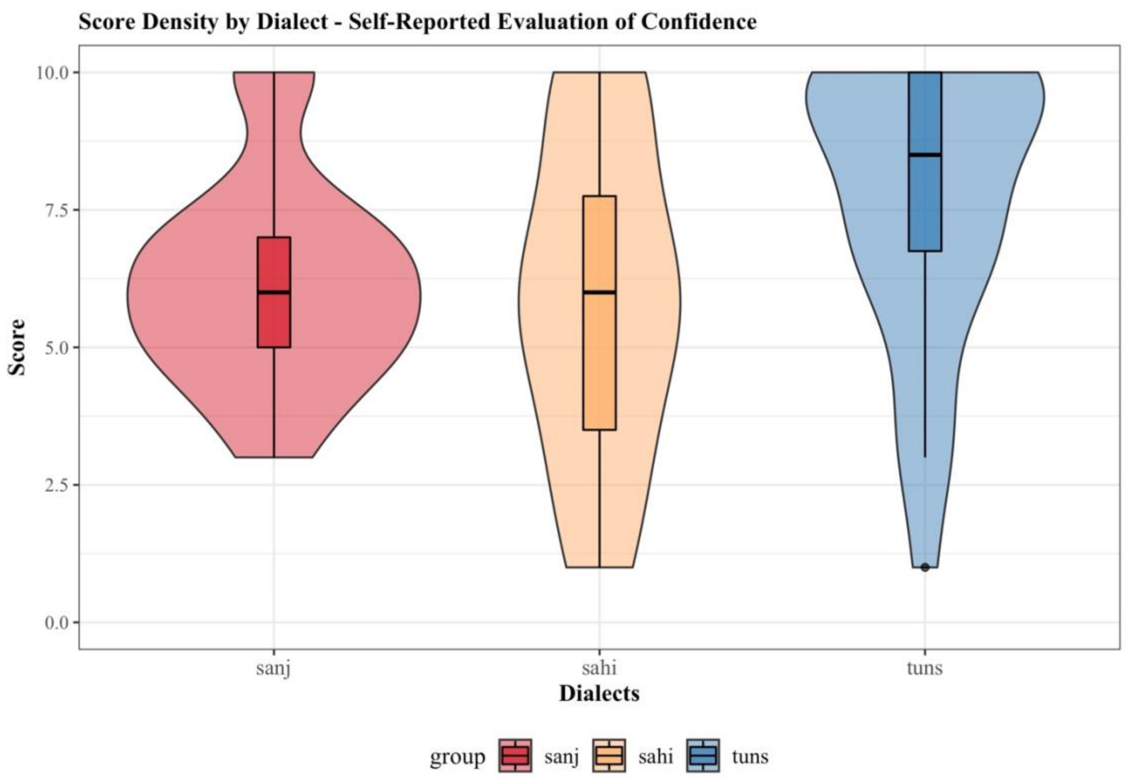


Figure 5-1 Self-reported MSA evaluation scores presented as boxplots to indicate the median and the quartiles range, and violin plots outlining the density.

In total, 69% participants self-reported a score above 5 in the (10:1) scoring scale, while 31% self-reported a score below 5. Tunisian speakers were the most confident while Hijazi speakers were the least confident. It is important to note that the metadata questionnaire was filled out by the participants before the recording started.

5.2.1.2 Researcher evaluation of MSA reading proficiency

Following data collection, and during the data analysis process, it was noticed that speakers in fact had varying degrees of proficiency in their MSA production. To the researcher, it seemed that their self-evaluation did not reflect their observed proficiency in MSA production. Therefore, I developed an evaluation system based on the Goodman Model of Reading (Miscue Analysis) and the Reading Miscue Inventory (RMI) developed by Ken and Yetta Goodman (Goodman, 2015).

Miscue analysis is a research tool developed by Ken Goodman and others (Brown, Goodman, & Marek, 1996). Goodman defines miscues as “mismatches between expected and observed responses” in the reading process (Goodman, 2005, p. 3), meaning an unexpected observed response produced by the reader to a written text.

Goodman argues that these mismatches are not random; rather they involve the use of existing cues available to the reader as expected responses to the text. This relationship provides a way to explain how the reader makes sense of the printed text (Goodman, 2005; Goodman, 2015).

Al-Fahid (2000) tested the Goodman Model on Arabic to investigate how readers of Arabic text construct meaning. As a Semitic language, Arabic represents vowels and syntactic markers only minimally in its orthographic form, by the use of the diacritics. Diacritics are markers that convey vowel sounds in Arabic, mainly short vowels, allowing accurate word pronunciation for a written text. Diacritics are not always present in a written text; they appear mostly to disambiguate homographic words (e.g., كَتَبَ /$k^{a}t^{a}b^{a}$/, *he wrote*, كُتِبَ /$k^{u}t^{i}b^{a}$/, *was written*). A common hypothesis is that displaying Arabic diacritics in a written text makes readers less efficient and less effective when reading a long passage aloud (Al-Fahid & Goodman, 2008; Hermena, Drieghe, Hellmuth, & Liversedge, 2015). More information and a thorough review of Al-Fahid's application of this model on reading of MSA text can be found in Al-Fahaid and Goodman (2008).

Research using Miscue Analysis has shown that readers tend to reveal their grammatical knowledge through their miscues (Goodman, 2015). To assess the level of grammatical knowledge of each participant when reading the MSA text for this study, I developed a rubric based on the five questions asked when evaluating any read text taken from the Reading Miscue Inventory (RMI):

1. The degree to which the sentence as finally read by the reader is syntactically acceptable in the sentence and the story.
2. Whether the sentence as finally read is semantically acceptable.
3. The degree to which meaning is changed. This question is only asked if the sentence, with all its miscues, is syntactically and semantically acceptable.
4. The degree to which substitution miscues are graphically similar.
5. The degree to which substitution miscues are phonologically similar.

Following Al-Fahid (2000), who devised a multi-tier extension the Goodman Model of reading to account for the diacritics in Arabic, the developed rubric, by the researcher, involved four main criteria:

1. The first is the reader's response to diacritics; this serves in finding miscues related to the first question about syntactic acceptability. As diacritics are often used in Arabic to mark grammatical case.
2. The second criterion is about graphical or phonological miscues, identified through mismatches in reading the lexical items in the text.
3. The third criterion evaluates semantic acceptability by assessing miscues in the intonation of reading. This was possible because there was an interrogative embedded in the passage.
4. Finally, the fourth criterion is an overall judgment on the observed confidence level. This last part is subjective, but as we shall see, both evaluators somewhat agreed that this can be evaluated by listening to the participants' performance in reading.

Goodman (2005) emphasises that miscue analysis should always involve authentic texts. Since the decision to evaluate the participants using this method happened post hoc, the MSA passage used in this study is not authentic as it was written specifically for the purpose of this study.

I used the Miscue analysis in addition to measuring the total reading time by participants for the MSA passage in the slow production speed[54] as an indication of fluency. The time taken to read the passage proved inadequate as a proxy for accuracy, as some participants' productions were slow and deliberate to ensure correctness, while others were fast, as if they were speeding up to mask any errors. The Miscue Analysis allowed me to categorize their production level based on their actual reading proficiency in MSA. This will serve later as a relevant factor to make sense of the VOT data.

[54] As seen in Appendix (D.3) column named (Time).

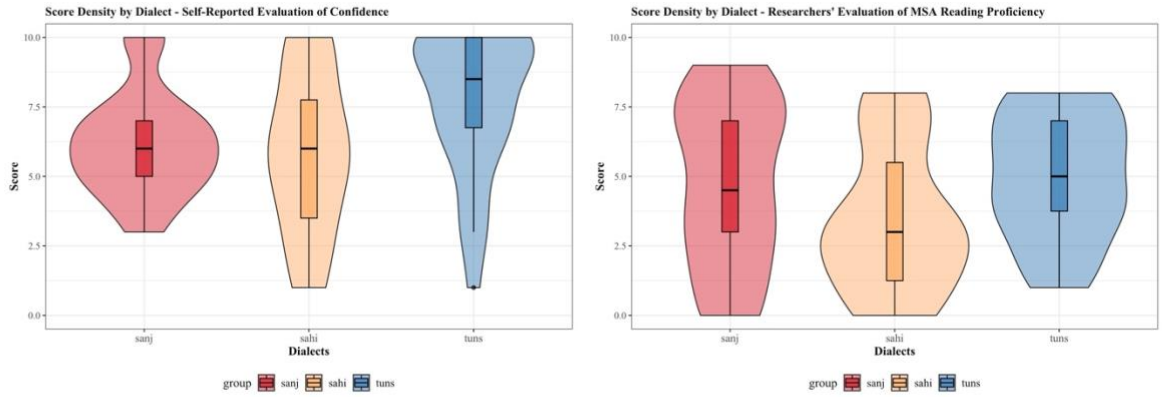


Figure 5-2 Scores for the researcher's evaluation (right) and the self-reported evaluation (left) presented as boxplots to indicate the median and the quartiles range, and violin plots outlining the density.

To avoid evaluation error or bias in judgment, in addition to my personal evaluation, I asked another native Arabic speaker to rate the same participants under the same guidelines. The inter-rater reliability percentage is over 84%. As seen in Figure 5-2 above, our results in the researchers' evaluation (right) mirror the self-reported confidence level in terms of the relative proficiency ranking among dialects (Tunis > Najd > Hijaz). In the agreed results between evaluators in three dialects, where a score of (10) is highly proficient and a score of (1) is not proficient, Tunisian speakers showed highest scores, while Saudi participants from Hijaz scored the least in MSA reading. More details on the self-reported and inter-rater evaluation results, in addition to the criteria of the rubric are available in Appendix (D.2) and Appendix (D.3).

In summary, the speakers were all determined to be sufficiently proficient in MSA to be included in this study. The use of the "score" from both evaluators will be included in Linear Mixed-Effects Models for the VOT dataset of MSA as a factor to control for this variation.

5.2.2 Materials

A reading passage in MSA, shown in Appendix (D.4), was presented to the participants. The short passage included target stimuli inserted in the passage, which are real words (N =12) of four contrastive plosives, two of which are voiced /d/ and /ḍ/, and two are voiceless /t/ and /ṭ/ in both word-initial and word-medial positions, followed by two vowels /i:/ and /a:/, as seen in Table 5-1.

Table 5-1 A list of the target stimuli embedded in the MSA reading passage.

Target Consonant	Item (IPA)	Gloss	Arabic	Position	Following Context
	da:r	house	دار	initial (onset)	/a:/
/d/	**d**i:n	religion	دين	initial (onset)	/i:/
	ʔa**d**.wija	medicines	أدوية	medial (coda)	/w/
	ḍa:r	harmful	ضار	initial (onset)	/a:/
/ḍ/	**ḍ**i:q	tightness	ضيق	initial (onset)	/i:/
	ʔa**ḍ**.wija	lights	أضوية	medial (coda)	/w/
	ta:h	got lost	تاه	initial (onset)	/a:/
/t/	**t**i:h	to stray	تيه	initial (onset)	/i:/
	ʔa**t**.riba	dust	أتربة	medial (coda)	/r/
	ṭa:b	to like	طاب	initial (onset)	/a:/
/ṭ/	**ṭ**i:n	mud	طين	initial (onset)	/i:/
	ʔa**ṭ**.ja:f	wraiths	أطياف	medial (coda)	/j/

In the design of the MSA reading passage, there was a trade-off while selecting the lexical items between naturalness of the reading passage, familiarity with the lexical items, having a minimal pair, and controlling the desired phonological environment. This trade-off resulted in medial position targets followed by different consonants. We also controlled for potential palatalization (in the context of /i:/ and /w, j/) in the data by including *following environment* as a variable in the Linear Mixed-effects Models used in R.

5.2.3 Procedure

Participants were first given the chance to read the short passage written in Modern Standard Arabic printed in a paper silently, to provide them with the confidence to perform the task, which most of them needed. Then, they were asked to read the passage aloud at a normal speaking rate (coded as "Slow"). They were instructed to read the passage as if they were reading it on a public radio or TV broadcast. Finally, the participants were asked to read the passage aloud for the second time, but this time faster (coded as "Fast"). The recordings were made directly to wav format at 44.1KHz 16bit, using a Marantz PMD660 and head-mounted Shure SM10 microphones.

5.3 Measurements and Segmentation

The acoustic measurement and segmentation methods followed in this chapter are similar to those in chapter 4 (see § 4.2). Praat version 6.1.37 was used to manually segment and analyse the MSA data. The audio files for this experiment were not force aligned; instead, Praat textgrids were created for each token with the help of a Praat script. The labelled textgrid for each token contains one “interval tier” and two “point tiers”. The interval tier was used to label the word duration, and the point tiers were used to label VOT, and following vowel start and end. The boundaries of segments were determined visually and labelled manually by inspecting waveforms and broadband spectrograms. VOT was measured for all plosives in milliseconds (ms). Negative VOT values were identified as the interval from the voice onset to the onset of the plosive burst. Both spectrograms and waveforms were also used visually to identify the onset of periodic striations as an indication of vocal fold vibration. As for positive VOT values, they were measured as the interval from the plosive burst to the onset of voicing. Overall word duration was also measured to indicate speech rate.

A number of acoustic correlates were manually labelled then automatically extracted using a Praat script. A total of 14 acoustic temporal and non-temporal variables were used, including VOT, word duration, closure duration, burst duration, percentage of voicing in voiced plosives, and several measurements from the vowels. The criteria and description of each acoustic variable is as discussed in chapter 4, § 4.2.

For this dataset, the total potential number of tokens is 1536 (64 subjects x 12 MSA target items x 2 speaking rates). This task generated 188 disfluent tokens, leaving a total of 1348 tokens in the MSA dataset to analyse.

5.4 Results 1: MSA overview

The structure of the results section is as follows: in § 5.4.1 we investigate the nature of the emphatic contrast (VOT and F2 lowering) of MSA in the three dialects. Then in § 5.4.2 we summarize the results of the MSA emphatic contrast in one speaking rate in the three dialects.

5.4.1 MSA Emphatic Contrast

In this section, I report the VOT results for all three dialects (Najdi, Hijazi, Tunisian) in their MSA production. In normal (slow) speech only, I investigate the VOT behaviour of the emphatic contrast (/t ~ /ṭ/ and /d ~ /ḍ/) in MSA compared to the VOT behaviour in the dialect (previous chapter 4).

5.4.1.1 MSA VOT

In the participants’ MSA production, for all three dialects, raw VOT values for the four plosives (/t ~ ṭ/ and /d ~ ḍ/) are illustrated in Figure 5-3 below. From the figure below, we can see a clear separation between the phonologically voiced and voiceless plosives in MSA, as in their dialects. The voiced plosives /d/ and /ḍ/ show voicing lead, and the voiceless plosives /t/ and /ṭ/ show voicing lag.

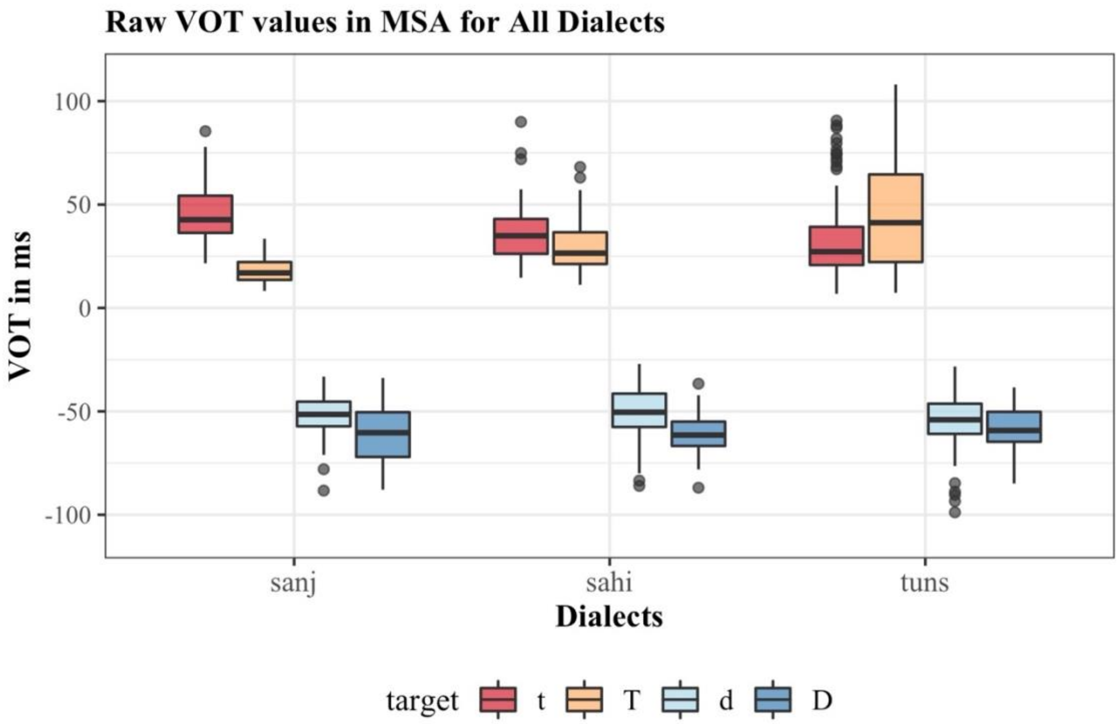


Figure 5-3 Median and interquartile range for raw VOT in MSA of Najdi, Hijazi, and Tunisian in /t/, /ṭ/, /d/ and /ḍ/ by dialect in slow speech-rate only.

As expected, the extent of voicing lag appears to vary among dialects. As we saw in the dialect production, the degree of VOT difference between MSA /t/ and /ṭ/ is not the same in all dialects. The Najdi MSA production mirrors their dialect production, where there is a three-way distinction of VOT (voiced, voiceless emphatic, voiceless plain). In contrast, the VOT of Hijazi and Tunisian MSA production is a little different. Hijazi speakers show more merged values of VOT in the voiceless subset /t/ and /ṭ/ compared to their dialect production, and Tunisian VOT in MSA production appears to be the opposite to what they produced in dialect speech. Although /t/ and /ṭ/ VOT values in Tunisian MSA are merged as well, the emphatic /ṭ/ shows greater spread of values than the plain /t/. Thus, Hijazi and Tunisian show a pattern of two VOT categories of voiced and voiceless in MSA. The gradual increase in overlap between values for /t/ and /ṭ/ starting from Najdi to Tunisian is accompanied by greater spread of values in the emphatic consonant /ṭ/, while the spread of values for the plain /t/ are visually consistent across the three dialects.

To investigate this variation statistically, we ran a model for the slow subset of the MSA data (N = 692) using linear mixed-effects models. The dependent variable in the selected model is *VOT*, and the fixed factors included *dialect*, *target*, *gender*, and *position*. The model accounted for the interaction between *dialect* * *target*. Random intercepts for *item* and *speaker* are considered, in addition to a random slope of *score*[55] by *speaker*[56]. The categorical factors in this model are sum-coded to center them around the mean. The coding of the categorical predictors with sum-coding is as follows: *gender* (female = 1, male = -1), *position* (initial = 1, medial = -1), *group* (two predictors of *sanj* ~ *tuns*, and *sahi* ~ *tuns*, in which *sanj* was set to 1, *tuns* to -1, and *sahi* to 0), and *target* (three predictors of /t/ ~ /ḍ/, /ṭ/ ~ /ḍ/, and /d/ ~ /ḍ/ in which the first listed segment was set to 1, /ḍ/ to -1, and all unlisted segments to 0).

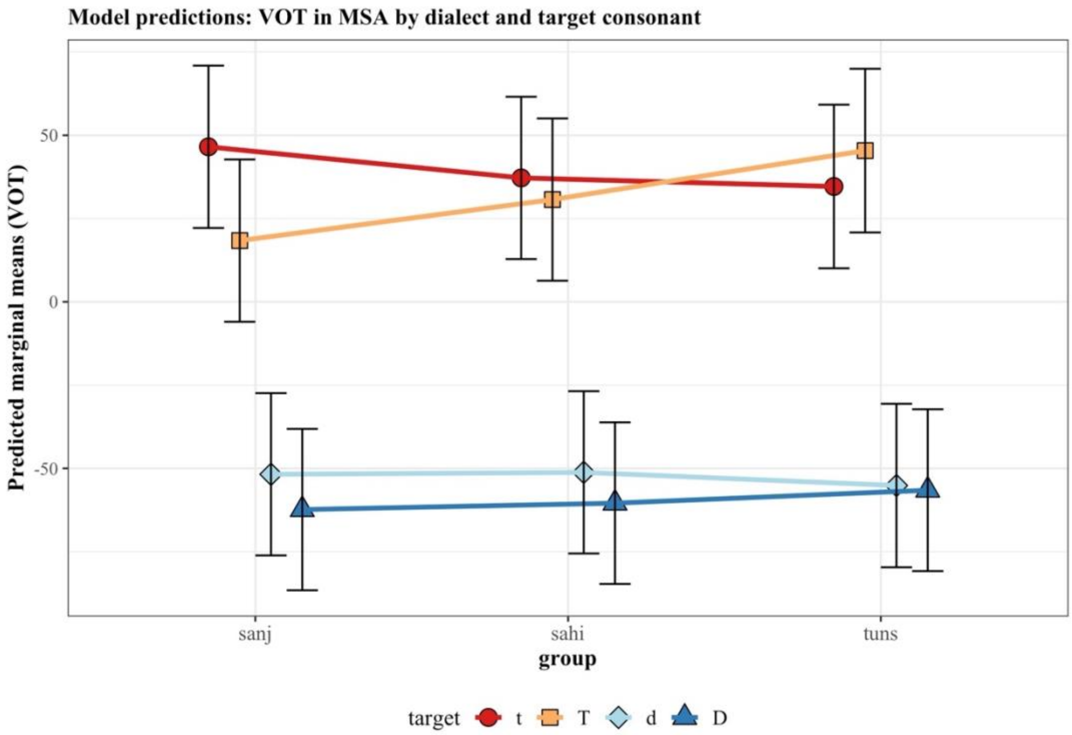


Figure 5-4 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian VOT in MSA by dialect and target.

The linear mixed-effects model on the slow subset revealed no main effect of *gender* (β = 0.86, *t* = 1.14, *p* = .258) nor of *position* (β = -1.55, *t* = -0.50, *p* = .635). The model

[55] *Score* refers to the results of the researchers' evaluation of the participants' reading proficiency in their MSA production discussed in § 5.2.1.2. Score is an evaluation scale from (1-10), it is a between participant and within item factor (score | speaker).
[56] Model used: vot ~ dialect * target + gender + position + (1 | item) + (1 + score | speaker), data = MSA (slow)

showed a significant two-way interaction of *dialect* (*group*) and *target* in some of the consonants. Among the voiceless plosives, Najdi dialect shows the most difference in values between /t/ (*sanj* x */t/*: $\beta = 9.00$, $t = 7.09$, $p < .001$) and /ṭ/ (*sanj* x */ṭ/*: $\beta = -11.19$, $t = -8.78$, $p < .001$). After Najdi, comes Tunisian, but the Tunisian distinction is in the opposite direction, as seen in Figure 5-4. Tunisian /t/ has significantly lower VOT than the average (*tuns* x */t/*: $\beta = -7.30$, $t = -6.33$, $p < .001$) while VOT in /ṭ/ in Tunisian is significantly higher than the average (*tuns* x */ṭ/*: $\beta = 11.44$, $t = 9.91$, $p < .001$). This translates to Tunisian /t/ having the shortest VOT values compared to the other dialects, while Najdi /t/ has the longest VOT values in MSA production. It was only Tunisian that showed a two-way interaction between *dialect* * *target* in the voiced plosives (*tuns* x */d/*: $\beta = -4.91$, $t = -4.21$, $p < .001$) in which VOT in Tunisian /d/ is longer (more negative values) than the other dialects. The interaction of Hijazi *dialect* with all four *target* consonants did not show any significant effect. The model results summary is provided in Appendix (D.5).

The results for the plain and emphatic voiceless plosives in Tunisian MSA are not as expected, so I was intrigued to investigate further. On closer inspection, this proved to be an effect of the target's environment, as shown in Figure 5-5 below. The raw values of VOT at slow rate only, grouped by the following vowel, show a clear effect of the target's environment. It is evident that the voiceless emphatic contrast /t/ and /ṭ/ behave as expected before the vowel /a:/; we see a tight spread of values for the emphatic /ṭ/, and obvious overlap between /ṭ/ and /t/, but the plain /t/ has higher values of VOT than the emphatic /ṭ/. Before the vowel /i:/, there is still overlap between /t/ and /ṭ/, but the spread of values is greater. However, in the last observed environment labelled (x), i.e., [__C], the targets were elicited in coda position followed by a consonant (/t/ = /ʔat.riba/, /ṭ/ = /ʔaṭ.ja:f/). This environment contributes most to the observed 'reversed' results of /t/ and /ṭ/.

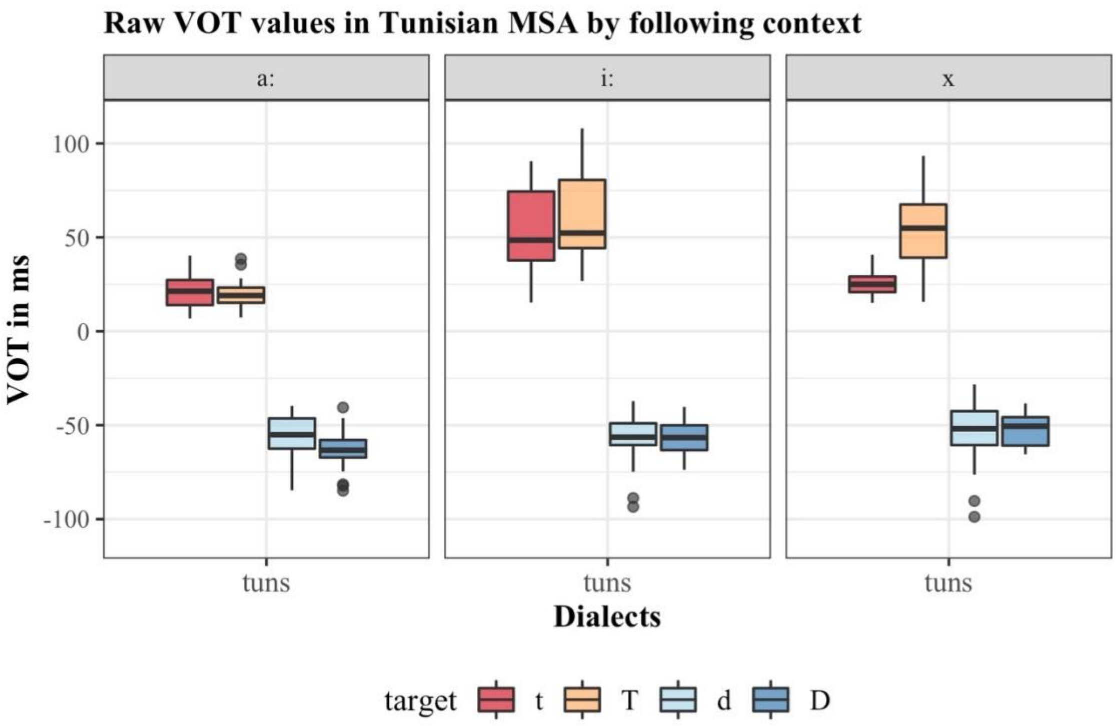


Figure 5-5 Median and interquartile range of raw VOT in MSA in Tunisian in /t/, /ṭ/, /d/, and /ḍ/ by the following context.

The greater spread of values before /i:/ and the reversal effect in the [_C] environment can be explained. In the context of /i:/, the production of VOT is confounded with palatalization for both /t/ and /ṭ/. This context is reported to trigger palatalization for Cairene Arabic (Youssef, 2015), which is also observed in the Tunisian data, resulting in the wide range of VOT values for both targets. In Figure 5-6 below, in two examples of segmentation of the voiceless emphatic /ṭ/ in the word /ṭi:n/, palatalization is clear in the context of a high front vowel /i:/ for the Tunisian speaker (top figure) not for the Najdi speaker (bottom figure).

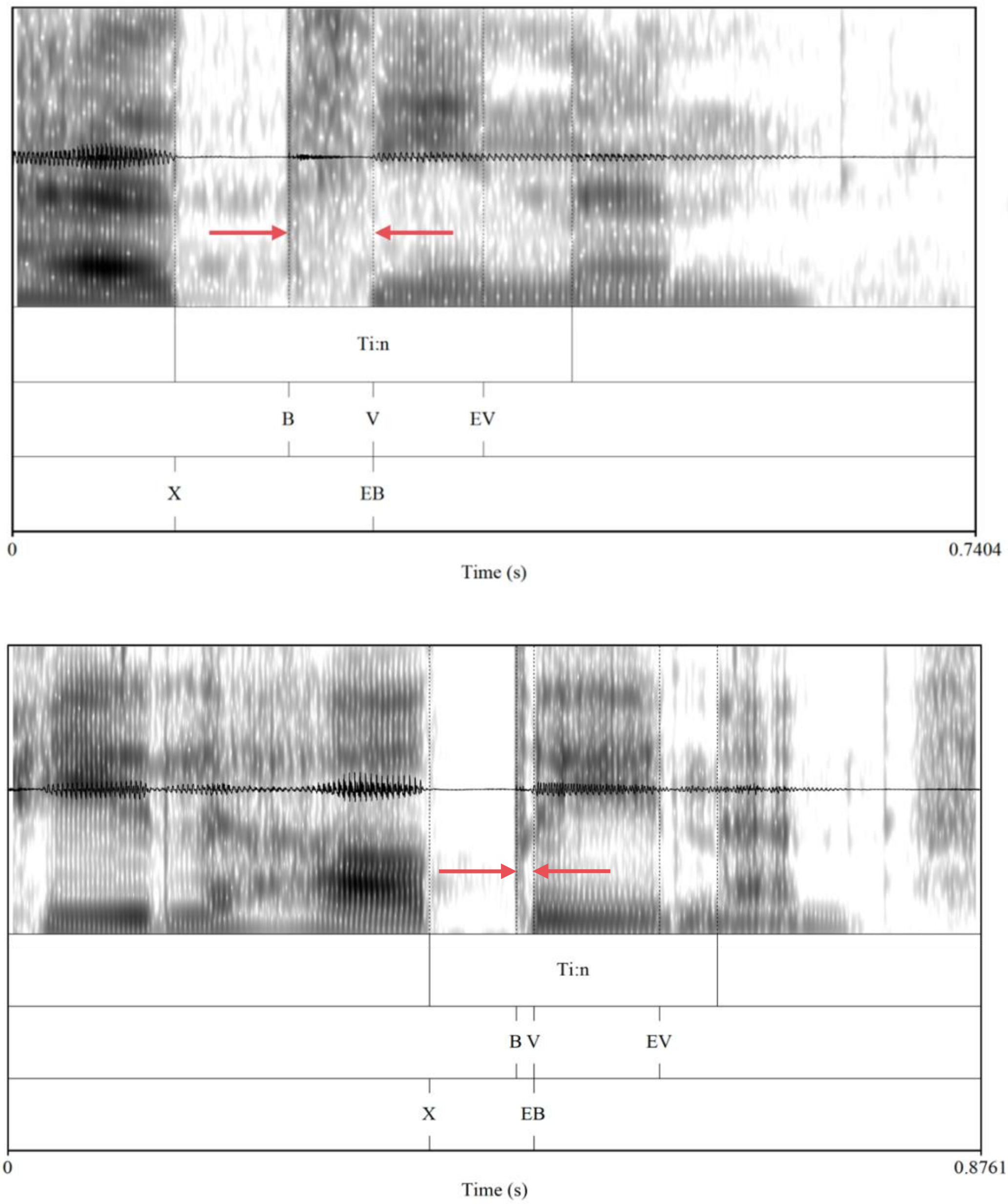


Figure 5-6 Segmentation of MSA production of /ṭi:n/ by a Tunisian female speaker (top) and a Najdi female speaker (bottom), with tiers (1) word, (2) VOT points and end of vowel "EV", and (3) closure duration. VOT between red arrows (Tunisian = 65 ms, Najdi = 16 ms).

In addition to the context before a front high vowel, Youssef (2015) notes palatalization in the context of a glide. Our data has this environment following /ṭ/ only, in /ʔaṭ.ja:f/, but not /t/. Direct comparison between the two targets /t/ and /ṭ/ in the [_C] context is thus invalid. The plain /t/ in /ʔat.riba/ is released into an alveolar trill /r/ which accounts for the tighter values of /t/ in the [_C] context. In contrast, the emphatic /ṭ/ in /ʔaṭ.ja:f/ is released into a palatal approximant /j/ which causes the palatalization effect that is translated into longer VOT values in this data. The apparent reversal in Tunisian, then, is an artefact of the items and differences in the context.

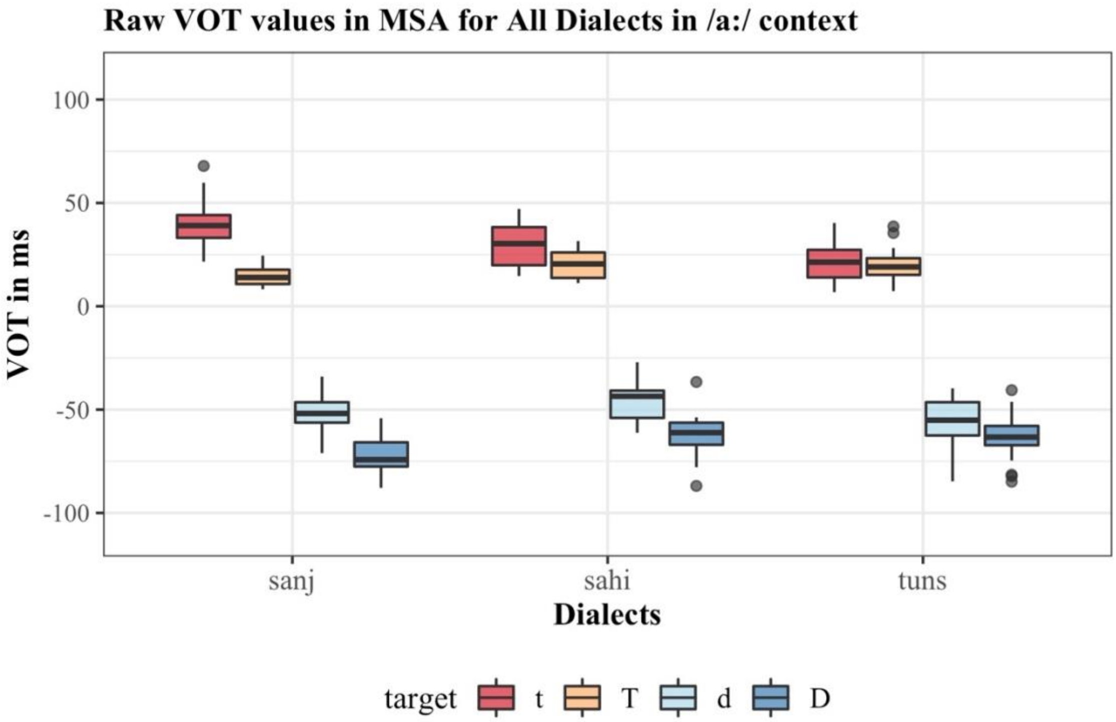


Figure 5-7 Median and interquartile range for raw VOT in MSA of Najdi, Hijazi, and Tunisian in /t/, /ṭ/, /d/ and /ḍ/ followed by /a:/, by dialect in slow speech-rate only.

As we saw in the chapter 4, and as seen in Figure 5-7, the conservative dialect (Najdi) shows a clear split between the short lag values of VOT in MSA, average values of /ṭ/ is 15 ms, and the long lag values of VOT in MSA, average values of /t/ is 36 ms. Whereas progressive dialects (Tunisian) display more merged VOT values of /t/ (21 ms) and /ṭ/ (19 ms), centring around 20 ms. The hybrid dialect (Hijazi) still shows intermediate values in MSA production, in that the average Hijazi MSA voicing lag values of /t/ is 28 ms and the average values of /ṭ / is 21 ms[57].

5.4.1.2 MSA F2 lowering /t/ ~ /ṭ/

In this section, we explore F2 values in vowels following the plain and emphatic contrast in the voiceless subset since it is the source of variation in VOT as we have seen in the previous section. We examined F2 in two vowels /i:/ and /a:/ following targets in word initial position, to confirm the lowering effect of F2 after an emphatic

[57] The average voicing lag values reported in this paragraph are from one environment, followed by /a:/, to avoid the palatalization effect present in the Tunisian dialect.

contrast, since VOT is no longer a consistent phonetic cue available to distinguish the plain from the emphatic in the voiceless subset for Hijazi and Tunisian MSA.

