Vowel duration and aspiration effects in Icelandic

Y3835994

This thesis is submitted in partial fulfilment of the
requirements for the degree of
Phonetics and Phonology
in the Department of Language and Linguistic Science

University of York
York, YO10 5DD, UK

08 September 2016

Abstract

This dissertation dealt with the relation between vowel duration and aspiration in consonants. Previous studies showed that vowels in Hindi are longer when followed by aspirated consonants (like in the word *sāth* "companionship"), while are shorter if a non-aspirated consonant follows (like in *sāt* "seven"). Such phenomenon has been called the "aspiration effect." No explanation has been given to this effect and the research conducted to clarify it only focussed on post-aspiration. The finding that post-aspiration in a consonant lengthens the preceding vowel implies two possible prediction for what concerns pre-aspiration: pre-aspiration could behave like post-aspiration, causing vowels to lengthen; or it could instead make them shorter. If the aspiration effect is caused by the relative phasing of glottal spreading, an early timing of glottal abduction (as in pre-aspiration) should have a shortening effect. I called this the "timing hypothesis." I carried out a data collection with 5 natives speakers of Icelandic, a language that has contrastive aspiration in stops and sonorants, to test the timing hypothesis. I then extracted the duration of vowels followed by aspirated versus non-aspirated consonants. The results showed that vowels before aspirated consonants (like in Icelandic *takka* 'key' [t^ha^hka]) were significantly shorter than vowels followed by non-aspirated consonants (like in *kagga* 'barrel' [k^hakka]). On average, vowels followed by aspirated consonants were 30 milliseconds longer than when followed by non-aspirated stops. While such difference can be attributed to the different duration of aspirated and non-aspirated sonorants, geminate stops showed precisely the predicted difference in timing of glottal spread. The beginning of the abduction gesture—necessary to sustain voicelessness in non-aspirated geminates and frication in pre-aspirated geminates—was timed significantly earlier in pre-aspirated stops. The results thus indicate that, at least in Icelandic, the differences in vowel duration could be caused by differences in timing of laryngeal spread, thus provisionally validating the timing hypothesis.

As more absolute universals “bite the dust,” we should not despair. The challenges posed by universal tendencies arc just as great, if not greater, than those posed by absolute universals. We must understand not only why most languages have a particular property, but why there are the rare exceptions that do not. By meeting these challenges, we will ultimately have a deeper understanding of sound patterns as they reflect human potential in articulation, perception, and general cognition.

– Blevins (2009)

Contents

List of Tables

List of Figures

1 Introduction

This study deals with the relationship between the presence versus the absence of aspiration in a post-vocalic consonant and the duration of the vowel preceding that consonant. Maddieson and Gandour (1976) and Maddieson (1976) found that, in Hindi and a few other Indic languages, vowels followed by aspirated consonants (like in the word *sāt* "seven") are longer than vowel followed by non aspirated consonants (like in *sāth* "companionship"). This phenomenon has been subsequently called the "aspiration effect" by Durvasula and Luo (2012), who further confirmed that the duration of vowel was related to aspiration. Even if the effect has been shown to exist, no explanation has been given for it. Moreover, these studies only dealt with post-aspiration. To fully understand the phenomenon is necessary to extend the enquiry to pre-aspiration.

Such step has been taken in conducting this research. Icelandic has been chosen as the language under study since its consonantal system can be interpreted as thoroughly based on aspiration contrasts. In particular, aspiration in Icelandic involves both stops and sonorants consonants. In certain contexts, non-aspirated geminate stops (/CC/) contrast with pre-aspirated stops (/^{h}C/). Similarly, voiced sonorants followed by a stop (/SC/) can contrast in the same contexts with voiceless sonorants followed by a stop (/S̥C/). Since voiceless sonorants can be classed together with pre-aspirated stops, it is possible to explore the effect the presence of aspiration has on the duration of the preceding vowel. The Icelandic phonological system thus constitutes an ideal locus of enquiry on the effects of aspiration on vowel duration in a more systematic way and in novel contexts than previously possible.

The studies on aspiration and vowel duration cited above showed that syllable-final *post*-aspirated stops lengthened the preceding vowel. These results are compatible with two opposite predictions with regard to *pre*-aspirated stops. In the simplest case, aspiration should consistently lengthen the vowel preceding the relevant consonant, independently of whether it is timed before or after the stop closure. In a case where aspiration behaves symmetrically in relation to its timing, pre-aspiration should show a pattern opposite to that of post-aspiration, where the vowel preceding the pre-aspirated stop should be shorter. If the second case is true, it could be asserted that the aspiration effect arises from the pattern of timing of the laryngeal spreading gesture in relation to the oral gestures. I call this view the “timing hypothesis,” which will be the focus of this study.

1.1 Research design

This research has been designed so as to answer two questions: (1) does Icelandic show the aspiration effect? and (2) is the aspiration effect caused by the relative timing of laryngeal gestures in relation to oral gestures? To access the existence of the aspiration effect in Icelandic, I collected data from native speakers of the language. They were recorded while reading a list of sentences which contained the target words. The list of target words was selected so as to include aspirated and non-aspirated stops and sonorants, both in word-final (in monosyllabic words) and word-medial position (in disyllabic words). To test the validity of the laryngeal timing hypothesis, I performed a statistical analysis to verify the following research hypothesis: vowels followed by aspirated consonants (pre-aspirated stops and voiceless sonorants) are shorter than vowels followed by non-aspirated consonants (plain stops and voiced sonorants).

1.2 Reproducibility

The term “reproducible research” was first coined by Prof. Jon Claerbout at Stanford University, around 2000 (Fomel and Claerbout, 2009). The concept of reproducible research stems from the idea the the product of scientific inquiry should not only consist in the dissemination of the research results in the form of an output document (like a journal paper, a dissertation or a book). Instead, such outcome should also include the environment in which the analysis that gave the results was performed. Such an environment consists of the data sets and of the computational operations (in the form of code) used in the analysis. A research is said to be reproducible when other people than the original authors can reproduce step by step the analysis that was conducted on the same data collected for that research. While the concept of reproducibility might at first sight be similar to that of replicability, a scientific discovery is replicated—and not reproduced—when "independent investigators, methods, data, equipment, and protocols" are used (Peng, 2009).

One way to satisfy the reproducibility criterion is using literate programming. In literate programming, the computer code used to generate the results of the research is embedded within the research document. The idea of literate programming has been developed by Donald Knuth (Knuth, 1984) as a way of simplifying the documentation process of computer programs. This solution has been subsequently applied to scientific research, as a means for ensuring reproducibility. Reproducibility and literate programming are new in linguistics, but some scholars are encouraging descriptive linguists to make their grammatical descriptions reproducible (Maxwell and Amith, 2005; Maxwell, 2013).

An interesting case in the area of phonology has been made by Maxwell (2013) regarding a theoretical debate based on data from Yokuts [GLOTTO: `yoku1256`].[1] As Weigel (2002) and Blevins (2004) pointed out, more than two thirds of the Yokuts lexical forms used as an argument in favour of theoretical claims turned out to be a construction of the researcher based on descriptions of the language. What is worse is that those constructed word forms were incorrect. More than thirty years of debate has been founded on false data. The moral of this story is that linguistics as a whole would benefit from the application of reproducibility. According to such spirit, this dissertation has been written in XeLaTeX and the analysis has been conducted using the R language (R Core Team, 2015). The code generating the analysis has been embedded in the dissertation source code using `knitr`, a package that enables literate programming support in R. The analysis source code can be found at `http://github.com/stefanocoretta/icelandic-preaspiration`. The dataset will soon be downloadable from `http://dx.doi.org/10.18710/7NLJSG`.

1.3 Dissertation structure

The dissertation is thus organised. Chapter 2 contains a review of the literature relevant for the present study. I will first introduce some background terminology and concepts

[1]For the major languages cited in this work, I give the linguistic code from the Glottolog database (Hammarström et al., 2016) as an alternative to the more conventional ISO 639 code.

of the phonetics of laryngeal activity, followed by an overview of the relation between presence versus absence of aspiration and vowel duration. Then, I will provide a brief description of the phonological system of Icelandic with a focus on the laryngeal contrasts in stops and sonorants. The chapter closes with the presentation of the research hypothesis. Chapter 3 deals with the methodology of the experiment. I will first discuss about the participants recruited for this study, the materials and the procedure used in the experimental task. I will then give particular attention to the annotation scheme that was applied to the data and to the criteria used to extract the measurements. In Chapter 4 I will present the results of the statistical analysis run on the extracted measurements. Separate sections will deal with the duration of the words, the Voice Onset to Release (VOR), the vowels, the stop closure, the glottal spread gesture and of the sonorants. Chapter 5 examines the results on the light of the research hypothesis. This chapter also deals with the linguistics aspects of the research, its limitations and challenges, and the possible future implementations of this study. Finally, Chapter 6 concludes the dissertation with a summary of this work.

2 Literature review

This chapter contains the literature review of this study. In Section 2.1, I briefly introduce some basic concepts and terminology related to laryngeal activity in speech production. Section 2.2 deals with the phenomenon known as “aspiration effect,” according to which vowels followed by aspirated consonants are longer than when they are followed by non-aspirated consonants. In Section 2.3, I discuss the major aspects of the phonology of Icelandic, focussing in particular on the types and contexts of laryngeal contrasts in stops and sonorants. Finally, in Section 2.4, I describe the research hypothesis of this study and the reasons behind it.

2.1 Phonation types and states of the glottis

The human voice is the product of the periodic noise generated in the larynx, the organ situated in the neck and involved mainly in breathing and speech. In particular, the vibration of the vocal folds generates oscillations of the air pressure which are translated into a percept of periodic noise, or voicing, by the auditory system. Even silence during speech is due to a particular configuration of the glottis (the empty space between the two vocal folds). The larynx, with the vocal folds, thus plays a major role in human communication.

Halle and Stevens (2002) review which configurations the vocal folds can assume. Such configurations are known as the states of the glottis. The four principal states of the glottis are (1) spread glottis, (2) constricted glottis, (3) stiff vocal folds, and (4) slack vocal folds. A spread glottis state corresponds to abducted vocal folds. This state is the one naturally used during normal breathing. If the vocal folds are adducted so as they tightly press against each other, the glottis assumes a constricted configuration. Independently from the first two states, the vocal folds can be either stiff or slack. When the vocal folds are stiff, the distance between the upper and the lower edges of the folds is greater. In the case of slack vocal folds, such distance is smaller and there is no tension in the vocal folds. The first two states of the glottis can combine with the other two to create several phonation types.

Ladefoged (1973) discusses the four main phonation types that are employed to create phonological contrast in the languages of the world. These are voicelessness, modal voice, breathy voice and creaky voice (Halle and Stevens, 2002). A voiceless sound is produced with a laryngeal configuration that prevents vibration of the vocal folds. The vocal folds can either be stiffened so as to prevent them vibrating, or abducted just enough to increase the threshold of pressure build up and favour voicing. Modal voicing, instead, is produced with slack vocal folds and/or with adducted vocal folds. When the transglottal pressure is sufficiently weak, the pressure of the airflow coming from the lungs can set the vocal folds to vibrate and produce a periodic sound, which is voicing. Such mechanism of generating voicing is not the only one. If the vocal folds are abducted enough to allow more airflow, but not as much as to prevent voicing, breathy voicing is produced. On the other hand, if the vocal folds are constricted, the rate of vocal folds vibration decreases and voicing is characterised by a creaky sound (hence the term creaky voice).

When glottal spread is combined with voiceleness in stop consonants, the result is what is commonly known as aspiration. Aspiration is, acoustically speaking, aperiodic noise in the higher frequency range generated by the substantial amount of airflow coming from the lungs through the abducted glottis. In aspirated stops, aspiration normally follows the stop release, even if in certain languages aspiration is timed before the occlusion is made. In the first case we talk about post-aspirated stops, while in the second of pre-aspirated stops.

2.2 The effect of aspiration on vowel duration

Vowel duration has been reported in the literature to correlate with the presence or absence of aspiration in the following consonant. In particular, Maddieson and Gandour (1976) and Durvasula and Luo (2012) found that vowels followed by aspirated

Table 2.1: Mean duration of vowels in Hindi when followed by dental and velar stops.[a]

Maddieson and Gandour (1976)		Lampp and Reklis (2004)	
consonant	vowel duration (msec)	consonant	vowel duration (msec)
t	160	k	188
d	184.5	g	187
tʰ	184.75	kʰ	217
dʱ	196	gʱ	221

[a] The values of the standard deviations were not reported in the studies, so it is not included here.

consonants in Hindi are longer than vowels followed by non-aspirated consonants. In the following paragraphs, I will briefly introduce the system of laryngeal oppositions of Hindi. I will then review some of the findings concerning the aspiration effect and the major theories regarding the cause of this phenomenon.

The consonantal system of Hindi [GLOTTO: `hind1269`] is based on a four-way opposition of laryngeal contrasts (Ohala, 1983). For each place of articulation, there is a voiceless unaspirated, a voiced unaspirated, a voiceless aspirated and (breathy) voiced aspirated stop: for example, [t], [d], [tʰ], [dʱ]. The voiceless aspirated stops (like [tʰ]) are similar to the aspirated stops of English: a relatively long VOT follows the release of the occlusion. The voiced counterpart (like [dʱ]) is normally voiced throughout the closure and the aspiration is characterised by breathy voicing. Maddieson and Gandour (1976) found that vowels followed by voiced and voiceless aspirated stops (like in [kaːd] 'embroider' and [kaːtʰ] 'wood') were of equal length but longer than vowels followed by voiceless stops (like in [kaːt] 'cut'). Moreover, vowels followed by voiced aspirated stops (like in [saːdʱ] 'balance') were even longer than voiced and voiceless aspirated stops. Table 2.1 shows the mean duration of vowels before the four alveolar stops as reported by Maddieson and Gandour (1976) and Lampp and Reklis (2004).

As it can be seen from Table 2.1, Lampp and Reklis (2004) could not replicate the findings in Maddieson and Gandour (1976). Durvasula and Luo (2012), however, performed a more controlled task and found clear evidence that aspiration lengthens the preceding vowel.[1] They also noted that an increase in closure duration of the stop fol-

[1] Voicing also played a role in the lengthening of vowels, where vowels followed by voiced stops were longer than vowels followed by voiceless stops, independently from aspiration. Voicing and aspiration interacted so that a vowel followed by a voiced aspirated consonant was roughly two times longer than a

lowing the vowel correlated with an increase of duration of the vowel, when controlling for other factors like voicing and aspiration. The positive correlation between vowel duration and closure duration found in this study is the opposite of what has been found by Maddieson and Gandour (1976) and Lampp and Reklis (2004). Aspirated consonants in Hindi have a shorter closure duration than non-aspirated consonants. Since vowels followed by aspirated consonants are longer, it is logical to assume that the shorter closure duration allows for the vowel to be longer, and, vice versa, the longer closure duration of non-aspirated stops makes the preceding vowel shorter. However, since—after controlling for confounds—the duration of the vowel was positively correlated with closure duration, such explanation must be discarded.

Maddieson and Gandour (1976) and Durvasula and Luo (2012) do not provide an explanation for the effect of aspiration on vowel duration. Instead, they review proposals from other studies that dealt with a homologous phenomenon, the voicing effect. The voicing effect states that vowels followed by voiced consonants are longer than vowels followed by voiceless consonants (House and Fairbanks, 1953; Chen, 1970; Hussein, 1994; Durvasula and Luo, 2012). The voicing effect has been found in several languages, including English, Swedish, French, Arabic and Hindi (Sóskuthy, 2013, p. 191). However, the strength and consistency of such effect vary depending on the language and on different contexts within one language. While English has been reported to have a strong and consistent voicing effect, other languages, like Spanish, show a weaker effect (Hussein, 1994). On the other hand, yet other languages, such as Italian, have been shown to have no voicing effect at all (Esposito, 2002).[2]

Several explanations for the voicing effect have been proposed in the literature. Sóskuthy (2013) listed them into two categories. One school of thought ascribes the voicing effect to articulatory properties of the consonant following the vowel (Belasco, 1953; Chen, 1970). For example, Belasco (1953) argues that the voicing effect depends on the articulatory force needed to produce voiceless consonants. Since, according to him, articulatory force is kept constant, vowels before voiceless stops are shortened to compensate for the higher articulatory force needed to produce them.