In this dataset, we obtained dynamic formant measurements at three points in the vowel. Before running linear mixed-effect models, these three points in the vowel (25-50-75) were plotted to visualize and assess which point showed the greatest degree of F2 lowering. The 25% point of the vowel had the greatest degree of lowering as expected and is thus reported here.

In the voiceless subset of the slow rate only in MSA data (N = 233), we explored the degree of F2 lowering in the following vowels using a linear mixed-effect model. The dependent variable in the model was the 25% point of F2 in the following vowel (*f2.25)* and the fixed factors included *dialect*, *target*, *following vowel*, and *gender,* as well as the interaction between *dialect* * *target* * *following vowel*. The categorical factors were sum-coded in this subset, and the model included a random intercept for *item*[58]. The model showed a significant main effect of *dialect* for Tunisian only (*tuns*: β = 1.22, *t* = 2.71, *p* = .007) which indicates that, generally, Tunisian MSA shows higher values of F2 compared to the other dialects. The model did not show a main effect of *gender* (β = 0.01, *t* = 0.28, *p* = .781), nor of *target* (β = 0.33, *t* = 0.63, *p* = .527) meaning that F2 following /t/ is not significantly higher compared to the overall average of values across all dialects. The *target* * *following vowel* interaction was not significant (β = -0.30, *t* = -0.59, *p* = .559). From Figure 5-8 of the predicted marginal mean, apparently there is an overall impact of lowering in the vowel /a:/, more than /i:/. The model also showed a significant three-way interaction between *dialect*, *target*, and *following vowel*, in one instance; F2 values of Tunisian vowel /i:/ following the plain /t/ are lower than the other dialects (*tuns* x */t/* x */i:/*: β = -9.57, *t* = -2.13, *p* = .035). The MSA dataset was based on N =233 tokens, which is a lot smaller than the dialect speech data in chapter 4 (N =812). The full summary of the model is provided in Appendix (D.6).

[58] Model used: f225 ~ group * target * folv + gender + (1 | item), data = MSA voiceless subset (slow)

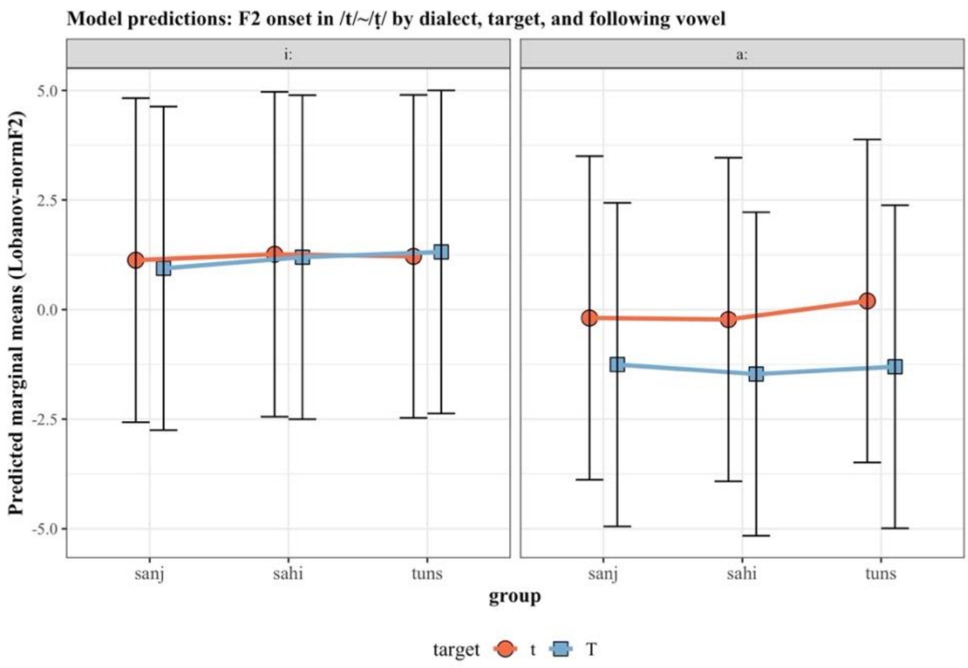


Figure 5-8 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian normalized F2 values of 25% point of vowels following MSA /t/ ~ /ṭ/ by following vowel and dialect.

5.4.2 Results 1 Summary

To summarize, in the slow speech rate subset of the MSA data, Najdi speakers maintained their dialect behaviour with respect to VOT, in which there is a distinction between the voiceless plain and emphatic /t/ and /ṭ/. Meanwhile, Hijazi and Tunisian speakers show a similar distribution of VOT values in /t/ and /ṭ/, but Tunisian values of the emphatic /ṭ/ appear longer than expected and show more spread values than the plain counterpart /t/. These patterns of VOT in Tunisian /t ~ ṭ/ were later clarified after detecting a palatalization effect in two contexts (with following /i:/ and /j/). When the effect of palatalization is removed, Tunisian dialect had overlapping values of /t/ and /ṭ/.

Considering the results for both the voiced and the voiceless plosives, we can conclude that Najdi MSA displays a pattern of three VOT categories of voicing lead /d ~ ḍ/, long lag /t/, and short lag /ṭ/. The results for Hijazi and Tunisian production of MSA indicate that they appear to be two VOT categories dialects, with voicing lead /d ~ ḍ/ and short lag /t ~ ṭ/ only. The Hijazi speakers' production of MSA does not reflect their production of VOT in their vernacular dialect speech. In Hijazi MSA, the values in the plain and emphatic voiceless contrast are more merged than in the dialect productions. The degree of F2 lowering in the vowels following the voiceless emphatic plosive /ṭ/

was also explored and the results showed a visual impact of lowering in the vowel /a:/ in all three dialects.

5.5 Results 2: MSA Speech-rate

In this section, before investigating speech-rate effects on VOT in MSA, it is required to confirm the validity of the speech-rate manipulation in MSA in § 5.5.1. In § 5.5.2 - 5.5.5 we start by exploring speech rate effects on the emphatic contrast in the MSA production of the three dialects collectively, then for each dialect separately.

5.5.1 MSA Word duration

Word duration in milliseconds is used as a proxy for speaking rate in the repeated measures. A linear mixed-effects model was run on the MSA dataset with *word duration* is the dependent variable while *speed* (fast, slow) and *dialect* (Najdi, Hijazi, Tunisian) as the fixed factors, with the interaction between them. The model included random slopes for *score* by *item*[59]. The categorical factors, *speed* and *dialect*, are sum-coded (*speed*: slow = 1, fast = -1). As expected, the model revealed a significant main effect of speech rate ($\beta = 34.21$, $t = 24.08$, $p < .001$), indicating that word duration in slow speech was significantly longer than it is in fast speech in all dialects, as we can see in Figure 5-9. The full model summary is provided in Appendix (D.7).

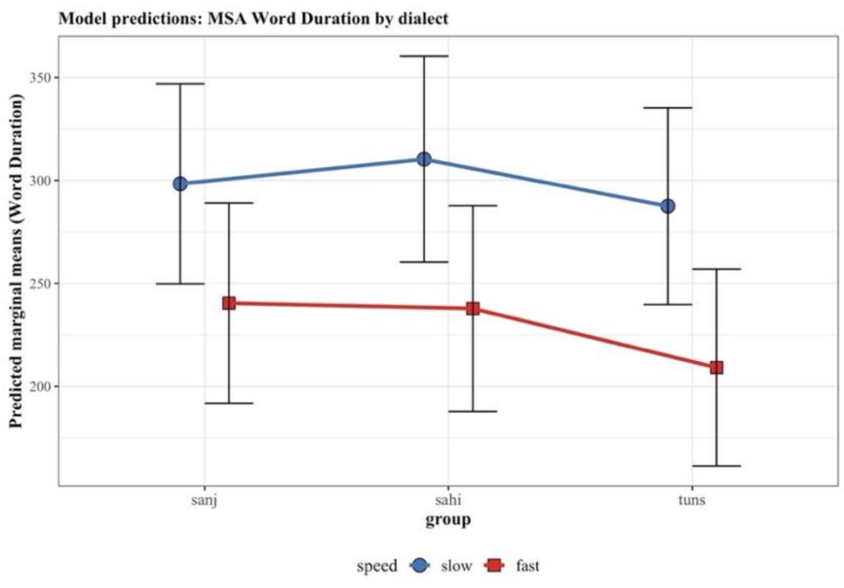


Figure 5-9 Predicted marginal mean (and 95% CI) for Najdi, Hijazi, and Tunisian word duration in MSA by dialect and speech rate.

59 Model used: worddur ~ speed * group + (1 + score | item), data = MSA

The strength of the speech rate effect in this task (MSA) is not as great as in the previous task (dialect: $\beta = 65.76$, $t = 75.73$, $p < .001$). To understand that difference, we calculated the ratio between the means of fast and slow word duration in milliseconds, in both tasks and in all dialects. This revealed a systematic speech rate relation in all dialect groups in the dialect task, i.e., a constant ratio of 1.5:1 between the means of slow and fast word duration in each dialect. However, in the MSA task the ratio between the means of slow and fast is generally lower than in the dialect task, and not consistent across dialects. Speakers do not speed up as much in MSA as they do in dialect production. Table 5-2 below shows the difference in ratio values.

Table 5-2 Ratio between the means of slow and fast word duration in milliseconds in MSA and Dialect tasks.

Dialect	**Task**	
	MSA (slow : fast)	Dialect (slow : fast)
Najdi	1.2 : 1	1.5 : 1
Hijazi	1.3 : 1	1.5 : 1
Tunisian	1.4 : 1	1.5 : 1

This difference is perhaps to be expected and might be due to many factors. There is a difference in the total number of tokens (MSA = 1347, Dialect = 7014), though it is not clear why this in itself might affect the degree of rate change. However, the nature of this current task (MSA) requires more concentration during production as it is a different register that is not used as often, especially in speech production. The speech rate difference was nevertheless deemed sufficient to allow investigation of the effect of speech rate on VOT in MSA.

5.5.2 Speech-rate effects on MSA emphatic contrast: Overview

In this section, we examine the effects of speech rate on VOT in MSA produced by speakers of all three dialects, since the speech rate manipulation attained the desired results overall. Figure 5-10 visualizes raw values of VOT by speech rate in MSA for all the dialects. Again, there is a clear separation between the phonologically voiced and voiceless plosives, with voicing lead in the voiced plosives /d ~ ḍ/ and voicing lag in the voiceless plosives /t ~ ṭ/ across the dialects in both speaking rates.

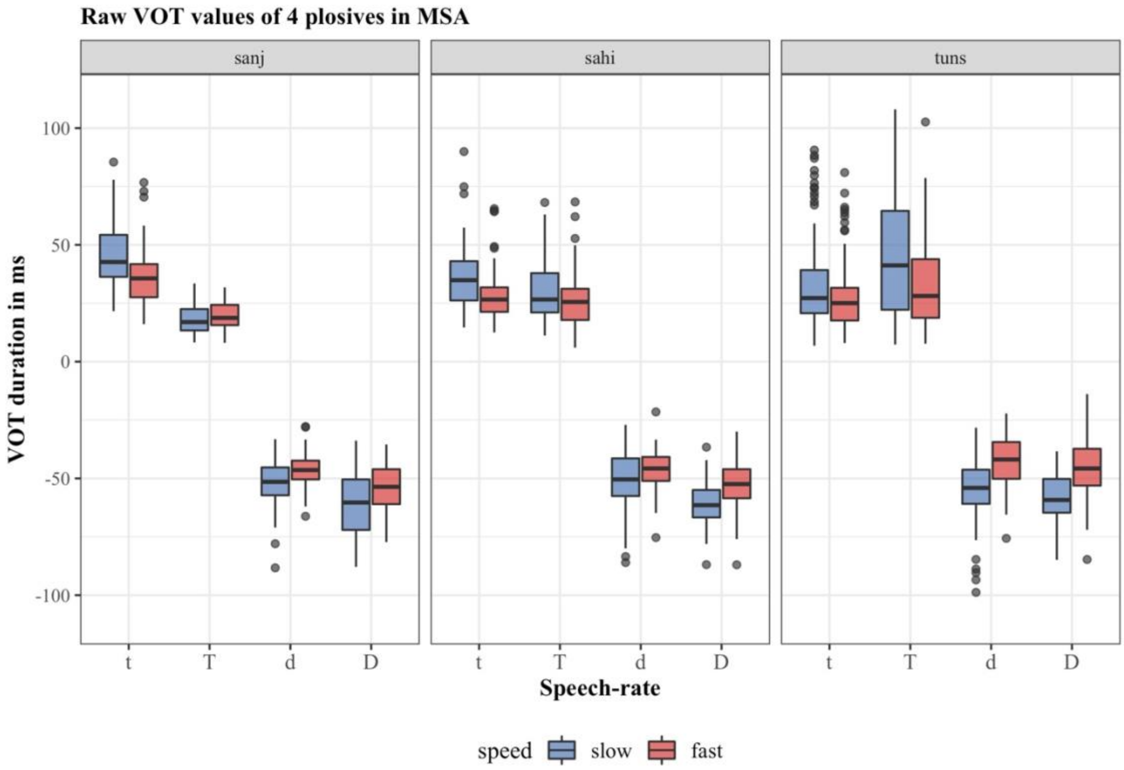


Figure 5-10 Median and interquartile range for raw effect of speech rate on VOT of MSA in Najdi, Hijazi, and Tunisian /t ~ ṭ/ and /d ~ ḍ/.

There appears to be an impact of speaking rate on VOT values for all the voiced plosives across all dialects; both the plain and emphatic voiced plosives /d ~ ḍ/ have longer voicing lead in slow speech. In the voiceless plosives, the impact of speaking rate on VOT in /t/ is also evident across the dialects (though smaller in Tunisian); voicing lag in /t/ is longer in slow speech. In contrast, VOT values in /ṭ/ display little or no impact of speech-rate in Najdi and Hijazi, but there appears to be some visible impact on VOT in the Tunisian /ṭ/. The spread of values of /ṭ/ in Najdi are compact in both speech rates, compared to more spread values of /ṭ/ in Hijazi and Tunisian in both speech rates.

In the coming sections, we will explore the effect of speech rate on the VOT of the emphatic contrasts /d ~ ḍ/ and /t ~ ṭ/ in MSA production for each dialect separately.

5.5.3 MSA in Najdi Arabic

We already established from the 'slow' dataset in § 5.4.1 that MSA productions by speakers of the Najdi variety of Saudi display a three VOT categories pattern. In that data subset, Najdi dialect showed the greatest separation in VOT values of /t/ and /ṭ/ compared to the production of MSA by speakers of the other dialects.

In the speeded dataset for speakers' MSA, we can see that the effect of speech rate on VOT in the MSA voiceless /t/ is the greatest compared to the other investigated dialects as seen in Figure 5-10 above. VOT values in MSA /ṭ/ appear to be less variable (compact spread of values) in both speech rates, in addition to the clear effect of speech rate on the voiced subset.

A linear mixed-effects model was run on the Najdi subset of the MSA data (N =379), to investigate this asymmetry statistically. The model considered *VOT* as the dependent variable with *speed*, *target*, *position*, and *gender* as fixed factors in addition to the interaction between *speed* * *target*. The model included a random slope for *score* by *item*[60]. The categorical factors are sum-coded in this model as follows: *voicing* (voiceless = 1, voiced = -1), *gender* (female = 1, male = -1), *position* (initial = 1, medial = -1), *target* (three predictors of /t/ ~ /ḍ/, /ṭ/ ~ /ḍ/, and /d/ ~ /ḍ/ in which the first listed segment was set to 1, /ḍ/ to -1, and all unlisted segments to 0).

The model did not show a main effect of either *gender* (β = -0.64, t = -1.27, p = .206) or *position* (β = -1.77, t = -0.79, p = .458). There was a significant interaction between *target* and *speed*, as expected, in three targets in Najdi MSA. Firstly, a significant interaction in the plain voiceless /t/ (*slow* x */t/*: β = 5.59, t = 6.68, p < .001) which means that voicing lag in slow /t/ in Najdi is significantly longer than the average. Secondly, a significant interaction in the plain voiced /d/ (*slow* x */d/*: β = -2.00, t = -2.37, p = .018) meaning that voicing lead in slow /d/ is significantly longer (more negative values) than the average. Finally, emphatic voiced /ḍ/[61] in Najdi MSA (*slow* x */ḍ/*: β = -3.85, t = -3.93, p < .001) has the longest voicing lead, compared to the average. However, values of /ṭ/ in Najdi MSA are not affected by speaking rate (β = 0.26, t = 0.32, p = .753). The full model summary is provided in Appendix (D.8).

[60] Model used: vot ~ speed * target + position + gender + (1 + score | item), data = Najdi MSA
[61] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

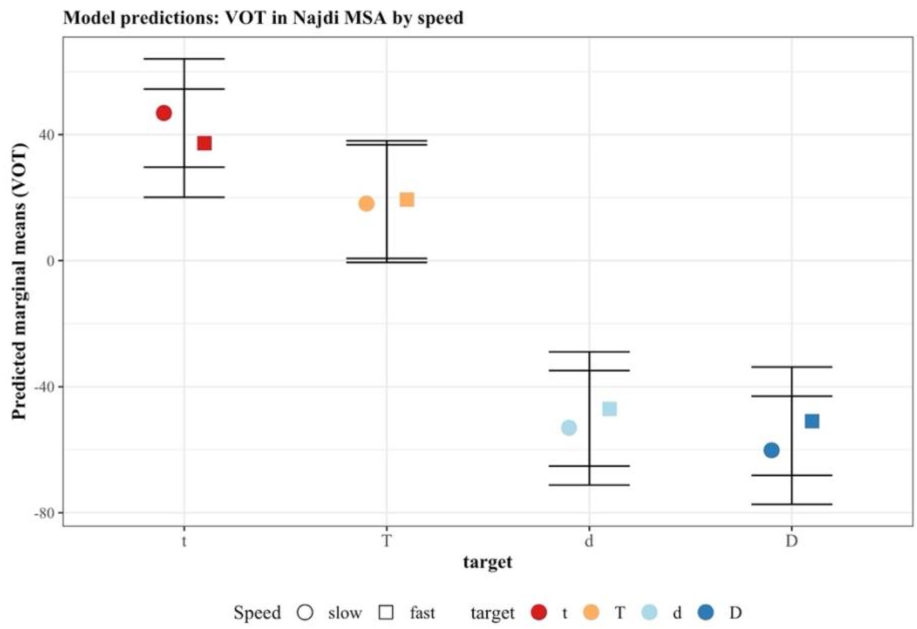


Figure 5-11 Predicted marginal mean (and 95% CI) of VOT in Najdi MSA production of /t ~ ṭ/ and /d ~ ḍ/ by speech rate.

The effect of speaking rate on these plosives is illustrated in Figure 5-11 above which shows the predicted marginal means for VOT in Najdi MSA. We can see that voiced plosives show an inverse relationship between speech rate and VOT; speech rate decreases, voicing lead lengthens by about 2 ms in /d/ and 3.8 ms in /ḍ/. As for the voiceless plosives, the significant two-way interaction between *speed* and *target* is only found for /t/; voicing lag in slow /t/ is estimated to be 5.6 ms longer than in fast /t/. The emphatic /ṭ/ is not affected by speaking rate.

5.5.4 MSA in Hijazi Arabic

For the Hijazi dialect, in § 5.4.1 we concluded that VOT in Hijazi MSA production revealed a greater overlap in values for /t/ and /ṭ/ than seen in the participants' vernacular production of this variety, leading us to conclude that Hijazi MSA has a pattern of two VOT categories of voiced and voiceless. In this section, we investigate the speech rate effect on VOT across the plain and emphatic contrast in the Hijazi MSA plosives.

In Figure 5-10, there appears to be a clear effect of speech rate on the voiced plosives of Hijazi MSA. In voiceless plosives, we can see an effect of speech rate on VOT in the MSA plain /t/, i.e., longer voicing lag in /t/ in slow speech; in contrast, the effect of speech rate on the MSA emphatic /ṭ/ is not so clear in this raw data visualization.

Following a similar statistical model structure used in the Najdi subset of MSA data above, the best fit model for the Hijazi MSA subset (N = 384) yielded a model that considered *VOT* as the dependent variable with *speed*, *target*, *position*, and *gender* as fixed factors in addition to the interaction between *speed* * *target*. The model included a random slope for *score* by *item*, without slope/intercept correlation [62]. The categorical factors are also sum-coded in this model.

This model showed no effect of *gender* (β = 0.55, *t* = 0.87, *p* = .385), or *position* (β = -1.65, *t* = -1.70, *p* = .092). A significant two-way interaction between *speed* and *target* is present in three plosives, albeit in varying degrees. The three plosives are: plain voiceless /t/ (*slow* x */t/*: β = 4.34, *t* = 4.17, *p* < .001), plain voiced /d/ (*slow* x */d/*: β = -2.38, *t* = -2.21, *p* = .028), and emphatic voiced /ḍ/ (*slow* x */ḍ/*: β = -3.76, *t* = -3.27, *p* = .001)[63]. As for the emphatic voiceless /ṭ/ in Hijazi MSA, the interaction of *speed* and *target* is not significant (β = 1.80, *t* = 1.72, *p* = .086). The full model results summary is provided in Appendix (D.9).

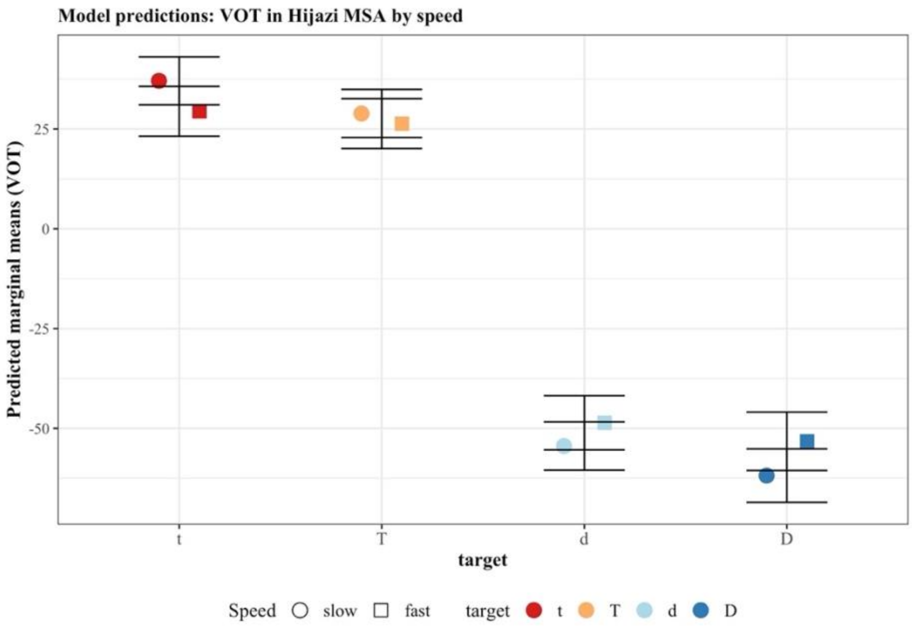


Figure 5-12 Predicted marginal mean (and 95% CI) of VOT in Hijazi MSA production of /t ~ ṭ/ and /d ~ ḍ/ by speech rate.

The model predictions in Figure 5-12 illustrate this varying effect more clearly. In the voiced set of plosives, voicing lead in /d/ is estimated to decrease by 2.4 ms in faster

[62] Model used: vot ~ speed * target + position + gender + (0 + score | item), data = Hijazi MSA
[63] To estimate the held-out factor /ḍ/, we rotated the levels of the model.

speech rate and 3.8 ms for /ḍ/ in faster speech. In the voiceless plosives of MSA, speech rate showed a positive relationship with VOT; as speech rate decreases, voicing lag lengthens by about 4.3 ms for /t/, while in /ṭ/ it is estimated to be only 1.7 ms longer in slow speech compared to fast speech.

This means that the effect of speech rate on VOT of the Hijazi MSA production is significant in three plosives, but the size of the effect varied. The effect of speech rate on the VOT values in the emphatic /ṭ/ is smaller compared to the other plosives in the Hijazi MSA dataset.

5.5.5 MSA in Tunisian Arabic

Earlier in § 5.4.1 we concluded that the Tunisian MSA production of plosives in the slow speech dataset displays a two VOT categories pattern: VOT values of /t/ and /ṭ/ in Tunisian MSA appear to be completely merged. Now in this section we explore the effect of the speech rate variable on plosives in the MSA productions of the Tunisian speakers.

Raw values of Tunisian MSA VOT in Figure 5-10 indicated an effect of speech rate on all of the plosives in Tunisian MSA. In the voiceless subset, it seems that the effect of speech rate on VOT is greater in the MSA emphatic /ṭ/ than in the plain /t/, i.e., longer voicing lag in /ṭ/ in slow speech. The effect of speech rate on the MSA plain /t/, from this raw visualization, seems to be less, which is surprising.

To investigate this statistically, we followed the same approach used in the Najdi and Hijazi subsets of the MSA data, with the categorical factors sum-coded. For the Tunisian subset of the MSA data (N =584), we ran several linear mixed-effect models resulting in the model that considered *VOT* as the dependent variable with *speed*, *target*, *position*, *following vowel*, and *gender* as fixed factors in addition to the interaction between *speed* * *target*. The model included a random slope for *score* by *item*, without slope/intercept correlation[64].

This model showed main effects of *gender* ($\beta = 2.01$, $t = 3.25$, p =. 001), *position* ($\beta = 10.05$, $t = 3.77$, $p < .001$), and *following vowel* ($\beta = -13.90$, $t = -4.54$, $p < .001$). This means that female speakers produced longer values than the average, initial position had longer values than the average, and VOT values following /a:/ are shorter than the

[64] Model used: vot ~ speed * target + position + gender + folv + (0 + score | item), data = Tunisian MSA

average (might be due to the palatalization effect in context of /i:/ and /j/). Most importantly, there was a significant interaction between *speed* and *target* in all four plosives as expected. However, the size of the interaction effect is greatest in the emphatic voiceless (*slow* x */ṭ/*: β = 6.35, *t* = 6.19, *p* < .001), and it is the least in the plain counterpart (*slow* x */t/*: β = 4.04, *t* = 3.96, *p* < .001), which is the opposite of what we might expect. In the voiced subset, both targets show a significant interaction: plain voiced (*slow* x */d/*: β = -5.33, *t* = -5.11, *p* < .001) and emphatic voiced (*slow* x */ḍ/*: β = -5.06, *t* = -4.04, *p* < .001)[65]; the effect was more prominent in the plain /d/. The full model summary is provided in Appendix (D.10).

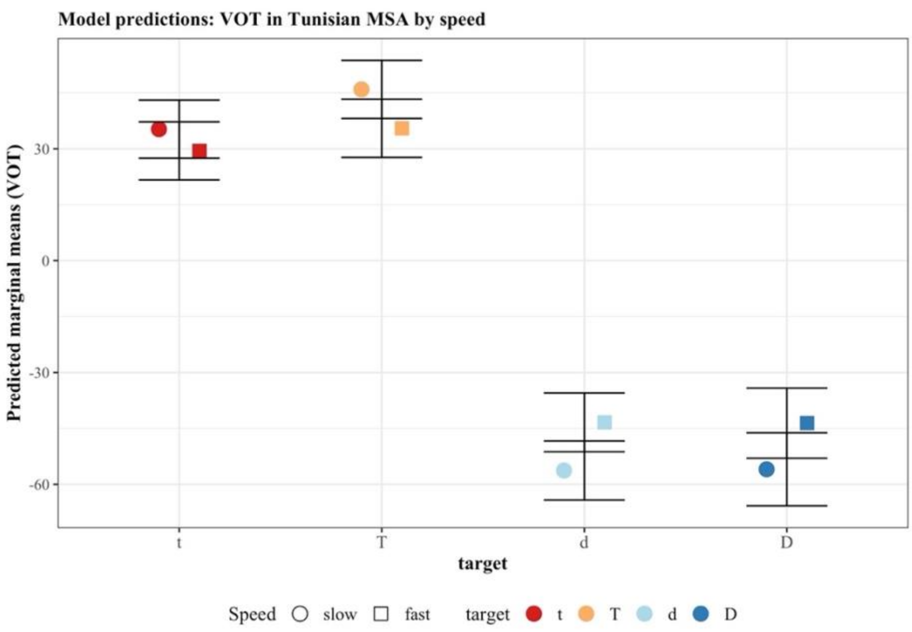


Figure 5-13 Predicted marginal mean (and 95% CI) of VOT in Tunisian MSA production of /t ~ ṭ/ and /d ~ ḍ/ by speech rate.

From the model predictions in Figure 5-13, speech rate showed an inverse relationship with VOT in the voiced plosives in the Tunisian subset of the MSA data: as speech rate decreases voicing lead lengthens by about 5.3 ms in /d/ and 5 ms in /ḍ/. In voiceless plosives, voicing lag in slow /t/ is estimated to be 4 ms longer than fast /t/, while in /ṭ/ it is estimated to be 6.4 ms longer in slow speech.

Knowing the potential effect of palatalization from § 5.4.1.1, we ran an additional model[66] to explore the interaction of speech-rate, target, and following context. The

[65] To estimate the held-out factor /ḍ/, we rotated the levels of the model.
[66] Model used: vot ~ speed * target * folv + position + gender + (0 + score | item), data = Tunisian MSA

purpose of this model is to see the effect of the interaction with the following context. Figure 5-14 shows predictions of this new model, to visualize the effect of following context on the production of VOT in Tunisian MSA /t/ and /ṭ/ in both speaking rates. The effect of speech rate on the voiceless emphatic contrast varies according to the following environment. There seems to be a minimal effect of rate on both voiceless plosives /t/ and /ṭ/ if followed by /a:/, while in the other two contexts there is an effect of speaking rate. The full model summary is reported in Appendix (D.11).

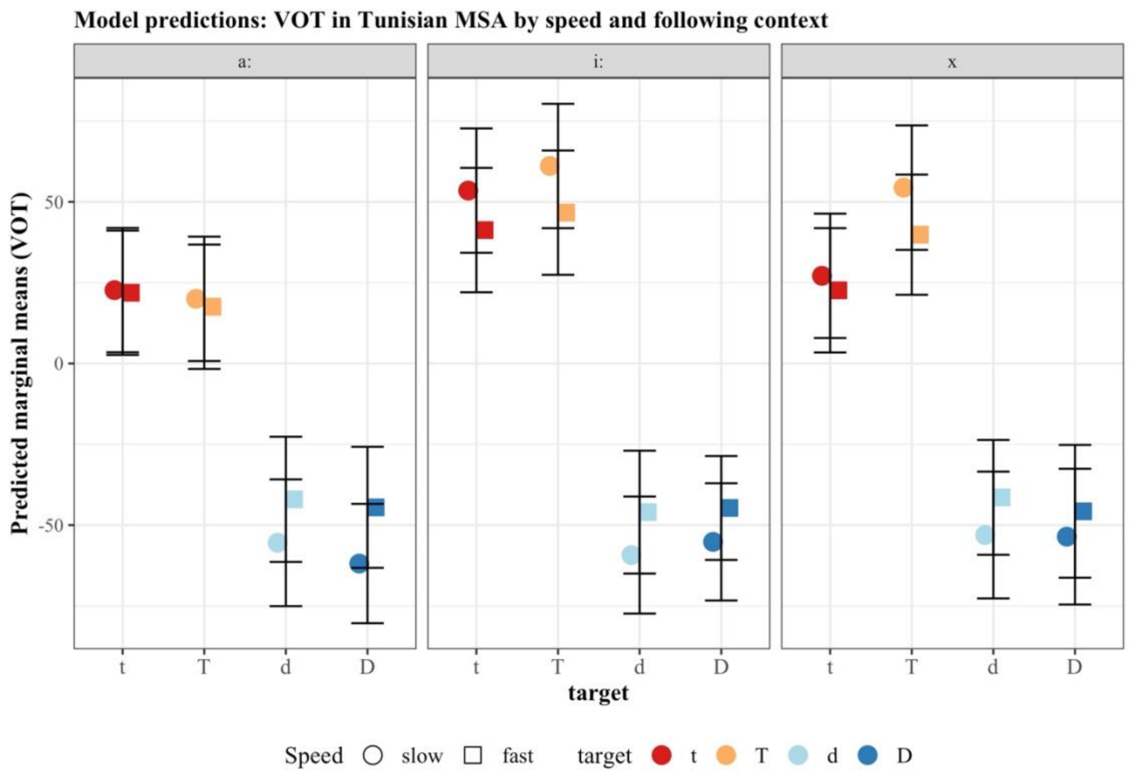


Figure 5-14 Predicted marginal mean (and 95% CI) of VOT in Tunisian MSA production of /t ~ ṭ/ and /d ~ ḍ/ by speech rate and following context.

The effect of speech rate on VOT was significant in all four plosives of the MSA production of the Tunisian dialect, but the effect size varied. The effect of speech rate on the VOT of the plain /t/ is smaller than the effect on its emphatic counterpart, and this result is not expected. This reversal behaviour is explained when we considered the following context; the palatalization effect when followed by a front vowel /i:/ and a glide /j/ caused this spread of VOT values in the emphatic /ṭ/.

5.5.6 Results 2 Summary

This section aimed to uncover potential asymmetric effects of speaking rate on VOT in the MSA production by speakers of the chosen dialects. This is relevant to the question of whether or not the phonological and/or phonetic properties of MSA are the same as those in a speakers' dialect. In this section, I summarize the results of rate effect on VOT across the emphatic contrast, in the MSA production in all three dialects.

Voiced and voiceless plosives in all three investigated Arabic dialects showed no overlap; the VOT values in voiced plosives in MSA display voicing lead with -13 ms as the smallest voicing lead value; the VOT values in voiceless plosives show clear voicing lag with the lowest value of 6 ms across all the dialects in both speaking rates.

The results confirmed that there is indeed an asymmetric effect of speaking rate on VOT in the voiced and voiceless plosives in the MSA production of speakers of different Arabic dialects. The effect of speaking rate is systematic in the voiced plosives, but smaller and more variable in the voiceless plosives. Both plosives in the voiced emphatic contrast /d/ ~ /ḍ/ in Najdi, Hijazi, and Tunisian MSA are affected by speech rate such that voicing lead in /d/ and /ḍ/ in MSA is longer in slower speech than in faster speech.

In contrast, in the voiceless plosives in MSA, the extent to which the emphatic /ṭ/ is affected by speech rate is different in the three dialects. Najdi is the only dialect where the MSA production of its voiceless emphatic is not affected at all by speech rate (*sanj slow* x */ṭ/:* $\beta = 0.26$, $t = 0.32$, $p = .753$), but the most affected among the dialects is Tunisian (*tuns slow* x */ṭ/:* $\beta = 6.35$, $t = 6.19$, $p < .001$) even more than its plain counterpart, in the full dataset, (*tuns slow* x */t/*: $\beta = 4.04$, $t = 3.96$, $p < .001$). The Hijazi production of the emphatic in MSA was in the middle (*sahi slow* x */ṭ/*: $\beta = 1.80$, $t = 1.72$, $p = .086$). MSA production of the plain /t/ in this full dataset showed a significant effect of speech rate across all three dialects; voicing lag in /t/ is longer in slower speech. Recall, however, that Tunisian results for the voiceless plosives in MSA showed a consistent effect of palatalization in two contexts, before /i:/ and /j/. Excluding these items affected by palatalization leaves us with one reliable following context, /a:/, to assess the true degree of speech rate effect on VOT. The results of the model predictions in the Tunisian subset in Figure 5-14 in § 5.5.5 show that, in the context of

/a:/ only, speech rate effects are only visible on the voiced plosives of the data. VOT in voiceless plosives /t/ and /ṭ/ preceding /a:/ are not affected by speech rate.

Excluding contexts that may lead to palatalization in some dialects (i.e., /i:/ and /j/ in the MSA dataset) makes it possible to explore the behaviour of VOT under the influence of speaking rate in the same plosives more clearly. Figure 5-15 illustrates the predictions of a model[67] run to explore speech-rate influence on VOT in the MSA production of /t/, /ṭ/, /d/, and /ḍ/ in only one context, with following /a:/, in the three investigated dialects.