A second line of reasoning, instead, relates it to auditory principles (Javkin, 1976; Kluender et al., 1988). Since the transition between vowels and voiced stops is gradual and does not have a clear boundary as with vowels followed by voiceless stops, Javkin

vowel followed by a non-aspirated voiceless stop. On the relation between vowel duration and voicing, see the following paragraphs.

[2]Laeufer (1992), however, argues that the differences in the strength of the effect are due to discrepancies in the materials and experimental design used in the various studies. Since this is not directly relevant to the present research, I refer the reader to Laeufer (1992) and references therein.

Table 2.2: Major consonantal allophones of Icelandic.[a]

	labial	**coronal**	**palatal**	**velar**	**glottal**
stops	p pʰ	t tʰ	c cʰ	k kʰ	ʔ
fricatives	f v	θ ð	ç ʝ	x ɣ	h
nasals	m m̥	n n̥	ɲ ɲ̊	ŋ ŋ̊	
laterals		l l̥			
trills		r r̥			

[a] Note that some of the segments in the table are debated in the literature as to whether they are phonemic or just allophones.

(1976) argues that the voiced portion of the consonant is perceived to be part of the previous vowel. In fact, Maddieson and Gandour (1976) and Durvasula and Luo (2012) argue that none of the suggested explanations is satisfactory. The majority of the studies make incorrect predictions for vowel length followed by aspirated consonants, as found in Hindi. In contrast, a promising line of enquiry emerges from the proposal by Chomsky and Halle (1968), who attributed the voicing effect to adjustments of the larynx when moving from a vowel to a consonant. In this work, I will take such an articulatory approach in defining a research hypothesis (see Section 2.4).

2.3 Icelandic

Icelandic [GLOTTO: `icel1247`] is spoken by the inhabitants of the Republic of Iceland (about 320,000 according to Árnason 2011), of which it constitutes the national language. Icelandic is a Germanic language and, together with Faroese [GLOTTO: `faro1244`], constitutes the Insular branch of the North Germanic group. Among the Nordic languages, Icelandic and Faroese are the most conservative (Harbert, 2006; König and Van der Auwera, 2013).

The modern phonological inventory of Icelandic has been subject to different analyses and it is still a matter of controversy (Thráisson, 1978; Jessen and Pétursson, 1998; Árnason, 2011). Table 2.2 reports the major consonantal allophones of Icelandic, as normally described in the literature (Árnason, 2011, p. 98). They do not necessarily represent the phonemic consonants of Icelandic and are rather surface segments.[3] The consonantal system of Icelandic includes stops, fricatives, nasals, laterals and trills. The

[3] On the quarrel about the phonemic status of pre-aspiration and voiceless sonorants, see Thráisson (1978); Jessen and Pétursson (1998); Berg (2001); Hansson (2003); Bombien (2006).

Table 2.3: Vocalic phonemes of Icelandic.

Monophthongs	**front**		**back**	
	plain	*rounded*	*plain*	*rounded*
high	i			u
high-mid	ɪ	ʏ		
low-mid	ɛ	œ		ɔ
low			a	
Diphthongs: /ai/, /ɛi/, /œi/, /ou/, /au/				

places of articulation encompass labial, coronal (alveolar), palatal, velar and glottal. Stops and fricatives have members in each place, nasals are labial, coronal, palatal and velar, while laterals and trills are coronal. For each combination of manner and place of articulation there are two segments, one voiceless and one (voiceless) aspirated or voiced depending on the manner. Length is distinctive in consonants, word-medially and word-finally. Of interest for the present study is that aspirated stops are pre-aspirated when they are geminates, while plain geminates are voiceless unaspirated, like the corresponding singletons: /$\mathrm{t^h}$akk$\mathrm{^h}$a/ 'key' [$\mathrm{t^h}$a$\mathrm{^h}$ka]; /$\mathrm{k^h}$akka/ 'barrel' [$\mathrm{k^h}$akka].

The vocalic phonemes of Icelandic are given in Table 2.3 (Árnason, 2011, 60). On the antero-posterior axis, two places of articulation are distinctive: front and back. On the vertical axis, vowels can be high, high-mid, low-mid and low. Vowels can be either unrounded or rounded, even if a true phonemic contrast is present only in front-mid vowels. Icelandic has the following diphthongs: the /i/-diphthongs /ai/, /ɛi/, /œi/, and the /u/-diphthongs /ou/, /au/. The vocalic system of this language does not exploit phonological length distinctively. Instead, vowels (both monophthongs and diphthongs) are predictably long when followed by two or more consonants, while they are short if one or no consonant follows. For example: *bað* /pað/ 'bath' [paːð], *kinn* /cɪnn/ 'cheek' [cɪnn], *baka* /pak$\mathrm{^h}$a/ 'to bake' [paːk$\mathrm{^h}$a] and *bagga* /pakka/ 'burden.ACC' [pakka]; *álnir* /aulnɪr/ 'wealth' [aulnɪr̥] 'wealth', *ál* /aul/ 'eel.ACC' [auːl̥], *álar* /aular/ 'eels' [auːlar̥].[4]

[4]The actual distribution of long and short vowels is, unsurprisingly, debated. For a review on the theoretical interpretations of the lengthening rule, see Booij (1986); Pind (1999), and Árnason (2011, pp. 160–173, 203–208).

2.3.1 Laryngeal contrasts in stops

As mentioned in Section 2.2, the contrastive system of Hindi consonants is built on the cross-cutting interaction between aspiration and voicing. Icelandic, on the other hand, contrasts only voiceless unaspirated with voiceless aspirated stops. Voicing is reported to be totally absent in Icelandic stops, and does not appear even as passive voicing in intervocalic position (Árnason, 2011). The actual phonetic realisation of the aspirated series though varies depending on the variety it is spoken and on the phonological context. In the so called "soft" variety (*linmæli*) of Icelandic (spoken everywhere in the island with the exception of the northern part), which is also the most widespread, aspiration is neutralised word-medially and word-finally, so that only non-aspirated stops can be found in those contexts. For example: /t^{h}ap^{h}a/ 'to lose' [t^{h}aːpa], /p^{h}at^{h}a/ 'to gesticulate' [p^{h}aːta] (Árnason, 2011, p. 104). In the "hard" variety (*harðmæli*), spoken in the northern area of the island, the same neutralisation pattern occurs, but the unmarked member in the neutralisation—the one that emerges phonetically—is the aspirated one. In this variety, stops are aspirated in every context, but they contrast with non-aspirated stops at the beginning of a word: /t^{h}ap^{h}a/ 'to lose' [t^{h}aːp^{h}a], /p^{h}at^{h}a/ 'to gesticulate' [p^{h}aːt^{h}a].

Moreover, aspirated geminate stops, both word-medially and word-finally, are realised as pre-aspirated stops. For example: /k^{h}ɔppɪ/ 'young seal' [k^{h}ɔppɪ], but /kɔppɪ/ 'chamber pot (dative)' [k^{h}ɔ^{h}pɪ].

2.3.2 Laryngeal contrasts in sonorants

The Lyon-Albuquerque Phonological Systems Database (Maddieson, 2012) reports that 51 out of 586 languages of the world use laryngeal settings contrastively in sonorants. The contrastive use of laryngeal distinctions in sonorants can thus be considered cross-linguistically rare (only 8.7% of languages in LAPSyD have this property). Unsurprisingly, sonorants produced with modal voicing are the most common type. If more than one phonation type is used in sonorants to accomplish phonological contrasts, at least one of those is always modal voicing. Icelandic is among those rare languages that show voicing contrasts in sonorants and it has voiced and voiceless nasals, laterals and trills. In the following paragraphs I will show how the voiceless sonorants of Icelandic can be though of as aspirated consonants. This interpretation makes it possible to group both the aspirated consonants and the voiceless sonorants under a single category "aspirated." Such move will prove to be useful for the development of the hypothesis in the present study.

Ladefoged and Maddieson (1996) discuss the possible combinations of phonation types and oral gestures in nasal consonants. Apart from modal voicing, three other types of phonations are found among nasals in the languages of the world: breathy voice, creaky voice and voicelessness. Breathy and creaky voice will not be dealt with here, since they are not immediately relevant for this study. Bhaskararao and Ladefoged (1991) state that voicelessness in nasals is implemented in two possible ways. The type known as the "Burmese type," which is also the most common, is constituted by an initial portion of voicelessness followed by a portion characterised by modal voicing (for example,bilabial [N̥͡m], and alveolar [N̥͡n]).[5] In the second type of voiceless nasals, the segment is voiceless through out its duration. However, when oral occlusion is released, the velum is still open, so that the nasal port is unobstructed. Since the velum is open, air can freely escape through the nasal cavity. The small burst produced by releasing the occlusion mimics the auditory features of an epenthetic stop (Bhaskararao and Ladefoged, 1991, p. 84). Such configuration could be represented as [N̥ᵖ] for the bilabial nasal and [N̥ᵗ] for the alveolar one.

Ladefoged and Maddieson (1996, p. 111) say that the nasal airflow of the voiceless nasals of Burmese is very high. Since air comes from the lungs and the oral cavity is closed by the tongue, the airflow rate is correlated with the amount of glottal width. A high nasal airflow seems thus to suggest that voiceless nasals are produced with abducted vocal folds. Due to this property, the researchers argue that voiceless nasals in general are perhaps better described as "aspirated." Kehrein (2002, p. 82), based on Ladefoged and Maddieson (1996), interprets voiceless nasals as phonologically aspirated.

As with nasals, laterals and trills can use phonation types contrastively, included voicelessness. Voiceless laterals can be realised as either approximants or fricatives (Ladefoged and Maddieson, 1996). These two ways of implementing voiceleness in laterals can be distinguished according to a variety of acoustic parameters. Given that in approximants the vocal tract is open while in fricatives it is more constricted, approximants are normally characterised by "a lower amplitude of noise, a greater tendency to anticipate the voicing of a following vowel, and a concentration of energy lower in the spectrum than voiceless fricative laterals" (Ladefoged and Maddieson, 1996, p. 198).

Trills are made up by a sequence of quickly repeated closures produced in the oral cavity. Voiced trills are characterised by the presence of voicing during the intervals between each closure. In voiceless trills, instead, voicing is absent during and between the closures that constitute the trill (Ladefoged and Maddieson, 1996, p. 236).

[5] Since place cues are not recoverable during the voiceless part of the nasal, the capital letter ⟨N⟩ is used to indicate placelessness. This solution is used by Silverman (2012).

Table 2.4: Possible phonation and gemination contrasts in Icelandic, ordered by context. These are, from left to right, word-initial, intervocalic, between a vowel and /l, r/, between a vowel and a stop, word-final.

	#_	V_V		V_(l/n)	V_C	_#	
stops	p	p		p		p	
			pp [a]				pp [a]
	p^h		^hp [a,b]	p^h			^hp [a,b]
sonorants	n	n			n	n̥	
	n̥				n̥		

[a] These units are phonological geminates. All other units are singletons.
[b] [^{h}p] is the phonetic realisation of the aspirated geminate /pp^h/.

Helgason (2002) notes, about Icelandic, how the phasing of the laryngeal gestures in pre-aspirated stops resembles the one in voiceless sonorants. In both cases, the peak of glottal spread is achieved before the stop closure. Moreover, the voice offset is simultaneous with the stop closure in plain geminates and voiced sonorants, while it occurs before the stop closure in both pre-aspirated geminates and voiceless sonorants. Such similarity in the relative timing of oral and laryngeal gestures strengthens the idea that pre-aspirated geminates and voiceless sonorants in Icelandic could be treated as members of a single class of aspirated or spread glottis consonants.

2.3.3 Phonotactics

Relevant to this study is the way phonation types are exploited to create contrast in stop and sonorant consonants in Icelandic. Even if the phonemic oppositions are ideally straightforward—aspirated consonants contrast with non-aspirated ones—the actual phonotactic distribution of contrastive consonantal segments is quite complex. There are restrictions on possible contrasts depending on the position within the word and the preceding or following segments. Except word-initially, both plain and aspirated stops can be geminated. Table 2.4 shows the distributions of contrasts in stops and sonorants in various contexts within the word. The labial stops and the alveolar nasals are used as an exemplification, respectively, of the distribution of stops and sonorants.

2.4 Research hypothesis

The studes by Maddieson and Gandour (1976) and Durvasula and Luo (2012) showed that syllable-final *post*-aspirated stops lengthened the preceding vowel. These results are compatible with two opposite predictions with regard to *pre*-aspirated stops: either aspiration consistently appears to lengthen the preceding vowel, independently of the relative timing with the stop closure; or pre-aspiration should show a pattern symmetric to that of post-aspiration, where pre-aspirated stops should be preceded by shorter vowels.

The hypothesis I propose in this study rests on the idea that the aspiration effect is the product of the relative alignment of the glottal spreading gesture and the oral gesture. Depending on how the spreading of the glottis is timed in relation to the tongue gestures in the oral cavity, modal voicing can be maintained for a certain amount of time or it is terminated earlier. An early timing of glottal spreading (like in pre-aspiration) would prevent voicing in the last portion of the preceding vocalic gesture, making the vowel shorter. On the contrary, if the spreading gesture is timed later relative to the oral gesture (like in post-aspiration), the voicing in the preceding vowel can be sustained, leading to a longer vowel. In a language like Hindi that contrasts non-aspirated with post-aspirated stops, a vowel followed by a post-aspirated stop would be relatively longer than a vowel followed by a non-aspirated stop since in the former case the spreading gesture that allows post-aspiration would allow longer voicing.

In Icelandic, the contexts where a contrast involving aspiration can be found post-vocalically is with geminates stops, RC clusters (/l/ or /r/ + stop) and NC clusters (nasal + stop) in word-medial and final position. As mentioned above, aspiration is realised as pre-aspiration in geminate stops and as (partial) voicelessness in sonorants. If the timing interpretation of the aspiration effect is correct, vowels before pre-aspirated geminates and voiceless sonorants should be shorter than vowels before plain geminates and voiced sonorants. Given that pre-aspirated stops and voiceless sonorants can be categorised under the umbrella term "aspirated consonants" (for the reasons above), the two separate statements can be merged into a single hypothesis:

(1) **Alternative hypothesis** (H_1)
Vowels followed by aspirated consonants are shorter than vowels followed by non-aspirated consonants.