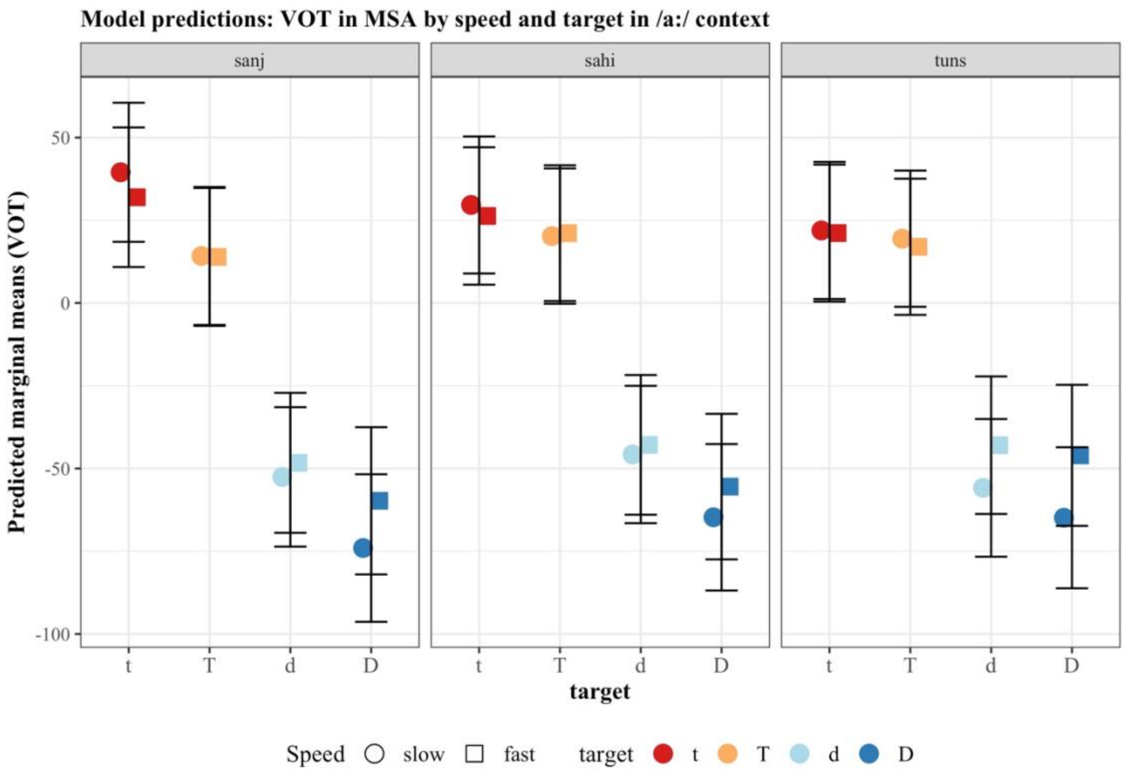


Figure 5-15 Predicted marginal mean (and 95% CI) of VOT in MSA production in Najdi, Hijazi, and Tunisian /t ~ ṭ/ and /d ~ ḍ/ by speech rate in one following context /a:/.

The prediction plots of VOT values with following /a:/ are behaving as expected. Both voiced plain and emphatic plosives /d/ ~ /ḍ/ are influenced by speaking rate in all three dialects, whereas the voiceless /t/ and /ṭ/ have varying degrees of speech-rate influence. The voiceless plain /t/ is clearly affected by speech rate in Najdi and Hijazi, however, the Tunisian /t/ is not influenced by speaking rate. The voiceless emphatic /ṭ/ does not show a clear influence of speaking rate in any of the three dialects. The full model summary is provided in Appendix (D.12).

[67] Model used: vot ~ dialect * speed * target + gender + (0 + score | item), data = MSAaa (N = 462)

In summary, the results demonstrated that the effect of speaking rate on the emphatic contrast in MSA is asymmetric as well. This effect is consistent for the voiced plain and emphatic plosives /d/ ~ /ḍ/ as well as the voiceless plain /t/ in all three dialects, whereas the voiceless emphatic plosive /ṭ/ had varying degrees of speech rate effect on VOT; Tunisian dialect showed the most effect (arising due to a palatalization effect) and Najdi and Hijazi dialects showed no effect of speaking rate. Putting the palatalization effect aside, the Tunisian results in one following context /a:/ revealed no effect of speaking rate on either of the voiceless plosives.

5.6 MSA General discussion and conclusion

Chapters 3 – 4 discussed the phonetic implementation of the laryngeal contrast in Arabic dialects. We concluded that there are differences in the mapping of the surface number of VOT categories to the phonological laryngeal contrast across Arabic dialects, in line with prior claims in the literature (see chapter 2). However, at the end of chapter 4, we arrived at the conclusion that Najdi, Hijazi, and Tunisian share the same phonological specifications, despite having different numbers of surface VOT categories, after inspecting speech rate effects on VOT in these three Arabic dialects. Speaking rate was used as a diagnostic tool to link the active acoustic behaviour to a phonological feature. Adopting the phonological representations of Honeybone's (2005) Laryngeal Realism, using privative rather than binary phonological features, we concluded that, regardless of having phonetically two (Tunisian) or three (Najdi and Hijazi) VOT categories, [voice] and [spread glottis] are both active phonological features in these three Arabic dialects. This over-specification of phonological features in the Tunisian dialect does not adhere to the 'principle of economy' in phonological representation (Clements, 2003). A similar over-specification effect is also observed in other languages e.g., Swedish (Beckman et al., 2011).

In this chapter, by following the same methodology of the previous chapter to collect parallel data in MSA, a different speech register of Arabic, the goal was to investigate the phonetic and phonological properties in MSA. We proposed two questions for this chapter in § 2.8: **3a)** Are the observed VOT patterns in dialect speech mirrored in MSA produced by the same speakers? **3b)** Do speakers of a dialect display the same active feature(s) in the two registers of Arabic? We addressed these research questions by investigating the effect of speech rate on VOT in MSA /d/ ~ /ḍ/ and /t/ ~ /ṭ/. The results in § 5.4 dealt with VOT behaviour across the emphatic contrast of MSA in three dialects at normal speaking rate. It showed that, phonetically, Najdi productions of MSA reflected their dialect speech with voicing lead in the phonetically voiced /d/ ~ /ḍ/ and a distinction between voiceless plain /t/ and emphatic /ṭ/ with long lag and short lag respectively (suggesting three VOT categories overall). The Hijazi data, however, did not show the same number of VOT categories in MSA as in dialect. Hijazi showed voicing lead in their phonetically voiced /d/ ~ /ḍ/ but, unlike the dialect speech, there was an overlapping distribution of VOT values for the voiceless plain and emphatic /t/ ~ /ṭ/ (suggesting two VOT categories overall). The Tunisian productions of MSA reflected their dialect speech, showing voicing lead in the phonetically voiced /d/ ~ /ḍ/, and an overlapping distribution of VOT values in the voiceless plain and emphatic /t/ ~ /ṭ/ (suggesting two VOT categories overall). The Tunisian MSA productions of emphatic /ṭ/ overlap with /t/ more than they do in Hijazi. All three dialects distinguish the plain and emphatic contrast in MSA through F2 lowering in the following vowel /a:/.

The section that followed, § 5.5, revealed that, as in dialect speech, the effect of speaking rate on VOT in MSA is asymmetric. In all the investigated dialects, speech rate affected VOT in MSA voiced plain /d/ and emphatic /ḍ/, which show longer voicing lead at slower speaking rate. This can be interpreted as evidence of an active [voice] feature in the Najdi, Hijazi, and Tunisian MSA register phonological systems. Speaking rate also appeared to affect the voiceless plain /t/ in all three dialects in the full dataset, with longer voicing lag at slower speaking rate, which would support the conclusion that there is an active [sg] feature in all three dialects also. However, the voiceless emphatic /ṭ/ showed varying degrees of speech rate effect on VOT. Najdi and Hijazi showed little or no effect of speaking rate on the voiceless emphatic /ṭ/ indicating it is unspecified with a laryngeal feature; Tunisian, however, appeared to show a very different pattern, but this was found to be caused by palatalization in the context of /i:/ and /j/. In Figure 5-15, we eliminated the noisy effect of palatalization by focusing on the emphatic contrast in following context /a:/, to have a true picture of the effects of speaking rate on these plosives in all three dialects. This resulted in the conclusion that Tunisian showed no effect of speaking rate on either of the voiceless plosives /t/ and /ṭ/. If we consider only the results in the context of /a:/, then Tunisian, which displays two surface VOT categories, would be specified with only one active phonological feature [voice]. Table 5-3 summarises and compares the results for both dialect speech and the MSA in all the dialects.

Table 5-3 Results Summary: comparison of the results in the two tasks (dialect and MSA) in terms of their phonetic number of VOT categories and the possible phonological representations. The asterisks * indicate potential phonological 'over-specification'.

Task	**Dialect**		**MSA**	
Dialect	Phonetics	Phonology	Phonetics	Phonology
Najdi	3	[voice] [Ø] [sg]	3	[voice] [Ø] [sg]
Hijazi	3	[voice] [Ø] [sg]	2	[voice] [Ø] [sg]*
Tunisian	2	[voice] [Ø] [sg]*	2	[voice] [Ø]

Differences between dialects in the number of surface VOT categories and active phonological features are detected in both dialect and MSA speech. However, we should be careful in the interpretation by noting some differences in the two tasks. Firstly, even though the speech rate manipulation achieved its intended purpose in MSA, the magnitude of rate difference in MSA vs. dialect speech is not the same. As was seen in Table 5-2 in § 5.5.1, the ratio of slow to fast word duration is greater in dialect than it is in MSA. Since the duration difference is smaller in MSA, this might have translated to a smaller influence of rate on VOT in MSA. Secondly, although the data used in chapters 4 and 5 are from the same speakers, the total number of tokens in the dialect task for all three dialects used in chapter 4 is (N = 7015) while the total number used in this chapter is (N = 1512), so the MSA speech dataset is smaller. Finally, recalling the diglossic situation in Arabic, speakers typically use MSA less in daily life hence they may be more careful in their MSA production, especially since the passage they read contained overt diacritics which are known to result in slower production (Al-Fahid & Goodman, 2008; Hermena et al., 2015).

Bearing these potential limitations in mind, and following the same procedure as in the discussion of chapter 4, Table 5-4 compares the mean VOT values of the two tasks, and provides a Ratio column to examine the rate effect as a proportion of slow to fast. It is important to be cautious in interpreting the data, due to the variability in the speaking rate in the two tasks, as we noted above. However, we notice some correspondence between the two tasks, at least in the short lag category, across the dialects. On average, short lag plosives in MSA (i.e., emphatic /ṭ/) vary across speech rates between 1 and 3 ms VOT; this amount is not considered as 'movement' in the previous chapter 4 nor by

Kessinger & Blumstein (1997), seen in Table 4-9. The mean difference of the, presumably, long lag category in MSA Najdi and Hijazi (i.e., plain /t/) stands on the threshold of difference with an average of 9 – 7 ms, which is relatively close to the dialect production of plain ‘mid lag’ /t/ in Hijazi and Tunisian.

The Ratio column in Table 5-4 is not very revealing for the MSA dataset. This might be a by-product of having many fewer tokens in the MSA task (see Appendix (D.13) for the number of tokens in each task reported in this table).

Table 5-4 Mean VOTs of different tasks (Dialect and MSA) as a function of speaking rate. Note that MSA Tunisian /t/ and /ṭ/ reported only in the environment of /a:/.

Task	**Language**	**Plosive**	**Slow (ms)**	**Fast (ms)**	**Difference (ms)**	**Ratio (fast/slow)**
Dialect	Najdi	/t/	67	47	20	0.70
		/ṭ/	18	16	2	0.89
		/d/	-66	-50	16	0.76
		/ḍ/	-75	-57	18	0.76
	Hijazi	/t/	42	32	10	0.76
		/ṭ/	24	23	1	0.96
		/d/	-66	-52	14	0.79
		/ḍ/	-81	-61	20	0.75
	Tunisian	/t/	35	26	9	0.74
		/ṭ/	28	25	3	0.96
		/d/	-65	-48	17	0.74
		/ḍ/	-64	-47	17	0.73
MSA	Najdi	/t/	46	37	9	0.80
		/ṭ/	18	20	-2	1.11
		/d/	-52	-46	6	0.88
		/ḍ/	-63	-54	9	0.86
	Hijazi	/t/	36	29	7	0.81
		/ṭ/	30	27	3	0.90
		/d/	-52	-46	6	0.88
		/ḍ/	-61	-52	9	0.85
	Tunisian	/t/	21	21	0	1.00
		/ṭ/	20	17	3	0.85
		/d/	-56	-43	13	0.77
		/ḍ/	-58	-45	13	0.78

Coming back to the main questions addressed in this chapter, since the differences between the three dialects also appear in their MSA production, this could indicate that, yes, these Arabic speakers share the same phonetic implementations in their dialect speech and their MSA, to a great degree. The phonological representations, however, differ. Both Najdi and Hijazi show the same phonological representations in dialect and MSA of having two phonological features active [voice] and [sg], even though Hijazi has one less VOT category in MSA, so it is over-specified in MSA register. Tunisian, however, appears not to mark [sg] as an active phonological feature in MSA production.

In conclusion, this experiment suggests that the phonetic and phonological patterns found in MSA production are not identical to those in dialect vernacular speech. One thing in common between MSA and dialect speech is that, in both registers we find differences between dialects in the number of surface phonetic VOT categories. This matches the expectation of transfer of the acoustic cue of VOT from an Arabic speaker's dialect to MSA. However, the phonological representations in the two Saudi dialects (Najdi and Hijazi) in MSA speech are the same as in dialect speech; they both show two active phonological features [voice] and [sg]. The Tunisian MSA production, however, seems to be not characterized by over-specification, as only the voiced plosives in Tunisian MSA are affected by differences in speaking rate [voice]. In MSA productions, regardless of having two or three VOT categories, the voiceless emphatic /ṭ/ is unspecified in all investigated dialects. Overall, we find that Najdi shows parallel behaviour in the two registers, with three surface VOT categories and two active phonological features in both MSA and dialect speech. The Hijazi pattern is different: VOT values in the voiceless plain and emphatic are more merged in MSA production than in dialect speech, and thus display a two-VOT-categories pattern in MSA. However, the phonological representations in Hijazi remain the same, with two active features, and this is a case of over-specification. Finally, the Tunisian surface VOT categories are the same in the two registers, with a two-VOT-categories pattern, but the phonological representations are not over-specified in MSA. In MSA production, Tunisian acts as a *true voicing language* with only one active phonological feature [voice].

Since Tunisian MSA appears not to mark [sg] as an active phonological feature in MSA productions, this might be due to the same idea proposed in chapter 3 namely of Tunisian possibly undergoing sound change. Interestingly, if we ascribe these patterns to sound change in progress, the process appears to be more advanced in MSA than in dialect speech, for both Tunisian and Hijazi. Why would MSA productions be further along a sound change trajectory than those of dialect speech? In the following general discussion, in chapter 6, we will explore the hypothesis that these patterns in Tunisian and Hijazi can be characterized as evidence of sound change in progress which has resulted in a different representation in the phonology.

Chapter VI

Discussion

6 Discussion

6.1 Summary of the results

In this thesis, I have reported results of three experiments in an attempt to answer questions pertaining to segmental typology in spoken Arabic, specifically in the realm of laryngeal contrasts and phonetic voicing and their interaction with emphasis.

Chapter 3 was concerned with examining the phonetic nature of variation between Arabic dialects in terms of their voicing contrasts. I also shed light on the classification of dialects as having either two or three VOT categories based on the literature on VOT and laryngeal contrasts (Bellem, 2007, 2014). Using a cross-dialectal dataset extracted from the IVAr corpus (Hellmuth & Almbark, 2019), I presented systematic acoustic analyses of multiple cues to the plain and emphatic contrast in Arabic plosives (/t/ ~ /ṭ/ and /d/ ~ /ḍ/). Examining eight Arabic dialects – Omani, Kuwaiti, Iraqi, Jordanian, Syrian, Egyptian, Tunisian, and Moroccan – resulted in a continuum of variation rather than a clear-cut dichotomy. This dialectal variation is mainly noticeable in the degree of voicing lag seen in VOT values of /t/ and /ṭ/. The claim of a sound change from a three-way to a two-way VOT distinction is initially motivated by the general assumption in the Arabic language literature that, diachronically, Arabic had a three-way VOT system (for more details see Bellem (2014)). Gender differences in two dialects in the middle of the continuum, in which male speakers seem to lead the change in merging the voiceless emphatic contrast /t/ ~ /ṭ/, lends further support to the claim that the observed variation can be interpreted as evidence of a sound change in progress.

Chapter 4 was concerned with examining the phonological representation of the voicing contrasts in three Arabic dialects as reflected in VOT of their plosives through the diagnostic tool of speech-rate manipulation. I attempt to compare Arabic dialects through a theory-oriented description of the laryngeal contrast within the Feature Geometry framework. Discovery of an asymmetric effect of speaking rate on VOT is used in previous cross-linguistic studies to argue for the presence of active laryngeal features in the examined languages. If a *true voice language* with two VOT categories (e.g., Spanish) has [voice] vs. [Ø] as a laryngeal feature, and an *aspirating language* with two VOT categories (e.g., English) has [spread glottis] vs. [Ø] as a feature, then faster speech rate only affects the phonetic cue of the ‘active’ phonological contrast (Allen & Miller, 1999; Beckman et al., 2011; Kessinger & Blumstein, 1997; Magloire

& Green, 1999; Miller et al., 1986; Pind, 1995; Summerfield, 1981). Recently, however, some studies reported an 'over-specification' of laryngeal features in a language (e.g., Swedish) that has two VOT categories, in which both laryngeal features [voice] and [sg] are affected by speech rate, thus are active (Beckman et al., 2011; Kulikov, 2020).

I examined multiple acoustic cues in the full Arabic plosive set (voiceless /t, ṭ, k, q/ and voiced /b, d, ḍ, g/), with a focus on VOT, at two speaking rates (fast ~ slow) in three Arabic dialects: Najdi (thee VOT categories), Hijazi (three VOT categories), and Tunisian (two VOT categories). The results indicate that the three dialects have the same number of active (underlying) phonological features while displaying a different (surface) phonetic pattern of VOT categories. In all investigated dialects, all voiced and some voiceless plosives are affected by speaking rate. The rate effect is consistent in voiced plosives across dialects, but varies in the voiceless plosives. Results for Najdi and Hijazi dialect speech are as expected, in that the dialects have a three-way surface VOT distinction and have two active phonological features [voice] and [sg]. In the Tunisian dialect, however, a pattern of 'over-specification' is found similar to that seen in Swedish in Beckman et al. (2011), in which a language/dialect with two surface VOT categories is argued to have two active phonological features [voice] and [sg]. Unlike Swedish, which has no unaffected categories, the Tunisian voiceless emphatic plosive /ṭ/ is not affected by speaking rate. This might indicate that Tunisian, as suggested, still has a three-way laryngeal contrast. Although Arabic dialects can have a three-way or a two-way surface VOT distinction, in all cases investigated here, the voiceless emphatic /ṭ/ is not affected by speaking rate, thus it is deemed to be laryngeally unspecified in all dialects with no phonological feature (represented as [Ø]).

Chapter 5, reporting the final experiment, was also concerned with examining the phonological nature of the voicing contrasts of different Arabic dialects looking at VOT in plosives under speech-rate manipulation. This chapter, however, examined a different register – Modern Standard Arabic (MSA). In a parallel dataset (i.e., from the same speakers), following the same analysis methods, we examined multiple cues to the emphatic contrast in MSA /d/ ~ /ḍ/ and /t/ ~ /ṭ/ at two speaking rates (fast ~ slow). The VOT values in the slow speech rate data in MSA showed that, phonetically, Najdi MSA patterns with Najdi dialect speech in showing a three-way VOT distinction of voiced (voicing lead), voiceless plain (long lag), and voiceless emphatic (short lag). The Hijazi speakers, however, showed a two-way VOT distinction in their MSA productions, with voiced (voicing lead) and voiceless (medium lag), which is unlike their dialect speech. Finally, the Tunisian speakers' MSA patterned with their dialect speech in showing a two-way VOT distinction of voiced (voicing lead) and voiceless (medium lag).

The effect of speaking rate on VOT in MSA in the three dialects showed interesting yet unexpected results. The voiced plosives are affected by speaking rate, and in slower speech have longer voicing lead, in all the investigated dialects. This indicates that the [voice] feature is active in all three dialects. The voiceless emphatic contrast /t/ ~ /ṭ/ had different results in the three dialects. In Najdi MSA, differences in speaking rate affected the plain voiceless plosive /t/ but not the emphatic one /ṭ/. Najdi MSA thus has a three-VOT-distinction surface pattern with two underlying active phonological features [voice] and [sg]. The Hijazi MSA showed similar results when exposed to speech-rate manipulation, i.e., effects of speaking rate on plain /t/ but not the emphatic /ṭ/. Hijazi MSA thus has a surface two-VOT-distinction with two underlying active phonological features [voice] and [sg], hence it is over-specified. Finally, in Tunisian MSA, we concluded that speech-rate manipulation did not affect the voiceless emphatic contrast /t/ ~ /ṭ/ after adjusting for the effect of palatalization in the /i:/ and /j/ environments. Tunisian MSA is thus analysed as having a two-way VOT distinction and only one active phonological feature [voice]. Table 6-1, reproduced from chapter 5, is a summary of the phonetic and phonological implications of the laryngeal contrast of the three investigated dialects in two Arabic registers: dialect speech and MSA.

Table 6-1 Results Summary: comparison of the results in the two tasks (dialect and MSA) in terms of their phonetic number of VOT categories and the possible phonological representations. The asterisks * indicate potential phonological over-specification.

Task	**Dialect**		**MSA**	
Dialect	Phonetics	Phonology	Phonetics	Phonology
Najdi	3	[voice] [Ø] [sg]	3	[voice] [Ø] [sg]
Hijazi	3	[voice] [Ø] [sg]	2	[voice] [Ø] [sg]*
Tunisian	2	[voice] [Ø] [sg]*	2	[voice] [Ø]

The following sections, § 6.2 - 6.5, discuss this summary of the overall results of the thesis. I will propose a theory-driven approach to segmental typology that ties the three empirical chapters together in a multi-layered analysis of surface (phonetic) and underlying (phonological) representations. I will also try to unpack the unexpected finding that a potential sound change in Arabic appears to be more advanced in the

MSA register. I will then discuss Arabic diglossia and its relation to bilingualism in light of the results. Finally, I will sketch a set of phonological representations to account for the observed laryngeal contrast in Arabic dialects.

6.2 Phonological Typology

The study of phonology is inherently typological. The study of laryngeal contrasts and voicing, specifically, is typological in nature; it provides a typology of the contrasts that are found in one system versus another (Hyman, 2014). Phonological theory and phonological typology are both concerned with how languages encode phonetic cues that are similar across languages into sound systems that are structured. In that sense, theory and typology are inseparable, since modern phonological theories are multi-layered and pluralistic in nature (Hyman, 2014; Kiparsky, 2018; Youssef, 2021). A cross-linguistic study, or cross-dialectal study for that matter, is a product of typological inferences. Arabic provides rich grounds for categorizing and typologizing the sound systems of its dialects through many aspects, which follow either geographical, lifestyle (Bedouin vs. sedentary), social, or religious bases. It is also noticeable that theoretically based typologies in Arabic usually treat suprasegmental features, such as syllabification processes and stress placement (Farwaneh, 2009; Kiparsky, 2003; Watson, 2011b).

Youssef (2021) introduced a theory-informed approach to segmental typology, whereby he classified Arabic dialects on the bases of consonant reflexes within the framework of Feature Geometry, and specifically the Parallel Structures Model (PSM) proposed by Morén (2003). Youssef (2021) analysed the varying surface reflexes of each cognate phoneme representationally through contrastive features; he based his theoretical typology on contrastive features for place and manner articulations[68] to describe the cognate phoneme reflexes in the dialects that have a contrastive phonemic status. Table 6-2 shows Youssef's representational typology for the major /q/ cognate phoneme reflexes, as an example of the approach.

[68] The laryngeal tier was accounted for in Youssef's (2021) representations by adopting the [voice] feature to differentiate voiced from voiceless obstruents (only).

Table 6-2 An example of Youssef's (2021, p. 6) representational typology: /q/ reflexes.

	C-place [dorsal]	C-manner [closed]	….. [voice]	**Geographical Distribution**
/q/	✓			Various sedentary: North Africa, Mesopotamian
/ʔ/		✓		Urban Egyptian and Levantine and sporadic Maghrebi
/k/	✓	✓		Ruralite Levantine dialects
/g/	✓	✓	✓	Bedouin(-origin) dialects

The relation of phonetics to phonology is commonly manifested through levels of representation, i.e., connecting the nature of underlying representations to the surface output. Some of the approaches to phonological typology adopt only a single-level, but it is argued to be more insightful to follow an approach that is concerned with how the underlying representations are brought to the surface (Hyman, 2014, p. 107). Hyman (2014) illustrates such an approach through the example of the Ik (Heine, 1999) and Kom (Hyman, 2005) tone systems. According to Hyman (2014), both Ik and Kom have underlying /H, L/ tones, in addition to a third mid tone [M] that surfaces in certain environments. Since these languages have two underlying-contrastive tones /H, L/ and three surface-contrastive tones [H, M, L], how can their height system be described? Typologizing based on the relation between underlying and surface contrastive elements is one way to describe their height system, i.e., Ik and Kom both have a 2→3 tone-height system (Hyman, 2014).

Similarly, an explanatorily adequate way to describe the different patterns found in the Arabic phonological system in terms of their laryngeal contrasts is thus through a theory-driven approach to segmental typology that ties the surface and underlying representations of these dialects. Within the framework of Feature Geometry, I will propose a feature-based typology of the variations found in the investigated dialects. The analysis adopted here, as mentioned earlier in § 2.3.1 and § 5.6, assumes privativity of laryngeal features defined by either the presence or the absence of the feature. I associate the underlying [voice] feature with the visible voicing lead found in voiced plosives and [spread glottis] with the aspiration or long lag found in voiceless plosives;

[Ø] in our analysis refers to short lag plosives, therefore to the non-existence of an underlying laryngeal specification (Beckman et al., 2011; Honeybone, 2005).

In this section, I offer a multi-layered segmental typology[69] based on non-linguistic (lifestyle: Bedouin – sedentary/rural – urban) and linguistic (surface – underlying) factors. Works that have typologized dialectal Arabic consonants according to non-linguistic geographical or lifestyle bases include Cadora (1992), Holes (2004), Versteegh (2014), Watson (2002, 2011a), and Bellem (2007, 2014). Although, as mentioned in § 2.6.1, pure classification of Bedouin versus sedentary is a somewhat simplified generalization, it is worth mentioning that this classification (Bedouin – Sedentary – Urban) refers to the linguistic features of the dialect rather than the current lifestyle of the speakers per se. Bedouin dialects are the dialects that have maintained the linguistic features that are closest to Classical Arabic. Sedentary dialects, however, developed through contact mainly through Bedouin migration and settlement outside the Peninsula before and after Islamic conquests. Urban dialects are the dialects that emerged in areas outside the Arabian Peninsula after the Islamic conquest; these dialects exhibited a "high rate of innovation" (Versteegh, 2014, p. 149) and later gained prestige due to their use in urban centres in the heart of the Islamic civilisation (Bellem, 2007; Cadora, 1992; Holes, 2006; Versteegh, 2014; Watson, 2011a). The lifestyle classification in particular has already been argued to show some typological similarities in terms of laryngeal contrasts (Bellem, 2007; Watson, 2011a). The consensus to date pertaining to laryngeal categories is that dialects of Bedouin origin mainly retain the three-way laryngeal system, whereas dialects that are more of a sedentary origin have shifted towards a two-way laryngeal system. Youssef (2021), to the best of my knowledge, was the first to typologize the place and manner of Arabic consonant reflexes in a theoretical feature-based approach while tying this segmental typology to non-linguistic (lifestyle/geographical) factors.

My findings in chapter 3 lay the foundation for a multi-layered laryngeal typology of Arabic dialects by analyzing the phonetic categories i.e., the surface structure and tying it to the suggested correlation with dialect type, or what we refer to as 'lifestyle'[70]. In

[69] It is important to note that the typology here is not a strict phonological typology, it is more of a substance-directed typology, unlike Youssef (2021) and Hyman (2014) who each offer a strict phonological typology that treats phonetics and phonology as distinct modules.
[70] Along the Bedouin – rural/sedentary– urban continuum, or "ecolinguistic" classification (Cadora, 1992). More in § 2.6.1.

terms of the surface laryngeal categories, as mentioned previously, Arabic dialects were shown not to fall into a dichotomy of an "either/or" two-way versus three-way phonetic implementation of voicing. Rather, the investigated dialects displayed a continuum that could reflect sound change in progress. Indeed, the dynamic nature of sounds in languages allows us to currently describe the phonetic surface level of Arabic dialects as either being (i) in the three-way voicing category (Bedouin), (ii) 'in transit' to the two-way voicing category (Bedouin descended or sedentary), or (iii) in the two-way voicing category (sedentary and urban). This, again, is based on the general assumption of the prototypical Semitic three-way laryngeal system.

In chapter 4, the investigation turned to the underlying laryngeal contrasts. The first step was to establish the facts on the surface layer through VOT values at normal speech rate in the three chosen dialects, with one dialect from each surface phonetic type, to confirm that they fall within this initial surface typology. We can note here the fit to the typology of non-linguistic factors i.e., lifestyle (Bedouin – sedentary – urban)[71] for the three dialects. The phonology of the laryngeal contrasts is then investigated at the underlying layer by observing the effect on VOT values of manipulation of speaking rate, as produced by the same speakers. The results for the three chosen dialects, as seen in Table 6-1, are in line with the observations regarding both non-linguistic and surface phonetic typology by Watson and Bellem (Bellem, 2007, 2014; Watson, 2002, 2011a) which assumed a Bedouin versus urban continuum.

However, the underlying representations, based on the inferred active phonological features, revealed that the three dialects are 'underlyingly' the same. All three dialects (Najdi (Bedouin), Hijazi (Bedouin-sedentary), and Tunisian (urban)) have two active features [voice] and [sg] despite having a three-way or two-way surface phonetic realization of the plosives. Typologizing the Laryngeal class node, then, unlike the Place and Manner, does not result in one-to-one mapping between surface and underlying representation. Youssef's (2021) typology of the Place and Manner class nodes is, as we mentioned earlier, based on contrastive features that map the surface to the underlying directly. Figure 6-1 offers a proposed multi-layered typology for the laryngeal system in Arabic dialects, which necessarily allows for the possibility of an

[71] The accepted generalization that the realization of /q/ as /g/ is associated with Bedouin or Bedouin descended populations (Watson, 2011a) is also apparent in our data in the absence of /g/ in our Tunisian tokens, hence, Tunisian is urban.

indirect mapping between the surface and underlying layers. The Najdi and Hijazi dialects seem to have direct mapping between the surface and underlying layers. Tunisian, however, does not result in one-to-one mapping between the surface and the underlying layers.

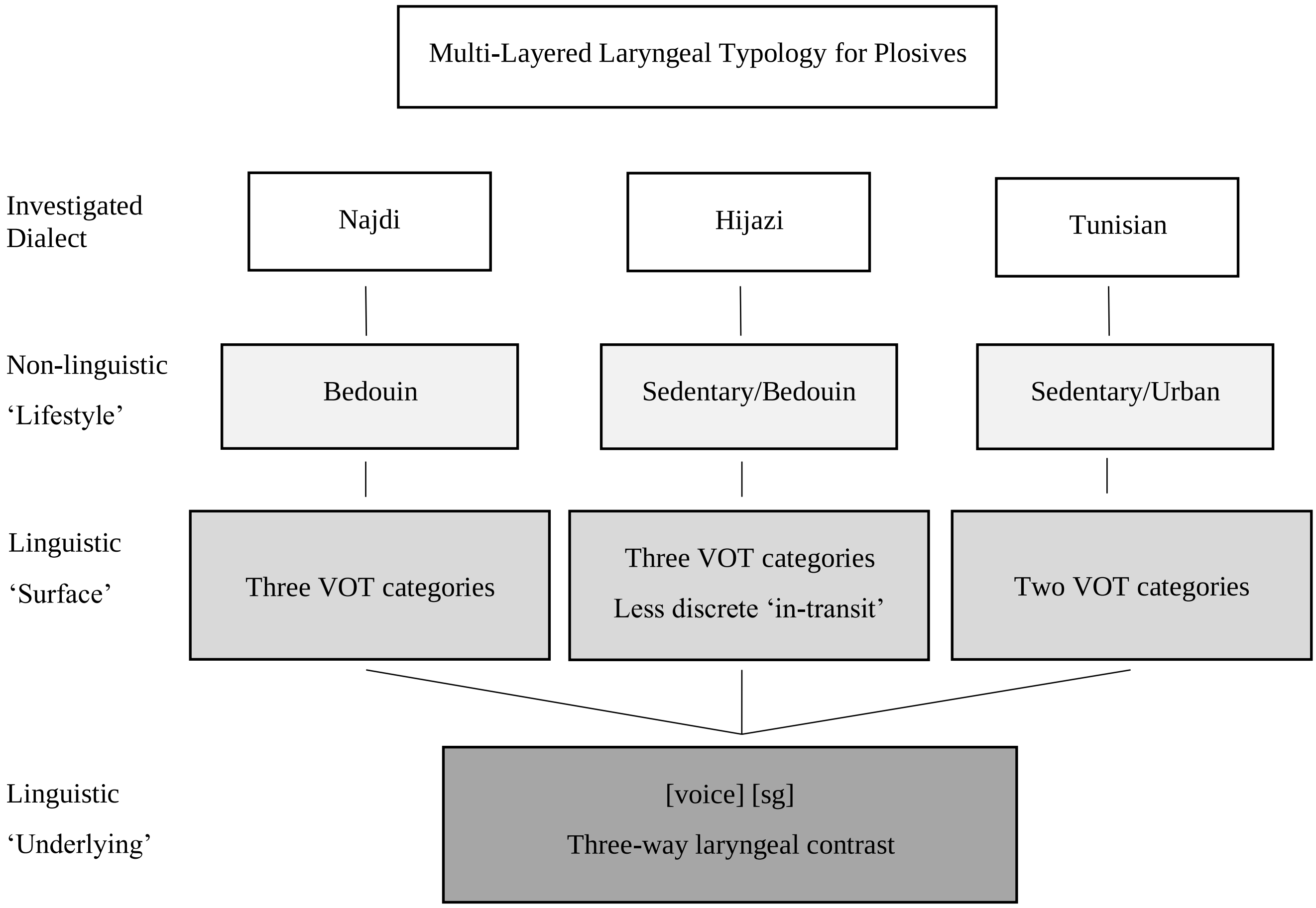


Figure 6-1 Representation of the proposed multi-layered typology of the laryngeal systems for plosives in Arabic dialect speech.

In chapter 5, I then investigated the underlying laryngeal contrasts of the same speakers following a similar methodology but in a different register – MSA. The investigation started with the surface structure by extracting the facts of VOT at normal speech rate. It revealed surface differences between the two registers (Dialect and MSA) in only one dialect: Hijazi. The Hijazi speakers display overlapping values of VOT in the plain and emphatic voiceless plosives in their MSA production, leading to the conclusion that there is a two-way phonetic VOT distinction in Hijazi MSA. This finding in MSA can still be considered in line with the initial surface typology described above, as Hijazi is a 'changing' or 'in-transit' dialect. To uncover the underlying layer of the MSA production, we also investigated values of VOT at a faster speaking rate. The underlying representations of the laryngeal contrast in MSA were different from these found in the dialect speech for one set of speakers: Tunisian. Underlyingly, Tunisians are the only speakers who display one active phonological feature in MSA [voice]. Tunisian MSA seems to have the same number of contrasts on both surface and underlying layers. MSA, unlike dialect speech, thus shows a one-to-one mapping between the surface and the underlying for speakers of two dialects only: Najdi and Tunisian. The evidence for only one active feature [voice] in Tunisian MSA, however, is not as strong as for the other dialects, as the conclusion was reached by investigating a smaller dataset in Tunisian, in the context of /a:/ only. Figure 6-2 offers a proposed multi-layered typology for the laryngeal system in Arabic dialects' MSA production, which also allows for the possibility of an indirect mapping between the surface and underlying layers, as in the Hijazi case.

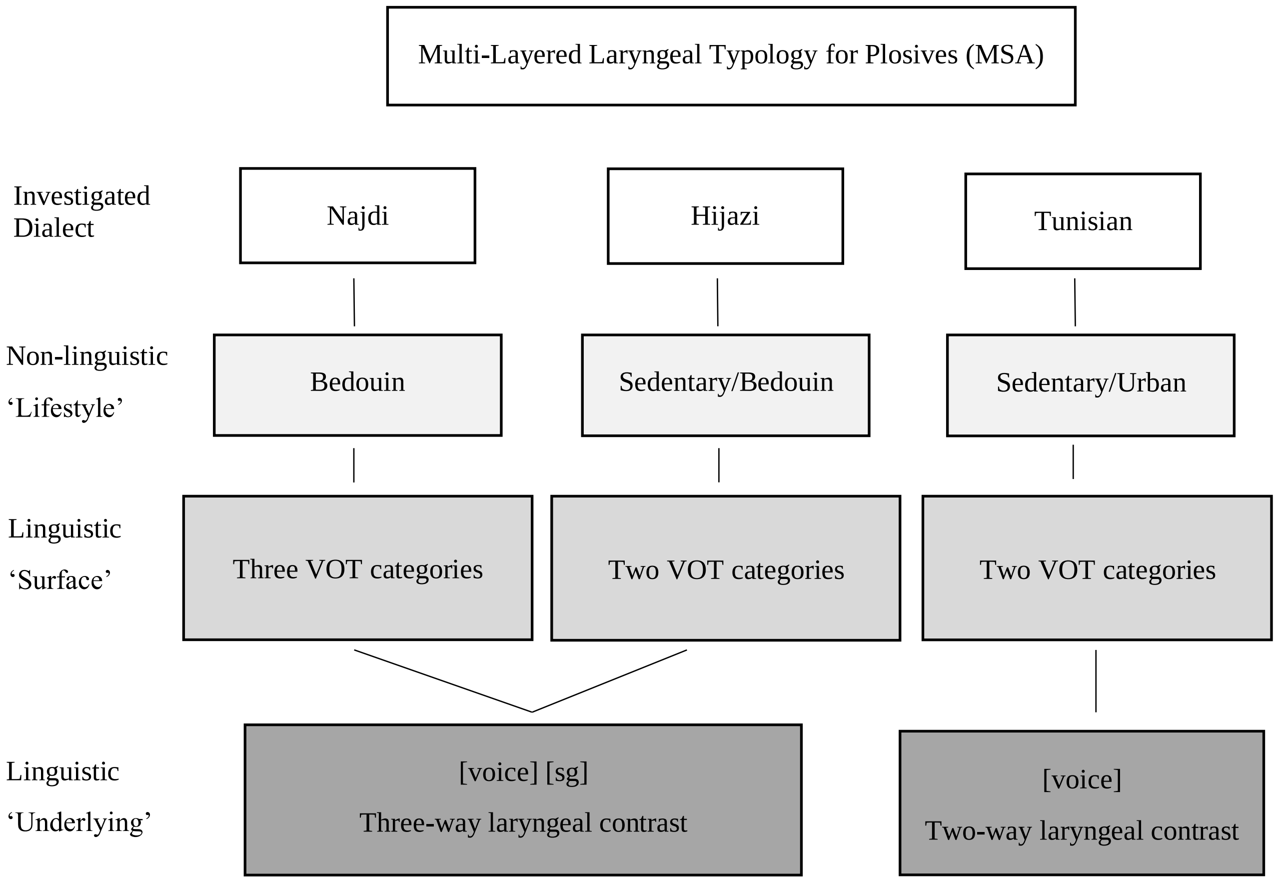


Figure 6-2 Representation of the proposed multi-layered typology of the laryngeal systems for plosives in Arabic MSA.

This section proposed a theory-driven typological approach for the laryngeal node, inspired by the segmental typology of Arabic cognate reflexes in the place and manner nodes suggested by Youssef (2021). The typology in this section is a multi-layered one that links the surface and underlying representations based on results from experimental data, which resulted in evidence of a ‘many-to-one’ mapping in some cases between the two layers, surface and underlying. These results hint at a sound change in progress, which is apparently more advanced in MSA. The direction and causality are still unknown, however, in the following section we will discuss the hypothesis of sound change in progress as a possible explanation of this non-simultaneous change in the phonetics and the phonology in dialect speech and MSA.

6.3 Sound change in progress

Let me start this section by clarifying that the putative analysis of ‘sound change in progress’ is not claimed to be based on sound change evidence from longitudinal or apparent-time data. It is rather based on prior claims in the literature, alongside indications of sound change in progress throughout the experimental studies in this thesis, including gender differences in the dialects described to be ‘in transition’ discussed in chapter 3, and the differences in the mapping of surface to underlying laryngeal representations of the dialects in both registers (MSA and dialect speech) investigated in chapters 4 and 5, and discussed in the typology section above (§ 6.2).

Bellem (2007) based her modern-day typology of emphatics on lifestyle bases or what she termed ‘dialect type’, namely a classification of Arabic dialects along the Bedouin – Rural sedentary – Urban sedentary continuum[72]. Bellem (2007) also argued on various grounds for historical changes to the Semitic emphatics from originally being glottalic (ejectives). She attributed her typology to the indication that Old Arabic would have had a three-way voicing category. In addition, Watson (2002), Holes (2006), Versteegh (2003, 2014), and Ingham (1994), as discussed above, have all attributed the conservativeness of an Arabic dialect to its preservation of Classical Arabic linguistic forms: this is a feature of Bedouin origin dialects, which are mainly dialects of the Arabian Peninsula. In contrast, Urban dialects are characterised by their progressiveness based on their ‘evolutive tendencies’ (Watson, 2002, p. 9). Describing Old or Classical Arabic as being conservative and as having a three-way voicing contrast implies that

[72] For Bellem, three-way voicing contrast dialects are Bedouin-origin, whereas two-way voicing contrast dialects are urban/sedentary.

modern-day varieties of Arabic dialects that report two-way voicing contrast, have plausibly undergone some sort of sound change processes.

At many stages throughout this thesis, hints of sound change in progress have been noticed. The following subsections will discuss two main instances supporting potential sound change. Firstly § 6.3.1 discusses the mismatches found in the surface and underlying representation of the laryngeal contrast in both dialect speech and MSA. Then § 6.3.2 discusses the gender differences noticed in some of the dialects that are described as being ‘in transit’ and are found the middle of the continuum of surface laryngeal contrasts from chapter 3.

6.3.1 Surface and underlying mismatch

What we have seen so far are cases in which phonetic change (as indicated in changes in the acoustics and articulation of plosives) occurs without obvious consequences in the phonological system of the dialects, though some consequences occur in another register of the same dialect. Firstly, in chapters 3, 4, and 5 we noticed variation between the dialects in the surface behaviour of the plosives. As mentioned, many times, when we investigated the surface laryngeal representations of eight Arabic dialects in chapter 3, some dialects displayed a two-way VOT distinction and others displayed a three-way VOT distinction while a few dialects displayed something in between. The close investigation of surface VOT patterns in chapter 4 also displayed the same pattern indicating sound change. Chapter 5 also revealed the same pattern of variation between dialects in the number of surface laryngeal categories in another register of Arabic, MSA. The pattern of phonetic variation in chapter 4 did not show a one-to-one mapping to the phonological representations. Then, in chapter 5, in one of the dialects’ phonological systems (Tunisian), the MSA productions differ from dialect speech which has surface to underlying mismatch, and instead shows one-to-one mapping between the surface and underlying representations. These patterns led us to adopt the working hypothesis that some dialects of Arabic are undergoing a process of sound change, which appears to be more advanced in MSA.

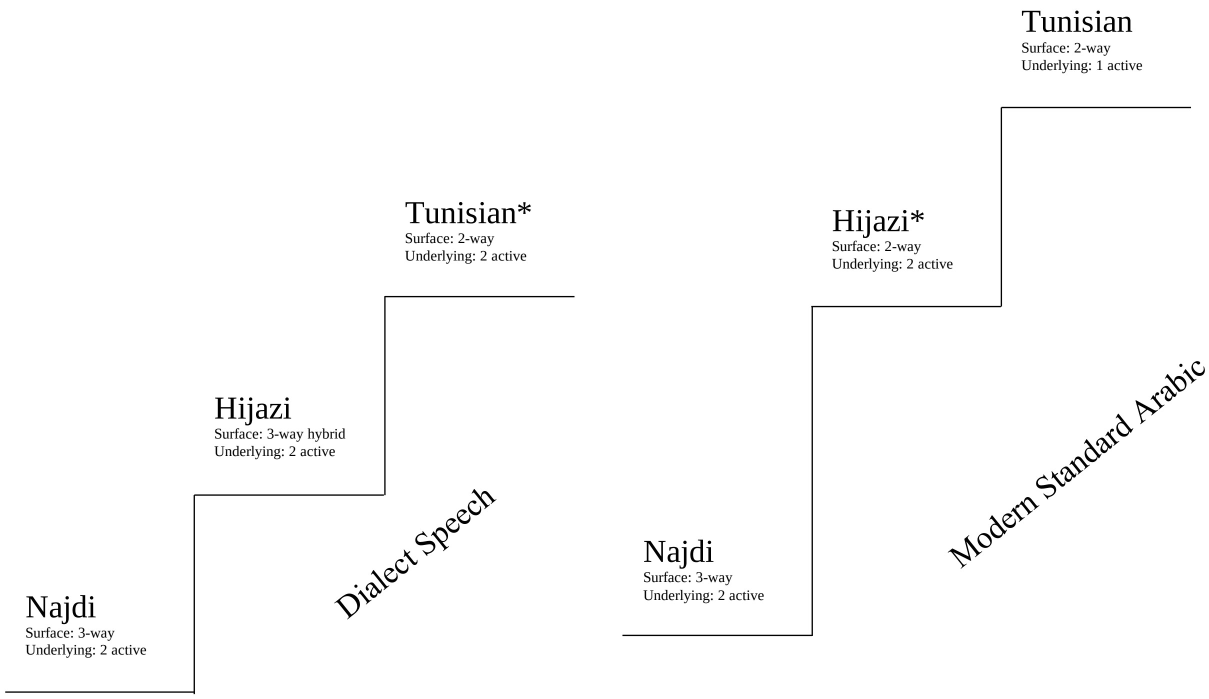


Figure 6-3 Stylised representation of the trajectory of the sound change process with MSA more advanced 'on the right'. The asterisks * indicate potential phonological 'over-specification' where the mismatch occurs.

Figure 6-3 is a schematic representation of my interpretation of the potential sound change trajectory. As I noted in the caption, evidence of potential phonological 'over-specification' indicates a phase in which phonetic change precedes phonological change. This phase appears to be where neutralization in the phonetics of underlyingly contrastive phonemes happens. In this phase, a merger has not yet occurred, as Hamann (2015) states "The term merger only applies if there is no indication that speakers retained a difference in their underlying phonological representation" (p. 3). Changes in the phonetics are often articulatory simplifications, which may translate here in the form of the intermediate lag in our data discussed in § 4.7, while changes in the phonology can be observed through 'structural simplicity' (Hamann, 2015, p. 254). Hamann (2015) defines structural simplicity as "the preference for acquiring simple, non-changing mappings between underlying and surface forms and as few phonological processes as possible" (p. 6). She then gives an example of this structural bias in the loss of the postnasal word-final [ɡ] described by Bermúdez-Otero & Trousdale (2012) in Late Modern English, in which underlying /ɡ/ emerged in the surface form only when a word with a vowel-initial followed (e.g., underlying /sɪŋɡ ɪt/ surface [sɪŋɡ ɪt], but underlying /sɪŋɡ / surface [sɪŋ]). They argued that since prevocalic cases with [ɡ] are less frequent than preconsonantal cases without [g], this has led following generations to produce underlying forms in the following way /sɪŋ/ (i.e., without /g/). The scenario in the Tunisian dialect is slightly different; it is a case of the more frequent surface phonetic form across the *population* – not across *positions* – mismatched with the underlying form. However, the same argument could hold that in order to have structural simplicity the phonological realization in MSA changes as a result. That is, the dissolving of the surface and underlying mismatch in Tunisian MSA results in transparency in the one-to-one mapping between the phonetics and the phonology.

Within the framework of Laryngeal Realism, Honeybone (2005) presented diachronic evidence to account for sound change processes. Honeybone sets out two types of phonological change: (i) processes of change that affect the number of segments or contrasts (e.g., merger and splits) and (ii) processes of sound change that do not affect the number of contrasts (i.e., shifts in the realizations of a segment and therefore the underlying form) (Honeybone, 2005, p.337). To explain the second process, relevant to our discussion here, Honeybone (2005) provided examples of what he termed 'Southern English Fricative Voicing' (p. 339). This process involved a shift from voiceless to voiced, presumably on the surface at first, then underlyingly in the segments resulting in voiced segments, in the underlying inventory. He further explained that this process of change is a shift from [sg] specification of the fricative to no feature specification [Ø], instead of [voice]. Honeybone suggested a term for these processes (both of (i) and (ii) mentioned above): 'Delaryngealisation' (p.345). Delaryngealisation happens when laryngeal features are lost and 'oral specifications'[73] remain (p. 321). Honeybone (2005) comments on this process:

> This is a positive result. If a model predicts the existence of a type of process, we should be able to find examples of it. (p. 345)

The process of change in our Tunisian data is an example of type (ii), in which the surface two-way VOT categories hold in both MSA and dialect speech productions, yet the underlying laryngeal representation of the segment /t/ shifted from [sg] to be non-specified [Ø] in MSA. If our MSA results are correct, then delaryngealisation has

[73] By oral specification, Honeybone (2005, p. 324) means that the plain voiceless stop is described by the absence of a laryngeal phonological feature; the segment, although not laryngeally active, can still be perceived through the "oral" characteristics of articulation.

occurred in Tunisian /t/, and as Honeybone explained, this change, under the model of Laryngeal Realism, is expected, natural, and simple.

This doesn't mean that the full trajectory of sound change is inevitable for all dialects. Dialect type or the 'lifestyle', as we mentioned, plays a role in this process, and sedentary 'progressive' dialects are more susceptible to change than Bedouin-origin 'conservative' dialects.

6.3.2 Gender differences

When discussing variation and potential sound change, we ought to also discuss social, non-linguistic factors, alongside the linguistic factors already discussed. In our data, we investigated gender as a social factor that might contribute to the observed variation. It was only in chapter 3, the corpus-based cross-dialectal examination, that gender differences were found. Chapters 4 and 5 did not show any effects of gender on the production of VOT. The pattern of cross-dialectal variation in chapter 3 (among eight Arabic dialects), found in the degree of voicing lag in the voiceless subset /t/ ~ /ṭ/, hints at sound change in progress due to the observed gender differences in only the two 'in-transition' dialects in the middle of the continuum (Omani and Tunisian). In this continuum, two dialects showed overlapping values of VOT in the plain and emphatic voiceless contrast (Egyptian, Syrian) and four dialects showed the opposite i.e., discrete values of VOT in the plain and emphatic voiceless plosives (Moroccan, Kuwaiti, Jordanian, Iraqi). The gender differences, found only in Omani and Tunisian, are manifested in having **less** overlap in the VOT values of the plain and emphatic voiceless plosives when produced by female speakers, in both dialects. Given the hypothesized direction of change, male speakers are behaving more 'progressively' hence leading the suggested sound change.

The directionality of sociolinguistic change and its interaction with the social factor, gender, is widely discussed in the sociolinguistic literature cross-linguistically and across Arabic dialects. It is often noted that women usually lead the change (Eckert, 1988; Milroy & Milroy, 1993). However, a cross-dialectal survey of sociolinguistic studies on Arabic dialects in Al-Wer et al. (2020) indicated that some phonological features exhibit change led by male speakers (Al-Hawamdeh, 2016; Alaodini, 2019) while other phonological features undergo change led by female speakers (Al-Essa, 2009; Alqahtani, 2015; Hussain, 2017). What matters here is that in both cases these gender differences are normally treated as an indication of sound change in progress. The general consensus here is as stated by Labov (1972):

> It would be a serious error to construct a general principle that women always lead in the course of linguistic change … The correct generalization then is not that women lead in linguistic change, but rather that the sexual differentiation of speech often plays a major role in the mechanism of linguistic evolution. (p. 303)

It is important to consider that many of the phonological patterns discussed in the sound change literature – affrication, deaffrication, and allophonic uses – are salient to the ear. In contrast, the linguistic variable of interest in our study (VOT) is probably less salient, so that speakers could be unaware of the change in its features. See also § 3.4 for further discussion on other gender studies concerning VOT as a variable produced below the level of consciousness.

6.3.3 Final remarks on the sound change analysis

In the absence, as yet, of longitudinal evidence from varieties of spoken Arabic at different periods of time, the best we can do is follow Ferguson's (1989) statement:

> …. the only satisfactory procedure is to investigate particular sets of phenomena on their own merits, piecing together the fragmentary evidence and calling upon general principles of language change when relevant. (p. 8)

In this spirit, in twenty years' time or more, this thesis will hopefully act as a base to assess this analysis of sound change properly with longitudinal or apparent-time evidence.

From what we have seen – if this is a case of sound change – then the change is not all at once. A key finding here is that, when the phonetics changes it is not necessarily the case that the phonology changes at the same time. However, if what we have found is sound change, and if it is in a certain direction, the question remains: why is the change more advanced in MSA? In the next section, § 6.4, we will discuss possible explanations of why MSA might be leading the sound change.

6.4 Why MSA?

By the end of chapter 5 we concluded that speech rate affected VOT in MSA in an interesting and unexpected pattern. A question then arises about MSA: Why would

MSA speech production be further along the sound change trajectory than the speech production of a dialect? Assuming this is a case of sound change, why would a variable (VOT) that is probably below the level of awareness show systematic (across dialects), yet different (within a dialect), results in two registers of the language?

6.4.1 The cognitive processing of Arabic

In order to lay out the context of the cognitive processing of Arabic diglossia in the brain, I shall briefly review relevant psycholinguistic and neurolinguistic[74] research. Most of these studies aim to answer a much-debated question, namely, whether dialect Arabic and Modern Standard Arabic are considered two distinct languages. For some authors, Arabic diglossia is a case of bilingualism (Eviatar & Ibrahim, 2000; Ibrahim, 1983; Ibrahim, 2006, 2009; Ibrahim & Aharon-Peretz, 2005). In their psycholinguistic investigations of child acquisition (Eviatar & Ibrahim, 2000), semantic priming (Ibrahim & Aharon-Peretz, 2005), lexical, and phonemic priming (Ibrahim, 2006, 2009), focusing mainly on Palestinian speakers of Arabic as a first language and Hebrew as a second language, the authors concluded that the cognitive system of an Arabic speaker processes dialect as the first language and MSA, like Hebrew, as a second language. However, Boudelaa & Marslen-Wilson (2013) argued that this conclusion is an unwarranted generalization given the limited context examined, that is, Palestinians in a Hebrew speaking environment. In a different context, Tunisian speakers of Arabic, Boudelaa & Marslen-Wilson (2013) examined the morphological processing of Southern Tunisian Arabic and MSA in an auditory-auditory priming experiment. They concluded that the two language varieties are processed in the same manner with similar priming effects for lexical roots and word patterns in dialectal Arabic and MSA. In fact, their results showed that the priming effect was faster and more accurate in MSA than it is in dialect.

Neurolinguistic investigations in the same language contact context, with bilingual speakers of Palestinian Arabic and Hebrew, has also provided further support to the argument for unity of processing MSA and dialectal Arabic as one language, however (Abou-Ghazaleh, Khateb, & Nevat, 2018; Froud & Khamis-Dakwar, 2018; Khamis-Dakwar, 2005; Khamis-Dakwar, Boudelaa, & Froud, 2009; Khamis-Dakwar & Froud,

[74] Including studies using two neuroimaging techniques, spatial resolution (fMRI) and temporal resolution (ERP) (Froud & Khamis-Dakwar, 2018).

2006, 2012; Nevat, Khateb, & Prior, 2014). Neurophysiological investigations using Event-Related Potential (ERP) methods to study neural representations in the Arabic diglossic situation include a series of studies by Khamis-Dakwar and Froud (Froud & Khamis-Dakwar, 2018; Khamis-Dakwar et al., 2009; Khamis-Dakwar & Froud, 2006, 2012). After their multiple investigations of ERP responses to lexical and phonological codeswitching between dialect and MSA, Khamis-Dakwar and Froud proposed a neurocognitive model of the diglossia of Arabic. This model suggests, considering the sociolinguistic situation of Arabic diglossia, a distinct representation for the two varieties at the lexical level, but a unified representation of the two varieties at the syntactic level. That is, the lexicon of the spoken dialect and the MSA varieties are differentiated into two separate linguistic representations, but the grammatical representations for the two varieties are most likely unified within one grammatical system.

In a functional Magnetic Resonance Imaging (fMRI) investigation, Nevat et al. (2014) examined the neural basis and cognitive processing of the diglossic situation of Arabic in the same context again, looking at MSA and Palestinian dialect along with Hebrew as an L2, in 26 female Arabic speakers. They used a semantic categorization task analysing the processing of MSA and dialect words presented visually, with comparison to their processing of Hebrew words as their L2. Their findings revealed that accuracy in the categorization task was highest in MSA words. In terms of brain activity, their analysis showed differences between MSA vs. dialect: words presented in dialect forms resulted in stronger activation in the left frontal and temporal areas. Their lateralization of activation analysis showed that the lateralization for MSA, however, was stronger than for dialect. Lateralization for Hebrew was found also to be marginally stronger than for dialect. The authors attribute these differences in the lateralization of activation between MSA and dialect to the possible effort made by participants in avoiding the interpretation of dialect words as MSA words. Nevat et al. (2014) discussed the findings and suggest that the variety dominance (MSA or dialect) is modality dependent. In the auditory modality, dialect is the L1 or the dominant variety and in the written visual modality MSA is the L1 that shows dominance.

Following this study, Abou-Ghazaleh, Khateb, and Nevat (2018) conducted another fMRI study to investigate the neural bases of diglossia in Arabic-Hebrew bilinguals, this time during language production in a picture naming task in MSA, Palestinian dialect, and Hebrew. The aim of their study is to compare brain activity during the processes of lexical search and production in the three language conditions – MSA, dialect, and L2 – then to assess the position these language conditions occupied in the cognitive system of native Arabic speakers, to determine whether or not MSA acts as an L2. They also explored the behavioural level by analysing response accuracy and reaction time (RT). Their results revealed that the performance in dialect speech production was faster and more accurate than in MSA and Hebrew, with no significant difference in RT between MSA and Hebrew. However, naming accuracy was higher in MSA than in Hebrew indicating that the L2 (Hebrew) was the weakest language condition overall. Results at the neural level through a whole-brain conjunction analysis revealed similar and complex dominance patterns of the left hemisphere in the three language conditions. Stronger activation in one region of interest (ROI), the cerebellum, was found in Hebrew production compared to dialect and MSA, and, according to Abou-Ghazaleh et al. (2018), cerebellar activation is normally attributed to "greater demands on articulatory control in L2" (p. 92). Abou-Ghazaleh et al. (2018) concluded that their findings suggest that dialect and MSA speech production behave similarly, yet competitively (due to differences in some linguistic features), and they are both considered to be first languages.

One important factor to consider here is that most of the studies on Arabic diglossia mentioned above, except for Boudelaa & Marslen-Wilson (2013), are from one common context (Arabic speakers in a Hebrew speaking environment). This context is not representative of the typical Arabic speaking world due to one major factor. According to Khamis-Dakwar, Taha, & Al-Khoshman (2021), this factor is the systematic subordination of Arabic as a language and Arab speakers in a context of occupation.

Some key findings in this review, such as faster priming effects in MSA than in Tunisian (Boudelaa & Marslen-Wilson, 2013) and higher accuracy in the semantic categorization task in MSA than in Palestinian (Nevat et al., 2014), seem relevant to our findings of MSA being ahead in the suggested sound change trajectory visualized in Figure 6-3, but not directly parallel. The reason for faster processing and higher accuracy of MSA in the abovementioned studies could be attributed to the higher exposure to MSA written forms by Arabic speakers compared to dialect, which is more commonly spoken than written[75]. The results in the present study, however, are at the phonological level measured in a phonetic reflex that is more than likely below the level of phonological awareness.

6.4.2 Predictability and 'Phonetic Uniformity'

Phonological, lexical, and structural differences between Arabic dialects are well established and documented (Versteegh, 2014), but documentation of cross-dialectal differences between the two registers produced by speakers of various dialects is comparatively scarce. This thesis offered an opportunity to examine the two registers in three contexts or dialects – Najdi, Hijazi, and Tunisian. This cross-comparison between dialect speech x MSA speech x context has revealed variation that is systematic and predictable. The variation is predictable in the sense that, if changes occur in both the phonetics and in the phonology of the dialect register, we can predict that changes in MSA production might have preceded, as we can see in Figure 6-3. It is also systematic in the sense that this pattern occurred in all the three investigated dialects.

This predictability adheres to the principle of 'Phonetic Uniformity' discussed extensively in Chodroff and Wilson (2017) and Chodroff et al. (2019) in which variability in the realization of speech sounds is highly structured. Variation in the values of VOT in plosives is known to be, for the most part, ranked ordinally by place of articulation, as values of VOT increase generally with more posterior place of articulation (Cho & Ladefoged, 1999). However, Chodroff et al. (2019) observed another stronger, and highly structured, linear relation within the laryngeal series. In their investigation of VOT values in over 100 languages, Chodroff et al. (2019) found that mean VOT of plosives in the same laryngeal contrast series shows a linear correlation across languages. This correlation is stronger with long lag and voicing lead categories, but still moderately strong within the short lag category. This indicates that, underlyingly, there is a uniformity constraint [76] (Keating, 2003) that ensures that, for instance, speakers with long lag means of VOT in a plosive in one voicing series are

[75] Although, based on the linguistic landscape literature (e.g., (Akbar et al., 2020)), Arabic dialects are increasingly used and seen in written forms in the landscapes of Arabic speaking world, besides its use in the discourse of texting and online blogging.
[76] The constraint or principle of uniformity discussed in Chodroff and Wilson (2017) is defined in Keating (2003) as 'uniform and parallel behaviour of members of a class' (p.375).

more likely to produce proportionally similar values of VOT in the other plosives in the same voicing series.

In our case, the observed structured variation is within laryngeal classes across dialects. Based on this, we can predict that variation in the VOT values of voiceless plosives in MSA. For instance, VOT values in the velar voiceless /k/, will reflect values of /t/ in each dialect, as we saw in the production of dialect speech (see Figure 4-46). Within a class of stops (by voicing category), we would expect phonetic uniformity in Najdi MSA productions of /t/ and /k/ (long lag class) and Najdi /ṭ/ and /q/ (short lag class), and similarly, uniformity in Hijazi MSA /t/ and /k/ (mid lag class) and Hijazi /ṭ/ and /q/ (short lag class). However, in Tunisian MSA, we would expect phonetic uniformity among the whole voiceless class /t/, /ṭ/, /k/, and /q/ (mid lag).

Phonetic Uniformity allows us to make predictions, therefore, and the prediction, as we mentioned, matches what was observed with Tunisian mid lag in the full voiceless set of plosives, even beyond the coronals. There is phonetic uniformity in the realization of underlying phonological categories, so if a feature is lost, this is expected to be reflected through the entire paradigm. In Tunisian MSA, the feature [sg] has dissolved in /t/ and accordingly it is expected to be non-existent in /k/ also. Phonetic Uniformity, in this sense, reinforces also the sound change argument discussed in § 6.3 above, as it indicates that any changes in a plosive of a given category will be reflected elsewhere in the same voicing category. Changes that appeared in Tunisian /t/ in the dialect (i.e., shortened values of VOT) did indeed also appear in plosives of the same category, i.e., /k/.

In short, although we did not expand our investigations beyond the coronals in MSA, given that the coronals are highly variable (Lahiri & Reetz, 2002)[77], we still suspect that the VOT of velar /k/ in Tunisian MSA would behave as it does in dialect speech and cluster around one mean (i.e., the same mean as /t/) based on the Phonetic Uniformity proposed by Chodroff et al. (2019; 2017). However, it is not yet known what the predictions are for the underlying representations in a system undergoing a process of change, like Arabic.

[77] Lahiri and Reetz's (2002) model of Featurally Underspecified Lexicon (FUL) discusses a central hypothesis about the feature [coronal], i.e., [coronal] as a feature is universally underspecified in the lexicon, so, is expected to be highly variable.

Finally, we can conclude that the results in this thesis support the position that diglossia involves ‘two varieties of the same language’. Despite the obvious and intensively researched lexical, structural, and functional differences between MSA and dialectal Arabic, the two registers appear to be cognitively processed in similar ways (Boudelaa & Marslen-Wilson, 2013). To summarize, in this section, we reviewed previous research on the cognitive processing of Arabic diglossia. We then discussed two views that argue for and against treating dialect speech and Modern Standard Arabic as two distinct languages. We also discussed the Phonetic Uniformity principle that allows us to make predictions about the behaviour of other members of the set of plosives in MSA production under the assumption of sound change processes discussed above. Finally, we discussed how our findings can be interpreted as support for the argument for Arabic diglossia as a case of one language and two registers. In the following final section of this discussion § 6.5, I will review the phonological representations used to account for the laryngeal contrast in Arabic dialects and discuss other phonological proposals.

6.5 Phonological Representations

This work is based on a Laboratory Phonology approach, so it is important to restate the chosen ‘mental representations’ adopted here to bridge the gap between phonetics and phonology. In chapter 2 of this thesis, we reviewed the hypothesis of ‘Laryngeal Realism’ as a phonological approach to represent the laryngeal contrast in two-way voicing languages. In this section, we will briefly revisit Laryngeal Realism as an approach to phonological feature representation and summarize our use of it in the representation of the thesis results. Finally, we shall review the phonological representation of the Arabic emphatics, with focus on Youssef’s (2013) representation by adapting Morén’s (2003) Parallel Structures Model (PSM) of feature geometry.

In § 2.3.1, I reviewed Laryngeal Realism as the phonological representation of laryngeal features adopted in this thesis. The term ‘Laryngeal Realism’ was coined by Honeybone (2005) based on the basic feature representation of the laryngeal contrast proposed by Iverson & Salmons (1995). Laryngeal Realism holds that in two-way voicing contrast languages, word-initial plosives that are fully voiced as in the case of French, are represented by a privative phonological feature [voice] and the corresponding plosives that are voiceless unaspirated are laryngeally neutral. In languages where the voiceless plosives are aspirated word-initially, like English, the voiceless plosives are represented privatively with the [spread glottis] feature, and the 'voiced' plosives in these languages are in fact unspecified or laryngeally neutral. Laryngeal Realism's basic representation was motivated through the examination of both the phonetic properties such as VOT as well as phonological phenomena (Honeybone, 2005; Iverson & Salmons, 1995, 2006). Issues related to the Laryngeal Realism approach, and other proposed approaches to laryngeal phonology such as Laryngeal Relativism (Cyran, 2011) and Onset Prominence (Schwartz, 2013), were also previously reviewed in § 2.3 of this thesis.

We adopted [voice] and [spread glottis] as the phonological features from the standard Laryngeal Realism approach to account for the laryngeal contrast in the 'underlyingly' three-way language, Arabic. The cross-dialectal examination here thus assumes the privativity of laryngeal features. In chapter 3, after analyzing the VOT of plain /t, d/ and emphatic /ṭ, ḍ/ in eight Arabic dialects, we arrived at the conclusion that Arabic dialects display differences in their laryngeal contrast. Some dialects display three-way voicing categories (voicing lead 'voiced'– short lag 'emphatic voiceless' – long lag 'plain voiceless') while other dialects have merged their plain and emphatic voiceless contrast and display a pattern of a two-way *true voice language*, like Spanish, with only two voicing categories (voicing lead 'voiced' – short lag 'voiceless'). However, there are a few Arabic dialects examined that display values of VOT in the middle ground between a three-way and a two-way contrast dialect. We initially assumed that thee-way voicing category dialects would have [voice] and [sg] as active phonological features in the dialects, while two-way dialects will only adopt [voice] as the active phonological feature, while we had no predictions for the dialects in the middle. Chapters 4 and 5 examined the validity of [voice] and [sg] as representative features through the manipulation of speech rate in the production of eight plosives in three dialects of Arabic: one dialect retains the three-way voicing category, another dialect displays a two-way voicing category, and a third dialect that is assumed to be in the middle. The results of chapters 4 and 5 that examined VOT in two speaking rates and in two registers of Arabic, MSA and dialect speech, revealed that two-way dialects of Arabic – as in the case with Swedish (Beckman et al., 2011) – can be over-specified with two active phonological features [voice] and [sg].

This over-specification analysis was inferred due to the varying degree to which the voiceless plosives show an effect of speaking rate, especially in the dialect that displays a two-way voicing category (Tunisian) and the dialect that is in the middle between two-way and three-way voicing categories (Hijazi). The effect of speaking rate on the VOT of the plain voiceless plosives was minimal[78] in the two-way dialects compared to the three-way dialect (Najdi). However, the voiceless emphatic /ṭ/ remained the only plosive that was not affected by speech-rate manipulation in all the three dialects, so we concluded that it is unspecified with a laryngeal phonological feature.

The view taken so far is a realist one within the Laryngeal Realism framework. It could also be possible to view the over-specification of [sg] in two-way contrast dialects, and even the feature [sg] contrasting /t/ ~ /ṭ/ in three-way contrast dialects, as evidence that [sg] is in fact phonologically redundant. This is because, as mentioned in § 2.6.1, in three-way contrast dialects, the voiceless emphatic plosive /ṭ/ is in fact not a true counterpart of the voiceless non-emphatic plosive /t/, because the contrast is not minimal; /t/ ~ /ṭ/ also contrast in emphasis, however represented (Bellem, 2007, p. 132). In his definition of feature economy, Clements (2003, p. 287) specifies that "Feature economy applies not only to distinctive feature values, but to redundant values of features that are distinctive or phonologically active elsewhere in the system." In this view, it is possible that [sg] is indeed a redundant feature, but one might further argue that the apparent over-specification in /t/ ~ /ṭ/ is spurious, since [sg] is active elsewhere in the system (such as /k/). This thesis displayed independent evidence that /k/ has an active [sg] feature because it shows the effects of speaking rate. These issues are not a challenge to the Laryngeal Realism view, but they do present a challenge to more economical views.

For the Arabic emphatic plosives, the phonological representation in the laryngeal node of the voiced /ḍ/ is [voice], and the voiceless /ṭ/ is not assigned a laryngeal feature, in all the investigated dialects. The property of emphasis has been widely debated to be either segmental or suprasegmental, and the articulation of the emphatics has also been debated, as either primarily a property of the consonant or the vowel; as a result, it is hard to capture the mental representation of the emphatics in a single feature. Proposals of non-laryngeal and mainly segmental and suprasegmental features in the literature include [flat] (Jakobson 1957), [F2 drop] (Card, 1983) [guttural] (Hayward & Hayward, 1989), [CP] Constricted Pharynx (Hoberman, 1987), [pharyngeal] (Herzallah, 1991;

[78] This minimal effect of speaking rate was in voiceless plosives that we considered in § 4.7 as having VOT values of an 'intermediate lag' or 'mid lag'. See the discussion of chapter 4.

McCarthy, 1994) [RTR] Retracted Tongue Root (Davis, 1995; Goad, 1991; Shahin, 1996; Zawaydeh, 1998).

6.5.1 Adaptation of Parallel Structures Model (PSM)

Following the Parallel Structures Model, Youssef (2013) proposes a feature representation for the emphatics that encapsulates both the consonant and the vowel place of articulation. The Parallel Structures Model (PSM) of feature geometry (Morén, 2003) is a minimalist model in which consonants (C-) and vowels (V-) are represented in parallel structures that have identical features in the place, manner, and laryngeal nodes. It is an integrated approach drawing on a number of proposals, including Element Theory (Harris & Lindsey, 1995), Unified Place Theory (Clements, 1991; Clements & Hume, 1995), and Dependency Phonology (Anderson & Ewen, 1987). In PSM, features are distinctive, and only present if there is a necessity to maintain a phoneme contrast or if they are active in the phonology (Morén, 2003; Youssef, 2021). In PSM features are hierarchical; the diagrams in Figure 6-4 show that a feature for place or manner can be represented using two separate tiers/nodes in which the V-node features depend on the C-node features. In this unified approach, consonants can have features from both the C-node and V-node, while vowels are only represented through the V-node.