The corresponding null hypothesis is:

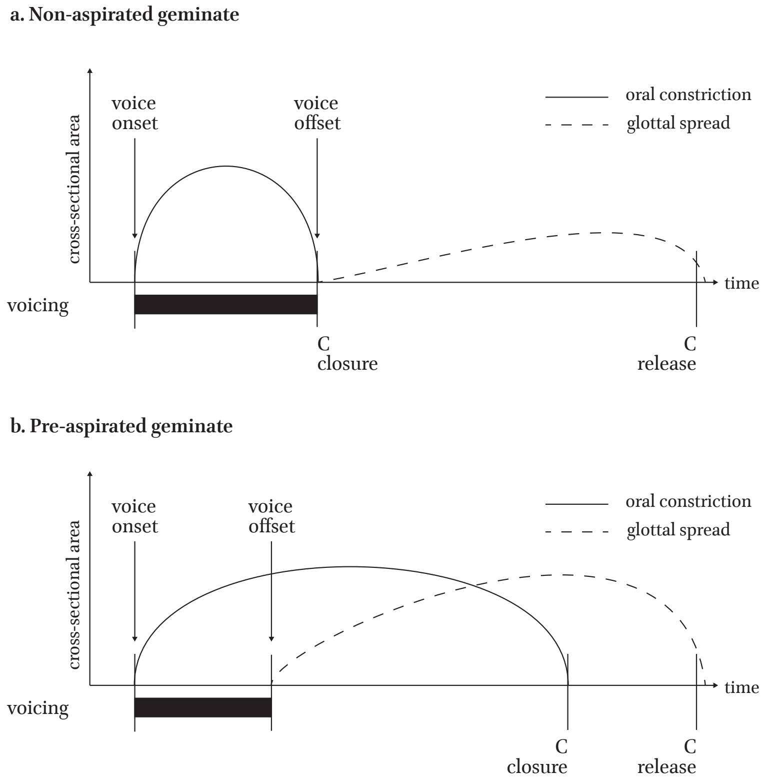


Figure 2.1: Visual representation of the research hypothesis. **a.** shows the articulatory events in words with non-aspirated geminates (like in *kagga* [k^hakka] 'barrel'), while **b.** for words with pre-aspirated geminates (*takka* [$t^h a^h$ka] 'key'). Both graphs are a zoom in of the -VCC- part of a word. The vertical axis represents the quotient of oral closure (solid line) and glottal abduction (dashed line). According to the research hypothesis of this study, voicing offset should be timed earlier in b. than in a. because of the earlier spreading of the vocal folds. Based on the design in Esposito (2002).

(2) **Null hypothesis** (H_0)
Vowels followed by aspirated consonants are of the same duration as vowels followed by unaspirated consonants.

The research hypothesis is illustrated schematically in Figure 2.1. The next chapters will describe the experimental design and the annotation scheme employed to test the research hypothesis (Chapter 3), followed by the results of the analysis (Chapter 4) and a general discussion of these (Chapter 5).

3 Methodology

In this chapter I will describe the methodology employed for this study. Section 3.1 contains information about the speakers that participated in the study. In Section 3.2, I detail the materials used in the task, which is outlined in Section 3.3. Section 3.4 discusses the annotation scheme, and the criteria employed in the annotation. Separate subsections deal with the labelling parameters for the detection of, respectively, word boundaries, voicing, glottal spread, nasality, laterality, rhoticity, and stop release. Section 3.5 defines the measurements that I extracted from the annotated files and describes the procedure used for data normalisation. Finally, Section 3.6 briefly mentions the features of the statistical analysis applied to the measurements and the software I used to implement it.

3.1 Participants

For this study, I recruited six Icelandic speakers who were living in York (UK) at the time the recordings were made. The methodologies of this research have gained the approval of the Ethics Committee and the subjects received an information sheet and they signed a consent form. Recruitment was done through University channels, the Icelandic Embassy in London and the York Anglo Scandinavian Society. All the participants were native speakers of Icelandic, above 18 years old and claimed to have normal hearing and speech abilities. The informants received compensation for their

Table 3.1: Information on participants.

id	sex	age	birthplace	languages	abroad
TT	F	24	Reykjavik	English, Danish, German	Yes
BRS	F	25	Höfn	Danish, English, Spanish	Yes
BTE	F	27	Reykjavik	English, Danish	Yes
JJ	F	46	Reykjavik (Kópavogur)	English, Danish	Yes
SHG	M	25	Selfoss	English	No
JR	M	66	Reykjavik (York)	English	Yes

time in the form of Amazon vouchers or chocolates for the value of £5.

The information on each participant is given in Table 3.1. The column labelled "birthplace" contains the city or town where the subjects were born; the eventual city or town in parenthesis is the place where they spent most of their life if different from their birthplace. The last column, "abroad", states if the subjects spent more that 6 months outside Iceland. The participant JR had to be excluded from the analysis since he misunderstood the task, while part of the participant SHG's task was lost due to technical fault in the recording equipment. All the speakers were from the Southern parts of Iceland so they are assumed to speak the "soft" variety of Icelandic (see Section 2.3).

3.2 Materials

The material used in the task consisted of a list of Icelandic words (the "target words") with the following forms: (C)VCC (monosyllabic) and (C)VCCV (disyllabic).[1] The list of target words is given in Table 3.3. I have chosen to use only real words to avoid the artificiality of constructed speech. The justification for such decision goes beyond the scope of this dissertation, but suffice it to say that under an exemplar model of speech production and perception (Johnson, 1997; Pierrehumbert, 2001; Bybee, 2002; Johnson, 2007) a word never experienced before would initially constitute an oddity in the cloud of exemplars and it would thus not be wise to include it in a study of phonetic nature. The target words were selected so as to control for as many of the following

[1] I will use the following conventions: V = any vowel, C = any consonant, KK = any geminate stop consonant, TT = any non-aspirated geminate, ^{h}T = any (pre-)aspirated geminate, S = any sonorant, N = any nasal, L = any lateral, R = any trill, -VCC = any monosyllabic word (either beginning with a vowel or a consonant), -VCCV = any bisyllabic word (either beginning with a vowel or a consonant). The terms "aspirated stop" will be a shortcut for "aspirated geminate stop." In case of ambiguities, the former label will be used.

Table 3.2: Number of tokens per class included in the analysis (discarded tokens not included).

	monosyllabic		disyllabic	
manner	non-asp	asp	non-asp	asp
stop	84	109	60	30
nasal	29	37	68	54
lateral	12	15	30	27
trill	-	-	15	15

aspects as possible: phonation, manner and place of articulation of consonants following the target vowel; height and frontness of the target vowel; phonation, manner and place of articulation of consonants preceding the target vowel; and height and frontness of the eventual word-final vowel.

Control over these parameters was prioritised according to the order in which they were presented here. Unfortunately, obtaining a well controlled word list proved to be extremely difficult and several compromises have been made. Respecting as many of the constraints as possible required the use of inflected forms. Thus, the final wordlist contained a total of 57 inflected Icelandic words (only real word forms were used). These were a mixture of nouns (25), verbs (22), adjectives (8) and adverbs (2).

The 57 words were equally divided in monosyllabic (28) and disyllabic (29) words. Of the monosyllabic words, 21 ended with a geminate stop (9 plain geminates and 11 pre-aspirated geminates); 5 with an /NC/ cluster (2 voiced and 3 voiceless nasals); 2 with an /lC/ cluster (one voiced, one voiceless). Of the disyllabic words, 14 had a word-medial geminate stop (8 plain and 6 pre-aspirated); 9 a /NC/ cluster (5 voiced and 4 voiceless); 4 an /lC/ cluster (2 voiced and 2 voiceless); and 2 had an /rC/ cluster (one voiced, one voiceless).

Table 3.3: List of target words.

word	**IPA**	**pos**	**gloss**
STOPS			
kokk	$\mathrm{k^ho^hk}$	noun	cook

word	**IPA**	**pos**	**gloss**
gogg	kokk	noun	beak
dökk	tœ^{h}k	adjective	dark
dögg	tœkk	noun	dew
kopp	k^{h}o^{h}p	noun	chamber pot
kubb	c^{h}ʏpp	noun	block of wood
vítt	vi^{h}t	adverb	far and wide
vídd	vitt	noun	width
þítt	θi^{h}t	verb	thaw
þíddi	θittɪ	verb	thaw
fætt	fai^{h}t	verb	feed
fæddi	faittɪ	verb	feed
ýtt	i^{h}t	verb	push
ydd	ɪtt	verb	sharpen
ótt	ou^{h}t	noun	point
odd	ott	adverb	fast
sets	sess	noun	sediment
sett	se^{h}t	verb	put
feits	feiss	adjective	fat
feitt	fei^{h}t	adjective	fat
vots	voss	adjective	wet
vott	vo^{h}t	adjective	wet
takka	t^{h}a^{h}ka	noun	key
kagga	k^{h}akka	noun	barrel
detta	te^{h}ta	verb	fall
gedda	cetta	noun	pike
NASALS			
kamp	k^{h}am̥p	noun	moustache
kamb	k^{h}amp	noun	comb
punt	p^{h}ʏn̥t	noun	decoration
pund	p^{h}ʏnt	noun	pound
kembt	c^{h}em̥t	verb	comb
kembdi	c^{h}emtɪ	verb	comb

word	**IPA**	**pos**	**gloss**
kampa	kʰam̥pa	noun	moustache
kamba	kʰampa	noun	comb
kempa	cʰem̥pa	noun	hero
kemba	cʰempa	verb	comb
punta	pʰʏn̥ta	noun	decoration
punda	pʰʏnta	noun	pound
vanta	van̥ta	verb	want
vanda	vanta	verb	do st carefully
fínn	fitn̥	adjective	smart
kinn	kʰɪnn	noun	cheek
LATERALS			
duld	tʏlt	adjective	complex
dult	tʏl̥t	adjective	reticent
gelta	cel̥ta	verb	bark
gelda	celta	verb	castrate
mjólka	mjoul̥ka	verb	milk
ólga	oulka	verb	foam
TRILLS			
orka	or̥ka	noun	energy
orga	orka	noun	scream

3.3 Procedure

The target words were embedded in the frame sentence *Segðu __ aftur*, 'Say __ again.' This sentence was chosen with the aid of one of the participants so as to control for naturalness, number of syllables and phonetic contexts preceding and following the target word, and phrase stress. The decision to use a single frame sentence for all the test words was justified by the wish to ensure that the metrical and intonational patterns were fixed. The participants were asked to read aloud the sentences with the target words shown on a computer screen. They were advised to speak as naturally as possible, while keeping the same volume and pace. They did not familiarise themselves with the word list before starting the task. The decision of not showing the words beforehand was made to reduce the speakers' control over their speech.

The task was presented through the software PyschoPy (Peirce, 2009), on a Apple MacBook Pro. Each sentences was shown three times consecutively and the order of appearance was randomised across subjects. The reading task was self-paced; the participant read a sentence shown on the screen and moved to the next sentence when ready by pressing the space bar. Four speakers were recorded in York in a meeting room at the travel agency they worked for, while one was recorded at the University of York. One speaker did not agree at coming to the University Campus, so he was recorded in his living room, at his house in York. The only subject who performed the task at the University of York was recorded in a sound-proof studio, using a Beyerdynamic OPUS 54 headset microphone (condenser, cardioid), plugged into a recording station. The software used for the recording was Adobe Audition, running on a Windows computer. The other speakers were recorded using the same headset microphone plugged into a Zoom H4n Handy Recorder. The audio files were encoded using the `.wav` format at a sampling rate of 44 kHz (16-bit). Even if the recording conditions differed between participants, the quality of the audio is comparable across files.

3.4 Annotation

The analysis of the audio files consisted of three phases: (1) conversion from stereo to mono, (2) annotation, and (3) extraction of the relevant measurements. I first converted the audio files from stereo to mono, without applying any sort of filtering. Since only one audio input was connected to the recorder, the two channels were actually a duplicate of one single channel, so that problems of differential phasing did not arise while converting the files. During the second phase, I annotated the files in PRAAT (Boersma and Weenink, 2015) using TextGrid files. The annotation files have four tiers. The tiers contain, respectively: (1) the graphemic transcriptions of the target words, (2) the voiced intervals within the relevant portion of the words, (3) the intervals within the words where laryngeal spread, nasality, laterality or rhoticity is present, and (4) the release of stops. Figure 3.1 shows an example of the TextGrid set-up.

3.4.1 Word boundaries

The first tier was segmented by target words. The left boundary of the word was considered to be the offset of voicing of the final vowel of *segðu*, which preceded the target word. The right boundary differed between consonant-final and vowel-final words. In consonant-final words, the right boundary coincided with the end of the friction following the burst of the release, as visible in the waveform and spectrogram. In vowel-final words, I used different criteria depending on the phonetic realisation. The right bound-

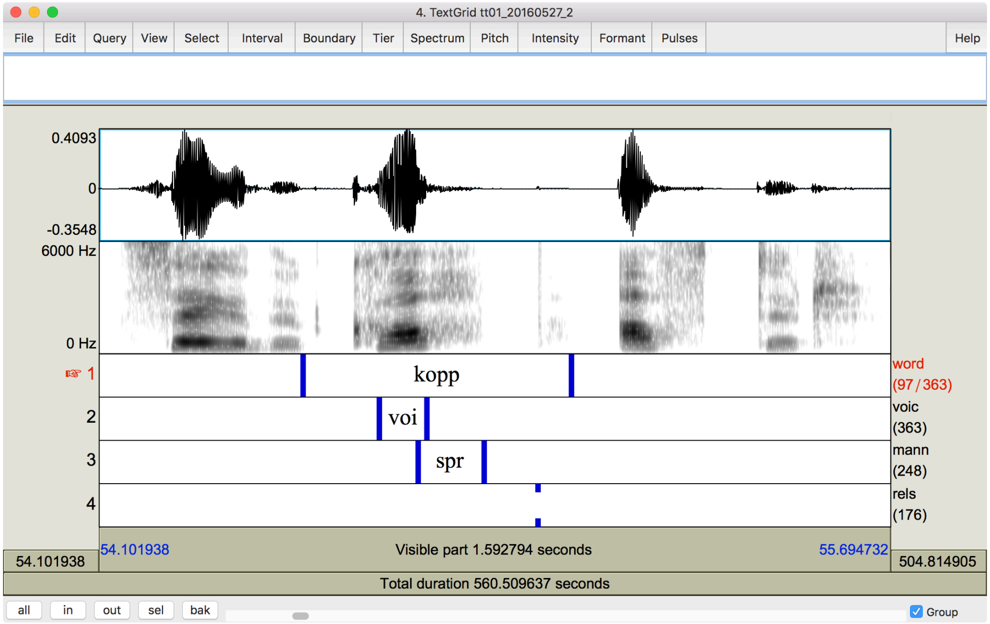


Figure 3.1: Screenshot of a PRAAT editor window showing the tier structure of the annotation files (PRAAT TextGrids). Tiers: 1. `word`, words; 2. `voic`, voicing; 3. `mann`, glottal spread and manner; 4. `rels`, release.

ary was placed at the offset of voicing of the word-final vowel if followed by a pause; if the word-final vowel differed from the following and there was no pause, I placed the boundary at the mid-point of the transition between the final vowel and the initial vowel of the following word (*aftur*); if a clear glottal stop separated the target word from the following, the boundary coincided with the onset of the glottal stop. In some cases, instead of a glottal stop, creaky voice was visible and the criterion of the transition mid-point was applied.

3.4.2 Voicing

The second tier was reserved for the interval in the word where vocal folds vibration (voicing) was active. The boundaries of the intervals in this tier were placed at the onset and offset of voicing around the target vowel. The onset of voicing was marked at the onset of periodicity of the waves in the waveform and/or at the beginning of the voice bar as visible on the spectrogram. In words starting with a voiceless consonant or with a vowel, the voicing onset coincided with the onset of voicing of the vowel. In words starting with one or more voiced continuant consonants, the voiced portion of those consonants was excluded from the interval and the left boundary was placed at

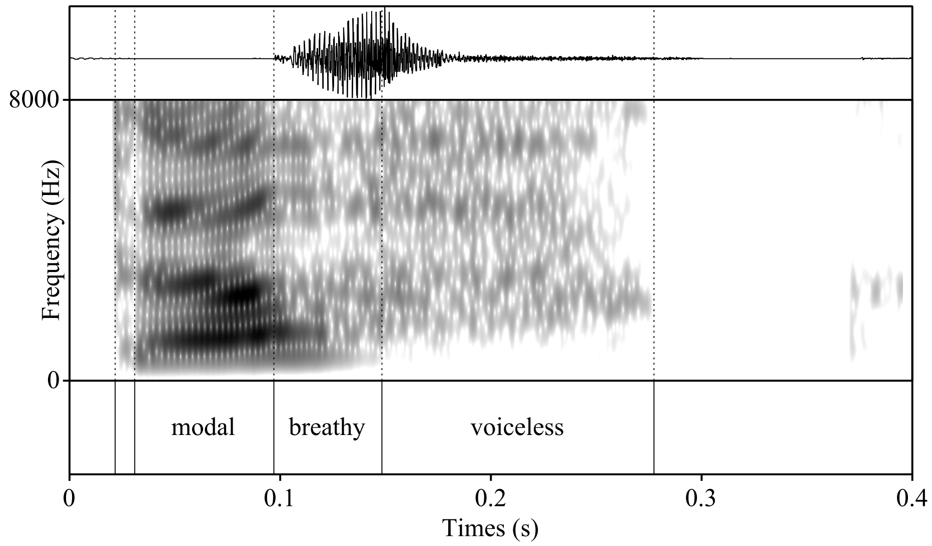


Figure 3.2: The word *dökk* /tœkk^{h}/ 'dark' [tœœ̤œ̥k] as an example of the annotation of glottal spread.

the beginning of the vowel following the word-initial consonants. The right boundary was in both cases the offset of voicing.

3.4.3 Glottal spread

A single tier was used to annotate glottal spread, nasal airflow, laterality and rhoticity. I opted for this solution since the intervals of each of the above classes, as defined below, were never overlapping. Of course, nasality, laterality and rhoticity never co-occur, and glottal spread can be inferred in these three classes by subtracting the voiced portion of the nasal, lateral or rhotic from the total duration of the consonant.

Marking the beginning of glottal spread proved to be particularly difficult. The common realisation of the combination "vowel + pre-aspiration" is structured as follows. The first portion of the vowel is accompanied by modal voicing: in the word *dökk*, shown in Figure 3.2, the vowel is initially realised as [œ]. Subsequently, the vocal folds start moving apart from each other in an abduction gesture while they still vibrate (breathy voice): [œ̤]. Finally, the vocal fold vibration stops while voiceless friction remains (at the glottis or at the oral cavity, depending on the place of articulation vowel): in this case, [œ̥].

As Khan (2012) and Nance and Stuart-Smith (2013) point out, breathy voice is expected to produce more round-shaped periodic waves. I took the onset of such more sinusoidal waves to coincide with glottal spread and I marked it as the left boundary of the spreading gesture. At times, however, the interpretation of the waveform was not straightforward. In these cases, I relied on the visual make-up of the spectrogram. According to Jones and Llamas (2006, cited in Nance and Stuart-Smith 2013, p. 134), breathy voice usually correlates with smeared off or totally absent higher formants. This is due to the presence of high-frequency noise produced by the increased amount of airflow coming from the abducted glottis. The right boundary was assumed to fall at the end of visible frication noise. See Figure 3.3 for an example.

3.4.4 Nasals, laterals and trills

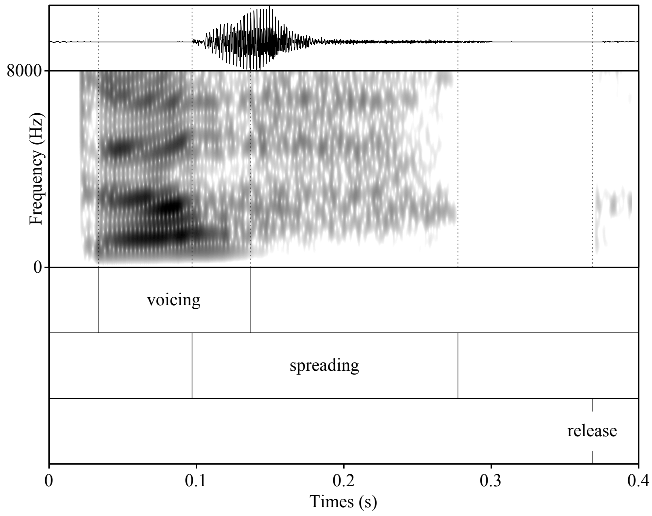
Following standard practice, I marked the beginning of nasality where a change in the shape of the waveform and in the amplitude of the spectrogram were visible. I applied the same principle to laterals and trills. I placed the right boundary of these intervals (nasal, lateral, trill) depending on the voicing of the segment. The voiceless nasal, lateral and trill consonants terminate with voiceless friction (nareal, lateral or central, respectively, see Figure 3.4a for an example of voiceless nasals). The end of friction in these consonants was used as the end of the interval. In the voiced counterparts of these, the end of voicing coincided with the right boundary (Figure 3.4b).

3.4.5 Stop release

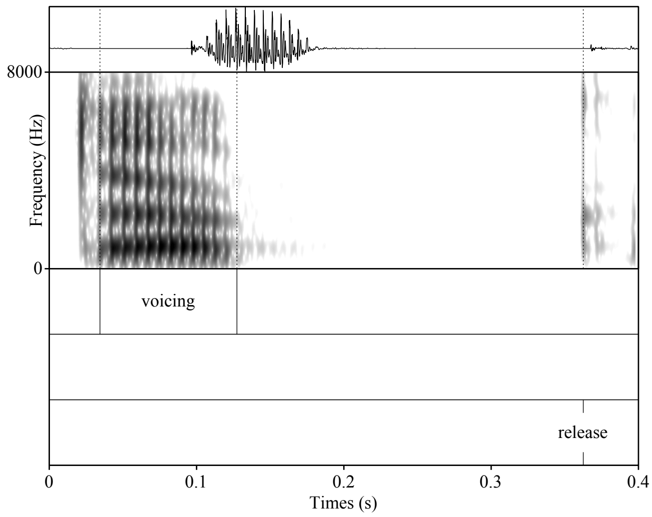
The last tier was dedicated to marking the consonant release of the stop following the target vowel (with or without an intervening sonorant). The release of the stop consonant was marked at the onset of the burst. This is normally visible on the waveform as one or more sudden peaks after the closure, or on the spectrogram as a short interval of low amplitude friction (Figure 3.5). If the burst was not identifiable from the waveform nor from the spectrogram, no release was marked.

3.5 Measurements

After the annotation was complete, I extracted the durational properties of the annotated intervals using an automated routine. The routine was performed in PRAAT with a script, specifically written for this study. The script with its documentation can be found in **??**. The output of the extraction procedure is a `.csv` file which contains the relevant

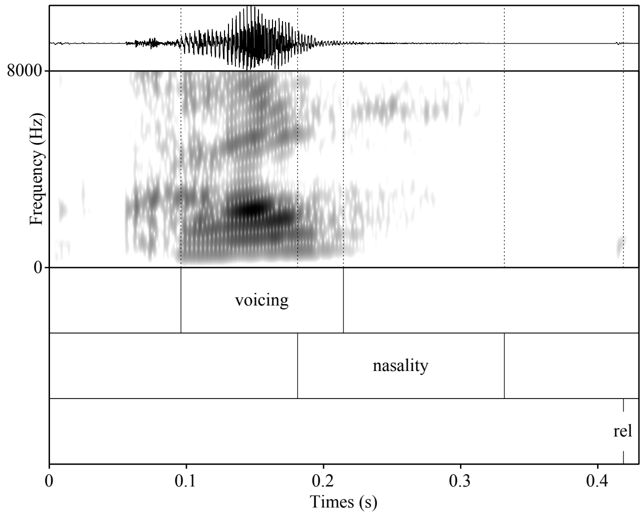


(a) /tøkk^h/ [tø^{h}k]

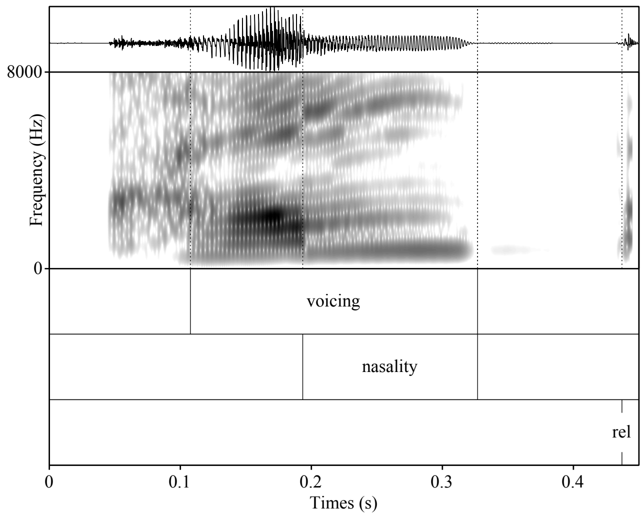


(b) /tøkk/ [tøkk]

Figure 3.3: Example of the annotation of glottal spread in the words *dökk* 'dark' and *dögg* 'dew'.



(a) /k^{h}amp^h/



(b) /k^{h}amp/

Figure 3.4: Example of the annotation of nasality in the words *kamp* ‘moustache’ and *kamb* ‘comb’.

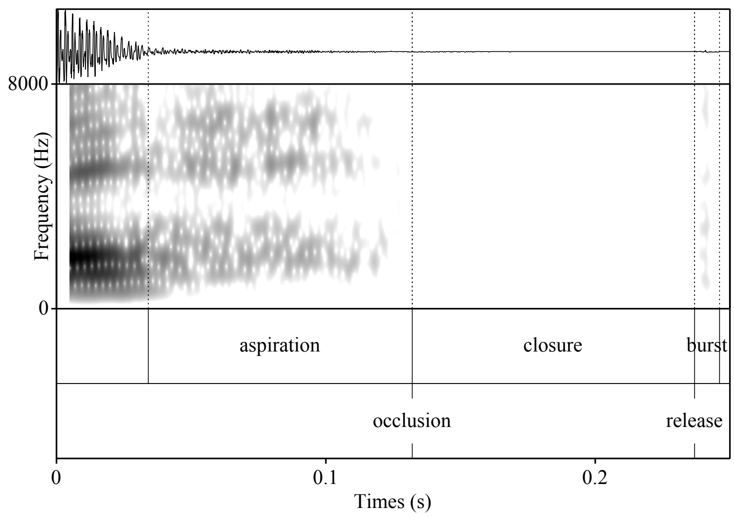


Figure 3.5: Release of /pp^h/ in the word *kopp* /k^{h}opp^h/ 'chamber pot' [k^{h}ɔ^{h}p].

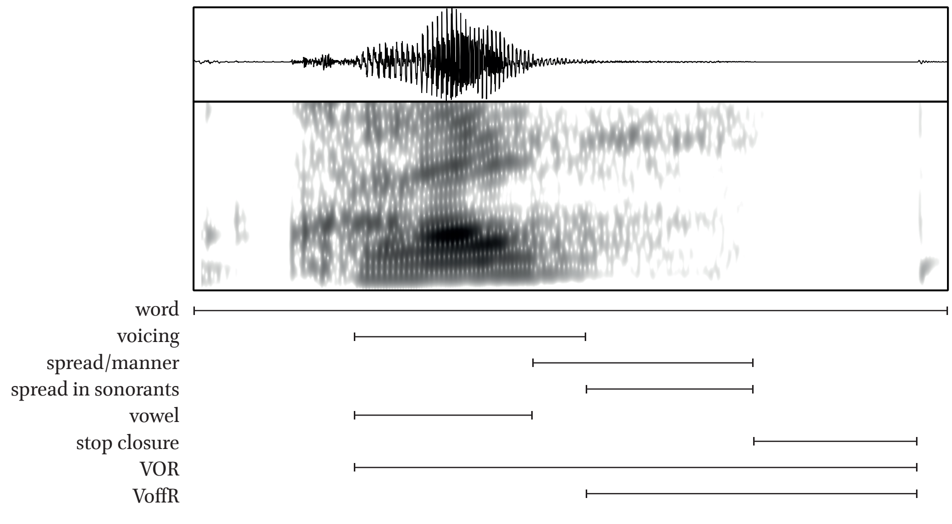


Figure 3.6: Schematic illustration of the measurements. The horizontal lines represent the interval for each of the measurements labelled on the left of the figure.

measurements. The following measures were taken in milliseconds. Figure 3.6 illustrates the measurements using the word *kamp* 'moustache' as an example.

- *Word duration*: the duration of the whole word as annotated on the word tier.
- *Duration of voicing*: the duration of the interval in the word containing vocal fold vibration (voicing tier).
- *Duration of glottal spread and consonant manner*: the duration of the portion characterised by glottal spreading, nasality, laterality or rhoticity (manner tier).
- *Duration of the vowel*. This was measured depending on the phonological form of the word. In words containing a pre-aspirated geminate or a sonorant (nasal, lateral or trill), as the duration of the interval between the onset of voicing and the onset of the interval on the manner tier. In words with non-aspirated geminates, the duration of the vowel is equal to the duration of voicing.
- *Stop closure duration*: the duration of the closure of the stop consonant, calculated as the duration of the interval between the offset of voicing, glottal spread or consonant manner, and the release.
- *CC duration*: the duration of geminate stops and "sonorant + stop" clusters. The interval between the onset of glottal spread or consonant manner and the release; in words with non-aspirated geminates, it is equal to the stop closure duration.
- *Voice Onset to Release* (VOR): the duration of interval between the onset of voicing and the release of the stop closure.
- *Voice Offset to Release* (VOffR): the duration of the interval between the offset of voicing and the release of the stop closure.
- *Duration of glottal abduction*. The duration of the interval calculated as the duration of spreading in stops and from the offset of voicing to the offset of manner in sonorants.

To control for differences in speech rate, I opted to normalise the measurements. Normalisation was achieved by dividing the relevant measure by the duration of the VOR. This operation results in a transformation from durations in milliseconds to ratios in percentages. I chose to use the VOR as the base for normalisation since this can be measured consistently across different word forms. Since the word list contained monosyllabic and disyllabic words, and both classes had either consonant-initial or vowel-initial words, the VOR is the only portion that constantly contains three segments (-VTT-, -V^{h}T-, -VRC-).[2]

3.6 Statistical analysis

I carried out the statistical analysis using the R programming language (R Core Team, 2015) in RStudio (RStudio Team, 2015). I performed independent sample two-tailed *t*-tests on parametric data and Mann-Whitney U-tests on non-parametric data. Normality was checked through Shapiro tests for normality. The R code used in the analysis is available at this address: `http://github.com/stefanocoretta/icelandic-preaspiration`.

[2]Remember that pre-aspirated stops are phonologically geminates: *detta* /tɛtt^{h}a/ 'to fall' [tɛ^{h}ta], *gedda* /cɛtta/ 'pike.ACC' [cɛtta].

4 Results

This chapter presents the results of the analysis on the durational data extracted from the recordings of the Icelandic informants. Section 4.1 and Section 4.2 deal with the duration of words and of the Voice Onset to Release (VOR) interval respectively. Section 4.3 treats the duration of vowels in the different contexts under study (geminate stops, nasals, laterals and trills). The last three sections discuss the duration of stop closure (Section 4.4), glottal spread (Section 4.5), and sonorant consonants (Section 4.6).

4.1 Word duration

The word list used in the reading task contains both mono- and disyllabic words. Both classes are further divided between consonant- and vowel-initial words. It is thus sensible to report some of the measures for each class separately. The mean duration of CVCC words is 441.36 msec (s.d. = 78.24), while for VCC is 338.71 msec (s.d. = 54.86). CVCCV words have a mean duration of 487.25 msec (s.d. = 59.94), while VCCV are on average 424.99 msec long (s.d. = 53.42).

The duration of vowels was not affected by the the number of syllables of the word. The boxplot in Figure 4.2 shows the duration of the vowel as a ratio of the Voice Onset to Release interval (VOR). According to a two-sample Mann-Whitney U-test, the mean duration of vowels in monosyllabic words (31%) does not differ significantly from the mean duration in disyllabic words (30%) [$U = 43349$, $p < 0.77$].