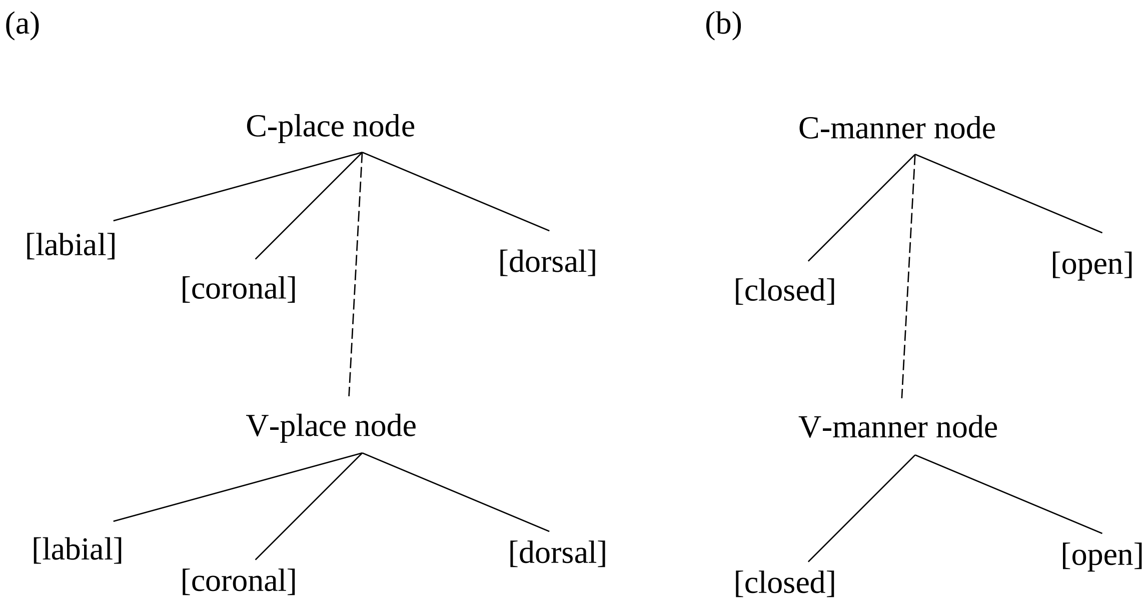


Figure 6-4 Basic PSM geometry from Youssef (2021, p. 3). (a) Place tier; (b) Manner tier.

The place tier is based on the articulatory features [labial], [coronal], [dorsal]. C-place features are used to represent simple consonants while both C- and V-place features are used to describe complex consonants, as in the case of Arabic emphatics in which a secondary articulation is involved (Youssef, 2021). The manner tier reflects the sonority of the sound segments using the loosely defined representations of [open] and [closed]. The same case as for place applies for the manner tier: features from both C- and V-nodes can be used to differentiate stops, fricatives, and sonorant segments (Morén, 2003; Youssef, 2021).

The laryngeal and tone tier, as seen in Figure 6-5, comprises [open], [closed], and [lax] features that correspond to the traditional features [spread glottis], [constricted glottis], and [voice]. The C-laryngeal node is for consonant features: plain plosives are unmarked, [open] is aspirated, [closed] is glottalized or ejective, and [lax] is voiced (Morén, 2003, p. 230). The V-laryngeal node is used to account for vowel features, which map to tone. Mid tone is unmarked, [open] is for low tone, [close] is for high, and [lax] corresponds to pitch or register (Morén, 2003, p. 231).

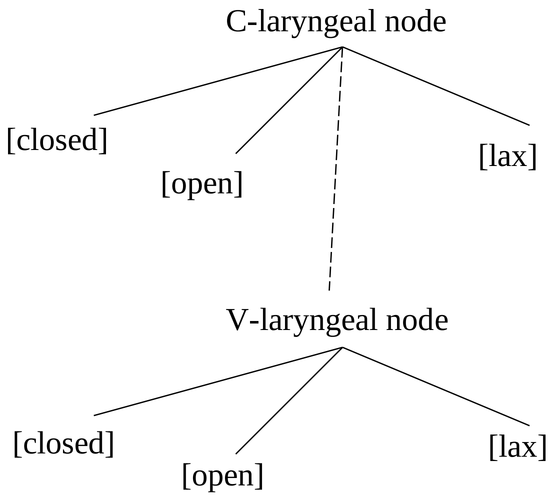


Figure 6-5 Basic PSM geometry of the Laryngeal tier (Morén 2003, p. 265).

Returning to Youssef's (2014) proposal for the phonological representation of the emphatics, he proposes that the emphatics are associated with V-place [dorsal] along with the primary consonantal representation of each emphatic. For instance, the plain /t, d/ and the emphatic /ṭ, ḍ/ are part of the C-place [coronal] class, and what distinguishes the two contrasts (plain ~ emphatic) is the additional V-place [dorsal] feature for the emphatic segments. However, the PSM laryngeal representation has been used to capture emphasis in different ways in Watson and Heselwood's (2016) competing proposal for the neighbouring Modern South Arabian languages. Watson and Heselwood use the features [open] and [closed] to describe the phonation categories in Mehri and San'aani Arabic consonants by grouping them into A (presence of voiceless turbulence) and B (absence of voiceless turbulence). Two daughter nodes [tense] for emphatics and [lax] for voiced are suggested to describe the phonetic interpretation under the laryngeal feature [closed], as illustrated in our visualization in Figure 6-6.

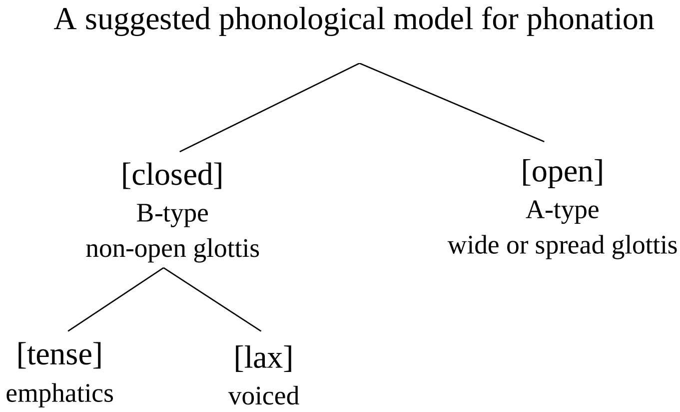


Figure 6-6 A diagram of the suggested phonological model for phonation from Watson and Heselwood (2016, p. 33).

Based on Youssef's (2014) proposal to represent the emphatics with the PSM additional V-place [dorsal], we attempt in the following diagrams to capture the possibility of over-specification by adding the laryngeal tier and proposing a transition phase which occurs during the sound change. The following diagrams in Figure 6-7 are the full PSM representation of the plain (left) and emphatic (right) voiceless contrast. The C-place and C-manner features are identical, yet the additional V-place feature is what differentiates emphasis. In addition, the laryngeal feature is unspecified in the emphatic /ṭ/, but present in the plain /t/ in the 'conservative' dialects of Arabic that retain a three-way system and are not susceptible to change (i.e., [open] for /t/).

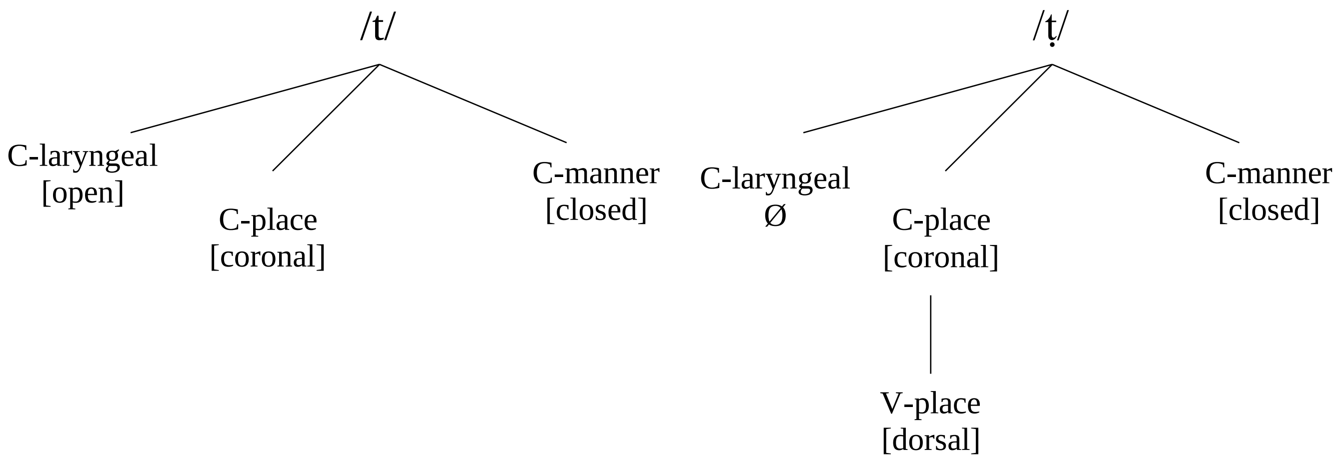


Figure 6-7 A diagram of the suggested PSM phonological representation of /t/ and /ṭ/ the Arabic underlying phonological system.

Figure 6-8 shows the suggested PSM phonological representation of /t/ in the temporary phase of change that caused the over-specification in Tunisian (dialect) and Hijazi (MSA); the dotted line is intended to represent the transitional phase found in dialects that are not Bedouin-origin and are likely to undergo change, and this is where over-specification of the laryngeal feature happens. As we discussed earlier, the over-specification is caused by a temporary phase during sound change that was observed in Tunisian dialect /t/ and Hijazi MSA /t/. Figure 6-9 shows the final phase of the sound change process, when the laryngeal feature dissolves and the system is no longer over-specified, as we have seen in the Tunisian MSA production of /t/.

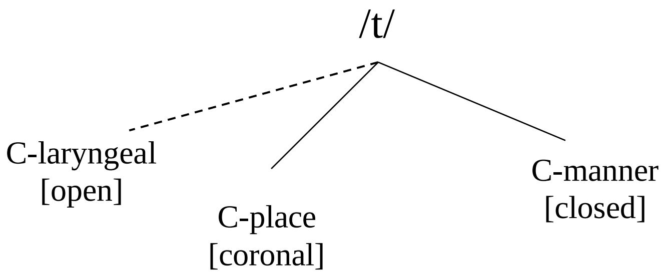


Figure 6-8 A diagram of the suggested PSM phonological representation of /t/ in the temporary phase of change that caused the over-specification in Tunisian (dialect) and Hijazi (MSA).

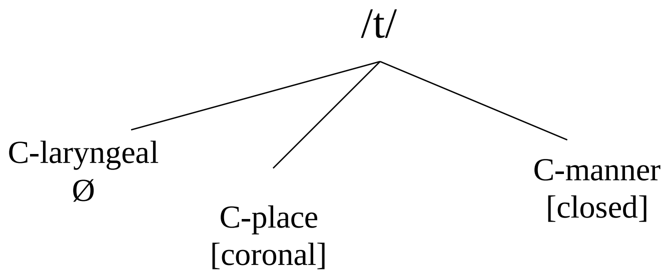


Figure 6-9 A diagram of the suggested PSM phonological representation of /t/ in the final phase of change that resulted in a two-way system in the Tunisian MSA.

Since the emphatic /ṭ/ showed consistent behaviour across all the dialects, it would be represented in the pattern shown in Figure 6-7 above (right) in all dialects/registers. In a three-way laryngeal contrast dialect, the PSM representations of /t/ and /ṭ/ are not minimally distinct. The plain /t/ does not contrast with the emphatic /ṭ/, it is however minimally distinct from the voiced /d/. In addition, the voiced plosives /d/ and /ḍ/ show the same effects and are consistent, and would have the same phonological representation, across all the dialects. In PSM, they would be represented as the following:

/d/ C-laryngeal [lax] – C-manner [closed] – C-place [coronal]

/ḍ/ C-laryngeal [lax] – C-manner [closed] – C-place [coronal] + V-place [dorsal]

To summarize, this section discussed the phonological representations used in the thesis drawn from widely used laryngeal features of the Laryngeal Realism. We then used Morén's (2003) Parallel Structures Model to review Youssef's (2013) phonological representation of the emphatics in Arabic and offered a representation of the over-specification analysis in the PSM framework.

6.6 Summary and conclusion

In this chapter, we discussed the combined results of three experiments from four different phonetic and phonological perspectives. In § 6.1 I started with a summary of the results of this thesis as a reminder. Section § 6.2 proposed a theory-driven typological analysis of the laryngeal node that was inspired by Youssef's (2021) work on theory-driven segmental typology for the place and manner nodes for varying Arabic reflexes of cognate consonants. The combination of this section and Youssef's proposal together offer a coherent phonological typology across Arabic dialects on three levels of phonological representation: place, manner, and laryngeal. After that, in § 6.3 I discussed the observed variation as a potential analysis of sound change in progress based on signs of mismatches found in the surface and underlying representation of the laryngeal contrast in two registers, dialect speech and MSA. I also discuss the sound change analysis based on the observed gender differences in the production of plosives in only some of the dialects in Arabic, namely the ones that are suspected to be undergoing change. Section § 6.4, seeks to understand why this sound change process might be further advanced in the MSA register. I reviewed previous literature on the cognitive processing of Arabic diglossia. In this section we also discussed two views of the nature of Arabic diglossia, processed either as one language or two distinct languages. Finally, in § 6.5 I revisited the Laryngeal Realism approach to phonological representations and summarized the use of its laryngeal features in this thesis. I also reviewed some phonological representations of emphasis in Arabic, specifically Youssef's (2013) representation by adapting Morén's (2003) Parallel Structures Model (PSM) of feature geometry to model a phonological representation of the empirical results of the thesis.

In conclusion, this research is multifaceted, and the findings can be discussed in various ways. This discussion chapter explored the interaction of phonology with phonetics specifically and, sociolinguistics, psycholinguistics, and neurolinguistics in a general sense. The multifaceted angles included aspects of sound change, bilingualism, diglossia, and phonological representations. In the next chapter 7, we will conclude this work with final remarks and mention the limitations, and the suggested future directions and implications.

Chapter VII

Conclusion

7 Conclusion

7.1 Contributions of the thesis

The first main contribution of this thesis is the proposal of a new characterization of laryngeal variation in Arabic; namely, that variation in realization of the Arabic voicing contrast among dialects is a continuum rather than a dichotomy between two types of voicing contrasts. This is an important addition to our understanding of the segmental typology of spoken Arabic.

Another main contribution of the thesis is the additional evidence to support the interpretation of the observed dialectal variation as a potential sound change in progress, based on (i) the observed gender differences in production of plosives, in only those dialects that are indicated to be in a transitional phase and might be currently undergoing change, and (ii) the mismatches found between the surface categories and underlying representation of the laryngeal contrast in the two registers (dialect speech and MSA).

The third main contribution of the thesis is new evidence relevant to theoretical claims regarding over-specification of phonological features in some two-way voicing contrast languages. The thesis offers one explanation of why a language might be in a state of 'temporary over-specification'. In Arabic, specifically, the observed over-specification of laryngeal features in some two-way voicing contrast dialects is a possible by-product of sound change in progress. In this thesis, the over-specification is suggested also to be possibly resolved over time. The working hypothesis that some Arabic dialects are undergoing sound change permits examination of linguistic phenomena in the context of change, which is an interesting testing ground for any given theory. Having said that, it is important to note that this explanation of over-specification is not necessarily applicable to all cases observed; for instance, this might not be the case in Swedish.

These contributions have been made possible through the richness and originality of approach in the interest of being comprehensive, which in itself is a methodological contribution. This thesis is a multifaceted investigation of the voicing contrast from different angles – comparing dialects (Najdi, Hijazi, Tunisian) x comparing speech production from two registers (dialect speech and MSA speech) x testing surface and underlying behaviours (two speaking rates) – and through examining multiple acoustic cues and target plosives.

In addition to the implications for linguistic typology, Arabic dialectology, and phonological representations, the observations and documentation here of the variation among dialects may be beneficial for practical uses in forensic or artificial intelligence (AI) applications, such as in accent/dialect detection and recognition.

7.2 Limitations of the thesis

The present thesis has a few limitations worth noting. Firstly, the sample size, as discussed earlier in § 5.5.1 and § 5.6 of this thesis, is smaller in MSA than in dialect speech (MSA: N = 1347, Dialect: N = 7014). In addition, the necessary exclusion of contexts that might have triggered a palatalization effect in some dialects (i.e., /iː/ and /j/ in the MSA dataset in § 5.5.6) led to forming generalizations based on even smaller numbers of tokens (N = 462). As a result, the findings, especially for MSA, should be interpreted with caution. It is possible that controlling for potential palatalization effects and including a larger number of data points in MSA speech production, in future research, might reveal different results.

Secondly, the choice of speaking rate as a diagnostic is only one method available to test active phonological features. Other methods that were not implemented in this thesis include examination of the percentage of voicing in word-medial plosive closures (Beckman et al., 2011; Beckman et al., 2013) and the extent to which voicing assimilation applies in across word boundary clusters (Al-Gamdi et al., 2022; Beckman et al., 2013; Iverson & Salmons, 1995) as mentioned in § 2.4.

Thirdly, the choice of Tunisian as a dialect that represents one end of the VOT continuum was due to difficulties in recording participants of a true two-way contrast dialect at the time of the study, as mentioned in chapter 4, p. 60. Investigations of the voicing contrast and speaking rate in a true two-way voicing contrast dialect, such as Egyptian or any of the urban Levantine dialects, rather than Tunisian which appears partly in transition, might reveal more insightful and straightforward results.

7.3 Directions for future work

The obvious direction for future work is to remedy those limitations of time and scope, by analysing more MSA data and/or through use of more diagnostics to test for active features.

This thesis builds on Bellem's (2007, 2014) initial observation of a typology within the Arabic voicing system. This thesis offers consistent and comprehensive methodological analyses from first-hand cross-dialectal data. Since the thesis is, thus, multidimensional, as we mentioned earlier, in the methods section of chapter 4 we noted that the target plosives used in the chapter for dialect speech production was the full set of plosives present in each dialect (i.e., voiced /b, d, ḍ, (g)/ ~ voiceless /t, ṭ, k, q/). The MSA dataset, however, as mentioned in the methods of chapter 5, was restricted to the emphatic contrast only (i.e., voiced /d, ḍ/ ~ voiceless /t, ṭ/), due to space and time limitations. Investigating the dialect speech, from three different varieties of Arabic, was thus thorough and the results can be considered insightful. In contrast, the MSA speech investigation, as we mentioned in the limitations above, was based on a smaller dataset, which leaves open for future research the need to investigate the voicing contrast of the MSA speech production by including the full set of plosives and a larger number of tokens, as we did for the dialect speech.

Finally, we concluded in many parts of the thesis that there are indications of potential sound change in progress in some of the dialects; in order to confirm the inference of sound change, longitudinal work and/or apparent-time data is needed.

Appendices

Appendix A (for chapter 2)

Appendix (A.1) Consonant cognate realizations in the Arab world from (Al-Essa, 2019; Youssef, 2021; Watson 2007).

MSA Phoneme	Arabic	Phoneme Reflex	Geographical Distribution
/q/	ق	/q/	Sedentary dialects: North Africa, Mesopotamian, parts of Oman, Yemen, and some Syrian
	أ/ئـ	/ʔ/	Urban, city dialects: Egyptian and Levantine and some Maghrebi
	ك	/k/	Rural Levantine dialects
	(ق)	/g/	Bedouin(-origin) dialects: Arabian Peninsula, Southern Iraq Rural dialects: Egypt, Libya, some North African
/dʒ/	ج	/dʒ/	Bedouin(-origin) dialects – Bedouin Levant
	ق	/g/	Lower Egyptian and sporadic Peninsular, Yemini
	(چ)	/ʒ/	Urban Levantine and most of Morocco
	ي	/j/	Gulf Arabic
/θ ~ ð/	ث ~ ذ	/θ ~ ð/	Bedouin(-origin) and few rural dialects
	ت ~ د	/t ~ d/	Sedentary dialects: Morocco, Egyptian, and Levantine. Also, Mecca, Jeddah, and Medina
	س ~ ز	/s ~ z/	Peripheral and Northern Mesopotamian
	ف ~ (ڤ)	/f ~ v/d/	Sporadic: Siirt, Tell Atlas, Palmyra, Qatif area in Saudi
/ḍ/	ض	/ḍ/	Sedentary dialects: Egyptian and Levantine Qatif area in Saudi
	(ظ)	/ẓ/	Sedentary dialects: Egyptian and Levantine
	ظ	/ð̣/	Bedouin(-origin) dialects

/ɣ/	غ	/ɣ/	All the other dialects
	ق	/q/	Some Gulf dialects
/k/	ك	/k/	Most Mashriqi and Maghrebi dialects
	(تش)	/tʃ/	Rural Levantine dialects, lower Iraqi dialects Gulf dialects vs. /k/
	(تس)	/ts/	Najdi Arabic

Appendix B (for chapter 3)

Appendix (B.1) List of stimuli and their Arabic gloss for targets in word-initial position from (Hellmuth & Almbark, 2019).

voicing	target	test word	gloss	Arabic script
voiced	/d/	di:n	religion	دين
		du:n	under	دون
		da:r	house	دار
		dubb	bear (animal)	دُب
		damm	blood	دَم
		darb	path	درب
		duru:b	paths	دروب
	/ḍ/	ḍi:q	hardship	ضيق
		ḍa:r	harmful	ضار
		ḍubb	collect/gather	ضُب
		ḍumm	included	ضَم
		ḍarb	hit	ضرب
		ḍuru:b	kinds	ضروب
voiceless	/t/	ti:n	fig	تين
		tu:b	repent	توب
		tifil	coffee grounds	تِفِل
		tubiʕ	followed by	تُبع
		tamir	dates	تمر
		tayyar	current	تيّار
		ta:bit	repent (f)	تابت
		tu:lik	**nonsense word**	تولك
		titbaʕ	she follows	تِتبع
	/ṭ/	ṭi:n	mud	طين
		ṭu:b	bricks	طوب
		ṭa:r	flew	طار
		ṭifil	child	طفل
		ṭubiʕ	was printed	طُبع
		ṭamir	covering	طَمر
		ṭayyar	pilot	طيّار
		ṭa:bt	healed (f)	طابت
		ṭu:lik	your height	طولك

Appendix (B.2) List of stimuli and their Arabic gloss for targets in word-medial position from (Hellmuth & Almbark, 2019).

voicing	target	test word	gloss	Arabic script
voiced	/d/	mudirr	passing of a liquid	مُدِر
		madru:s	studied	مدروس
		taxdi:rhum	anesthetize	تخديرهم
		taʕdi:lat	modifications	تعديلات
		ʔdwija	medicines	أدوية
	/ḍ/	maḍu:n	**nonsense word**	ماضون
		muḍirr	harmful	مُضِر
		maḍru:b	hit	مضروب
		taxḍi:rhum	to make sth green	تخضيرهم
		taʕḍi:lat	**nonsense word**	تعضيلات
		ʔḍwija	lights	أضوية
voiceless	/t/	taʕti:lat	**nonsense word**	تعتيلات
	/ṭ/	taʕṭi:lat	obstructions	تعطيلات
		tiṭbaʕ	to print (f)	تِطبَع

Appendix (B.3) List of carrier phrases that translate to ‘write …… twice’ and their Arabic gloss for each chosen dialect in the IVAr corpus (Hellmuth & Almbark, 2019).

code	**dialect**	**carrier phrase**	**Arabic script**
moca	Moroccan (Casablanca)	ktib …… ʒu:ʒ marra:t	اكتب .... زوج مرات
tuns	Tunisian (Tunis)	iktib …… marti:n	اكتب .... مرتين
egca	Egyptian (Cairo)	iktib …… kama:n marra	اكتب .... كمان مرة
joka	Jordanian (Karak)	uktub …… kama:n marra	اكتب .... كمان مرة
syda	Syrian (Damascus)	ktu:b …… kama:n marra	اكتب .... كمان مرة
irba	Iraqi (Muslim Baghdadi)	uktub …… baʕad marra	اكتب .... بعد مرة
kwur	Kuwaiti (Urban)	iktib …… baʕad marra	اكتب .... بعد مرة
omba	Omani (Buraimi)	iktib …… marra θa:nja:	اكتب .... مرة ثانية

Appendix (B.4) Section § 3.3.2 Model summary for the voiced subset.

Formula: vot ~ dialect * target * gender + position + (1 | item) + (1 | subject)
Data: voiced

REML criterion at convergence: 14224.6
Scaled residuals:

Min	1Q	Median	3Q	Max
-7.5140	-0.4920	0.0200	0.5033	4.3030

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	41.21	6.420
item	(Intercept)	21.12	4.596
Residual		408.11	20.202