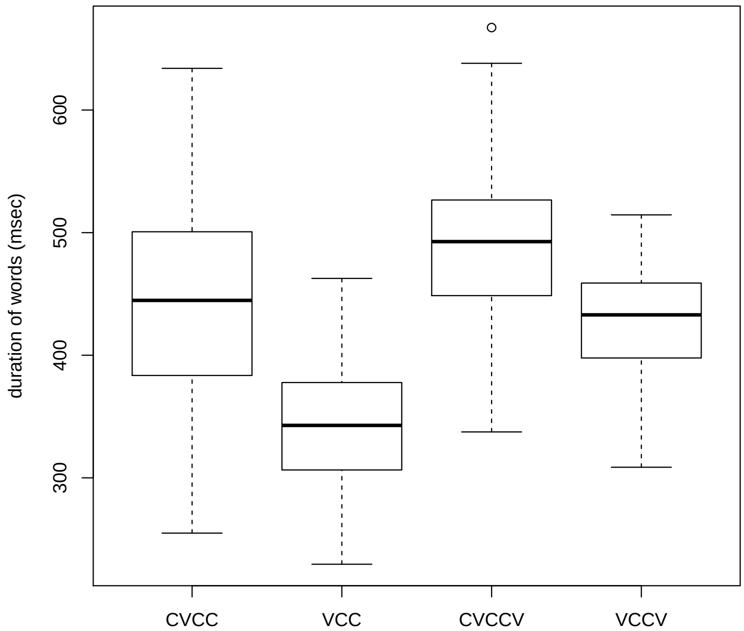


Figure 4.1: Duration of CVCC, VCC, CVCCV, and VCCV words in milliseconds.

4.2 Voice onset to release (VOR)

The Voice onset to Release (VOR) measures the interval between the voice onset of the critical syllable to the release of the next consonant. According to tests for the difference between means, the VOR in all classes of words, except for -VKK words (words ending with a geminate stop, either non-aspirated or pre-aspirated), did not have a significant difference in mean duration in both the non-aspirated and aspirated condition. Figure 4.3 and Figure 4.4 show box plots with the durations in milliseconds of the VOR in various conditions.

4.3 Vowel duration

The following sections report the results of the statistical analysis of the duration of vowels in the various word classes. I deal first with the duration of vowels before geminate stops, followed by vowels before nasals, laterals and trills. Figure 4.5 and Figure 4.6

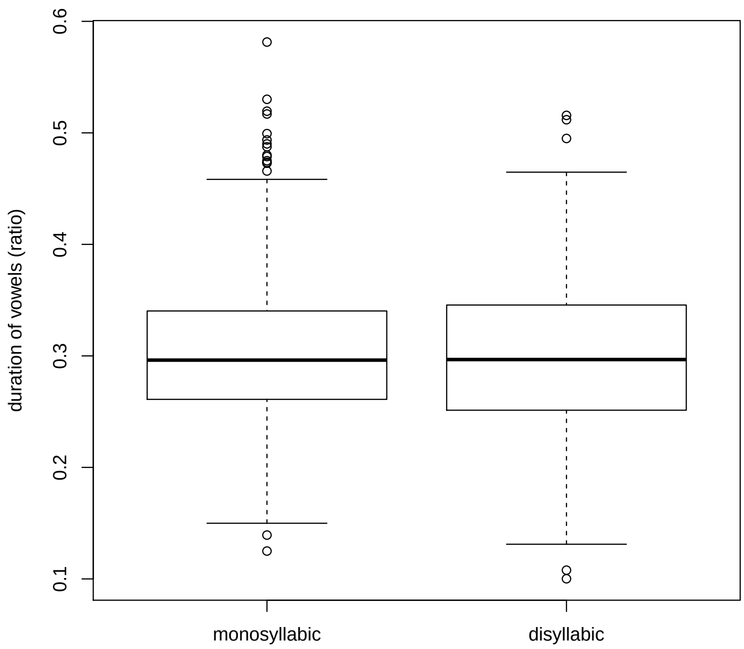


Figure 4.2: Duration of vowels in monosyllabic and disyllabic words, as ratios of the VOR.

show boxplots of the duration of vowels in various conditions as ratios of the VOR (in percentages).

4.3.1 Duration of vowels before geminate stops /KK/

The mean duration of vowels in milliseconds was 98 msec (s.d. = 19.1) in monosyllabic words ending in a non-aspirated geminate stop, while it was 84.47 msec (s.d. = 22.71) if followed by a pre-aspirated geminate stop. Figure 4.5 shows the ratio duration of the vowel in the two conditions. The difference in the mean of the vowel ratio between the non-aspirated (37%, s.d. = 0.07) and the aspirated class of words (28%, s.d. = 0.05) was significant [U = 7793, $p < 0.001$].

In disyllabic words, vowels are 86.97 (s.d. = 13.28) msec long if followed by a non-aspirated geminate, and 78.04 (s.d. = 20.53)msec long if followed by a pre-aspirated geminate. In ratio terms, the vowel makes up the 33.61% (s.d. = 0.05) of the VOR in word with plain geminates, and the 28.27% (s.d. = 20.53) in words with pre-aspirated

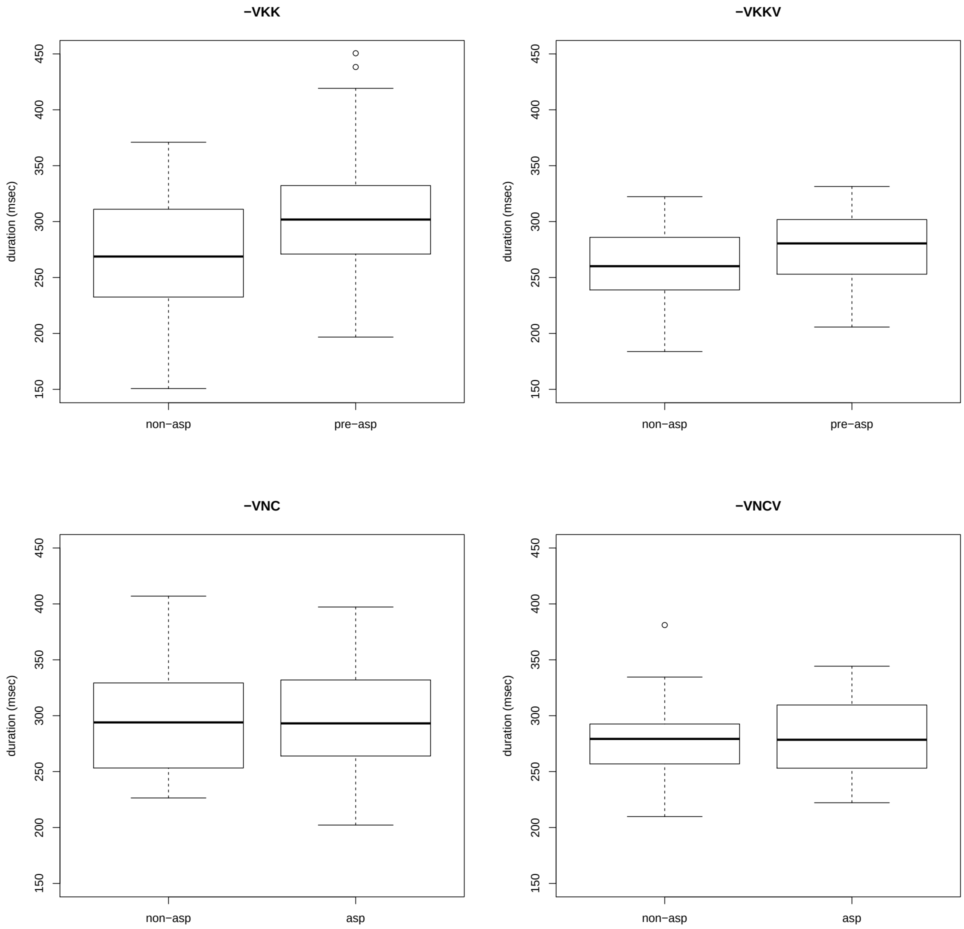


Figure 4.3: Duration in milliseconds of the Voice Onset to Release (VOR). The plots on the left hand side show the durations of the VOR in monosyllabic words. On the right hand side, the durations of the VOR in disyllabic words are displayed. From top to bottom, respectively, there are the plots for words with geminate stops and nasals (for laterals and trills, see Figure 4.4). In each plot, the duration of the VOR in the non-aspirated and aspirated condition is reported.

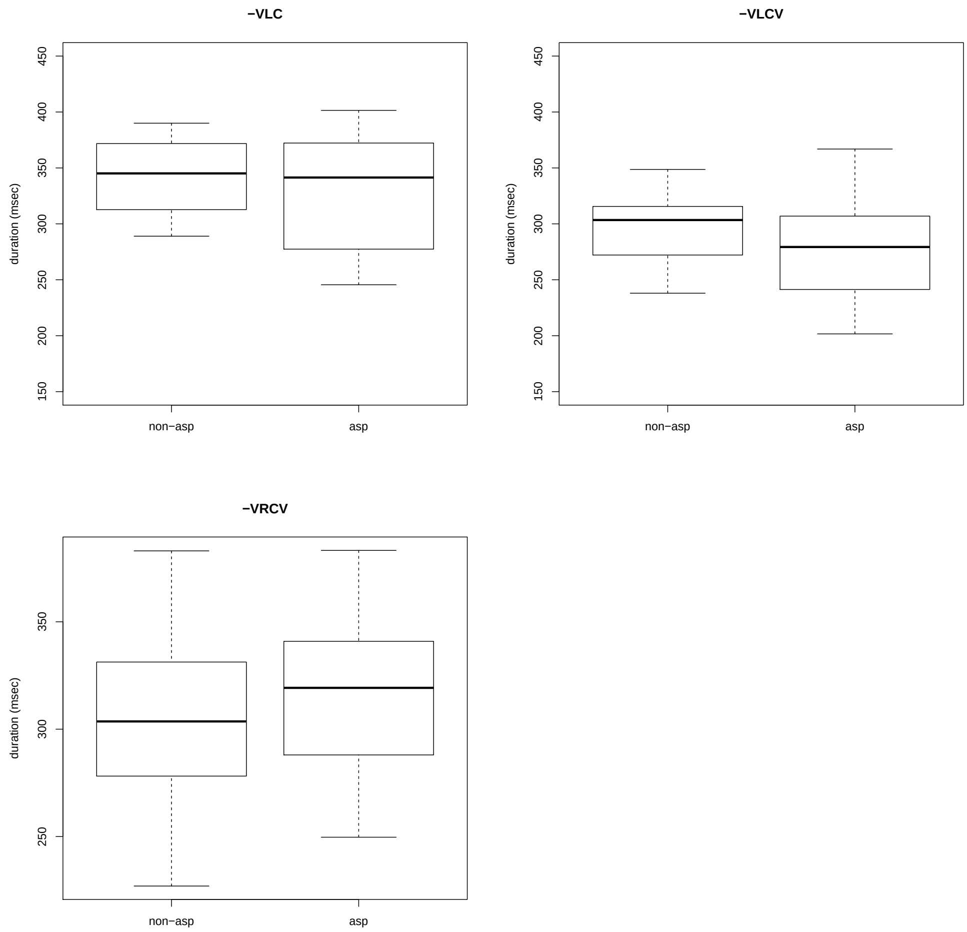


Figure 4.4: Duration in milliseconds of the Voice Onset to Release (VOR) of -VLC- (mono- and disyllabic) VRCV words (disyllabic words only). In each plot, the box on the left side reports the duration of the VOR in words with a voiceless lateral or trill is reported, while the box on the right the one in words with a voiced lateral or trill.

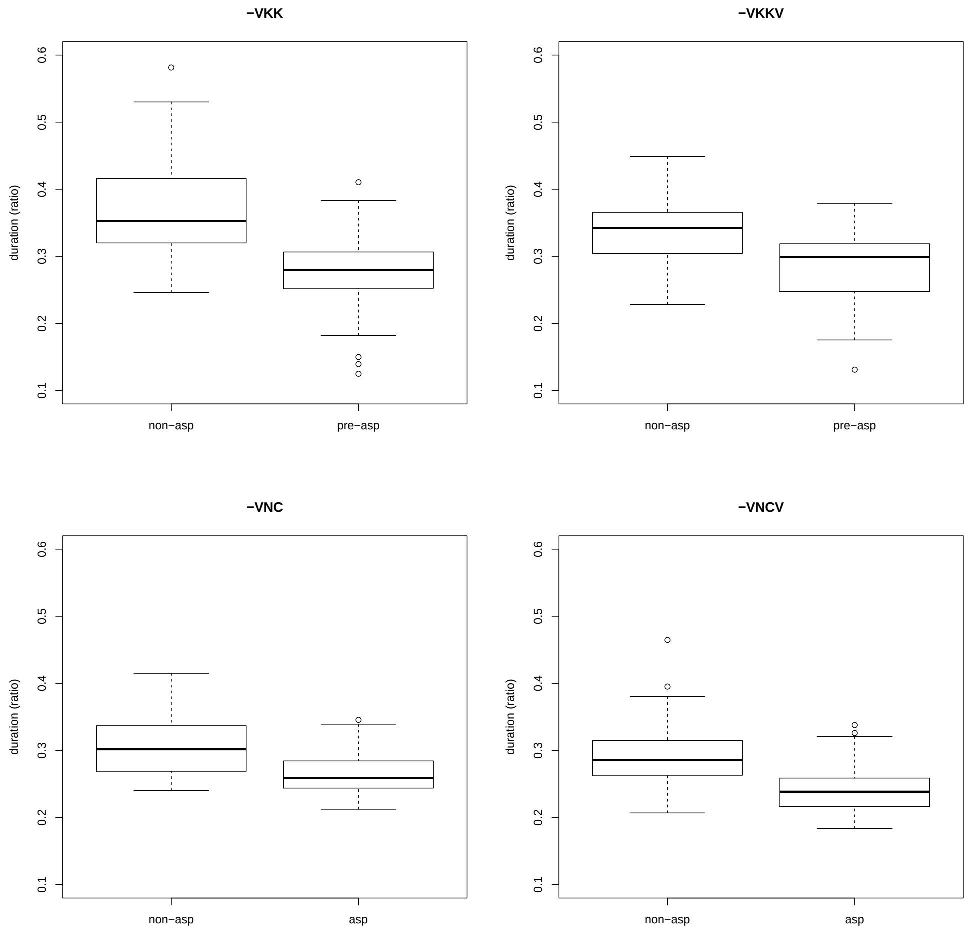


Figure 4.5: Duration of vowels as a ratio of the VOR (in percentages) depending on the presence *vs*. absence of pre-aspiration in the following consonant. The plots on the left hand side show the durations of vowels in monosyllabic words. On the right hand side, the durations in disyllabic words are displayed. From top to bottom, respectively, there are the plots for words with geminate stops and nasals (for laterals and tills, see Figure 4.6). In each plot, the duration of vowels in the non-aspirated and aspirated condition is reported.