Number of obs: 1607, groups: subject, 88; item, 24

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-60.72007	1.30762	40.43459	-46.436	< 2e-16	***
dialect1	1.73927	2.16533	65.98325	0.803	0.424724	
dialect2	5.51830	2.18236	68.42249	2.529	0.013763	*
dialect3	0.20307	2.26113	79.50381	0.090	0.928664	
dialect4	8.01067	2.55986	88.17208	3.129	0.002376	**
dialect5	-5.26173	2.32984	88.82146	-2.258	0.026369	*
dialect6	0.62605	2.37518	97.10750	0.264	0.792662	
dialect7	-8.68203	2.19755	67.31938	-3.951	0.000189	***
dialect8	-2.15360	2.99408	72.42333	-0.719	0.47428	
target1	-0.41551	1.09879	22.99942	-0.378	0.708785	
gender1	-0.81291	0.90588	77.68217	-0.897	0.372293	
position1	-3.23150	1.07925	21.01997	-2.994	0.006908	**
dialect1:target1	-2.33353	1.23737	1491.28583	-1.886	0.059505	.
dialect2:target1	-2.49692	1.27593	1495.38310	-1.957	0.050539	.
dialect3:target1	1.17782	1.42533	1524.81966	0.826	0.408735	
dialect4:target1	-1.60031	1.67169	1527.09575	-0.957	0.338566	
dialect5:target1	0.56578	1.52836	1534.23313	0.370	0.711293	
dialect6:target1	-3.25029	1.60079	1513.64835	-2.030	0.042488	*
dialect7:target1	3.80409	1.27250	1489.99930	2.989	0.002840	**
dialect8:target1	4.13336	1.80229	1505.84696	2.293	0.02196	*
dialect1:gender1	2.56752	2.16477	65.91516	1.186	0.239861	
dialect2:gender1	-0.53115	2.18207	68.38073	-0.243	0.808411	
dialect3:gender1	-0.09408	2.26145	79.53254	-0.042	0.966921	
dialect4:gender1	0.33823	2.56133	88.37013	0.132	0.895242	
dialect5:gender1	-0.22210	2.32865	88.64407	-0.095	0.924231	
dialect6:gender1	0.99182	2.37302	96.73088	0.418	0.676906	
dialect7:gender1	-3.20604	2.19730	67.28900	-1.459	0.149195	
dialect8:gender1	0.15580	2.99320	72.35249	0.052	0.95863	
target1:gender1	-1.28352	0.56253	1511.78415	-2.282	0.022647	*
dialect1:target1:gender1	1.61930	1.23643	1489.64234	1.310	0.190515	
dialect2:target1:gender1	3.81485	1.27539	1494.93599	2.991	0.002825	**
dialect3:target1:gender1	-2.01758	1.42577	1526.01358	-1.415	0.157247	
dialect4:target1:gender1	-0.34242	1.67391	1529.88108	-0.205	0.837939	
dialect5:target1:gender1	-0.06521	1.52668	1532.34139	-0.043	0.965937	
dialect6:target1:gender1	-2.26329	1.59738	1510.45836	-1.417	0.156726	
dialect7:target1:gender1	0.20232	1.27208	1489.62778	0.159	0.873651	
dialect8:target1:gender1	-0.94797	1.80130	1504.71891	-0.526	0.59878	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (B.5) Section § 3.3.3 Model summary for the voiceless subset.

```
Formula: vot ~ dialect * target * gender + position + (1 | item) + (1 | subject)
   Data: unvoiced
REML criterion at convergence: 12838
Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5528 -0.5629 -0.0845  0.3954  5.3812
Random effects:
Groups     Name         Variance   Std.Dev.
subject    (Intercept)  17.61      4.197
item       (Intercept)  53.58      7.320
Residual                125.32     11.195

Number of obs: 1661, groups:  subject, 88; item, 23

Fixed effects:
                          Estimate  Std. Error          df  t value   Pr(>|t|)
 (Intercept)              32.35538     1.97164    34.64113   16.410    < 2e-16  ***
 dialect1                  5.15025     1.35340    72.17524    3.805   0.000294  ***
 dialect2                  2.33925     1.35765    73.99853    1.723   0.089064  .
 dialect3                 -0.95416     1.34533    71.84130   -0.709   0.480473
 dialect4                 -0.20871     1.47824    71.64499   -0.141   0.888119
 dialect5                 -7.18410     1.34899    72.67321   -5.326   1.08e-06  ***
 dialect6                  2.88567     1.36108    75.28757    2.120   0.037290  *
 dialect7                  1.16261     1.36630    72.77724    0.851   0.397606
 dialect8                 -3.19082     1.87392    79.93699   -1.703   0.092501  .
 target1                   9.41523     1.56818    20.69304    6.004   6.23e-06  ***
 gender1                  -0.10599     0.54561    73.83188   -0.194   0.846504
 position1                -4.66727     1.52695   105.96062   -3.057   0.002833  **
 dialect1:target1         12.32251     0.70062  1543.64983   17.588    < 2e-16  ***
 dialect2:target1          7.51204     0.71897  1543.07317   10.448    < 2e-16  ***
 dialect3:target1          3.18703     0.70235  1541.09547    4.538   6.13e-06  ***
 dialect4:target1          2.36540     0.77088  1540.02533    3.068   0.002189  **
 dialect5:target1         -4.82371     0.71024  1558.38147   -6.792   1.57e-11  ***
 dialect6:target1         -6.09706     0.73357  1557.62160   -8.312    < 2e-16  ***
 dialect7:target1         -6.95502     0.71801  1542.97818   -9.687    < 2e-16  ***
 dialect8:target1         -7.51118     1.04372  1544.68857   -7.197   9.60e-13  ***
 dialect1:gender1         -1.18394     1.35214    71.89758   -0.876   0.384160
 dialect2:gender1          0.09797     1.35677    73.80749    0.072   0.942634
 dialect3:gender1          1.45577     1.34474    71.71264    1.083   0.282628
 dialect4:gender1         -3.03698     1.47746    71.49376   -2.056   0.043480  *
 dialect5:gender1          1.26780     1.34325    71.44366    0.944   0.348440
 dialect6:gender1          1.03862     1.35587    74.14801    0.766   0.446098
 dialect7:gender1          0.44025     1.36423    72.34726    0.323   0.747845
 dialect8:gender1         -0.07948     1.87189    79.58999   -0.042   0.966240
 target1:gender1           1.10884     0.28944  1540.81442    3.831   0.000133  ***
 dialect1:target1:gender1 -2.00901     0.69806  1541.17046   -2.878   0.004057  **
 dialect2:target1:gender1  0.33991     0.71726  1541.24798    0.474   0.635639
 dialect3:target1:gender1 -0.52172     0.70114  1539.44932   -0.744   0.456924
 dialect4:target1:gender1 -1.01354     0.76932  1537.97162   -1.317   0.187884
 dialect5:target1:gender1  1.37824     0.69916  1538.58232    1.971   0.048869  *
 dialect6:target1:gender1  1.94852     0.72373  1542.35914    2.692   0.007172  **
 dialect7:target1:gender1 -0.10402     0.71411  1540.02524   -0.146   0.884205
 dialect8:target1:gender1 -0.01838     1.04012  1543.39297   -0.018   0.985907

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix (B.6) Section § 3.3.4 Model summary for the F2 lowering in the voiceless subset.

Formula: normf2 ~ dialect * target * gender + position + fol.v + (1 |item) + (1|subject)
Data: formants.unvoiced
REML criterion at convergence: 1773.8
Scaled residuals:

Min	1Q	Median	3Q	Max
-7.1641	-0.5088	0.0104	0.5164	5.6051

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	0.007648	0.08745
item	(Intercept)	0.109438	0.33081
Residual		0.158507	0.39813

Number of obs: 1505, groups: subject, 88; item, 22

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-1.974e-01	1.580e-01	1.341e+01	-1.250	0.23269	
dialect1	1.128e-01	3.608e-02	6.613e+01	3.126	0.00263	**
dialect2	-7.101e-02	3.586e-02	6.858e+01	-1.980	0.05167	.
dialect3	7.114e-03	3.573e-02	6.848e+01	0.199	0.84278	
dialect4	-1.229e-01	3.866e-02	6.463e+01	-3.179	0.00227	**
dialect5	8.484e-02	3.568e-02	6.862e+01	2.377	0.02022	*
dialect6	3.660e-02	3.620e-02	7.251e+01	1.011	0.31527	
dialect7	-6.121e-02	3.658e-02	7.113e+01	-1.673	0.09865	.
dialect8	1.380e-02	4.998e-02	7.748e+01	0.276	0.78320	
target1	2.771e-01	7.173e-02	1.358e+01	3.863	0.00181	**
gender1	-3.283e-03	1.447e-02	6.950e+01	-0.227	0.82112	
position1	7.098e-02	1.670e-01	1.324e+01	0.425	0.67762	
fol.v1	1.676e+00	2.084e-01	1.326e+01	8.042	1.85e-06	***
fol.v2	-1.106e+00	1.577e-01	1.332e+01	-7.014	8.01e-06	***
fol.v3	-2.514e-01	1.581e-01	1.345e+01	-1.590	0.13502	
fol.v4	2.148e-01	1.612e-01	1.435e+01	1.332	0.20352	
fol.v5	-5.958e-01	2.086e-01	1.332e+01	-2.856	0.01323	*
fol.v6	6.290e-02	1.576e-01	1.328e+01	0.399	0.69612	
dialect1:target1	-8.416e-02	2.652e-02	1.393e+03	-3.173	0.00154	**
dialect2:target1	-1.264e-02	2.658e-02	1.392e+03	-0.476	0.63443	
dialect3:target1	4.099e-02	2.650e-02	1.389e+03	1.547	0.12213	
dialect4:target1	-2.289e-02	2.834e-02	1.384e+03	-0.808	0.41936	
dialect5:target1	7.594e-02	2.648e-02	1.384e+03	2.867	0.00420	**
dialect6:target1	-1.196e-01	2.719e-02	1.403e+03	-4.400	1.17e-05	***
dialect7:target1	3.284e-02	2.732e-02	1.392e+03	1.202	0.22949	
dialect8:target1	8.957e-02	3.796e-02	1.392e+03	2.359	0.01844	*
dialect1:gender1	-2.574e-02	3.598e-02	6.541e+01	-0.715	0.47688	
dialect2:gender1	8.261e-02	3.582e-02	6.825e+01	2.307	0.02411	*
dialect3:gender1	-3.720e-02	3.570e-02	6.824e+01	-1.042	0.30111	
dialect4:gender1	1.976e-02	3.862e-02	6.432e+01	0.512	0.61060	
dialect5:gender1	-2.536e-02	3.559e-02	6.793e+01	-0.712	0.47868	
dialect6:gender1	6.022e-04	3.585e-02	6.985e+01	0.017	0.98665	
dialect7:gender1	-5.874e-02	3.647e-02	7.035e+01	-1.610	0.11180	
dialect8:gender1	4.406e-02	4.991e-02	7.706e+01	0.883	0.38017	
target1:gender1	-1.341e-02	1.076e-02	1.389e+03	-1.246	0.21286	
dialect1:target1:gender1	6.970e-03	2.639e-02	1.392e+03	0.264	0.79173	
dialect2:target1:gender1	-1.333e-02	2.652e-02	1.391e+03	-0.503	0.61526	
dialect3:target1:gender1	-2.480e-03	2.646e-02	1.389e+03	-0.094	0.92535	
dialect4:target1:gender1	-7.413e-02	2.828e-02	1.383e+03	-2.622	0.00884	**
dialect5:target1:gender1	-2.804e-02	2.636e-02	1.383e+03	-1.064	0.28767	
dialect6:target1:gender1	5.154e-02	2.673e-02	1.391e+03	1.928	0.05407	.
dialect7:target1:gender1	3.165e-02	2.718e-02	1.391e+03	1.164	0.24444	
dialect8:target1:gender1	2.783e-02	3.788e-02	1.391e+03	0.735	0.46267	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix C (for chapter 4)

Appendix (C.1) List of stimuli and their Arabic gloss for targets in word-initial position.

voicing	target	test word	gloss	Arabic script
voiced	/b/	bi:r	well	بير
		bu:t	boot	بوت
		ba:b	door	باب
		bin	son of	بِن
		bunn	coffee beans	بُن
		baṭ	duck	بَط
	/d/	di:n	religion	دين
		du:n	under	دون
		da:r	house	دار
		dubb	bear	دُب
		damm	blood	دَم
		darb	path	دَرب
	/g/	gi:ran	a name	قيران
		gu:ṭi:	a can	قوطي
		ga:m	stood up	قام
		gird	monkey	قِرد
		gubgub	crab	قبقب
		garaʕ	pumpkin	قرع
	/ḍ/	ḍi:q	hardship	ضيق
		ḍa:r	harmful	ضار
		ḍubb	collect/gather	ضُب
		ḍumm	included	ضَم
		ḍarb	hit	ضرب
		ḍuru:b	kinds	ضروب
voiceless	/t/	ti:n	fig	تين
		tu:b	repent	توب
		ta:r	nonsense word	تار
		tifil	coffee grounds	تِفِل
		tubiʕ	followed by	تُبع
		tamir	dates	تمر
		tayyar	current	تيّار
	/k/	ki:s	bag	كيس
		ku:b	mug	كوب
		ka:n	was	كان

	kitab	book	كتاب
	kumm	sleeve	كُمْ
	kam	how much	كَمْ
/q/	qi:l	has been said	قيل
	qu:t	sustenance	قوت
	qa:rrah	continent	قارة
	qimma	summit	قِمة
	qumʕ	funne	قُمع
	qazam	suppress	قزم
/ṭ/	ṭi:n	mud	طين
	ṭu:b	bricks	طوب
	ṭa:r	flew	طار
	ṭifil	child	طفل
	ṭubiʕ	was printed	طُبع
	ṭamir	covering	طَمر
	ṭayyar	pilot	طيّار

Appendix (C.2) List of stimuli and their Arabic gloss for targets in word-medial position.

voicing	target	test word	gloss	Arabic script
voiced	/d/	mudirr	passing of a liquid	مُدِر
		madru:s	studied	مدروس
		taxdi:rhum	anesthetize	تخديرهم
		taʕdi:lat	modifications	تعديلات
		ʔdwija	medicines	أدوية
	/ḍ/	maḍu:n	nonsense word	ماضون
		muḍirr	harmful	مُضِر
		maḍru:b	hit	مضروب
		taxḍi:rhum	to make sth green	تخضيرهم
		taʕḍi:lat	nonsense word	تعضيلات
		ʔḍwija	lights	أضوية
voiceless	/t/	taʕti:lat	nonsense word	تعتيلات
	/ṭ/	taʕṭi:lat	obstructions	تعطيلات
		tiṭbaʕ	to print (f)	تِطبَع

Appendix (C.3) Section § 4.4.3 Model summary for the voiced subset.

Formula: vot ~ group * target * gender * position + (1 | item) + (1 + target | speaker)
Data: v3 (data is subset of voiced + slow + only 4 targets)
REML criterion at convergence: 11131.9
Scaled residuals:

Min	1Q	Median	3Q	Max
-8.2537	-0.4851	0.0745	0.5467	4.2614

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	80.78	8.987	
	target1	14.92	3.863	-0.69
item	(Intercept)	33.52	5.789	
Residual	404.32	20.108		

Number of obs: 1249, groups: speaker, 64; item, 23

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-68.54722	1.77528	44.28524	-38.612	< 2e-16	***
group1	-1.92850	1.88905	54.51102	-1.021	0.31182	
group2	-3.06160	1.88572	54.12833	-1.624	0.11028	
target1	3.54150	1.43769	23.65259	2.463	0.02144	*
gender1	-4.29249	1.29765	55.40694	-3.308	0.00166	**
position1	-5.87959	1.34693	18.81524	-4.365	0.00034	***
group1:target1	0.83550	1.11723	55.03026	0.748	0.45775	
group2:target1	3.70025	1.11196	53.99610	3.328	0.00158	**
group3:target1	-4.53575	1.05732	64.11373	-4.290	6.15e-05	***
group1:gender1	0.10937	1.88880	54.48658	0.058	0.95404	
group2:gender1	-2.71349	1.88545	54.10212	-1.439	0.15586	
target1:gender1	0.74232	0.77409	57.12686	0.959	0.34162	
group1:position1	-2.30194	0.84630	1108.56553	-2.720	0.00663	**
group2:position1	-1.60952	0.83789	1108.77732	-1.921	0.05500	.
target1:position1	1.22555	1.34682	18.80965	0.910	0.37436	
gender1:position1	0.78718	0.58879	1115.84223	1.337	0.18151	
group1:target1:gender1	-0.41024	1.11681	54.98549	-0.367	0.71478	
group2:target1:gender1	1.29641	1.11150	53.94524	1.166	0.24860	
group1:target1:position1	0.47064	0.84568	1110.88023	0.557	0.57797	
group2:target1:position1	-0.33773	0.83775	1109.31049	-0.403	0.68692	
group1:gender1:position1	0.62460	0.84598	1108.32715	0.738	0.46048	
group2:gender1:position1	-0.37806	0.83762	1108.58450	-0.451	0.65182	
target1:gender1:position1	-0.27495	0.58859	1116.79970	-0.467	0.64050	
group1:target1:gender1:position1	-0.02201	0.84539	1110.58338	-0.026	0.97923	
group2:target1:gender1:position1	1.07660	0.83746	1109.14209	1.286	0.19887	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (C.4) Section § 4.4.4 Model summary for the voiceless subset.

Formula: vot ~ group * target * gender * position + (1 | item) + (1 | speaker)
Data: uv3 (data is subset of voiceless + slow + only 4 targets)
REML criterion at convergence: 8639.4
Scaled residuals:

Min	1Q	Median	3Q	Max
-2.4422	-0.6025	-0.0740	0.4386	8.8900

Random effects:

Groups	Name	Variance	Std.Dev.
speaker	(Intercept)	29.61	5.442
item	(Intercept)	97.12	9.855
Residual		175.33	13.241

Number of obs: 1069, groups: speaker, 64; item, 17

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	40.18241	3.41590	14.22710	11.763	1.01e-08	***
group1	3.77450	1.32417	88.23403	2.850	0.005434	**
group2	-3.79530	1.33746	91.68948	-2.838	0.005594	**
group3	0.0208	1.1956	90.2625	0.017	0.986159	
target1	14.94783	3.34372	13.07308	4.470	0.000622	***
gender1	3.36761	0.91024	90.02377	3.700	0.000371	***
position1	-6.33935	3.34371	13.07299	-1.896	0.080294	.
group1:target1	10.19057	0.84273	974.29103	12.092	< 2e-16	***
group2:target1	-3.41873	0.86347	975.83629	-3.959	8.06e-05	***
group3:target1	-6.7718	0.7672	974.9381	-8.827	< 2e-16	***
group1:gender1	-2.38988	1.32417	88.23548	-1.805	0.074516	.
group2:gender1	-0.58281	1.33737	91.66648	-0.436	0.664015	
group3:gender1	2.9727	1.1955	90.2329	2.487	0.014740	*
target1:gender1	2.37545	0.58364	975.01267	4.070	5.08e-05	***
group1:position1	4.19646	0.84271	974.25447	4.980	7.53e-07	***
group2:position1	1.33401	0.86318	975.31577	1.545	0.122559	
group3:position1	-5.5305	0.7674	975.4108	-7.206	1.15e-12	***
target1:position1	-3.01005	3.34376	13.07366	-0.900	0.384297	
gender1:position1	-2.01086	0.58362	974.94922	-3.445	0.000594	***
group1:target1:gender1	0.93535	0.84274	974.29834	1.110	0.267322	
group2:target1:gender1	-1.22975	0.86334	975.79911	-1.424	0.154646	
group3:target1:gender1	0.2944	0.7670	974.8972	0.384	0.701201	
group1:target1:position1	2.53262	0.84295	974.67837	3.004	0.002728	**
group2:target1:position1	0.09436	0.86392	976.61008	0.109	0.913050	
group3:target1:position1	-2.6270	0.7674	975.3344	-3.423	0.000645	***
group1:gender1:position1	1.20465	0.84273	974.26030	1.429	0.153191	
group2:gender1:position1	1.05924	0.86305	975.29068	1.227	0.219996	
group2:gender1:position1	-2.2639	0.7673	975.3603	-2.951	0.003248	**
target1:gender1:position1	-1.61262	0.58385	975.54668	-2.762	0.005852	**
group1:target1:gender1:position1	0.57674	0.84296	974.68273	0.684	0.494024	
group2:target1:gender1:position1	0.32584	0.86378	976.56447	0.377	0.706092	
group2:target1:gender1:position1	-0.9026	0.7672	975.2903	-1.176	0.239727	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (C.5) Section § 4.4.5 Model summary for the F2 lowering in the voiceless subset.

Formula: f225 ~ group * target + gender + folv + (1 | item)
Data: uf2 (data is subset of voiceless + slow + only 4 targets)
REML criterion at convergence: 1057.6
Scaled residuals:

Min	1Q	Median	3Q	Max
-8.9374	-0.4769	0.0120	0.5122	3.8337

Random effects:

Groups	Name	Variance	Std.Dev.
speaker	(Intercept)	0.00000	0.0000
item	(Intercept)	0.09914	0.3149
Residual		0.19515	0.4418

Number of obs: 812, groups: speaker, 64; item, 14
Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-0.01330	0.08852	6.96078	-0.150	0.884866	
group1	-0.08403	0.02258	793.08508	-3.721	0.000212	***
group2	0.04181	0.02263	793.10595	1.848	0.064996	.
group3	0.04222	0.02149	793.52241	1.965	0.049802	*
target1	0.36132	0.08560	6.96336	4.221	0.003979	**
gender1	-0.02870	0.01556	792.97862	-1.844	0.065557	.
folv1	0.26436	0.20558	7.03976	1.286	0.239148	
folv2	0.04049	0.15793	6.96645	0.256	0.805040	
folv3	-0.67256	0.20519	6.98677	-3.278	0.013566	*
folv4	1.77514	0.20469	6.91893	8.672	5.81e-05	***
folv5	-0.28533	0.20471	6.92172	-1.394	0.206480	
group1:target1	0.07929	0.02253	793.26975	3.519	0.000457	***
group2:target1	0.13627	0.02255	793.25490	6.043	2.33e-09	***
group3:target1	-0.21556	0.02148	794.21176	-10.035	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Correlation of Fixed Effects:

	(Intr)	group1	group2	targt1	gendr1	folv1	folv2	folv3	folv4	folv5
group1	0.008									
group2	0.009	-0.548								
target1	-0.001	0.002	0.003							
gender1	0.001	0.066	-0.081	0.001						
folv1	0.042	-0.01	-0.009	-0.001	0.001					
folv2	-0.254	0.000	-0.003	0.000	0.001	-0.155				
folv3	0.04	-0.003	-0.002	-0.001	0.000	-0.221	0.155			
folv4	0.038	0.005	0.004	0.000	0.000	-0.221	-0.153	-0.220		
folv5	0.038	0.004	0.005	0.001	-0.001	-0.221	-0.154	-0.220	-0.219	
grop1:trgt1	0.002	-0.011	-0.007	0.008	-0.001	0.004	-0.002	0.002	-0.001	-0.002
grop2:trgt1	0.003	-0.008	-0.009	0.008	-0.003	0.005	-0.002	0.000	-0.002	-0.001

	grp1:1
grop1:trgt1	
grop2:trgt1	-0.546

Appendix (C.6) Section § 4.5.1.1 Model summary for effects of speed on word duration.

Formula: worddur ~ speed * group + (1 | item)
Data: vot3

REML criterion at convergence: 81868.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9802	-0.5597	-0.0640	0.4563	12.4593

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	11218	105.92
Residual		6559	80.99

Number of obs: 7015, groups: item, 64

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	345.8221	13.2786	62.9536	26.044	< 2e-16	***
speed1	66.2662	0.9831	6946.1117	67.408	< 2e-16	***
group1	20.2336	1.4376	6947.8240	14.075	< 2e-16	***
group2	9.7786	1.4360	6947.1251	6.810	1.06e-11	***
group3	-30.0122	1.3350	6952.0473	-22.480	< 2e-16	***
speed1:group1	3.9488	1.4291	6946.0718	2.763	0.00574	**
speed1:group2	2.3770	1.4298	6946.0168	1.663	0.09645	.
speed1:group3	-6.3258	1.3075	6946.1426	-4.838	1.34e-06	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

	(Intr)	speed1	group1	group2	spd1:1
speed1	-0.001				
group1	0.004	-0.009			
group2	0.005	-0.009	-0.568		
speed1:grp1	-0.001	0.079	-0.016	0.019	
speed1:grp2	-0.001	0.081	0.016	-0.017	-0.582

Appendix (C.7) Section § 4.5.3 Model summary for effects of speech rate on Najdi /t ~ ṭ/ and /d ~ ḍ/.

Formula: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker)
Data: sanjVOT

REML criterion at convergence: 11178.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.8045	-0.4078	-0.0033	0.5006	7.1689

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	30.19	5.494
speaker	(Intercept)	15.79	3.973
Residual		274.86	16.579

Number of obs: 1315, groups: item, 40; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-10.7007	1.4320	37.1414	-7.472	6.56e-09	***
speed1	-1.8830	0.4618	1264.5302	-4.077	4.84e-05	***
target1	71.8607	1.9024	33.6221	37.774	< 2e-16	***
target2	30.4920	1.7671	33.0501	17.255	< 2e-16	***
target3	-47.0502	1.6855	34.2594	-27.914	< 2e-16	***
target4	-55.3025	1.6749	35.4833	-33.018	< 2e-16	***
position1	-4.7567	1.0685	44.8997	-4.452	5.58e-05	***
gender1	-1.2643	1.0488	16.0362	-1.205	0.245521	
speed1:target1	11.0198	0.8391	1276.3286	13.133	< 2e-16	***
speed1:target2	2.8034	0.7993	1257.0096	3.507	0.000469	***
speed1:target3	-6.4592	0.7746	1262.5271	-8.338	< 2e-16	***
speed1:target4	-7.3640	0.7810	1259.9609	-9.429	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	speed1	targt1	targt2	targt3	postn1	gendr1	spd1:1	spd1:2
speed1	-0.033								
target1	0.150	-0.017							
target2	0.054	0.000	-0.344						
target3	-0.096	0.006	-0.384	-0.335					
position1	-0.290	0.081	-0.268	-0.101	0.178				
gender1	0.079	0.000	0.002	0.001	0.002	0.001			
sped1:trgt1	-0.027	0.087	-0.034	-0.014	0.024	0.105	-0.002		
sped1:trgt2	0.020	-0.002	0.010	0.001	-0.004	-0.048	-0.001	-0.367	
sped1:trgt3	0.001	-0.052	0.012	0.006	-0.028	-0.023	-0.002	-0.348	-0.314

Appendix (C.8) Section § 4.5.4 Model summary for effects of speech rate on Hijazi /t ~ ṭ/ and /d ~ ḍ/.

```
Formula: vot ~ speed * target + position + gender + (1|item) + (1 + target|speaker)
   Data: sahiVOT

REML criterion at convergence: 11360.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-10.2380  -0.4379  -0.0246   0.4856   3.8918

Random effects:
 Groups    Name         Variance   Std.Dev.   Corr
 item      (Intercept)    19.396     4.404
 speaker   (Intercept)    46.119     6.791
           target1        29.290     5.412    -0.04
           target2        14.788     3.846    -0.87    0.51
           target3         8.787     2.964     0.32   -0.36   -0.59
 Residual                239.451    15.474

Number of obs: 1352, groups:  item, 40; speaker, 18

Fixed effects:
                  Estimate   Std. Error          df    t value    Pr(>|t|)
 (Intercept)      -14.7344       1.8394     20.2345     -8.010    1.05e-07   ***
 speed1            -3.0509       0.4268   1263.6595     -7.149    1.47e-12   ***
 target1           57.4663       2.0353     36.5367     28.235     < 2e-16   ***
 target2           41.5827       1.7293     37.6511     24.046     < 2e-16   ***
 target3          -43.0983       1.5603     35.6966    -27.621     < 2e-16   ***
 target4          -55.9504       2.1584     30.3742    -25.923     < 2e-16   ***
 position1         -6.2498       0.9011     42.5324     -6.936    1.69e-08   ***
 gender1            0.7661       1.2633     19.0491      0.606       0.551
 speed1:target1     7.2173       0.7866   1271.6648      9.176     < 2e-16   ***
 speed1:target2     3.7106       0.7524   1258.7009      4.932    9.23e-07   ***
 speed1:target3    -4.1017       0.7077   1261.5871     -5.796    8.57e-09   ***
 speed1:target4    -6.8273       0.7036   1262.9682     -9.703     < 2e-16   ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Correlation of Fixed Effects:

	(Intr)	speed1	targt1	targt2	targt3	postn1	gendr1	spd1:1	spd1:2
speed1	-0.019								
target1	0.055	-0.015							
target2	-0.361	-0.002	-0.062						
target3	0.066	0.008	-0.371	-0.396					
position1	-0.189	0.068	-0.213	-0.094	0.163				
gender1	-0.078	0.007	0.000	-0.003	0.001	0.001			
sped1:trgt1	-0.015	0.111	-0.026	-0.006	0.019	0.081	-0.002		
sped1:trgt2	0.008	0.032	0.008	-0.002	-0.002	-0.033	0.007	-0.387	
sped1:trgt3	0.004	-0.074	0.011	0.006	-0.020	-0.030	-0.002	-0.353	-0.319

Appendix (C.9) Section § 4.5.5 Model summary for effects of speech rate on Tunisian /t ~ ṭ/ and /d ~ ḍ/.

Formula: vot ~ speed * target + position + gender + folv + (1 | item) + (1 | speaker)
Data: tunsVOT

REML criterion at convergence: 16469.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.5180	-0.5068	-0.0253	0.5380	5.6061

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	68.44	8.273
speaker	(Intercept)	19.36	4.400
Residual		244.15	15.625

Number of obs: 1964, groups: item, 40; speaker, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-13.66369	1.78059	46.81579	-7.674	8.05e-10	***
speed1	-2.72083	0.36118	1918.72362	-7.533	7.58e-14	***
target1	45.37149	2.64697	28.19699	17.141	< 2e-16	***
target2	40.86842	2.45868	27.45005	16.622	7.40e-16	***
target3	-43.75581	2.31980	27.50575	-18.862	< 2e-16	***
target4	-42.48409	2.37750	29.82182	-17.869	< 2e-16	***
position1	-2.70026	1.63421	112.99632	-1.652	0.10124	
gender1	0.10803	0.90371	25.70962	0.120	0.90578	
folv1	-0.88911	3.95135	28.13457	-0.225	0.82359	
folv2	-1.93670	3.26339	29.68415	-0.593	0.55737	
folv3	-2.80044	3.55381	30.98758	-0.788	0.43668	
folv4	11.56599	2.79513	29.32506	4.138	0.00027	***
folv5	0.01526	4.02046	29.00012	0.004	0.99700	
folv6	-3.47016	3.90139	27.38490	-0.889	0.38150	
speed1:target1	6.80421	0.65162	1910.66583	10.442	< 2e-16	***
speed1:target2	4.40809	0.61815	1902.78555	7.131	1.41e-12	***
speed1:target3	-5.86284	0.57903	1904.87030	-10.125	< 2e-16	***
speed1:target4	-5.34946	0.65120	1905.45809	-8.215	3.88e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.10) Section § 4.6.3.1 Model summary for effects of speech rate on Najdi VOT for /t, ṭ, k, q/ and /b, d, ḍ, g/.

Formula: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker)
Data: sanjVOT

REML criterion at convergence: 17204.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.9324	-0.4456	-0.0038	0.5577	7.2598

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	28.07	5.298
speaker	(Intercept)	15.64	3.954
Residual		265.45	16.292

Number of obs: 2035, groups: item, 64; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-9.8351	1.4227	48.2227	-6.913	9.78e-09	***
speed1	-1.3902	0.3834	1956.6537	-3.626	0.000296	***
target1	71.0129	1.9975	52.4414	35.550	< 2e-16	***
target2	29.6330	1.8977	52.3344	15.615	< 2e-16	***
target3	76.1075	2.2716	52.2472	33.504	< 2e-16	***
target4	34.8144	2.4385	65.3520	14.277	< 2e-16	***
target5	-62.1436	2.2821	53.2128	-27.231	< 2e-16	***
target6	-47.9363	1.8640	55.3402	-25.717	< 2e-16	***
target7	-56.2316	1.8576	57.6866	-30.271	< 2e-16	***
target8	-45.2564	2.3523	59.8375	-19.239	< 2e-16	***
position1	-4.7992	1.0364	70.3310	-4.631	1.62e-05	***
gender1	-1.3871	1.0061	15.9710	-1.379	0.187016	
speed1:target1	10.5200	0.9269	1982.7124	11.350	< 2e-16	***
speed1:target2	2.3074	0.8738	1948.3447	2.641	0.008340	**
speed1:target3	12.1787	1.0359	1947.8627	11.756	< 2e-16	***
speed1:target4	0.9483	1.2536	1962.5958	0.756	0.449468	
speed1:target5	-7.7663	1.0586	1948.7942	-7.336	3.21e-13	***
speed1:target6	-6.9650	0.8417	1957.9614	-8.274	2.36e-16	***
speed1:target7	-7.8608	0.8496	1952.9024	-9.252	< 2e-16	***
speed1:target8	-3.3622	1.1883	1951.7568	-2.830	0.004709	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (C.11) Section § 4.6.3.2 Model summary for effects of speech rate on Najdi closure duration for /t, ṭ, k, q/ and /b, d, ḍ, g/.

Formula: closdur ~ target * speed + gender + position + (1 | item) + (1 |speaker)
Data: sanjVOT (CD)

REML criterion at convergence: 17658.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.0359	-0.6230	-0.0752	0.4903	9.4738

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	30.72	5.542
speaker	(Intercept)	34.91	5.909
Residual		331.91	18.218

Number of obs: 2035, groups: item, 64; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	58.8752	1.8074	34.3149	32.575	< 2e-16	***
target1	-10.2690	2.1216	51.1481	-4.840	1.24e-05	***
target2	0.7249	2.0155	51.0360	0.360	0.720584	
target3	-9.5019	2.4125	50.9174	-3.939	0.000250	***
target4	2.8584	2.6046	64.5807	1.097	0.276523	
target5	13.7818	2.4248	51.9618	5.684	6.07e-07	***
target6	-1.0375	1.9825	54.0283	-0.523	0.602894	
target7	7.0736	1.9778	56.4640	3.576	0.000723	***
target8	-3.6304	2.5073	59.1189	-1.448	0.152908	
speed1	9.2296	0.4287	1955.6093	21.528	< 2e-16	***
gender1	1.4679	1.4594	16.0381	1.006	0.329422	
position1	4.3813	1.1084	67.4019	3.953	0.000188	***
target1:speed1	0.5696	1.0359	1980.3777	0.550	0.582429	
target2:speed1	2.6988	0.9770	1947.2281	2.762	0.005794	**
target3:speed1	-1.0302	1.1584	1946.7422	-0.889	0.373908	
target4:speed1	2.8606	1.4015	1962.1680	2.041	0.041373	*
target5:speed1	-0.1420	1.1837	1947.6454	-0.120	0.904523	
target6:speed1	-0.8700	0.9411	1957.6880	-0.924	0.355378	
target7:speed1	0.1415	0.9500	1952.0196	0.149	0.881636	
target8:speed1	-4.2283	1.3287	1950.4703	-3.182	0.001485	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.12) Section § 4.6.3.3 Model summary for effects of speech rate on Najdi f0 on vowels following /t, ṭ, k, q/ and /b, d, ḍ, g/.

Formula: f025 ~ speed * target + gender + folv + (1 | item) + (1 | speaker)
Data: f0sanj

REML criterion at convergence: 4897

Scaled residuals:

Min	1Q	Median	3Q	Max
-12.5936	-0.5102	-0.0337	0.5615	4.2552

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.1945	0.441
speaker	(Intercept)	5.4987	2.345
Residual		1.1791	1.086

Number of obs: 1556, groups: item, 50; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	8.51840	0.56060	16.42940	15.195	4.23e-11	***
speed1	-0.24486	0.02812	1484.10226	-8.707	< 2e-16	***
target1	0.54919	0.17142	34.21238	3.204	0.002933	**
target2	0.09475	0.17162	34.36755	0.552	0.584446	
target3	0.26006	0.18487	35.57353	1.407	0.168173	
target4	0.43410	0.19319	41.27534	2.247	0.030046	*
target5	-0.27797	0.18373	34.82797	-1.513	0.139320	
target6	-0.20637	0.18512	34.44653	-1.115	0.272645	
target7	-0.54128	0.19489	36.23398	-2.777	0.008624	**
target8	-0.31248	0.18741	37.63855	-1.667	0.103751	
gender1	4.63091	0.55684	15.99539	8.316	3.35e-07	***
folv1	-0.55952	0.13909	34.84965	-4.023	0.000294	***
folv2	-0.40205	0.15596	35.32826	-2.578	0.014265	*
folv3	0.41863	0.18071	36.49442	2.316	0.026255	*
folv4	0.12332	0.15913	37.54845	0.775	0.443202	
folv5	0.23318	0.14935	35.55259	1.561	0.127301	
speed1:target1	0.05154	0.06613	1480.93300	0.779	0.435906	
speed1:target2	0.23312	0.06658	1480.81102	3.502	0.000476	***
speed1:target3	-0.11871	0.07422	1487.20227	-1.600	0.109900	
speed1:target4	0.19454	0.08562	1489.89026	2.272	0.023225	*
speed1:target5	-0.20052	0.07265	1482.18612	-2.760	0.005854	**
speed1:target6	0.10125	0.07134	1480.81015	1.419	0.156032	
speed1:target7	-0.17795	0.07658	1484.86554	-2.324	0.020270	*
speed1:target8	-0.08328	0.08012	1482.10621	-1.039	0.298758	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.13) Section § 4.6.4.1 Model summary for effects of speech rate on Hijazi VOT for /t, ṭ, k, q/ and /b, d, ḍ, g/.

Formula: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker)
Data: sahiVOT

REML criterion at convergence: 17292.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.1501	-0.5167	0.0067	0.5364	6.3091

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	15.63	3.953
speaker	(Intercept)	36.41	6.034
Residual		286.33	16.921

Number of obs: 2029, groups: item, 64; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-10.6848	1.6878	26.0161	-6.331	1.05e-06	***
speed1	-1.9086	0.4084	1946.2407	-4.674	3.16e-06	***
target1	53.6315	1.6429	52.8719	32.644	< 2e-16	***
target2	37.8012	1.5610	52.7848	24.217	< 2e-16	***
target3	66.4476	1.8727	53.0262	35.482	< 2e-16	***
target4	47.0118	2.0864	78.6254	22.533	< 2e-16	***
target5	-60.6314	1.8626	51.9119	-32.553	< 2e-16	***
target6	-46.8939	1.5270	53.8355	-30.709	< 2e-16	***
target7	-60.0557	1.5255	56.3507	-39.368	< 2e-16	***
target8	-37.3111	2.0100	70.0163	-18.563	< 2e-16	***
position1	-6.3289	0.8694	61.6079	-7.280	7.27e-10	***
gender1	-1.1807	1.4809	15.9704	-0.797	0.436968	
speed1:target1	6.1496	0.9742	1962.9084	6.313	3.38e-10	***
speed1:target2	2.4702	0.9250	1943.0161	2.670	0.007641	**
speed1:target3	9.4159	1.1037	1943.3548	8.531	< 2e-16	***
speed1:target4	4.8224	1.3972	1945.3918	3.452	0.000569	***
speed1:target5	-5.1711	1.0864	1942.0110	-4.760	2.08e-06	***
speed1:target6	-5.2390	0.8591	1944.8130	-6.098	1.29e-09	***
speed1:target7	-8.0128	0.8521	1949.2904	-9.404	< 2e-16	***
speed1:target8	-4.4352	1.3110	1945.6107	-3.383	0.000731	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.14) Section § 4.6.4.2 Model summary for effects of speech rate on Hijazi Closure duration for /t, ṭ, k, q/ and /b, d, ḍ, g/.

Formula: closdur ~ target * speed + gender + position + (1 | item) + (1 | speaker)
Data: sahiVOT (CD)

REML criterion at convergence: 17825.8

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.9467	-0.6130	-0.0881	0.4876	8.5470

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	28.22	5.312
speaker	(Intercept)	54.24	7.365
Residual		370.71	19.254

Number of obs: 2029, groups: item, 64; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	62.8681	2.0809	27.4766	30.213	< 2e-16	***
target1	-6.3098	2.0941	51.1069	-3.013	0.004016	**
target2	3.7829	1.9898	51.0269	1.901	0.062937	.
target3	-10.6397	2.3866	51.2407	-4.458	4.53e-05	***
target4	10.2417	2.6096	71.3235	3.925	0.000198	***
target5	11.4532	2.3763	50.3735	4.820	1.36e-05	***
target6	-4.9055	1.9443	52.1625	-2.523	0.014720	*
target7	7.6279	1.9378	54.2248	3.936	0.000238	***
target8	-11.2506	2.5276	64.2507	-4.451	3.48e-05	***
speed1	8.6096	0.4647	1944.3541	18.526	< 2e-16	***
gender1	2.4344	1.7996	15.9934	1.353	0.194952	
position1	3.6731	1.0981	61.2876	3.345	0.001409	**
target1:speed1	-1.2341	1.1093	1964.8758	-1.112	0.266075	
target2:speed1	2.7357	1.0526	1940.8142	2.599	0.009422	**
target3:speed1	-1.9842	1.2560	1940.9994	-1.580	0.114315	
target4:speed1	2.7690	1.5900	1942.9102	1.742	0.081748	.
target5:speed1	-0.9902	1.2362	1939.8933	-0.801	0.423223	
target6:speed1	-1.6913	0.9776	1942.4118	-1.730	0.083783	.
target7:speed1	1.2152	0.9698	1946.2773	1.253	0.210321	
target8:speed1	-0.8202	1.4920	1943.0171	-0.550	0.582557	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (C.15) Section § 4.6.4.3 Model summary for effects of speech rate on Hijazi f0 on vowels following /t, ṭ, k, q/ and /b, d, ḍ, ɡ/.

Formula: f025 ~ speed * target + gender + folv + (1 | item) + (1 | speaker)
Data: f0sahi

REML criterion at convergence: 5205.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-8.9791	-0.4377	-0.0475	0.3702	19.1458

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.1601	0.4002
speaker	(Intercept)	2.4558	1.5671
Residual		1.6728	1.2934

Number of obs: 1499, groups: item, 50; speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	7.361e+00	3.779e-01	1.681e+01	19.479	5.70e-13	***
speed1	-1.297e-01	3.446e-02	1.425e+03	-3.762	0.000175	***
target1	3.410e-01	1.647e-01	3.415e+01	2.070	0.046051	*
target2	1.464e-01	1.647e-01	3.413e+01	0.889	0.380183	
target3	3.774e-01	1.789e-01	3.657e+01	2.109	0.041832	*
target4	5.432e-01	1.909e-01	4.678e+01	2.845	0.006566	**
target5	-2.389e-01	1.760e-01	3.431e+01	-1.358	0.183432	
target6	-2.862e-01	1.766e-01	3.341e+01	-1.620	0.114566	
target7	-4.310e-01	1.873e-01	3.618e+01	-2.301	0.027231	*
target8	-4.520e-01	1.844e-01	4.127e+01	-2.452	0.018541	*
gender1	4.306e+00	3.732e-01	1.600e+01	11.537	3.62e-09	***
folv1	-3.147e-01	1.346e-01	3.549e+01	-2.338	0.025143	*
folv2	-2.431e-01	1.504e-01	3.563e+01	-1.616	0.114862	
folv3	4.244e-01	1.740e-01	3.670e+01	2.440	0.019652	*
folv4	-5.321e-02	1.520e-01	3.685e+01	-0.350	0.728331	
folv5	1.797e-01	1.449e-01	3.678e+01	1.241	0.222610	
speed1:target1	2.601e-03	8.068e-02	1.426e+03	0.032	0.974291	
speed1:target2	1.596e-02	8.047e-02	1.425e+03	0.198	0.842802	
speed1:target3	1.361e-01	9.175e-02	1.425e+03	1.483	0.138289	
speed1:target4	-6.685e-02	1.100e-01	1.425e+03	-0.608	0.543361	
speed1:target5	-2.215e-01	8.663e-02	1.424e+03	-2.557	0.010672	*
speed1:target6	1.216e-01	8.399e-02	1.424e+03	1.448	0.147877	
speed1:target7	1.508e-02	9.089e-02	1.426e+03	0.166	0.868282	
speed1:target8	-2.978e-03	1.014e-01	1.426e+03	-0.029	0.976565	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.16) Section § 4.6.5.1 Model summary for effects of speech rate on Tunisian VOT for /t, ṭ, k, q/ and /b, d, ḍ/.

Formula: vot ~ speed * target + position + gender + (1 | item) + (1 | speaker)
Data: tunsVOT

REML criterion at convergence: 24928.3

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.1654	-0.4810	-0.0349	0.5072	7.0490

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	71.01	8.427
speaker	(Intercept)	19.97	4.469
Residual		258.13	16.066

Number of obs: 2951, groups: item, 58; speaker, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-6.8312	1.6914	84.0419	-4.039	0.000118	***
speed1	-1.1641	0.3046	2879.2284	-3.821	0.000136	***
target1	41.0546	2.8740	50.0294	14.285	< 2e-16	***
target2	35.7297	2.7280	49.8910	13.097	< 2e-16	***
target3	44.7383	3.2614	50.3098	13.718	< 2e-16	***
target4	37.5065	3.2642	50.4838	11.490	1.06e-15	***
target5	-62.3057	3.2619	50.3434	-19.101	< 2e-16	***
target6	-48.8126	2.6215	52.4405	-18.620	< 2e-16	***
target7	-47.9108	2.6100	56.0741	-18.357	< 2e-16	***
position1	-3.8540	1.3237	109.0593	-2.911	0.004363	**
gender1	-0.6937	0.8951	25.8598	-0.775	0.445370	
speed1:target1	5.1042	0.7252	2901.9704	7.039	2.41e-12	***
speed1:target2	2.8405	0.6875	2863.7546	4.131	3.71e-05	***
speed1:target3	5.8605	0.8041	2861.9885	7.288	4.05e-13	***
speed1:target4	5.0937	0.8157	2862.5209	6.245	4.87e-10	***
speed1:target5	-4.7487	0.8064	2862.1514	-5.889	4.35e-09	***
speed1:target6	-7.4082	0.6342	2864.7020	-11.681	< 2e-16	***
speed1:target7	-6.7421	0.7313	2869.8371	-9.219	< 2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.17) Section § 4.6.5.1 Model summary for effects of speech rate on Tunisian VOT for the voiceless subset /t, ṭ, k, q/.

Formula: vot ~ speed * target + position + folv + (1 | item) + (1 | speaker)
Data: tunsUV (Tunisian + voiceless)

REML criterion at convergence: 12291.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2575	-0.6040	-0.1099	0.5116	4.9117

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	36.98	6.081
speaker	(Intercept)	25.95	5.094
Residual		127.81	11.305

Number of obs: 1586, groups: item, 29; speaker, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	31.9261	1.9854	53.3816	16.081	< 2e-16	***
speed1	3.5191	0.2920	1536.5418	12.050	< 2e-16	***
target1	0.2568	1.9824	18.9584	0.130	0.898298	
target2	-3.5630	1.9936	19.3134	-1.787	0.089609	.
target3	5.3006	2.1818	19.0351	2.429	0.025190	*
target4	-1.9944	2.1833	19.0862	-0.913	0.372383	
position1	-5.7940	1.4943	445.1463	-3.877	0.000121	***
folv1	1.4251	3.0454	19.8350	0.468	0.644933	
folv2	-3.6470	2.6270	20.3219	-1.388	0.180079	
folv3	-1.2324	3.0463	19.8542	-0.405	0.690135	
folv4	17.8320	2.5819	19.2133	6.906	1.3e-06	***
folv5	-1.8890	3.0446	19.8134	-0.620	0.542040	
folv6	3.1204	3.0458	19.8451	1.024	0.317932	
speed1:target1	0.1586	0.4951	1543.8543	0.320	0.748735	
speed1:target2	-1.7898	0.4658	1530.1358	-3.842	0.000127	***
speed1:target3	1.1724	0.5265	1529.2892	2.227	0.026101	*
speed1:target4	0.4587	0.5326	1529.4828	0.861	0.389257	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.18) Section § 4.6.5.2 Model summary for effects of speech rate on Tunisian Closure duration for /t, ṭ, k, q/ and /b, d, ḍ/.