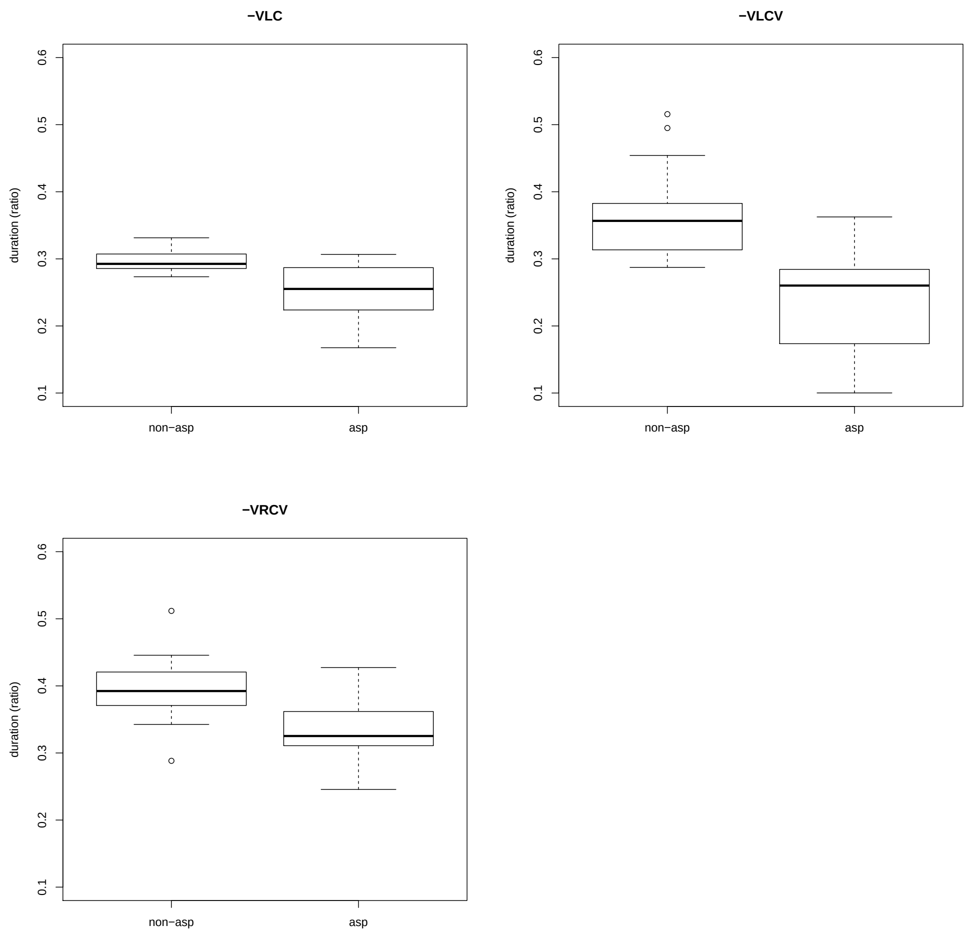


Figure 4.6: Duration of vowels as a ratio of the VOR in words with an /lC/ or /rC/ cluster, when followed by a voiced (left box of each plot) or a voiceless consonant (right box in each plot).

geminates. According to a independent two-sample *t*-test, the difference in ratio is significant [$t = 4.39$, $df = 50.46$, $p < 0.001$].

4.3.2 Duration of vowels before /NC/ clusters

Words of the form (C)VNC showed a mean vowel duration of 89.96 msec if the nasal is voiced and 77.73 msec if voiceless. The ratio duration was, respectively, 31% (s.d. = 0.05) and 26% (s.d. = 0.03). The difference in ratio duration was significant [$U = 836$, $p < 0.001$].

CVNCV words have a mean vowel duration of 81.72 msec (s.d. = 17.66) if the nasal is voiced and 67.48 msec (s.d. = 0.03) if it is voiceless. As percentages, vowels are 29% of the VOR if followed by a voiced nasal, and 24% when the nasal is voiceless. A *t*-test showed that this difference is significant [$t = 7.04$, $df = 119.99$, $p < 0.001$].

4.3.3 Duration of vowels before /lC/ clusters

In the case of monosyllabic words ending in a /lC/ cluster, the mean duration of the vowels were 102.05 (s.d. = 11.77) and 80.7 msec (s.d. = 13.02) when followed, respectively by a voiced and voiceless lateral. These correspond to the 30% and the 25% of the VOR. A *t*-test shows that the difference is significant [$t = 4.1$, $df = 20.43$, $p < 0.001$].

The duration of vowels in -VLCV words is on average 107.68 msec long (s.d. = 22.78) when the lateral is voiced and 66.66 msec (s.d. = 0.07) when it is voiceless. The vowel takes the 36% of the VOR if followed by a voiced lateral, and 23% when the lateral is voiceless. A *t*-test reveals the difference to be significant [$t = 7.47$, $df = 50.72$, $p < 0.001$].

4.3.4 Duration of vowels before /rC/ clusters

With words containing an /rC/ cluster, the mean duration of vowels is 121.07 msec long (s.d. = 0.05) when the rhotic is voiced and 104.78 msec (s.d. = 0.05) when it is voiceless. The vowel takes the 40% of the VOR if followed by a voiced rhotic, and 33% when the rhotic is voiceless. A *t*-test gave a significant result [$t = 3.4$, $df = 27.98$, $p < 0.001$].

4.4 Duration of stop closure

Stop closure was measured from the closure of the stop (with an intervening sonorant or not) to the release. The ratio duration of stop closure in words containing a /SC/ cluster

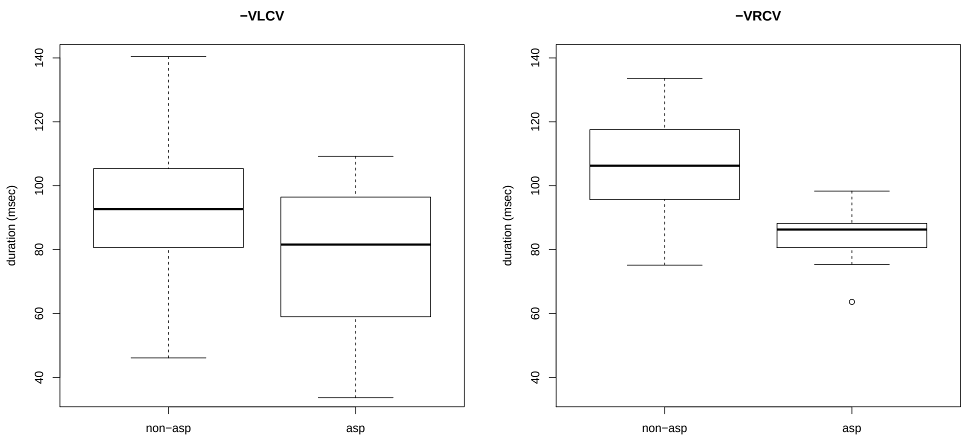


Figure 4.7: Duration of stop closure in -VLCV (on the left) and -VRCV words (on the right) in milliseconds. For both graph, the box on the left side shows the stop closure duration of non-aspirated liquids, while the box on the right the closure duration of aspirated liquids.

was not significantly different in the non-aspirated and aspirated condition, except in disyllabic words with an /lC/ [$t = 2.51$, $df = 54.7$, $p < 0.001$] and /rC/ cluster [$t = 5.62$, $df = 27.82$, $p < 0.001$]. Figure 4.7 shows the duration of the stop closure in disyllabic words with the form -VLCV and -VRCV. Unsurprisingly, the duration of stop closure in words with geminate stops was significantly different in the two conditions, since pre-aspirated stops have a later closure.

4.5 Duration of glottal spread

The duration of glottal spread was measured as the duration of vocal folds abduction in stops and as the duration of the voiceless portion in sonorants. According to a Kruskal-Wallis rank sum test, the duration of glottal spread in the stop, nasal, lateral and trill classes are different at the level of significance [Kruskal-Wallis $\chi^2 = 44.11$, $df = 3$, $p < 0.001$]. A post-hoc Games-Howell test for unequal variances showed that, while the duration of the spreading gesture did not differ between laterals, trills and stops, the duration of glottal abduction in nasals was significantly different from that one in laterals [$t = 3.79$, $df = 75.99$, $p = 0.002$] and stops [$t = 6.53$, $df = 137.2$, $p < 0.001$] (but not from the one in trills). Figure 4.8 shows the durations of glottal spread depending on the manner of the consonant.

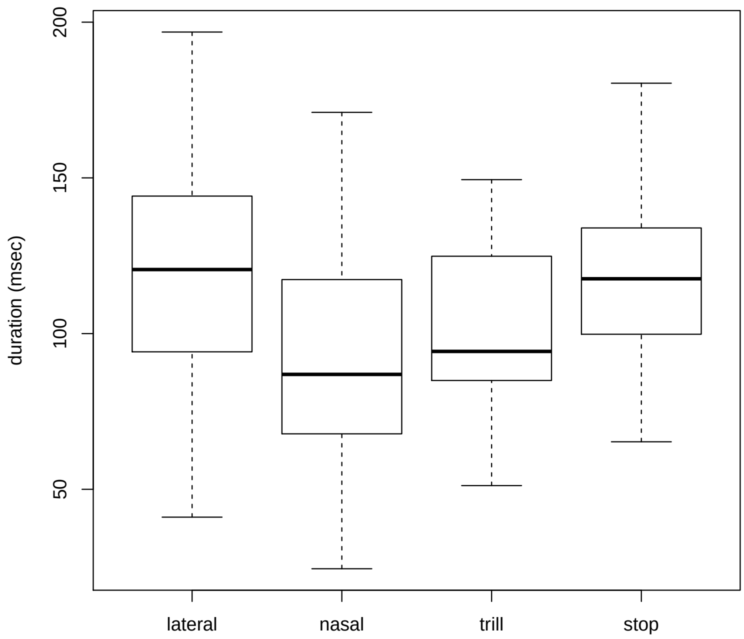


Figure 4.8: Duration (in milliseconds) of the glottal spread gesture in sonorants and stops.

Table 4.1: Mean duration and standard deviation of sonorants (in milliseconds).

	voiced		voiceless	
	mean	sd	mean	sd
nasals	120.41	27.19	135.09	32.03
laterals	106.13	26.78	142.78	31.24
trills	75.86	21	125.57	32.54

4.6 Duration of sonorant consonants

The duration of voiced sonorant consonants was significantly different from the duration of voiceless sonorants. On average, voiceless sonorants are 24 milliseconds longer than voiced sonorants [$t = 8.238$, $df = 297$, $p < 0.001$]. Table 4.1 gives the mean duration and standard deviation of voiced and voiceless sonorants.

5 Discussion

The research hypothesis of this study is based on the idea that the aspiration effect could be the product of the relative timing of the glottal spread gesture in relation to the tongue gestures in the oral cavity. An early timing of glottal spread would shorten the vowel, while a later timing would lengthen it. In particular, the hypothesis states that, in Icelandic, vowels followed by aspirated consonants (pre-aspirated geminates and voiceless sonorants) should be shorter than vowels followed by plain consonants. As shown in the result chapter (Chapter 4), I found a significant difference between the duration of vowels in words with aspirated consonants and their duration in words with non-aspirated consonants. This finding seems to initially support the timing hypothesis. The following sections will discuss the results further on the light of the hypothesis.

5.1 Duration of vowels before geminates and duration of geminates

We have seen that the duration of vowels as a ratio of the VOR is smaller when they are followed by a pre-aspirated geminates than when they are followed by a plain geminate. It is worth noting that the VOR of the two classes is different, as said in Section 4.2. However, the fact that the vowel is shorter in words with pre-aspirated geminates cannot be attributed to differences in the over-all duration of the interval between the onset of voicing and the release of the stop. This is because the VOR is *longer* in words with

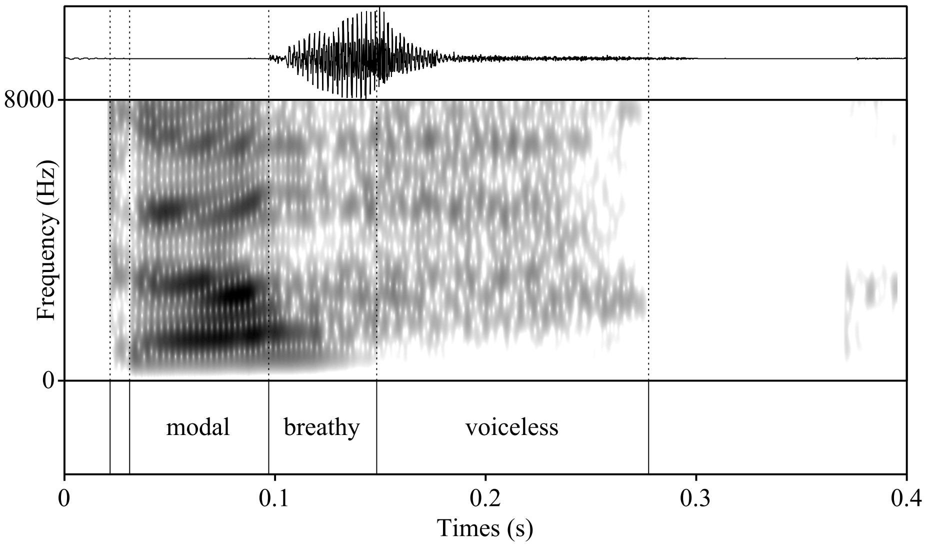


Figure 5.1: The word *dökk* /tœkkʰ/ ‘dark’ [tœœ̤œ̥k] as an example of the how phonation is coordinated in words with pre-aspirated geminates. Modal voice during the vowel is followed by breathy phonation (voicing plus glottal spread). Voicing eventually stops, leaving voiceless aspiration.

aspiration while the ratio duration of the vowel is *smaller*. Other things being equal, we would expect vowels—i.e. the interval between voice onset and the onset of glottal abduction—in words with pre-aspiration to take the same ratio of the VOR as the vowels in words with plain geminates. Instead, laryngeal spreading starts earlier in words with pre-aspirated stops, while voicing is still active even after glottal spread is initiated (see Figure 5.1). Instead, vowels are the 28% of the VOR if followed by pre-aspirated geminates and the 36% if followed by non-aspirated stops.

Even if we look at absolute durations, vowels followed by aspirated stops are shorter, as Figure 5.3 shows. However, the hypothesis rested on the idea that voicing would cease earlier in vowels followed by pre-aspirated stops, resulting in a shorter vowel. What I found rather points to a longer voicing portion in words with pre-aspirated stops (Figure 5.4). The duration of voicing is, in fact, greater in words with pre-aspirated geminates than in words with plain geminates [Mann-Whitney $U = 5160$, $p < 0.001$]. Figure 5.2 schematically shows this finding. This is not surprising, since glottal spread is not incompatible with voicing and could actually allow voicing to be sustained longer.

A possible explanation for the longer duration of voicing is that, simply put, the absence of a closure at the VC transition allows for voicing to continue. In words with

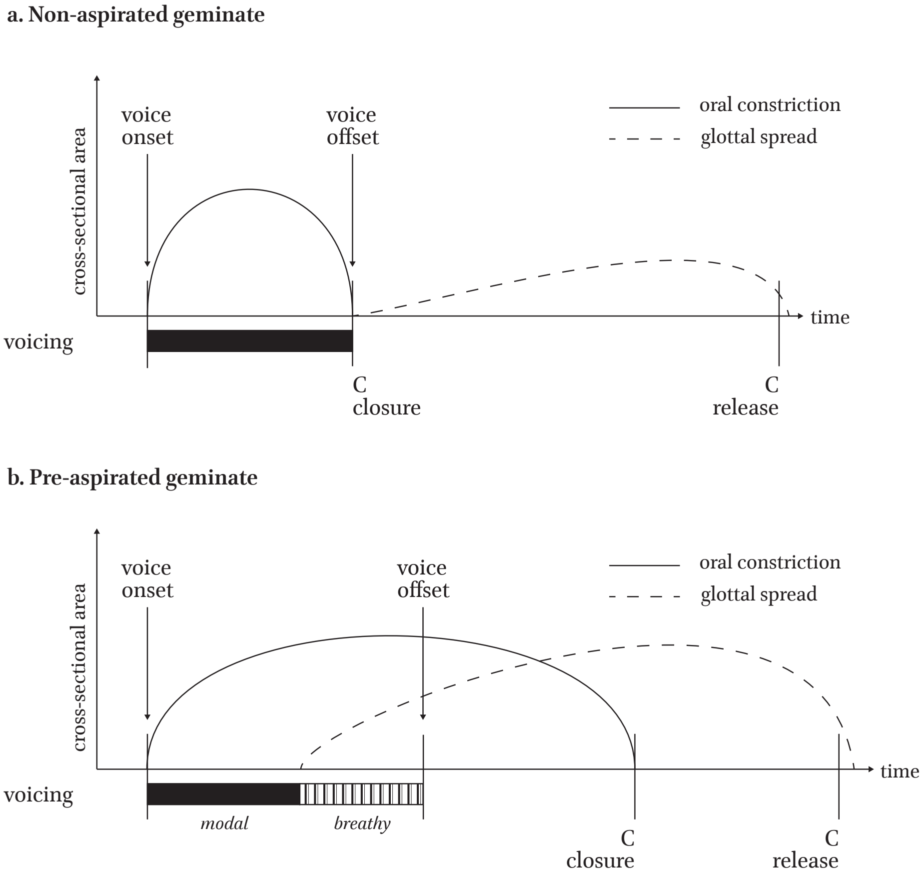


Figure 5.2: Schematic representation of the phasing of glottal spread and oral gestures in word with geminate stops. a. shows the articulatory events in words with non-aspirated geminates (like in *kagga* [k^hakka] 'barrel'), while b. for words with pre-aspirated geminates (*takka* [$t^h a^h$ka] 'key'). Both graphs are zoomed in the -VCC- part of the words. The vertical axis represents the quotient of oral closure (solid line) and glottal abduction (dashed line). Based on the design in Esposito (2002).

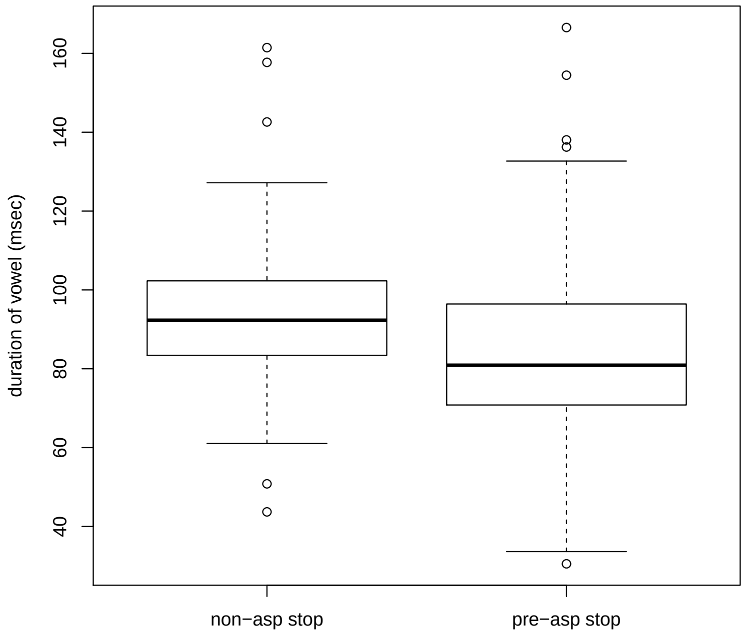


Figure 5.3: Duration of vowels in milliseconds before non-aspirated and pre-aspirated geminate stops.

non-aspirated geminates, the closure of the consonant—which is realised immediately after the vowel—quickly produces an increase in oral pressure, until the point is reached when voicing can no longer be sustained. In pre-aspirated geminates, instead, since the stop closure is executed later, there is no pressure build-up and vocal folds vibration can be maintained. At the same time, though, the variance of the duration of voicing in the two classes is not the same. As it can be seen from Figure 5.4, the duration of the voicing portion in words with pre-aspirated geminates has greater variance. Indeed, in some of the tokens from the pre-aspirated condition, the duration of voicing is as short—or in a few cases even shorter—than the duration in non-aspirated tokens.

Ladefoged and Maddieson (1996, pp. 70–71) stated that the glottal width in Icelandic pre-aspirated stops is wider than in non-aspirated stops, but they argued that the timing of glottal spread is the same. Glottal spread, according to their description, starts at the beginning of the geminate and ends at its release, independently of the presence of pre-aspiration. This remark does not fit with the results found in this study.

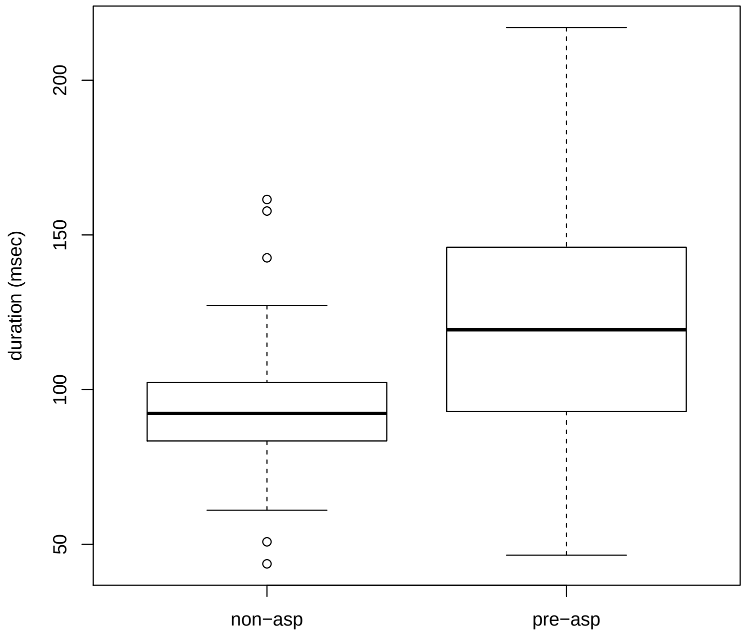


Figure 5.4: Duration in milliseconds of voicing in mono- and disyllabic words containing geminate stops, depending on presence or absence of glottal spread in the geminate. The duration of voicing in words with pre-aspirated geminates is, on average, longer than the duration in words with non-aspirated stops, and it also has greater variance.

Although I could not measure glottal width in a direct way, if the description given in Ladefoged and Maddieson (1996) is correct, glottal width in non-aspirated geminates can be inferred from the duration of the interval between the offset of voicing of the vowel preceding the geminate and the release of the geminate. In the data available, the duration of the pre-aspirated geminates was on average longer (212.17 msec) than the duration of the non-aspirated counterparts (172.17 msec). The difference was significant as per an unpaired-samples *t*-test [t = -8.9048, df = 279.04, p < 0.001]. Figure 5.5 shows the duration of pre-aspirated *vs.* non-aspirated geminates.

These facts do not fit well with the idea that voicing should cease earlier due to glottal spread. However, the assumption that glottal spread is timed earlier in pre-aspirated geminates than in non-aspirated geminates relative to the onset of voicing was correct (at least for Icelandic). Given than glottal spread starts earlier, modal voicing is indeed shorter in words with pre-aspirated stops. It can then be argued that the relative earlier

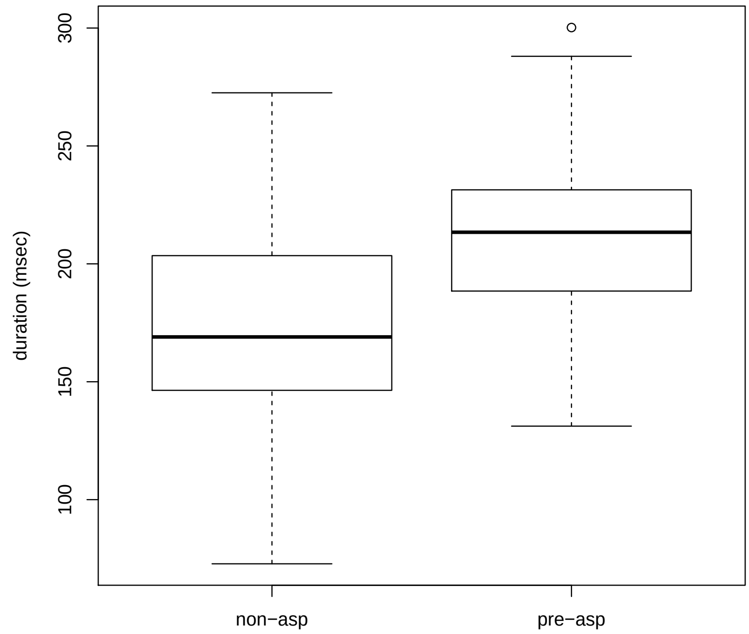


Figure 5.5

timing of glottal abduction reduces the vocalic portion where modal voicing is maintained. This kind of alignment results in a shorter vowel, if we consider the vowel to be the modal voiced interval of the vocalic gesture. Moreover, when breathy voicing is visible from the spectrogram, it should be assumed that the spreading gesture has already reached the critical point in which the abduction is enough to produce breathiness. This would imply that the gesture might have started even earlier, strengthening the central idea proposed in this research. In conclusion, we could say that, even if some adjustments are required, the Icelandic data seem to indicate that the timing hypothesis is the right direction for further exploration.

5.2 Duration of vowels followed by sonorants

The class of sonorant consonants (nasals, laterals and trills) seem to pattern together for what concerns durational properties. As discussed above, all sonorants showed to produce a difference in the ratio duration of voicing depending on whether they were aspirated or not. Differently from geminate stops, the VOR in words containing sonorant consonants was constant in both the aspirated and non-aspirated classes, except in disyllabic words containing laterals. The duration of the stop closure after the sonorant was the same in both the aspirated and the non-aspirated condition, except with laterals and rhotics in disyllabic words. The stop consonants in words with /NC/ and /N̥C/ clusters had identical closure durations, thus the difference in vowel duration cannot be attributed to closure properties. Moreover, the presence versus absence of aspiration did not affect the duration of stop closure.

On the contrary, the ratio duration of sonorant consonants (nasals, laterals and trills) varied between the aspirated and non-aspirated conditions. The aspirated sonorants are longer than the non-aspirated ones. This characteristic can be attributed to perceptual factors. An aspirated sonorant carries less cues for manner and place of articulation, given that most of it is produced with voiceless noise. One possible mechanism to enhance manner and place cues and ensure that they are salient enough to be audible is to lengthen the total duration of the sonorant consonant. This solution would allow more time for sustaining enough modal voicing—which carries better cues than voiceless noise—during the first portion of the sonorant.

In the case of sonorants, it is reasonable to say that the possible cause of shorter vowels in words with aspirated sonorants is the longer duration of the former. While with stops the VOR is longer with pre-aspirated geminates, with sonorants its duration is the same with aspirated and non-aspirated consonants. If the speaker intention is to maintain the same VOR while warranting a longer duration of the sonorant, this can be done at detriment of the vowel. The vowel preceding the aspirated sonorant is thus compressed and it becomes relatively shorter than a vowel followed by a voiced sonorant. The timing hypothesis presented in this study, thus, seems not to be straightforwardly compatible with the data found in words containing sonorants, even if the effects predicted by it were found.

5.3 Duration of sonorant consonants

As reported in Section 4.6, voiceless sonorant consonants were significantly longer than voiceless sonorants. This is not surprising, since voiceless sonorants are less salient from a perceptual point of view. Silverman (1997) argues that, in terms of perceptual optimality, the phasing of non modal voicing in sonorants should involve what he calls truncation of the contrastive laryngeal gesture in respect to the oral ones. This solution involves the sequencing of non-modal phonation before modal phonation. According to this principle, an optimal phasing of voicelessness in nasals is achieved when the glottal abduction gesture is terminated before the oral occlusion is released. In the case the abduction gesture is sequenced before modal voicing, the first portion of the nasal consonant is voiceless (with nareal friction noise), then modal voicing follows. If, on the contrary, glottal spread follows modal voicing, the last portion of the nasal is characterised by breathy voice.

With laterals, instead, there are two possible optimal phasings of glottal spreading. In the first type, the abduction of the vocal folds and the lateral tongue gesture are simultaneous. A voiceless lateral produced in this way normally results in a voiceless lateral fricative more than an approximant. In a second type of phasing, the laryngeal abduction is truncated in a similar way as in nasals. Silverman (1997) ascribes such difference in the ways nasals and laterals achieve an optimal phasing of non-modal phonations to a combination of two factors: (1) all laterals are phonetically similar and (2) lateral consonants are easily distinguishable from non-laterals. The first factor explains why laryngeal distinctions in laterals are cross-linguistically rare, while the second accounts for the higher variability of realizations of laterals compared to that of nasals.

Moreover, these results are similar to those in Jessen and Pétursson (1998), Bombien (2006) and Silverman (2012). Bombien (2006), in fact, reports that voiceless sonorants in Icelandic have the following acoustic properties: (i) voiceless sonorants are longer in duration than their voiced counterparts; (ii) they have greater H1-H2 (first and second harmonic) difference than the voiced ones (which indicates breathy phonation in the former); (iii) they show greater zero crossing rate values (which indicates higher frication compared to the voiced sonorants).[1]

5.4 Phonological aspects

This study implicitly assumed that the ultimate origin of the aspiration effect is to be sought in physiological and aerodynamic constraints. Such constraints could be related to the ways the timing of laryngeal spread is phased in relation to other gestures, as hinted by the results of this research. However, we cannot exclude the possibility that the speaker is actively varying articulatory gestures either to reduce articulatory effort, to enhance contrast, or both. Kingston and Diehl (1994, p. 423) point out that "speakers and listeners employ extensive and subtle phonetic knowledge."

[1] Frication has been proved to create higher frequencies compared to non-fricated sounds and hence it involves higher rate of zero crossing (Weigelt et al., 1990).

In classical models of phonological competence, the articulatory output of the speech stream can be predicted on phonetic grounds from the form of the phonological representation. This kind of conceptualisation gives the phonetic constraints predictive power. Knowing which phonological representation underlies a linguistic unit allows us to anticipate the phonetic realisation of that unit. Kingston and Diehl, however, argue that the phonetic constraints normally called for in the literature as the cause of most phonological behaviours do not have predictive power Rather, they say they limit the possible phonetic outcomes of a phonological abstract unit. This approach leaves space for the idea that speakers can operate an implicit control over their productions. The authors call “phonetic knowledge” the knowledge underlying such ability.

There are two types of phonetic knowledge. The speaker-oriented knowledge deals with the possibilities speakers have to minimise the articulatory effort necessary to produce a particular sequence of gestures (and sounds). The other type of knowledge, the listener-oriented knowledge, is instead related to the ways speakers can maximise the distinctive cues of contrastive sounds. Such two kinds of phonetic knowledge work together to create an optimal solution between minimum articulatory effort and maximum contrast. Borrowing a term from linguistic typology, they can be thought of as “competing motivation”, opposing forces that shape the phonological systems of the languages of the world (Haiman, 1983; Croft, 2002).

We have seen that pre-aspirated stops in Icelandic are preceded by shorter vowels, if we consider the vowel to be the modally voiced vocalic gesture preceding the aspiration. The magnitude of the difference in duration between vowels followed by pre-aspirated or plain consonants was around 30–40 milliseconds, enough for the auditory system to perceive it. It is reasonable, thus, to assume that such difference is deliberately used by Icelandic speakers to enhance the contrast between words containing a plain geminate and words with a pre-aspirated geminate. Even if the original cause of the aspiration effect caused by stops in present day Icelandic could have been related to the phasing of the laryngeal spreading gesture, the synchronic situation we can observe now could instead well be the product of the controlled use of such differences on the part of the speaker.

In the case of sonorants, even if the same shortening effect is seen in vowels followed by aspirated sonorants, we cannot claim that the effect is caused by how laryngeal gestures are timed, as explained above. Instead, the aspiration effect could be attributed here to the fact that voiceless sonorants are generally longer than voiced sonorants so as to enhance cuing. To maximise even more the difference between voiced and voiceless sonorants, speakers could decrease the duration of the vowel to magnify the longer duration of the sonorant.