Formula: closdur ~ target * speed + gender + position + (1 | item) + (1 | speaker)
Data: tunsVOT (CD)

REML criterion at convergence: 28442.8

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.400	-0.348	-0.054	0.262	39.608

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	25.21	5.021
speaker	(Intercept)	34.21	5.849
Residual		886.24	29.770

Number of obs: 2951, groups: item, 58; speaker, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	55.7802	1.6134	54.9712	34.572	< 2e-16	***
target1	-9.8953	2.1288	47.5200	-4.648	2.68e-05	***
target2	-5.9426	2.0252	48.3077	-2.934	0.00510	**
target3	-6.5223	2.4192	47.5481	-2.696	0.00967	**
target4	-3.6244	2.4321	48.5633	-1.490	0.14263	
target5	23.6556	2.4217	47.7429	9.768	5.74e-13	***
target6	3.0769	1.9761	49.3307	1.557	0.12586	
speed1	8.4478	0.5630	2872.1328	15.006	< 2e-16	***
gender1	2.3562	1.2340	25.9168	1.909	0.06735	.
position1	-0.2886	1.1779	63.6114	-0.245	0.80723	
target1:speed1	-4.2448	1.3254	2893.8502	-3.203	0.00138	**
target2:speed1	-0.3947	1.2731	2864.9973	-0.310	0.75654	
target3:speed1	-1.5423	1.4894	2862.4717	-1.036	0.30051	
target4:speed1	0.5880	1.5105	2864.9395	0.389	0.69709	
target5:speed1	3.5498	1.4936	2862.9824	2.377	0.01753	*
target6:speed1	2.4329	1.1741	2866.2004	2.072	0.03834	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.19) Section § 4.6.5.3 Model summary for effects of speech rate on Tunisian f0 on vowels following /t, ṭ, k, q/ and /b, d, ḍ/.

Formula: f025 ~ speed * target + gender + folv + (1 | item) + (1 | speaker)
Data: f0tuns

REML criterion at convergence: 6446.6

Scaled residuals:

Min	1Q	Median	3Q	Max
-11.1825	-0.5401	0.0276	0.5565	4.1510

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.1735	0.4165
speaker	(Intercept)	8.7724	2.9618
Residual		1.2093	1.0997

Number of obs: 2033, groups: item, 44; speaker, 28

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	1.051e+01	5.641e-01	2.672e+01	18.639	< 2e-16	***
speed1	-1.399e-01	2.474e-02	1.957e+03	-5.655	1.79e-08	***
target1	1.447e-01	1.603e-01	3.034e+01	0.903	0.373761	
target2	2.705e-01	1.615e-01	3.115e+01	1.675	0.103904	
target3	3.473e-01	1.702e-01	2.983e+01	2.041	0.050197	.
target4	4.418e-01	1.692e-01	2.934e+01	2.611	0.014084	*
target5	-4.274e-01	1.682e-01	2.866e+01	-2.541	0.016739	*
target6	-3.673e-01	1.695e-01	2.843e+01	-2.167	0.038751	*
target7	-4.097e-01	1.821e-01	3.175e+01	-2.249	0.031560	*
gender1	5.062e+00	5.603e-01	2.600e+01	9.035	1.68e-09	***
folv1	-5.474e-01	1.377e-01	3.035e+01	-3.975	0.000403	***
folv2	-3.604e-01	1.537e-01	2.872e+01	-2.345	0.026158	*
folv3	3.180e-01	1.888e-01	3.382e+01	1.685	0.101267	
folv4	1.930e-01	1.536e-01	2.864e+01	1.257	0.219037	
folv5	1.555e-01	1.489e-01	3.086e+01	1.044	0.304530	
speed1:target1	-3.484e-03	5.891e-02	1.960e+03	-0.059	0.952842	
speed1:target2	9.362e-02	6.022e-02	1.955e+03	1.555	0.120179	
speed1:target3	-2.554e-02	6.131e-02	1.967e+03	-0.417	0.677000	
speed1:target4	9.897e-02	5.979e-02	1.954e+03	1.655	0.098018	.
speed1:target5	-6.679e-02	5.753e-02	1.954e+03	-1.161	0.245791	
speed1:target6	1.116e-02	5.704e-02	1.954e+03	0.196	0.844880	
speed1:target7	-1.079e-01	6.853e-02	1.956e+03	-1.575	0.115417	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.20) Section § 4.6.6 Model results of the three-way interaction of speed*dialect*target in the voiceless subset. Followed by the full model summary. To estimate the held-out factors (/q/ for *target*, and Tunisian for *dialect*), we rotated the levels of the model.

Interactions of Interest	***Estimate***	***Std. Error***	***t.value***	***p.value***	
slow : Najdi : t	2.4261	0.5145	4.715	<.001	***
slow : Najdi : ṭ	-1.0915	0.962	-2.200	0.028	*
slow : Najdi : k	2.1633	0.5633	3.840	<.001	***
slow : Najdi : q	-3.4979	0.6520	-5.365	<.001	***
slow : Hijazi : t	-1.1496	0.5214	-2.205	0.028	*
slow : Hijazi : ṭ	-0.0900	0.5038	-0.179	0.858	
slow : Hijazi : k	0.1959	0.5730	0.342	0.732	
slow : Hijazi : q	1.0437	0.6775	1.541	0.124	
slow : Tunisian : t	-1.2765	0.4585	-2.784	0.005	**
slow : Tunisian : ṭ	1.1815	0.4441	2.660	0.008	**
slow : Tunisian : k	-2.3592	0.5036	-4.684	<.001	***
slow : Tunisian : q	2.4542	0.5517	4.448	<.001	***

```
Formula: vot ~ speed * group * target + folv + gender + (1 | item) + (1 | speaker)
   Data: votuv

REML criterion at convergence: 27228.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-8.5716 -0.6158 -0.0998  0.4844 10.2073

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept)  26.35    5.133
 item     (Intercept)  19.29    4.392
 Residual             144.56   12.023

Number of obs: 3464, groups:  speaker, 64; item, 29

Fixed effects:
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	33.8469	1.2184	35.4192	27.780	< 2e-16	***
speed1	4.2742	0.2215	3354.7953	19.301	< 2e-16	***
group1	4.7696	1.0146	61.2575	4.701	1.51e-05	***
group2	0.2852	1.0171	61.8674	0.280	0.780097	
group3	-5.0548	0.9085	60.6760	-5.564	6.31e-07	***
target1	6.6909	1.4342	18.8325	4.665	0.000172	***
target2	-12.2697	1.4339	18.8169	-8.557	6.56e-08	***
target3	14.0146	1.5758	18.7991	8.894	3.67e-08	***
target4	-8.4358	1.5946	19.7068	-5.290	3.71e-05	***
folv1	-2.6785	2.1688	18.6849	-1.235	0.232133	
folv2	-4.2539	1.8567	18.6829	-2.291	0.033766	*
folv3	-2.1246	2.1693	18.7023	-0.979	0.339878	
folv4	12.0534	1.8605	18.8297	6.479	3.45e-06	***
folv5	-2.5516	2.1700	18.7249	-1.176	0.254385	
folv6	2.3397	2.1742	18.8645	1.076	0.295435	
gender1	0.5890	0.6759	59.8155	0.871	0.387062	
speed1:group1	0.9257	0.3232	3354.7007	2.864	0.004203	**
speed1:group2	-0.3266	0.3307	3355.2558	-0.987	0.323526	
speed1:group3	-0.5991	0.2837	3354.7356	-2.112	0.034759	*
speed1:target1	1.9439	0.3528	3354.2691	5.510	3.86e-08	***
speed1:target2	-3.2063	0.3409	3354.2336	-9.406	< 2e-16	***

speed1:target3	3.4073	0.3872	3354.5907	8.801	< 2e-16	***
speed1:target4	-2.1448	0.4450	3355.9592	-4.820	1.50e-06	***
group1:target1	10.7264	0.5162	3358.0977	20.780	< 2e-16	***
group2:target1	-4.6827	0.5231	3359.1595	-8.951	< 2e-16	***
group3:target1	-6.0437	0.4598	3357.9083	-13.144	< 2e-16	***
group1:target2	-9.8773	0.4979	3358.1179	-19.839	< 2e-16	***
group2:target2	0.4911	0.5057	3359.6743	0.971	0.331518	
group3:target2	9.3861	0.4455	3357.9995	21.069	< 2e-16	***
group1:target3	8.6156	0.5646	3356.7111	15.260	< 2e-16	***
group2:target3	0.6460	0.5736	3356.2250	1.126	0.260140	
group3:target3	-9.2616	0.5046	3356.2207	-18.354	< 2e-16	***
group1:target4	-9.4647	0.6629	3373.5764	-14.279	< 2e-16	***
group2:target4	3.5456	0.6865	3373.6251	5.164	2.55e-07	***
group3:target4	5.9191	0.5612	3373.1763	10.548	< 2e-16	***
speed1:group1:target1	2.4261	0.5145	3354.2035	4.715	2.51e-06	***
speed1:group2:target1	-1.1496	0.5214	3354.3954	-2.205	0.027523	*
speed1:group3:target1	-1.2765	0.4585	3354.2543	-2.784	0.005395	**
speed1:group1:target2	-1.0915	0.4962	3354.1413	-2.200	0.027894	*
speed1:group2:target2	-0.0900	0.5038	3354.5318	-0.179	0.858236	
speed1:group3:target2	1.1815	0.4441	3354.1941	2.660	0.007843	**
speed1:group1:target3	2.1633	0.5633	3354.2220	3.840	0.000125	***
speed1:group2:target3	0.1959	0.5730	3355.0689	0.342	0.732472	
speed1:group3:target3	-2.3592	0.5036	3354.2535	-4.684	2.92e-06	***
speed1:group1:target4	-3.4979	0.6520	3355.9371	-5.365	8.66e-08	***
speed1:group2:target4	1.0437	0.6775	3356.6462	1.541	0.123500	
speed1:group3:target4	2.4542	0.5517	3355.7523	4.448	8.94e-06	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (C.21) The number of tokens used to measure the mean VOT values in each dialect/speaking condition featured in Table 4-9.

Dialect	Plosive Type	Slow (N)	Fast (N)
Najdi	/t/	143	141
	/ṭ/	161	162
	/d/	189	170
	/ḍ/	184	165
Hijazi	/t/	141	136
	/ṭ/	157	155
	/d/	195	182
	/ḍ/	198	188
Tunisian	/t/	222	222
	/ṭ/	245	241
	/d/	302	295
	/ḍ/	181	256

Appendix D (for chapter 5)

Appendix (D.1) The language background questionnaire provided to participants.

UNIVERSITY of York

DEPARTMENT OF
LANGUAGE AND
LINGUISTIC SCIENCE
Heslington, York, YO10 5DD, UK
Phone number +44 7452792204
lmaa501@york.ac.uk

Language Background Questionnaire

Participant Code: ______________________

Nationality: __________________________

Age: ________________________________

Gender: ______________________________

1. Where were you born?

   Country: _____________ State/Province: ______________ City: ________________

2. Where are your parents from? Mother: _______________ Father: _______________

3. What is your native language? ____

4. What other languages (besides Arabic) do you know/speak? ___________________

5. How old were you when you first began to learn the second language? ____________

6. How did you learn your second language? (academically or naturalistically)

7. Have you ever lived in a non-Arabic speaking country? If yes, where and how long?

   ____

8. When were you first introduced to Modern Standard Arabic? _____________________

9. Where were you first introduced to Modern Standard Arabic? _____________________

10. In what context do you use Modern Standard Arabic? ___________________________

11. How confident do you feel when you talk in Modern Standard Arabic?

    Very confident 10○ 9○ 8○ 7○ 6○ 5○ 4○ 3○ 2○ 1○ Not confident

12. Do you have any hearing or speaking problems? ______________________________

Appendix (D.2) Rubric followed in Researcher Evaluation of MSA reading performance.

	(100%)	(66.6%)	(33.3%)	(0%)
التشكيل اللغوي **Diacritics**	خال من الأخطاء - خطأين none - two miscues 25	ثلاثة أخطاء three miscue 16.65	أربعة أخطاء four miscue 8.33	أكثر من أربع أخطاء more than 4 miscues 0
المفردات اللغوية **Lexical Usage**	خال من الأخطاء no miscue 25	خطأ واحد one miscue 16.65	خطأين two miscue 8.33	أكثر من خطأين more than 2 miscues 0
علامات الترقيم في القراءة **Intonation**	أخذ بالاعتبار علامة الاستفهام وكل الفاصلات rising intonation in (?) and pause in (,) throughout 25	أخذ بالاعتبار علامة الاستفهام ونقص فاصلة rising intonation in (?) and pause in (,) mostly 16.65	لم يأخذ بالاعتبار علامة الاستفهام rising intonation in (?) and pause in (,) somewhat 8.33	لم يُأخذ بالاعتبار علامة الاستفهام وفاصلة rising intonation in (?) and pause in (,) none 0
طلاقة القراءة **Confidence**	واثق وجهوري loud and confident 25	واثق confident 16.65	متلعثم stuttering 8.33	متلعثم وبطيء stuttering and slow 0

Appendix (D.3) Participants' performance list.

(**Time**) time took to read MSA passage in seconds, (**SRE**) Self-Reported Evaluation, (**RE**) Researcher's Evaluation, (**AE**) Arabic speaking External Evaluator, (**Difference**) difference between the two evaluators, and (**Final**) final score agreed upon.

#	Group	Participant	Time	SRE	RE	AE	Difference	Final
1	tuns	tuns-f01	44	10	58	67	-8	7
2	tuns	tuns-f02	44	10	75	75	0	8
3	tuns	tuns-f03	69	10	58	50	8	6
4	tuns	tuns-f04	63	10	58	67	-8	7
5	tuns	tuns-f05	51	10	58	50	8	6
6	tuns	tuns-f06	45	10	58	58	0	6
7	tuns	tuns-f07	47	6	17	8	8	2
8	tuns	tuns-f08	61	10	17	8	8	2
9	tuns	tuns-f09	53	8	50	50	0	5
10	tuns	tuns-f10	50	9	58	58	0	6
11	tuns	tuns-f11	44	6	8	8	0	1
12	tuns	tuns-f12	42	7	83	83	0	8
13	tuns	tuns-f13	51	1	42	33	8	4
14	tuns	tuns-f14	44	3	83	83	0	8
15	tuns	tuns-m01	48	7	50	50	0	5
16	tuns	tuns-m02	48	10	75	75	0	8
17	tuns	tuns-m03	47	7	67	67	0	7
18	tuns	tuns-m04	39	9	42	42	0	4
19	tuns	tuns-m05	54	10	8	8	0	1
20	tuns	tuns-m06	61	10	25	25	0	3
21	tuns	tuns-m07	56	9	42	33	8	4
22	tuns	tuns-m08	54	8	17	17	0	2
23	tuns	tuns-m09	49	7	25	25	0	3
24	tuns	tuns-m10	44	9	75	75	0	8
25	tuns	tuns-m11	47	3	42	42	0	4
26	tuns	tuns-m12	44	7	58	58	0	6
27	tuns	tuns-m13	40	5	42	42	0	4
28	tuns	tuns-m14	44	4	42	33	8	4
29	sahi	sahi-f01	58	10	33	33	0	3
30	sahi	sahi-f02	56	3	58	58	0	6
31	sahi	sahi-f03	49	1	67	67	0	7
32	sahi	sahi-f04	58	6	8	8	0	1
33	sahi	sahi-f05	72	5	17	17	0	2
34	sahi	sahi-f06	56	6	8	8	0	1
35	sahi	sahi-f07	58	7	25	25	0	3
36	sahi	sahi-f08	50	10	83	83	0	8
37	sahi	sahi-f09	53	5	33	33	0	3

38	sahi	sahi-f10	43	6	25	25	0	3
39	sahi	sahi-m01	47	9	67	67	0	7
40	sahi	sahi-m02	64	5	42	42	0	4
41	sahi	sahi-m03	57	1	0	0	0	0
42	sahi	sahi-m04	62	10	33	33	0	3
43	sahi	sahi-m05	61	3	0	0	0	0
44	sahi	sahi-m06	45	3	75	75	0	8
45	sahi	sahi-m07	52	8	25	25	0	3
46	sahi	sahi-m08	46	7	8	8	0	1
47	sanj	sanj-f01	85	5	0	0	0	0
48	sanj	sanj-f02	49	5	67	67	0	7
49	sanj	sanj-f03	41	7	75	75	0	8
50	sanj	sanj-f04	52	10	33	33	0	3
51	sanj	sanj-f05	51	6	33	33	0	3
52	sanj	sanj-f06	54	4	42	42	0	4
53	sanj	sanj-f07	54	6	50	50	0	5
54	sanj	sanj-f08	46	8	83	83	0	8
55	sanj	sanj-m01	44	4	67	67	0	7
56	sanj	sanj-m02	61	5	0	0	0	0
57	sanj	sanj-m03	53	10	92	92	0	9
58	sanj	sanj-m04	56	7	67	58	8	7
59	sanj	sanj-m05	46	7	42	42	0	4
60	sanj	sanj-m06	58	5	8	8	0	1
61	sanj	sanj-m07	48	6	67	67	0	7
62	sanj	sanj-m08	60	7	75	75	0	8
63	sanj	sanj-m09	74	6	8	8	0	1
64	sanj	sanj-m10	51	3	25	25	0	3

Appendix (D.4) MSA passage used in chapter 5. Underlined words in **bold** are intended to be the stimuli.

ذات ليلةٍ ..
أسدل جفنيه متعباً و**تاه** في البعيد. وجد نفسه في **دارٍ** بعيدةٍ راح يطوف في أرجائها يلمس جدرانها ويلثُمُ عتباتها التي تغطيها **أتربةٌ** مضمخةٌ بالحنين. سالت دموعه عليها وكادت أن تحيلها **طيناً**. كانت **أضوية** خفيفة تنبعث من حجراتها، فدلف إليها و**طاب** له المقام فيها. شعر **بضيق** في صدره ما كان **لأدوية** العالم كله أن تزيله. حاول أن يعود من **تيهٍ** تبعثر فيه حتى أصبح **ضارا** بكل جوارحه. تساءل في نفسه مستغربا أكان **ديناً** عليّ قضاؤه ألما؟ وعندما فتح عينيه وجد الإجابة ماثلة أمامه، فما كان ذلك إلا **أطياف** وطنٍ مرتْ به في حلم.

[ða:ta lajlatin ..
ʔsdala d͡ʒəfnajhi mutʕaban wa **ta:ha** filbaʕi:d. wad͡ʒada nafsahu fi: **da:rin** baʕi:datin ra:ħa jaṭu:fu fi: ʔard͡ʒa:ʔiha jalmisu d͡ʒudra:naha wajalθumu ʕataba:tiha ʔallati tuɣaṭi:ha **ʔatribatun** muḍaxxamatun bilħani:n. sa:lat dumuʕuhu ʕalajha waka:dat ʔan tuħi:laha **ṭi:nan.** ka:nat **ʔaḍwijatun** xafi:fatun tanbaʕiθu min ħud͡ʒura:tiha, fadalafa ʔilajha: wa **ṭa:ba** lahu ʔalmuqa:mu fi:ha. ʃaʕara bi **ḍi:qin** fi: ṣadrihi ma: ka:na li **ʔadwijati** lʕa:lami kullihi ʔan tuzi:lah. ħa:wala ʔan jaʕu:da min **ti:hin** tabaʕθara fi:hi ħatta ʔaṣbaħa **ḍa:rran** bikul d͡ʒawa:riħihi. tasa:ʔala fi: nafsihi mustaɣriban ʔaka:na **di:nan** ʕalaja qaḍa:ʔuhu ʔalaman ? wa ʕindama: fataħa ʕajnajhi wad͡ʒada lʔid͡ʒa:bata ma:θilatan ʔamamahu, fama: ka:na ðalika ʔilla: **ʔaṭja:fa** waṭanin marrat bih fi: ħulumin]

One night ..
He shut his eyes and got **lost** with his thoughts. He found himself in a **house** far away, he started wandering around, touching its walls, kissing its steps. Those steps that are covered with **dust** that made him so nostalgic where he shed tears turning dust into **mud**. Dim **lights** were released from its rooms. He went into the rooms, he **liked** the feeling, yet he felt **anguished**, there existed no **medicine** to cure this feeling. He tried to come back from this **stray** feeling that **harmed** his soul. He wondered, was this a **debt** I had to pay in pain? When he opened his eyes, he found the answer! These were only **wraiths** of a home passing by in a daydream.

Appendix (D.5) Section § 5.4.1.1 Model summary for VOT in MSA Najdi, Hijazi, and Tunisian in one speech rate only.

Formula: vot ~ group * target + gender + position + (1 | item) + (1 + score | speaker)
Data: MSAS (Subset: MSA + slow + all dialects)

REML criterion at convergence: 5584.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.5525	-0.6238	0.0086	0.5818	4.0392

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	51.378	7.168	
	score	2.037	1.427	-0.91
item	(Intercept)	100.812	10.041	
Residual		176.632	13.290	

Number of obs: 692, groups: speaker, 64; item, 12

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-10.3883	3.1768	7.5320	-3.270	0.0123	*
group1	-1.9224	1.1539	59.1267	-1.666	0.1010	
group2	-0.5468	1.1451	53.9913	-0.478	0.6349	
group3	2.4693	1.0186	58.3270	2.424	0.0185	*
target1	49.8397	5.0950	7.0095	9.782	2.45e-05	***
target2	41.8762	5.0953	7.0113	8.219	7.60e-05	***
target3	-42.3183	5.0961	7.0155	-8.304	7.09e-05	***
target4	-49.3977	5.1217	7.1517	-9.645	2.36e-05	***
gender1	0.8614	0.7531	55.2062	1.144	0.2577	
position1	-1.5499	3.1224	7.0305	-0.496	0.6348	
group1:target1	9.0024	1.2690	614.7657	7.094	3.60e-12	***
group2:target1	-1.7074	1.2646	614.2270	-1.350	0.1775	
group3:target1	-7.2950	1.1520	616.2571	-6.333	4.65e-10	***
group1:target2	-11.1906	1.2742	614.6120	-8.782	< 2e-16	***
group2:target2	-0.2450	1.2659	614.1858	-0.194	0.8466	
group3:target2	11.4355	1.1534	616.2092	9.914	< 2e-16	***
group1:target3	2.8534	1.2778	615.4602	2.233	0.0259	*
group2:target3	2.0594	1.2691	614.4398	1.623	0.1051	
group3:target3	-4.9129	1.1667	617.1063	-4.211	2.92e-05	***
group1:target4	-0.6652	1.4647	630.5077	-0.454	0.6499	
group2:target4	-0.1070	1.4392	628.6330	-0.074	0.9408	
group3:target4	0.7723	1.3906	636.2806	0.555	0.5788	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (D.6) Section § 5.4.1.2 Model summary for the F2 lowering in the MSA slow speech rate voiceless subset.

Formula: f225 ~ group * target * folv + gender + (1 | item)
Data: uf2

REML criterion at convergence: 391.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-6.4419	-0.2932	0.0988	0.4793	2.3134

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	1.0782	1.0384
Residual		0.2573	0.5072

Number of obs: 233, groups: item, 4

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.233953	0.520376	220.000000	0.450	0.6535	
group1	-0.078315	0.051256	220.000000	-1.528	0.1280	
group2	-0.043591	0.053692	220.000000	-0.812	0.4177	
group3	1.219e-01	4.494e-02	2.200e+02	2.712	0.00721	**
target1	0.329778	0.520375	220.000000	0.634	0.5269	
folv1	0.942012	0.520377	220.000000	1.810	0.0716	.
gender1	0.009336	0.033607	220.000000	0.278	0.7814	
group1:target1	-0.017417	0.051061	220.000000	-0.341	0.7334	
group2:target1	-0.002788	0.053435	220.000000	-0.052	0.9584	
group3:target1	2.020e-02	4.495e-02	2.200e+02	0.449	0.65352	
group1:folv1	-0.063565	0.051050	220.000000	-1.245	0.2144	
group2:folv1	0.096415	0.053405	220.000000	1.805	0.0724	.
target1:folv1	-0.304890	0.520376	220.000000	-0.586	0.5585	
group1:target1:folv1	0.085238	0.051051	220.000000	1.670	0.0964	.
group2:target1:folv1	0.010407	0.053419	220.000000	0.195	0.8457	
group3:target1:folv1	-9.565e-02	4.496e-02	2.200e+02	-2.127	0.03452	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (D.7) Section § 5.5.1 Model summary for effects of speed on word duration in MSA.

Formula: worddur ~ speed * group + (1 + score | item)
Data: MSA

REML criterion at convergence: 14454.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1929	-0.6065	-0.0956	0.4535	6.6637

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
item	(Intercept)	4821.008	69.433	
	score	7.388	2.718	-0.87
Residual		2602.833	51.018	

Number of obs: 1347, groups: item, 12

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	246.5034	12.9140	11.6322	19.088	3.86e-10	***
speed1	34.2075	1.4206	1320.0222	24.080	< 2e-16	***
group1	5.1550	2.0863	1329.4159	2.471	0.0136	*
group2	9.8988	2.1107	1319.8618	4.690	3.02e-06	***
group3	-15.054	1.884	1330.145	-7.988	2.94e-15	***
speed1:group1	-5.1950	2.0770	1319.7001	-2.501	0.0125	*
speed1:group2	0.5767	2.0711	1320.7374	0.278	0.7807	
speed1:group3	4.618	1.873	1320.539	2.466	0.0138	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (D.8) Section § 5.5.3 Model summary for effects of speech rate on Najdi MSA /t ~ ṭ/ and /d ~ ḍ/.

```
Formula: vot ~ speed * target + position + gender + (1 + score | item)
   Data: sanj

REML criterion at convergence: 2797

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.1219 -0.5176 -0.0461  0.6029  3.0417

Random effects:
Groups    Name         Variance  Std.Dev.  Corr
item      (Intercept)  45.1530   6.7196
          score        0.1291    0.3593    0.25
Residual               92.9904   9.6432

Number of obs: 379, groups:  item, 12

Fixed effects:
```

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-11.0254	2.2621	7.0568	-4.874	0.00177	**
speed1	-0.9951	0.5064	352.8938	-1.965	0.05018	.
target1	54.0817	3.6653	6.8418	14.755	1.94e-06	***
target2	29.7592	3.6693	6.8741	8.110	9.22e-05	***
target3	-39.0859	3.6653	6.8418	-10.664	1.64e-05	***
target4	-44.7537	3.7524	7.4719	-11.927	3.94e-06	***
position1	-1.7728	2.2594	7.0124	-0.785	0.45835	
gender1	-0.6351	0.5011	354.4878	-1.267	0.20585	
speed1:target1	5.5876	0.8366	352.2463	6.679	9.42e-11	***
speed1:target2	0.2645	0.8390	351.9914	0.315	0.75272	
speed1:target3	-2.0048	0.8447	351.9330	-2.373	0.01816	*
speed1:target4	-3.8472	0.9803	354.6805	-3.925	0.000104	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

	(Intr)	speed1	targt1	targt2	targt3	postn1	gendr1	spd1:1	spd1:2
speed1	-0.015								
target1	-0.011	0.002							
target2	-0.009	0.004	-0.325						
target3	-0.011	0.002	-0.325	-0.325					
position1	-0.335	-0.007	0.001	0.003	0.001				
gender1	0.038	-0.010	0.004	0.008	0.004	0.002			
sped1:trgt1	0.000	-0.082	-0.017	0.003	0.004	0.004	-0.014		
sped1:trgt2	0.002	-0.077	0.003	-0.014	0.003	0.007	-0.007	-0.269	
sped1:trgt3	0.000	-0.065	0.004	0.002	-0.017	0.004	-0.021	-0.274	-0.276

Appendix (D.9) Section § 5.5.4 Model summary for effects of speech rate on Hijazi MSA /t ~ ṭ/ and /d ~ ḍ/.

Formula: vot ~ speed * target + position + gender + (0 + score | item)
Data: sahi

REML criterion at convergence: 2995.8

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.6418	-0.5923	-0.0128	0.5567	3.6604

Random effects:

Groups	Name	Variance	Std.Dev.
item	score	1.13	1.063
Residual		145.84	12.076

Number of obs: 384, groups: item, 12

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-12.0456	0.9822	130.4834	-12.264	< 2e-16	***
speed1	-0.5308	0.6238	364.1290	-0.851	0.3954	
target1	45.2941	1.5709	129.8321	28.834	< 2e-16	***
target2	39.6598	1.5709	129.4684	25.246	< 2e-16	***
target3	-39.4654	1.6196	126.8414	-24.368	< 2e-16	***
target4	-45.4885	1.7275	112.4686	-26.332	< 2e-16	***
position1	-1.6521	0.9738	128.9861	-1.697	0.0922	.
gender1	0.5464	0.6279	364.2118	0.870	0.3848	
speed1:target1	4.3439	1.0413	363.5386	4.172	3.79e-05	***
speed1:target2	1.7975	1.0446	363.4709	1.721	0.0862	.
speed1:target3	-2.3820	1.0794	364.1530	-2.207	0.0280	*
speed1:target4	-3.7594	1.1510	365.5845	-3.266	0.00119	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	speed1	targt1	targt2	targt3	postn1	gendr1	spd1:1	spd1:2
speed1	-0.078								
target1	-0.047	0.023							
target2	-0.047	0.023	-0.288						
target3	0.001	-0.029	-0.310	-0.310					
position1	-0.292	0.004	-0.025	-0.025	-0.007				
gender1	-0.078	0.043	0.005	0.006	-0.034	0.012			
sped1:trgt1	0.023	-0.063	-0.055	0.000	0.030	-0.002	-0.007		
sped1:trgt2	0.023	-0.057	0.000	-0.054	0.030	-0.004	0.001	-0.286	
sped1:trgt3	-0.025	0.000	0.031	0.030	-0.111	-0.017	0.040	-0.309	-0.311

Appendix (D.10) Section § 5.5.5 Model summary for effects of speech rate on Tunisian MSA /t ~ ṭ/ and /d ~ ḍ/.

Formula: vot ~ speed * target + position + gender + folv + (0 + score | item)
Data: tuns

REML criterion at convergence: 4793.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1085	-0.6262	-0.0245	0.5650	4.2355

Random effects:

Groups	Name	Variance	Std.Dev.
item	score	1.772	1.331
Residual		217.224	14.739

Number of obs: 584, groups: item, 12

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-10.1226	1.5727	114.5618	-6.437	2.95e-09	***
speed1	-1.1271	0.6291	563.8504	-1.792	0.073718	.
target1	38.9878	2.1333	129.6212	18.276	< 2e-16	***
target2	47.3513	2.1343	129.9324	22.186	< 2e-16	***
target3	-43.1954	2.1519	127.7748	-20.073	< 2e-16	***
target4	-43.1438	2.4162	97.9581	-17.856	< 2e-16	***
position1	10.0504	2.6665	127.2943	3.769	0.000249	***
gender1	2.0133	0.6186	570.8155	3.254	0.001203	**
folv1	-13.9015	3.0597	128.5993	-4.543	1.26e-05	***
speed1:target1	4.0445	1.0209	563.6427	3.962	8.40e-05	***
speed1:target2	6.3471	1.0248	563.6534	6.193	1.14e-09	***
speed1:target3	-5.3297	1.0433	563.8092	-5.108	4.45e-07	***
speed1:target4	-5.0619	1.2521	564.2809	-4.043	6.02e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Correlation of Fixed Effects:

	(Intr)	speed1	targt1	targt2	targt3	postn1	gendr1	folv1	spd1:1	spd1:2
speed1	0.010									
target1	-0.053	-0.010								
target2	-0.052	-0.012	-0.283							
target3	-0.039	-0.007	-0.289	-0.290						
position1	-0.575	0.004	0.008	0.007	0.002					
gender1	0.099	-0.012	0.019	0.022	0.005	0.007				
folv1	0.480	-0.007	0.004	0.004	0.012	-0.861	-0.008			
sped1:trgt1	-0.006	-0.113	0.006	0.007	0.004	-0.002	0.007	0.004		
sped1:trgt2	-0.009	-0.106	0.007	0.004	0.005	-0.001	0.000	0.004	-0.244	
sped1:trgt3	0.003	-0.075	0.004	0.005	0.010	-0.015	0.004	0.015	-0.257	-0.260

Appendix (D.11) Section § 5.5.5 Model summary for effects of speech rate on Tunisian MSA /t/ ~ /ṭ/ and /d/ ~ /ḍ/ by following context.

Formula: vot ~ speed * target * folv + position + gender + (0 + score | item)
Data: tuns (by following context)

REML criterion at convergence: 4695.4
Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2616	-0.6231	0.0118	0.5693	4.2380

Random effects:

Groups	Name	Variance	Std.Dev.
item	score	0.2566	0.5065
Residual		204.0281	14.2838

Number of obs: 584, groups: item, 12

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-7.2545	0.8835	3.7470	-8.211	0.00158	**
speed1	-1.0462	0.6143	548.7257	-1.703	0.08915	.
target1	38.7731	1.4738	3.5486	26.309	3.39e-05	***
target2	47.1841	1.4759	3.5777	31.970	1.58e-05	***
target3	-42.2824	1.4877	3.6495	-28.421	2.04e-05	***
target4	-43.6747	1.6646	4.4808	-26.238	4.45e-06	***
folv1	-7.9632	1.2314	3.6895	-6.467	0.00388	**
folv2	6.9304	1.2497	4.0830	5.546	0.00487	**
folv3	1.0328	1.2602	3.7360	0.820	0.46155	
gender1	1.9443	0.5950	556.4943	3.268	0.00115	**
speed1:target1	3.9636	0.9923	547.4089	3.994	7.37e-05	***
speed1:target2	6.2860	0.9962	547.4291	6.310	5.76e-10	***
speed1:target3	-5.3442	1.0138	548.2050	-5.271	1.95e-07	***
speed1:target4	-4.9054	1.2343	550.6969	-3.974	8.00e-05	***
speed1:folv1	-2.4006	0.8461	547.7771	-2.837	0.00472	**
speed1:folv2	1.4087	0.8841	550.1740	1.593	0.11168	
target1:folv1	-1.2597	2.0747	3.5034	-0.607	0.58086	
target2:folv1	-13.1862	2.0762	3.5180	-6.351	0.00481	**
target3:folv1	8.7528	2.0999	3.5576	4.168	0.01789	*
target1:folv2	8.9306	2.0856	3.6297	4.282	0.01579	*
target2:folv2	7.0254	2.0871	3.6450	3.366	0.03244	*
target3:folv2	-10.0144	2.1026	3.7587	-4.763	0.01035	*
speed1:target1:folv1	-0.1180	1.3893	547.0139	-0.085	0.93235	
speed1:target2:folv1	-1.6135	1.3921	547.0234	-1.159	0.24695	
speed1:target3:folv1	2.0727	1.4172	547.4970	1.463	0.14416	
speed1:target4:folv1	-0.3413	1.6473	549.0028	-0.207	0.83595	
speed1:target1:folv2	1.7866	1.4129	548.0406	1.265	0.20656	
speed1:target2:folv2	0.5710	1.4156	548.0478	0.403	0.68685	
speed1:target3:folv2	-1.6557	1.4446	548.5909	-1.146	0.25226	
speed1:target4:folv2	-0.7019	1.8153	553.1853	-0.387	0.69915	
speed1:target1:folv3	-1.6686	1.4076	547.1670	-1.185	0.23634	
speed1:target2:folv3	1.0425	1.4185	547.2060	0.735	0.46271	
speed1:target3:folv3	-0.4170	1.4396	548.3482	-0.290	0.77216	
speed1:target4:folv3	1.0432	1.7702	549.1767	0.589	0.55588	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Appendix (D.12) Section § 5.5.6 Model results of the three-way interaction of speed * dialect * target in the MSA subset of one following context /a:/.

Formula: vot ~ group * speed * target + gender + (0 + score | item)
Data: MSAaa (MSA + All dialects + only aa context)

REML criterion at convergence: 3330.7
Scaled residuals:

Min	1Q	Median	3Q	Max
-3.8096	-0.5623	-0.0011	0.6203	3.3271

Random effects:

Groups	Name	Variance	Std.Dev.
item	score	0.04973	0.223
Residual		89.12558	9.441

Number of obs: 462, groups: item, 4

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-15.2506	0.6283	3.0306	-24.274	0.000143	***
group1	-1.1701	0.6675	436.6190	-1.753	0.080311	.
group2	1.7673	0.6761	408.8456	2.614	0.009282	**
group3	-0.5972	0.5964	436.9755	-1.001	0.317238	
speed1	-2.0507	0.4546	434.1172	-4.511	8.30e-06	***
target1	44.4765	1.0550	3.1213	42.156	2.09e-05	***
target2	33.0917	1.0620	3.0446	31.161	6.49e-05	***
target3	-32.7944	1.0798	3.0767	-30.371	6.47e-05	***
target4	-44.7738	1.1504	3.0207	-38.921	3.52e-05	***
gender1	-0.4904	0.4438	436.9907	-1.105	0.269699	
group1:speed1	0.3325	0.6625	433.8978	0.502	0.615975	
group2:speed1	1.0819	0.6680	434.1584	1.620	0.106043	
group1:target1	8.6563	1.1133	436.9918	7.775	5.45e-14	***
group2:target1	-2.2007	1.1231	405.3102	-1.959	0.050741	.
group1:target2	-2.0944	1.1212	436.7684	-1.868	0.062419	.
group2:target2	0.9486	1.1256	398.3877	0.843	0.399888	
group1:target3	-0.9627	1.1250	436.5077	-0.856	0.392618	
group2:target3	1.7468	1.1626	424.5602	1.503	0.133704	
speed1:target1	3.8682	0.7605	434.0492	5.087	5.44e-07	***
speed1:target2	2.3408	0.7630	433.7562	3.068	0.002292	**
speed1:target3	-1.3502	0.7786	435.0257	-1.734	0.083576	.
group1:speed1:target1	1.3745	1.1115	433.7109	1.237	0.216887	
group2:speed1:target1	-1.3701	1.1147	433.7344	-1.229	0.219695	
group3:speed1:target1	-0.0044.	0.9957	433.8629	-0.004	0.996515	
group1:speed1:target2	-0.4800	1.1186	433.6305	-0.429	0.668071	
group2:speed1:target2	-1.8700	1.1166	433.8320	-1.675	0.094693	.
group3:speed1:target2	2.3500	0.9977	433.6923	2.356	0.018941	*
group1:speed1:target3	0.8541	1.1239	433.6341	0.760	0.447700	
group2:speed1:target3	1.0065	1.1575	434.8874	0.870	0.385044	
group3:speed1:target3	-1.8606	1.0173	434.8429	-1.829	0.068092	.
group1:speed1:target4	-1.7486	1.2320	434.5699	-1.419	0.156524	
group2:speed1:target4	2.2337	1.2368	434.1876	1.806	0.071614	.
group3:speed1:target4	-0.4851	1.1099	433.6747	-0.437	0.662300	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix (D.13) The number of tokens used to measure the mean VOT values in each dialect/speaking condition featured in section § 5.6, Table 5-4.

Task	MSA	Target	Slow (N)	Fast (N)
Dialect	Najdi	/t/	143	141
		/ṭ/	161	162
		/d/	189	170
		/ḍ/	184	165
	Hijazi	/t/	141	136
		/ṭ/	157	155
		/d/	195	182
		/ḍ/	198	188
	Tunisian	/t/	222	222
		/ṭ/	245	241
		/d/	302	295
		/ḍ/	181	256
MSA	Najdi	/t/	54	51
		/ṭ/	53	51
		/d/	53	49
		/ḍ/	37	31
	Hijazi	/t/	54	51
		/ṭ/	54	50
		/d/	54	42
		/ḍ/	41	38
	Tunisian	/t/	28	28
		/ṭ/	28	28
		/d/	79	78
		/ḍ/	45	48

References

Abou-Ghazaleh, A., Khateb, A., & Nevat, M. (2018). Lexical Competition between Spoken and Literary Arabic: A New Look into the Neural Basis of Diglossia Using fMRI. *Neuroscience, 393*, 83-96. doi:10.1016/j.neuroscience.2018.09.045

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75-86.

Abudalbuh, M. (2010). *Effects of gender on the production of emphasis in Jordanian Arabic: A sociophonetic study.* (Doctoral dissertation). University of Kansas,

Akbar, R., Taqi, H., & Al-Gharabally, N. (2020). The Linguistic Landscape on the Streets of Kuwait: A Challenge to the Concept of Diglossia. *Studies in Linguistics and Literature, 4*(4), 65-83. doi:10.22158/sll.v4n4p65

Al-Ani, S. (1970). *Arabic Phonology: An Acoustical and Physiological Investigation* (Vol. 61): De Gruyter Mouton.

Al-Ani, S., & El-Dalee, M. (1983). *Tafkhim in Arabic: The Acoustic and Physiological Parameters.* Paper presented at the Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht, The Netherlands.

Alaodini, H. (2019). *A Sociolinguistic study of the Dawāsir dialect in Dammam, Eastern Arabia: fortition of /j/ and unrounding of /a:/.* (Doctoral dissertation). University of Essex,

Albdairat, I. S. (2021). *Sociolinguistic Study of Phonological Variation in Jordanian Arabic in Al-Karak.* (Doctoral dissertation). Indiana University,

AlDahri, S. S. (2013). A study for the effect of the Emphaticness and language and dialect for Voice Onset Time (VOT) in Modern Standard Arabic (MSA). *Signal & Image Processing: An International Journal, 4*(2), 11-29.

Al-Essa, A. (2009). When Najd meets Hijaz: dialect contact in Jeddah. In E. Al-Wer & R. de Jong (Eds.), *Arabic Dialectology* (pp. 203-222). Amsterdam: Brill.

Al-Fahid, J. M. (2000). *The Goodman psycholinguistic model of English reading and its applicability to Semitic languages.* (Doctoral dissertation). The University of Arizona,

Al-Fahid, J. M., & Goodman, K. S. (2008). The Reading Process In Arabic Making Sense of the Arabic Print. In A. D. Flurkey, E. J. Paulson, & K. S. Goodman (Eds.), *Scientific Realism in Studies of Reading* (pp. 155-168): Routledge.

Al-Gamdi, N., Al-Tamimi, J., & Khattab, G. (2022). *Voicing contrast in Najdi Arabic stops: Implication for Laryngeal Realism*. Paper presented at the Colloquium of the British Association of Academic Phoneticians, University of York.

Al-Hawamdeh, A. M. (2016). *A Sociolinguistic Investigation of Two Hōrāni Features in Sūf, Jordan.* (Doctoral dissertation). University of Essex,

Ali, L. H., & Daniloff, R. G. (1972). A Contrastive Cinefluorographic Investigation ofthe Articulation of Emphatic–Non Emphatic Cognate Consonants. *Studia Linguistica, 18*, 81-105.

Al-Jallad, A. (2018). What is Ancient North Arabian? In D. Birnstiel & N. a. Pat-El (Eds.), *Re-engaging Comparative Semitic and Arabic Studies* (1st ed., pp. 1-44): Harrassowitz Verlag.

Alkalesi, Y. M. (2001). *Modern Iraqi Arabic: A Textbook*. Washington D. C.: Georgetown University Press.

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America, 106*(4), 2031-2039.

Al Malwi, I. M. (2017). *The effects of gender and age on voice onset time by Abah Arabic speakers.* (Master's thesis). California State University,

Al-Masri, M., & Jongman, A. (2004). *Acoustic Correlates of Emphasis in Jordanian Arabic: Preliminary Results.* Paper presented at the Proceedings of the 2003 Texas Linguistics Society Conference, Somerville, MA.

Almbark, R. (2008). *A sociophonetic study of emphasis in Syrian Arabic.* (Master's thesis). The University of York,

Almbark, R., & Hellmuth, S. (2016). *Variation in quantity and quality of Arabic vowels: Comparison of 8 dialects*. Paper presented at the Arabic Linguistics Forum, University of York.

Almurashi, W., Al-Tamimi, J., & Khattab, G. (2020). Static and dynamic cues in vowel production in Hijazi Arabic. *The Journal of the Acoustical Society of America, 147*(4), 2917-2927.

Al-Nuzaili, A. M. (1993). *Experimental study of emphasis and voicing in the plosives of Yemeni Spoken Arabic with some implications for foreign language teaching and learning.* (Doctoral dissertation). University of Leeds,

Al-Qenaie, S. (2011). *Kuwaiti Arabic: A Socio-Phonological Perspective.* (Doctoral dissertation). Durham University,

Alqahtani, K. (2015). *A sociolinguistic study of the Tihami Qahtani dialect in Asir, Southern Arabia.* (Doctoral dissertation). University of Essex,

Al-Tamimi, F., Alzoubi, F., & Tarawnah, R. (2009). A videofluoroscopic study of the emphatic consonants in Jordanian Arabic. *Folia Phoniatrica et Logopaedica, 61*(4), 247-253.

Al-Tamimi, F., Tarawneh, R., & Howell, P. (2021). Development of voice onset time in Arabic. *Lingua*, 1-13.

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics, 71*, 306-325.

Altoma, S. (1969). *The Problem of Diglossia in Arabic: A Comparative Study of Classical and Iraqi Arabic*: Harvard University Press.

Al-Wer, E., Horesh, U., AlAmmar, D., Alaodini, H., Al-Essa, A., Al-Hawamdeh, A., Al-Qahtani, K., Hussain, A. (2020). Probing linguistic change in Arabic vernaculars: a sociohistorical perspective. *Language in Society, 51*, 29-50.

Amayreh, M. M. (2003). Completion of the Consonant Inventory of Arabic. *Journal of speech, language, and hearing research, 46*(3), 517-529. doi:10.1044/1092-4388(2003/042)

Amer, F. H., Adaileh, B. A., & Rakhieh, B. A. (2011). Arabic Diglossia: A Phonological Study. *Argumentum, 7*, 19-36.

Anderson, J. M., & Ewen, C. J. (1987). *Principles of Dependency Phonology* (Vol. 47). Cambridge: Cambridge University Press.

Backley, P. (2011). *Introduction to element theory*. Edinburgh: Edinburgh University Press.

Badawi, E.-S. (1973). *Mustawayat al-lugha al-ʾarabiyya al-muʿaṣira fi miṣr*. Cairo: Dar Al-maʿarif.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics, 39*(1), 39-49.

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics, 49*(2), 259-284.

Bellem, A. (2007). *Towards a Comparative Typology of Emphatics Across Semitic and into Arabic Dialect Phonology.* (Doctoral dissertation). University of London,

Bellem, A. (2014). Triads, emphatics and interdentals in Arabic sound system typology. *Journal of Semitic studies supplement, 34*, 9-41.

Ben Hamadi, H. M. (2019). *The Linguistic Landscape in Tunisia.* Paper presented at the Functional Aspects of Intercultural Communication. Translation and Interpreting Issues.

Benkirane, T. (1998). Intonation in Western Arabic (Morocco). In D. Hirst & A. Di Cristo (Eds.), *In Intonation systems: a survey of twenty languages* (pp. 348-362). Cambridge: Cambridge University Press.

Bentahila, A., & Davies, E. (1991). Standards for Arabic: One, two, or many. *Indian journal of applied linguistics., 17*(1), 69-88.

Bermúdez-Otero, R., & Trousdale, G. (2012). Cycles and continua: on unidirectionality and gradualness in language change. In T. Nevalainen & E. Traugot (Eds.), *The Oxford handbook on the history of English* (pp. 691-720): Oxford University Press.

Blau, J. (1977). *The Beginnings of the Arabic Diglossia: A Study of the Origins of Neo-Arabic*. Malibu, CA: Undena.

Boersma, P., & Weenink, D. (2016). Praat, a system for doing phonetics by computer. Retrieved from http://www.praat.org

Boudelaa, S., & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes, 28*(10), 1453-1473. doi:10.1080/01690965.2012.719629

Brown, J., Goodman, K. S., & Marek, A. M. (1996). *Studies in miscue analysis: An annotated bibliography*. Newark, DE: International Reading Association.

Bukshaisha, F. A. (1985). *An experimental phonetic study of some aspects of Qatari Arabic.* (Doctoral dissertation). University of Edinburgh,

Cadora, F. J. (1992). *Bedouin, Village, and Urban Arabic: an Ecolinguistic Study*. Leiden: E.J. Brill.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*(3), 239-245.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421-428.

Card, E. A. (1983). *A phonetic and phonological study of Arabic emphasis.* (Doctoral dissertation). Cornell University,

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*(2), 207-229.

Chodroff, E., Golden, A., & Wilson, C. (2019). Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America, 145*(1), EL109-EL115.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30-47.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. (1985). The Geometry of Phonological Features. *Phonology, 2*(1), 225-252.

Clements, G. N. (1991). Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory, 5*, 77-123.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology, 20*(3), 287-333. doi:10.1017/S095267570400003X

Clements, G. N., & Hallé, P. (2010). Phonetic bases of distinctive features": Introduction. *Journal of Phonetics, 1*(38), 3-9.

Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (Vol. 7, pp. 245-306). Oxford: Wiley-Blackwell.

Cyran, E. (2011). Laryngeal Realism and Laryngeal Relativism: Two Voicing Systems in Polish? *Studies in Polish Linguistics, 6*, 45-80.

Cyran, E. (2017). Voice' languages with no [voice]? Some consequences of Laryngeal Relativism. *Acta Linguistica Academica, 64*(4), 477-511. doi:10.1556/2062.2017.64.4.1

Davis, S. (1995). Emphasis Spread in Arabic and Grounded Phonology. *Linguistic Inquiry, 26*(3), 465-498.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics, 49*, 77-95.

Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents* (Vol. 9): De Gruyter Mouton.

Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society, 17*(2), 183-207.

El Zarka, D., & Hellmuth, S. (2009). Variation in the intonation of Egyptian Formal and Colloquial Arabic. *Langues et Linguistique, 22*, 73-92.

Elgibali, A. (1993). Stability and language variation in Arabic: Cairene and Kuwaiti dialects. In C. Holes & M. Eid (Eds.), *Perspectives on Arabic linguistics V: Papers from the 5th Annual Symposium on Arabic Linguistics* (Vol. 4, pp. 75-96). Amsterdam / Philadelphia: John Benjamins.

Eviatar, Z., & Ibrahim, R. (2000). Bilingual is as bilingual does: Metalinguistic abilities of Arabic-speaking children. *Applied Psycholinguistics, 21*(4), 451-471.

Farwaneh, S. (2009). Toward a typology of Arabic dialects: The role of final consonantality. *Journal of Arabic and Islamic Studies, 9*, 82-109.

Fathi, R. (2013). *Vowel length in Egyptian Arabic: a different view.* (Doctoral dissertation). Université Paris Diderot (Paris 7),

Ferguson, C. (1989). Grammatical Agreement in Classical Arabic and the Modern Dialects. *Al-'Arabiyya: journal of the American Association of Teachers of Arabic - Al-'Arabīyah: majallat Rābiṭat Asātidhat al-Lughah al-'Arabīyah, 22*(1), 5-17.

Ferguson, C. A. (1959). Diglossia. *Word, 15*(2), 325-340.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research, 92*, 233-277.

Flemming, E. (2005). Deriving natural classes in phonology. *Lingua, 115*(3), 287-309.

Foulkes, P., Scobbie, J. M., & Watt, D. (2010). Sociophonetics. In W. Hardcastle, J. Laber, & F. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 703-754): Wiley-Blackwell.

Froud, K., & Khamis-Dakwar, R. (2018). Neurophysiological investigations in studies of Arabic linguistics: The case of Arabic diglossia. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge Handbook of Arabic Linguistics* (pp. 249-262): Routledge.

Ghazeli, S. (1977). *Back consonants and backing coarticulation in Arabic.* (Doctoral dissertation). University of Texas at Austin,

Gibson, M. (2002). Dialect Levelling in Tunisian Arabic: Towards a New Spoken Standard. In A. Rouchdy (Ed.), *Language Contact and Language Conflict in Arabic: Variation on a Sociolinguistic Theme* (pp. 24-40): Routledge.

Gibson, M. (2008). Tunis Arabic. *Encyclopedia of Arabic Language and Linguistics, 4*, 563-571.

Goad, H. (1991). *[ATR] and [RTR] are Different Features.* Paper presented at the Proceedings of the 10th West Coast Conference on Formal Linguistics, Somerville.

Goodman, K. (2005). Making Sense of Written Language: A Lifelong Journey. *Journal of Literacy Research, 37*(1), 1-24.

Goodman, Y. (2015). Miscue Analysis: A Transformative Tool for Researchers, Teachers, and Readers. *Literacy Research: Theory, Method, and Practice, 64*(1), 92-111.

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics, 39*(3), 192-193.

Hale, M., & Reiss, C. (2000). Phonology as Cognition. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological Knowledge: Conceptual and Empirical Issues* (pp. 161-184).

Hamann, S. (2015). Phonological changes. In C. Bowern & B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics* (pp. 267-281): Routledge.

Harrell, R. (1960). A Linguistic Analysis of Egyptian Radio Arabic. In C. Ferguson (Ed.), *Contributions to Arabic Linguistics* (pp. 3-67). Cambridge & Massachusetts: Harvard University Press.

Harris, J. (1990). Segmental Complexity and Phonological Government. *Phonology, 7*(2), 255-300.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Harris, J., & Lindsey, G. (1995). The elements of phonological representation. In J. Durand & F. Katamba (Eds.), *Frontiers of Phonology: Atoms, Structures, Derivations* (pp. 34-79): Routledge.

Hassan, Z. M. (2005). *Acoustic evidence of the prevalence of the emphatic feature over the word in Arabic.* Paper presented at the Proceedings FONETIK, Göteborg University.

Hayes-Harb, R., & Durham, K. (2016). Native English speakers' perception of Arabic emphatic consonants and the influence of vowel context. *Foreign Language Annals, 49*(3), 557-572.

Hayward, K. M., & Hayward, R. J. (1989). Guttural': Arguments for a new distinctive feature. *Transactions of the Philological Society, 87*(2), 179-193. doi:10.1111/j.1467-968X.1989.tb00626.x

Heine, B. (1999). *Ik Dictionary* (Vol. 15). Köln: Rüdiger Köppe Verlag.

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics, 36*(4), 607-628.

Hellmuth, S., & Almbark, R. (2019). *Intonational Variation in Arabic Corpus*. In: Retrieved from Colchester, Essex http://reshare.ukdataservice.ac.uk/852878/.

Hermena, E. W., Drieghe, D., Hellmuth, S., & Liversedge, S. P. (2015). Processing of Arabic diacritical marks: Phonological–syntactic disambiguation of homographic verbs and visual crowding effects. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 494-507.

Herzallah, R. S. (1991). *Aspects of Palestinian Arabic phonology: A nonlinear approach.* (Doctoral dissertation). Cornell University,

Heselwood, B. (1996). Glottal states and emphasis in Baghdadi and Cairene Arabic: synchronic and diachronic aspects. In J. Dickins, B. Heselwood, & J. Watson (Eds.), *Three topics in Arabic phonology: some speculations on the vowel system of classical Arabi* (Vol. 53, pp. 20-44). Durham: Centre of Middle Eastern and Islamic Studies.

Hetzron, R. (2013). *The Semitic Languages*: Routledge.

Hoberman, R. D. (1987). Emphasis (pharyngealization) as an autosegmental harmony feature. *Parasession on Autosegmental and Metrical Phonology, 23*, 167-181.

Holes, C. (1989). Towards a dialect geography of Oman. *Bulletin of the School of Oriental and African Studies, 52*(3), 446-462.

Holes, C. (1990). *Gulf Arabic*: Routledge.

Holes, C. (2004). *Modern Arabic: Structures, Functions, and Varieties*. Washington, DC: Georgetown University Press.

Holes, C. (2006). *The Arabic dialects of Arabia.* Paper presented at the Proceedings of the Seminar for Arabian Studies, London.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In M. van Oostendorp & J. van de Weijer (Eds.), *The Internal Organization of Phonological Segments* (pp. 319-354). Berlin: Mouton de Gruyter.

Hussain, A. (2017). *The sociolinguistic correlates of dialect contact and koineisation in Medini Arabic: Lenition and resyllabification.* (Doctoral dissertation). University of Essex,

Hyman, L. M. (2005). Initial vowel and prefix tone in Kom: Related to the Bantu Augment? In K. Bostoen & J. Maniacky (Eds.), *Studies in African comparative linguistics with special focus on Bantu and Mande: Essays in honour of Y. Bastin and C. Grégoire* (pp. 313-341). Köln: Rüdiger Köppe Verlag.

Hyman, L. M. (2014). What is Phonological Typology? *UC Berkeley PhonLab Annual Report*, 101-118. doi:10.5070/P73hr9r9cd

Ibrahim, M. H. (1983). Linguistic distance and literacy in Arabic. *Journal of Pragmatics, 7*(5), 507-515.

Ibrahim, R. (2006). Morpho-phonemic similarity within and between languages: A factor to be considered in processing Arabic and Hebrew. *Reading and Writing, 19*(6), 563-586.

Ibrahim, R. (2009). The cognitive basis of diglossia in Arabic: Evidence from a repetition priming study within and between languages. *Psychology Research and Behavior Management, 2*, 93-105.

Ibrahim, R., & Aharon-Peretz, J. (2005). Is literary Arabic a second language for native Arab speakers?: Evidence from semantic priming study. *Journal of Psycholinguistic Research, 34*(1), 51-70.

Ibrahim, Z. (2009). *Beyond Lexical Variation in Modern Standard Arabic: Egypt, Lebanon and Morocco*: Cambridge Scholars Publishing.

Ingham, B. (1994). *Najdi Arabic: Central Arabian*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Israel, A., Proctor, M., Goldstein, L., Iskarous, K., & Narayanan, S. (2012). *Emphatic segments and emphasis spread in Lebanese Arabic: a real-time magnetic resonance imaging study.* Paper presented at the Thirteenth Annual Conference of the International Speech Communication Association, Portland, OR, USA.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology, 12*(3), 369-396.

Iverson, G. K., & Salmons, J. C. (2006). On the typology of final laryngeal neutralization: Evolutionary Phonology and laryngeal realism. *Theoretical Linguistics, 32*(2), 205-216.

Jabeur, M. (1987). *A Sociolinguistic Study in Rades: Tunisia.* (Doctoral dissertation). University of Reading,

Jakobson , R. (1957). Mufaxxama: The 'emphatic' phonemes in Arabic. *Volume I Phonological Studies*, 510-522.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch.* (Doctoral dissertation). University Library Groningen,

Jesry, M. (1996). *Some cognitively controlled coarticulatory effects in Arabic and English, with particular reference to voice onset time.* (Doctoral dissertation). University of Essex,

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology, 19*(2), 189-218.

Johnstone, T. (1967). *Eastern Arabian Dialect Studies*. London / New York: Oxford University Press.

Jongman, A., Herd, W., Al-Masri, M., Sereno, J., & Combest, S. (2011). Acoustics and perception of emphasis in Urban Jordanian Arabic. *Journal of Phonetics, 39*(1), 85-95.

Kahn, M. (1975). Arabic emphatics: The evidence for cultural determinants of phonetic sex-typing. *Phonetica, 31*(1), 38-50.

Keating, P. (1984). Phonetic and Phonological Representation of Stop Consonant Voicing. *Language, 60*(2), 286-319.

Keating, P. (2003). *Phonetic and other influences on voicing contrasts.* Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.

Kehoe, M., & Kannathasan, K. (2021). Development of voice onset time in monolingual and bilingual French-speaking children. *Lingua, 252*, 1-19.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*(2), 143-168.

Khamis-Dakwar, R. (2005). Children's attitudes towards the diglossic situation in Arabic and its impact on learning. *Languages, Communities, and Education*, 75-86.

Khamis-Dakwar, R., Boudelaa, S., & Froud, K. (2009). *Lexical processing in diglossic code switching between Modern Standard Arabic and Palestinian Colloquial Arabic: An event-related brain potential study of Arabic native speakers*. Paper presented at the Poster presented at Cognitive Neuroscience Symposium, San Francisco.

Khamis-Dakwar, R., & Froud, K. (2006). Lexical Processing In Two Language Varieties. In M. Mughazy (Ed.), *Perspectives on Arabic Linguistics XX: Papers from the 20th annual symposium on Arabic linguistics* (Vol. 290, pp. 153-166). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Khamis-Dakwar, R., & Froud, K. (2012). Neurocognitive modeling of the two language varieties in Arabic Diglossia. In R. Khamis-Dakwar & K. Froud (Eds.), *Perspectives on Arabic Linguistics XXVI: Papers from the annual symposium on Arabic Linguistics* (Vol. 2, pp. 285-301). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Khamis-Dakwar, R., & Froud, K. (2019). Diglossia and language development. In E. Al-Wer & U. Horesh (Eds.), *The Routledge Handbook of Arabic Sociolinguistics* (1 ed., pp. 300-313): Routledge.

Khamis-Dakwar, R., Taha, H., & Al-Khoshman, A. (2021). How Arabs can win back their language. https://mondoweiss.net/2021/12/how-arabs-can-win-back-their-language/

Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. In D. B. Parkinson & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics XIII-XIV: Papers from the 13th and 14th Annual Symposia on Arabic Linguistics* (pp. 95-122). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Khattab, G., Al-Tamimi, F., & Heselwood, B. (2006). Acoustic and Auditory Differences in the /t/-/ṭ/ Opposition in Male and Female Speakers of Jordanian Arabic. In S. Boudelaa (Ed.), *Perspectives on Arabic Linguistics XVI: Papers from the 16th Annual Symposium on Arabic Linguistics* (Vol. 266, pp. 131-160). Amsterdam / Philadelphia: John Benjamins Publishing.

Kim, C.-W. (1970). A Theory of Aspiration. *Phonetica, 21*(2), 107-116.

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language, 70*(3), 419-454.

Kiparsky, P. (2003). Syllables and moras in Arabic. In C. Fery & R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 147-182). Cambridge: Cambridge University Press.

Kiparsky, P. (2018). Formal and Empirical Issues in Phonological Typology. In *Phonological Typology* (pp. 54-106): De Gruyter Mouton.

Kriba, H. A. (2010). *Acoustic parameters of emphasis in Libyan Arabic.* (Doctoral dissertation). Newcastle University,

Kulikov, V. (2020). Laryngeal Contrast in Qatari Arabic: Effect of Speaking Rate on Voice Onset Time. *Phonetica, 77*(3), 163-185.

Kulikov, V. (2021). Voice and Emphasis in Arabic Coronal Stops: Evidence for Phonological Compensation. Language and Speech, 65(1), 73 –104.

Labov, W. (1972). *Sociolinguistic Patterns*: University of Pennsylvania Press.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology, 7*, 637-675.

Laufer, A., & Baer, T. (1988). The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Language and Speech, 31*(2), 181-205.

Lehn, W. (1963). Emphasis in Cairo Arabic. *Language, 39*(1), 29-39.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America, 49*(2B), 606-608.

Lombardi, L. (1994). *Laryngeal Features and Laryngeal Neutralization* (1st ed.). London: Routledge.

Lombardi, L. (1995). Laryngeal features and privativity. *The Linguistic Review, 12*(1), 35-60.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica, 56*(3-4), 158-185.

McCarthy, J. J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (Vol. 86, pp. 191-233): Cambridge University Press.

Mester, R. A., & Itô, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language*, 258-293.

Midtlyng, P. J. (2011). *The effects of speech rate on VOT for initial plosives and click accompaniments in Zulu.* Paper presented at the Selected Proceedings of the 40th Annual Conference on African Linguistics, Somerville, MA.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*(1-3), 106-115.

Milroy, J., & Milroy, L. (1993). Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics, 3*(1), 57-77.

Mitleb, F. (2009). *Voice Onset Time of Jordanian Arabic Stops.* Paper presented at the 3rd International Conference on Arabic Language Processing, Rabat, Morocco.

Morén, B. (2003). The Parallel Structures Model of Feature Geometry. *Working Papers of the Cornell Phonetics Laboratory, 15*, 194-270.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1), 1-18.

Nevat, M., Khateb, A., & Prior, A. (2014). When first language is not first: an functional magnetic resonance imaging investigation of the neural basis of diglossia in Arabic. *European Journal of Neuroscience, 40*(9), 3387-3395. doi:10.1111/ejn.12673

Norlin, K. (1987). *A phonetic study of emphasis and vowels in Egyptian Arabic.* (Doctoral dissertation). Lund University,

Owens, J. (2013). *The Oxford handbook of Arabic linguistics*. Oxford: Oxford University Press.

Palva, H. (1991). Is there a North West Arabian Dialect Group? In M. Forstner (Ed.), *Festgabe für Hans-Rudolf Singer* (pp. 151-166). Frankfurt am Main: Peter Lang.

Parkinson, D. (1993). Knowing Standard Arabic: Testing Egyptians' MSA Abilities. In C. Holes & M. Eid (Eds.), *Perspectives on Arabic Linguistics V: Papers from the 5th Annual Symposium on Arabic Linguistics* (Vol. 4, pp. 47-73). Amsterdam / Philadelphia: John Benjamins.

Parkinson, D. (1996). Variability in Standard Arabic Grammar Skills. In A. Elgibali & E.-S. M. Badawi (Eds.), *Understanding Arabic: Essays in Contemporary Arabic Linguistics in Honor of El-Said Badawi* (pp. 91-101). Cairo: American University in Cairo Press.

Petrova, O., Plapp, R., Ringen, C., & Szentgyörgyi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review, 23*(1), 1-35.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics, 57*(3), 291-304.

Plumlee, M. (2017). The linguistic landscape of Cairo from the Rosetta Stone to the Ring Road billboards. In A. Gebril (Ed.), *Applied Linguistics in the Middle East and North Africa: Current Practices and Future Directions* (pp. 115-160): John Benjamins Publishing Company.

Prochazka, T. (1988). *Saudi Arabian Dialects*. London / New York: Kegan Paul International: Distributed by Routledge.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rabin, C. (1955). The Beginnings of Classical Arabic. *Studia Islamica*(4), 19-37.

Raphael, L. J., Tobin, Y., Faber, A., Most, T., Kollia, H. B., & Milstein, D. (1995). Intermediate Values of Voice Onset Time. In F. Bell-Berti & L. Raphael (Eds.), *Producing Speech: Contemporary Issues. for Katherine Safford Harris* (pp. 117-127): AIP Press.

Raphael, L. J., Tobin, Y., & Most, T. (1983). Atypical VOT categories in Hebrew and Spanish. *The Journal of the Acoustical Society of America, 74*.

Rifaat, K. (2003). *Voice Onset Time in Egyptian Arabic: A case where phonological categories dominate.* Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences.

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics, 35*(3), 439-443.

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics, 20*(2), 269-286.

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics, 41*(6), 479-490.

Royal, A.-M. (1985). *Male-Female pharyngealization patterns in Cairo Arabic: A sociolinguistic study of two neighborhoods.* (Doctoral dissertation). University of Texas at Austin,

Saiegh-Haddad, E. (2012). Literacy Reflexes of Arabic Diglossia. In M. Leikin, M. Schwartz, & Y. Tobin (Eds.), *Current Issues in Bilingualism* (pp. 43-55): Springer.

Sayahi, L. (2011). Introduction: Current perspectives on Tunisian sociolinguistics. *International Journal of the Sociology of Language*(211), 1-8.

Schwartz, G. (2010). Phonology in the speech signal-Unifying cue and prosodic licensing. *Poznań Studies in Contemporary Linguistics, 46*(4), 499-518.

Schwartz, G. (2013). A representational parameter for onsetless syllables. *Journal of Linguistics, 49*(3), 613-646. doi:10.1017/S0022226712000436

Schwartz, G., & Arndt, D. (2018). Laryngeal Realism vs. Modulation Theory–evidence from VOT discrimination in Polish. *Language Sciences, 69*, 98-112.

Shahin, K. N. (1996). Accessing Pharyngeal Place in Palestinian Arabic. In D. Parkinson & M. Eid (Eds.), *Perspectives on Arabic Linguistics IX: Papers from the 9th Annual Symposium on Arabic Linguistics* (pp. 131-150). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Siemund, R., Heuft, B., Choukri, K., Emam, O., Maragoudakis, E., Tropf, H., . . . Rodriguez, A. N. (2002). *OrienTel—Arabic speech resources for the IT market.* Paper presented at the LREC (Arabic workshop).

Summerfield, Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception, 2*(4), 61-72.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074-1095.

Tamim, N. (2017). *Voicing Contrast of Stops in the Palestinian Arabic Dialect.* (Master's thesis). Universiteit van Amsterdam,

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. Retrieved from http://ncslaap.lib.ncsu.edu/tools/norm/

Traunmüller, H. (1994). Conventional, biological and environmental factors in speech communication: A modulation theory. *Phonetica, 51*(1-3), 170-183.

Van Putten, M. (2017a). The Archaic Feminine Ending *-at* in Shammari Arabic. *Journal of Semitic Studies, 62*(2), 357-369.

Van Putten, M. (2017b). The Status of Quaestionis of 'Arabiyyah, Pre-Islamic Poetry and the Quran. Retrieved from https://iqsaweb.wordpress.com/2017/11/06/the-status-quaestionis-of-%ca%bfarabiyyah-pre-islamic-poetry-and-the-quran/

Vaux, B., & Samuels, B. (2005). Laryngeal Markedness and Aspiration. *Phonology, 22*(3), 395-436.

Versteegh, K. (2003). Pidginization and creolization revisited: The case of Arabic. In K. Versteegh, M. Haak, & R. de Jong (Eds.), *Approaches to Arabic dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday* (pp. 343-357): Brill.

Versteegh, K. (2014). *The Arabic Language* (2nd ed.): Edinburgh University Press.

Wahba, K. (1996). Linguistic variation in Alexandrian Arabic: The feature of emphasis. In A. Elgibali & E.-S. M. Badawi (Eds.), *Understanding Arabic: Essays in Contemporary Arabic Linguistics in Honor of El-Said Badawi*. Cairo: American University in Cairo Press.

Watson, J. (1999). The Directionality of Emphasis Spread in Arabic. *Linguistic Inquiry, 30*(2), 289-300.

Watson, J. (2002). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Watson, J. (2011a). Arabic Dialects (general article). In S. Weninger, G. Khan, M. Streck, & J. Watson (Eds.), *The Semitic Languages: An International Handbook*. Berlin: Mouton De Gruyter.

Watson, J. (2011b). Word Stress in Arabic. In M. Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 2990 - 3019). Oxford: Wiley-Blackwell.

Watson, J., & Heselwood, B. (2016). Phonation and glottal states in Modern South Arabian and San'ani Arabic. In Y. A. Haddad & E. Potsdam (Eds.), *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics* (pp. 3-37). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Wetzels, W. L., & Mascaró, J. (2001). The Typology of Voicing and Devoicing. *Language, 77*(2), 207-244.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed.): Springer International Publishing.

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation (Version R package version 0.8.3). Retrieved from https://CRAN.R-project.org/package=dplyr

Wojtkowiak, E., & Schwartz, G. (2018). *Sandhi*-Voicing in Dialectal Polish: Prosodic Implications. *Studies in Polish Linguistics, 13*(2), 123-143.

Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*(1), 35-48.

Younes, M. (1982). *Problems in the Segmental Phonology of Palestinian Arabic.* (Doctoral dissertation). University of Texas at Austin,

Youssef, I. (2013). *Place Assimilation in Arabic: Contrasts, Features, and Constraints.* (Doctoral dissertation). University of Tromsø, Tromsø, Norway.

Youssef, I. (2014). *Emphasis Spread in Cairene Arabic: A Reassessment.* In O. Durand, A. Langone, and G. Mion (Eds.), Alf lahǧa wa lahǧa: Proceedings of the 9th AIDA Conference, (pp. 455-464). Münster: LIT Verlag

Youssef, I. (2015). Palatalization in educated Cairene Arabic. *Nordlyd, 42*, 21-31.

Youssef, I. (2021). Contrastive Feature Typologies of Arabic Consonant Reflexes. *Languages*, *6*(3), 141-158. https://doi.org/10.3390/languages6030141

Zaba, A. M. (2007). Native English speakers and Arabic pharyngealization contrasts: Perceptual and acoustic data. *The Journal of the Acoustical Society of America, 121*(5), 3071-3071. https://doi.org/10.1121/1.4781871

Zawaydeh, B. A. (1998). Gradient uvularization spread in Ammani-Jordanian Arabic. In E. Benmamoun, M. Eid, & N. Haeri (Eds.), *Perspectives on Arabic Linguistics XI: Papers from the 11th Annual Symposium on Arabic Linguistics* (pp. 117-141). Amsterdam / Philadelphia: John Benjamins Publishing Company.