5.5 Challenges and further studies

This section discusses the challenges encountered in this study that generated limitations. In those cases where I applied a specific solution to overcome such limitations, I describe how this was carried out. The section concludes with some ideas for further studies on the relation between vowel duration and aspiration from an articulatory point of view.

Some of the limitations of this study concerned the participants recruited for the reading task. The first difficulty was finding a sensible number of Icelandic speakers in York. Due to a restriction of time and resources, I could not recruit people from outside York. This significantly restrained the number of speakers and increased the possibilities that idiosyncrasies could creep into the data. Moreover, the age range of the participants was quite large (24-66 years old). The average, though, was biased towards the younger age: 4 out of 6 were within the 24-27 range. This made conducting age related analyses impossible. The same argument can be done for sex: 4 out of 6 speakers were female. There was heterogeneity also in the foreign languages they were confident with. While Danish was quite common among the youngest speakers, two of them spoke a third language as well (German and Spanish).

Another potential issue with the participants is that all of them, except one, lived in the UK for more than 6 months. It is reasonable to assume that this had a significant impact on their productions of Icelandic. Even if the informants allegedly spoke Icelandic on a regular basis, with their family, friends and colleagues, the chances that English influenced Icelandic (and—although not of a concern for the present study—vice versa) are quite high. The research conducted by Sancier and Fowler (1997) studied how strong the influence was from one language to the other in a Brazilian Portuguese and English bilingual speaker who frequently travelled between Brazil and the United States of America. The study showed that the speaker produced a shorter voice onset time in both Portuguese and English after spending more than 6 months in Brazil than when she was in the United States. Thus, it is worth stressing that the results presented here should be considered preliminary.

A further challenge was finding words that would satisfy all of the parameters. The choice to use real words came at the cost of making the experiment design less balanced. As mentioned in Section 3.2, building the word list and controlling for all the criteria while avoiding nonce-words was impossible. Such gaps in the lexicon of Icelandic show that the functional load of the aspiration contrast is somewhat constrained. Even if from a more restricted contextual point of view, aspiration contrasts can be found everywhere in the lexicon, true minimal pairs were harder to find.

Some design and technical problems arose during the reading task. Since the participants did not have the chance to familiarise themselves with the words before starting the task, a few speakers read the first token of some words with uncertainty. In some cases, they said the wrong word altogether. One of the speakers (JR) did not recognise most of the inflected words which were not in the standard citation form and decided to read them as if they were English words. Of course, I had to exclude all of these tokens from the analysis.

The sound equipment and the recording conditions varied between subjects. Although the quality of the audio was comparable across recordings, I expect to be minor differences in the way echoes and background noises affected them. This might have perturbed some speech stretches, in principle rendering visual annotation less reliable. However, as described in Section 3.4, the criteria used during annotation have been devised so as to be consistently applicable. Such solution substantially reduced possible errors due to the presence of echoes and noises.

The annotation scheme presented a few further challenges. In particular, the visual inspection of the spectrogram was, at times, not unequivocal. Especially with glottal abduction, I have encountered some difficult tokens where the shape of the waveform did not clearly indicate the onset of breathiness. In such cases I had to rely on a change in the look of the spectrogram. In particular, the results concerning the duration of the spreading gestures are to be taken with caution. In fact, it is possible that the difference found between stops and sonorants (where the duration of glottal abduction was higher in the former) could be due to the way the measurements was taken. With sonorants, identifying the beginning of glottal spread by inspecting the waveform or the spectrogram is not straightforward. Relying on the voiceless portion of the sonorant could potentially leave out the initial part of the spreading gesture. If one includes such hypothetical part to the duration of spreading, we can expect that the total duration would not be much different in sonorants and stops.

A second aspect of the annotation procedure that presented some difficulties was marking the left and right word boundaries. Since the frame sentence was kept the same in every trial, words starting and/or ending in a vowel posed problems in identifying the word boundaries. In such cases, it would not be safe to say that the word duration measurements are unequivocal and reliable. Moreover, a recurrent problem with studies of phonetic duration is that speech tempo is not stable and it is difficult to control. This is why I decided to use ratio durations to overcome the shortcomings of varying tempo. Using the VOR instead of the word duration as the base for normalisation guaranteed that errors in the marking of word boundaries would not affect normalisation.

Some of the challenges just discussed can be overcome in future studies if articulatory data are collected together with acoustic data. Laryngeal measures can be taken using photoglottography or a laryngograph. While the first tool is invasive, the second is not and its portability make its use simple and ideal even in fieldwork conditions. A laryngograph measures the impedance of the electricity running through the neck. The amplitude of the impedance is known to be correlated with glottal abduction. This kind of data would give a better picture of the actions of the glottis and it would allow to retrieve the start of glottal spread with better accuracy. An ultrasound scanner would provide an account of the gestures of the tongue. Knowing about tongue movements could make it easier to test for hypotheses that rely on articulatory factors.

Combined with electropalatography, ultrasound scanning can also improve the accuracy of closure detection. As mentioned above, at times closure was not clearly visible from the spectrogram, especially in cases when voicing is maintained after it. Tracking the movements of the tongue through ultrasound scanning and palatography can improve the temporal detection of the oral occlusion. Finally, measuring the oral and the nasal airflows could allow better measurements related to nasals. The boundary of nasal consonants are normally difficult to discern on the spectrogram. Voiceless nasals are even more difficult to see, since the friction created at the nostrils is usually very low in amplitude. Being able to detect nasal airflow allows to clearly define at which time point the velar port has been opened and when it has been closed during speech.

6 Conclusions

This dissertation dealt with the phenomenon known as the "aspiration effect," according to which vowels followed by aspirated consonants are longer than vowels followed by plain consonants (Maddieson and Gandour, 1976; Maddieson, 1976). Such relation between the presence of aspiration and lengthening in vowels is paralleled by the so called "voicing effect." The voicing effect consists in the shortening of vowels when they are followed by voiceless stops. Even if several explanations have been proposed in the literature for the voicing effect, no one has been given for the aspiration effect. In fact, none of the explanations for the voicing effect is compatible with the aspiration effect, since they usually predict the opposite of what has been empirically found. Moreover, the studies about the aspiration effect only dealt with post-aspiration.

In this study, I proposed that the aspiration effect has its origin in the way the laryngeal spreading gesture (that causes aspiration) is timed relative to oral gestures. I called this the "timing hypothesis." To verify the validity of this hypothesis I extended the focus of enquiry to pre-aspiration in Icelandic. In fact, the timing hypothesis predicts that vowels in Icelandic should be shorter when followed by a pre-aspirated consonant. As discussed in Chapter 4, the data collected from five speakers of Icelandic confirmed that vowels followed by pre-aspirated geminates or voiceless sonorants are shorter than vowels followed by plain geminates or voiced sonorants. Such finding seems to confirm the timing hypothesis.

In Chapter 5, I revised the central hypothesis of this thesis on the light of the present findings. Even though the prediction of shorter vowels before aspirated consonants is supported by the data, the underlying idea that voicing should cease earlier when aspirated consonants follow could not be confirmed. On the contrary, the duration of voicing in words with pre-aspirated stops was significantly longer than its duration in words with plain geminates. This indicates that pre-aspiration in the collected data was normally realised phonetically as a portion of breathy voicing followed by voiceless aspiration, before a closure is created in the oral cavity.

In the case of sonorants, the first portion of the sonorant was voiced, followed by voiceless friction in the oral cavity (in laterals and trills) or at the nostrils (in nasals). In sonorants voicing does not cease before the oral gesture of the sonorant is achieved. However, I showed that voiceless sonorants are longer than their voiced counterparts, due to perceptual enhancement (cf **??**). I thus argued that the shorter duration of vowels followed by voiceless sonorants could be attributed to perceptual reasons. Decreasing the duration of vowels when followed by voiceless sonorants makes the difference in duration of the former even more salient, if compared with the durations in words with voiced sonorants (that are shorter than the voiceless ones).

Although this research focussed more on the phonetics of the aspiration effect, in **??** I briefly discussed of its phonological aspects. In that section, I argued that, even if the effects of aspiration on vowel duration can have a physiological cause, it is possible that the speakers of Icelandic "actively" exploited and inflated those effects to enhance the phonological contrast.

To sum up, the results of this study indicated that vowels in Icelandic are shorter when followed by pre-aspirated geminates and voiceless sonorants, while they are longer when followed by plain geminates and voiced sonorants. The data on the duration of glottal spread in geminate stops supported the idea that the duration of vowels could be affected by the fact that glottal abduction is phased earlier in words with pre-aspirated stops. Even if the details of the timing hypothesis as presented in Section 2.4 need to be reworked, such hypothesis seems to be worth exploring further.

Bibliography

Árnason, K. (2011). *The phonology of Icelandic and Faroese*. Oxford: Oxford University Press.

Belasco, S. (1953). The influence of force of articulation of consonants on vowel duration. *The Journal of the Acoustical Society of America* 25(5), 1015–1016.

Berg, T. (2001). An experimental study of syllabification in Icelandic. *Nordic Journal of Linguistics* 24(1), 71–105.

Bhaskararao, P. and Ladefoged, P. (1991). Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(02), 80–88.

Blevins, J. (2004). A reconsideration of yokuts vowels. *International Journal of American Linguistics* 70(1), 33–51.

Blevins, J. (2009). Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics* 48(1), 264–273.

Boersma, P. and Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.08. Available at: `http://www.praat.org/` [Accessed 2013-08-11].

Bombien, L. (2006). Voicing alterations in icelandic sonorants–a photoglottographic and acoustic analysis. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* 37, 63–82.

Booij, G. E. (1986). Icelandic vowel lengthening and prosodic phonology. In H. Bennis and W. U. S. Van Lessen Kloeke, (Eds.). *Linguistics in the Netherlands*. Dordrecht: Foris, pp. 9–18.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24(02), 215–221.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22(3), 129–159.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. New York, Evanston, and London: Harper & Row.

Croft, W. (2002). *Typology and universals*. Cambridge: Cambridge University Press.

Durvasula, K. and Luo, Q. (2012). Voicing, aspiration, and vowel duration in Hindi. *Proceedings of Meetings on Acoustics* 18, 1–10.

Esposito, A. (2002). On vowel height and consonantal voicing effects: Data from italian. *Phonetica* 59(4), 197–231.

Fomel, S. and Claerbout, J. (2009). Guest editors' introduction: Reproducible research. *Computing in Science and Engineering* 11(1), 5–7.

Haiman, J. (1983). Iconic and economic motivation. *Language* 59, 781—819.

Halle, M. and Stevens, K. N. (2002). A note on laryngeal features. In M. Halle, (Ed.). *From Memory to Speech and Back: Papers on Phonetics and Phonology, 1954-2002*. Berlin: Mouton De Gruyter, pp. 45–61.

Hammarström, H., Forkel, R., Haspelmath, M. and Bank, S. (2016). Glottolog 2.7, Jena: Max Planck Institute for the Science of Human History. Available at: `http://glottolog.org` [Accessed 2016-08-30].

Hansson, G. Ó. (2003). Laryngeal licensing and laryngeal neutralization in Faroese and Icelandic. *Nordic Journal of Linguistics* 26(01), 45–79.

Harbert, W. (2006). *The Germanic Languages*. Cambridge: Cambridge University Press.

Helgason, P. (2002). *Preaspiration in the Nordic languages: synchronic and diachronic aspects*, PhD thesis, Institutionen för lingvistik.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1), 105–113.

Hussein, L. (1994). *Voicing-dependent vowel duration in Standard Arabic and its acquisition by adult American students*, PhD thesis, The Ohio State University.

Javkin, H. R. (1976). The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1, 78–92.

Jessen, M. and Pétursson, M. (1998). Voiceless nasal phonemes in Icelandic. *Journal of the International Phonetic Association* 28(1-2), 43–53.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson and J. W. Mullenix, (Eds.). *Talker variability in speech processing*. San Diego, California: Academic Press, pp. 145–165.

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. *Experimental approaches to phonology* pp. 25–40.

Jones, M. and Llamas, C. (2006). Voice source variation in vowel–consonant sequences in Middlesbrough English, Presented at British Association of Academic Phoneticians Colloquium.

Kehrein, W. (2002). *Phonological representation and phonetic phasing: Affricates and laryngeals*. Walter de Gruyter.

Khan, S. u. D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40(6), 780–795.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language* pp. 419–454.

Kluender, K. R., Diehl, R. L. and Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16, 153–169.

Knuth, D. E. (1984). Literate programming. *The Computer Journal* 27(2), 97–111.

König, E. and Van der Auwera, J. (2013). *The Germanic languages*. London and New York: Routledge.

Ladefoged, P. (1973). The features of the larynx. *Journal of Phonetics* 1(1), 73–83.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics* 20(4), 411–440.

Lampp, C. and Reklis, H. (2004). Effects of coda voicing and aspiration on Hindi vowels. *The Journal of the Acoustical Society of America* 115(5), 2540–2540.

Maddieson, I. (1976). Further studies on vowel length before aspirated consonants, Paper presented at the 91st meeting of the Acoustical Society of America.

Maddieson, I. (2012). Lyon-Albuquerque Phonological Systems Database (LAPSyD). `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/index.php`.

Maddieson, I. and Gandour, J. (1976). Vowel length before aspirated consonants, *UCLA Working papers in Phonetics*, Vol. 31, pp. 46–52.

Maxwell, M. (2013). A system for archivable grammar documentation, In G. Rehm, C. Mahlow and M. Piotrowski, (Eds.). *Systems and Frameworks for Computational Morphology. Third International Workshop*, Springer, pp. 72–91.

Maxwell, M. and Amith, J. D. (2005). Language documentation: the Nahuatl grammar. In A. Gelbukh, (Ed.). *Computational Linguistics and Intelligent Text Processing*. Berlin Heidelberg: Springer-Verlag, pp. 474–485.

Nance, C. and Stuart-Smith, J. (2013). Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. *Journal of the International Phonetic Association* 43(02), 129–152.

Ohala, M. (1983). *Aspects of Hindi phonology*. Motilal Banarsidass Publisher.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2(10).

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics* 10(3), 405–408.

Pierrehumbert, J. B. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. L. Bybee and P. J. Hopper, (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam Philadelphia: John Benjamins Publishing Company, pp. 137–157.

Pind, J. (1999). Speech segment durations and quantity in icelandic. *The Journal of the Acoustical Society of America* 106(2), 1045–1053.

R Core Team (2015). R: A language and environment for statistical computing. Available at: `https://www.R-project.org` [Accessed 11/04/2016].

RStudio Team (2015). Rstudio: Integrated Development Environment for R. Available at: `http://www.rstudio.com/` [Accessed 11/04/2016].

Sancier, M. L. and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25(4), 421–436.

Silverman, D. (1997). Phasing and recoverability. *UCLA Occasional Papers in Liguistics*, Vol. 15 of *Dissertasion Series 1*.

Silverman, D. (2012). Voiceless nasals in auditory phonology, *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, Vol. 22, pp. 364–374.

Sóskuthy, M. (2013). *Phonetic biases and systemic effects in the actuation of sound change*, PhD thesis, University of Edinburgh.

Thráisson, H. (1978). On the phonology of Icelandic prespiration. *Nordic Journal of Linguistics* 1(01), 3–54.

Weigel, W. F. (2002). The Yokuts canon: A case study in the interaction of theory and description, *Annual meeting of the Linguistics Society of America. San Francisco*.

Weigelt, L. F., Sadoff, S. J. and Miller, J. D. (1990). Plosive/fricative distinction: The voiceless case. *The Journal of the Acoustical Society of America* 87(6), 2729–2737.