

Vowel duration and aspiration effects in Icelandic

Stefano Coretta

August 28, 2016

Contents

1	Literature review	1
1.1	Phonation types and states of the glottis	1
1.2	The effect of aspiration on vowel duration	2
1.3	Icelandic	4
1.3.1	Laryngeal contrasts in stops	6
1.3.2	Laryngeal contrasts in sonorants	6
1.4	Research hypothesis	8
2	Methodology	10
2.1	Participants	10
2.2	Materials	11
2.3	Procedure	13
2.4	Annotation	14
2.4.1	Word boundaries	15
2.4.2	Voicing	16
2.4.3	Glottal spreading	16
2.4.4	Nasals, laterals and trills	17
2.4.5	Stop release	17
2.5	Measurements	17
2.6	Statistical analysis	20
3	Results	21
3.1	Word duration	21
3.2	Voice onset to release (VOR)	22
3.3	Vowel duration	23
3.3.1	Geminate stops	23
3.3.2	Nasals	23

3.3.3	Laterals	23
3.3.4	Rhotics	25
3.4	Duration of stop closure	25
3.5	Duration of glottal spread	27
3.6	Duration of sonorant consonants	29
4	Discussion	31
4.1	Geminate stops	31
4.2	Sonorants	33
4.3	Linguistic aspects	34
4.4	Diachrony and aspiration effects	36
4.5	Limitations and challenges	36
4.6	Further studies	38
	Bibliography	39

Chapter 1

Literature review

This chapter contains the literature review of this study. In Section 1.1, I briefly introduce some basic concepts and terminology related to laryngeal activity in speech production. Section 1.2 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 1.3, I discuss the major aspects of the phonology of Icelandic, focussing in particular on the laryngeal contrasts and contexts in stops and sonorants. Finally, in Section 1.4, I describe the hypothesis of this study and the reasons behind it.

1.1 Phonation types and states of the glottis

The periodic noise characteristic of the human voice is produced in the larynx. In particular, the vibration of the vocal folds generates oscillations of the air pressure which is translated as 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 state of the glottis are (1) spread glottis, (2) constricted glottis, (3) stiff vocal folds, (4) slack vocal folds. A spread glottis 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 to vibrate, or abducted just enough to not allow enough pressure to build up and set off the vibration. 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 contrary, 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 voicelens 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.

1.2 The effect of aspiration on vowel duration

Vowel duration has been reported in the literature to correlate with the presence vs. absence of aspiration in the following consonant. In particular, Maddieson and Gandour (1976) and Durvasula and Luo (2012) found that vowels followed by aspirated 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 is based on a four-way opposition of laryngeal

Table 1.1: Mean duration of vowels in Hindi before stops.

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
/t ^h /	184.75	/k ^h /	217
/d ^h /	196	/g ^h /	221

contrasts (Ohala, 1983). For each place of articulation, there are a voiceless unaspirated, a voiced unaspirated, a voiceless aspirated and (breathy) voiced aspirated stop: for example, [t], [d], [t^h], [d^h]. The voiceless aspirated stops (like [t^h]) are similar to the aspirated stops of English: a relatively long VOT follows the release of the occlusion. The voiced counterpart (like [d^h]) 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 [ka:d] ‘embroider’ and [ka:t^h] ‘wood’) were of equal length but longer than vowels followed by voiceless stops (like in [ka:t] ‘cut’). Moreover, vowels followed by voiced aspirated stops (like in [sa:d^h] ‘balance’) were even longer than voiced and voiceless aspirated stops. Table 1.1 shows the mean duration of vowels before the four alveolar stops as reported by Maddieson and Gandour (1976) and Lampp and Reklis (2004).

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. He also noted that an increase in closure duration of the stop following 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 is positively correlated with closure duration, such explanation is not compatible.

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, 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).

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, 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. In fact, most 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 attribute the voicing effect to adjustments of the larynx when moving from a vowel to a consonant.

1.3 Icelandic

Icelandic [GLOTTO: ice11247¹] is nowadays spoken by the inhabitants of the Republic of Iceland (about 320,000 according to Árnason 2011), of which it constitutes the na-

¹Entry for “Icelandic” at Glottolog: <http://glottolog.org/resource/languoid/id/ice11247>.

Table 1.2: Major consonantal allophones of Icelandic.

	labial	coronal	palatal	velar	glottal
stops	p p ^h	t t ^h	c c ^h	k k ^h	ʔ
fricatives	f v	θ ð	ç ʝ	x ɣ	h
nasals	m m̥	n n̥	ɲ ɲ̥	ŋ ŋ̥	
laterals		l l̥			
trills		r r̥			

Table 1.3: Vowels of Icelandic.

	front	back
high	i	u
high-mid	ɪ ʏ	
low-mid	ɛ œ	ɔ
low		a

tional language. Icelandic is a Germanic language and, together with Faroese [GLOTTO: *farø1244*²], 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 1.2 reports the major consonantal segments of Icelandic, as normally employed in the literature (Árnason, 2011, 98). They do not represent necessarily the phonemic consonants of Icelandic and are rather surface allophones. The vocalic phonemes are given in Table 1.3 (Árnason, 2011, 60).

Relevant to this study is the exploitation of phonation types in stop and sonorant consonants in Icelandic.

²Entry for “Faroese” at Glottolog: <http://glottolog.org/resource/languoid/id/farø1244>.

1.3.1 Laryngeal contrasts in stops

As mentioned in Section 1.2, 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.

1.3.2 Laryngeal contrasts in sonorants

The Lyon-Albuquerque Phonological Systems Database (LAPSyD) phonological database (Maddieson, 2012) states 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 thought of as aspirated consonants. This interpretation allows 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) review the possible combinations of laryngeal settings 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) states 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. In the second type of voiceless nasals, the segment is voiceless through out its duration. However, the release of the oral occlusion is accompanied by a lowered velum. Since the velum is open, air freely escapes through the nasal cavity. The small burst produced by releasing the occlusion mimics the auditory

features of an epenthetic stop (Bhaskararao and Ladefoged, 1991, 84).

Ladefoged and Maddieson (1996, 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 these consonants are perhaps better described as “aspirated.” Kehrein (2002, 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 way of implementing voicelens 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, 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 between the closures that constitute the trill (Ladefoged and Maddieson, 1996, 236).

Helgason (2002) notes 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 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.

Bombien (2006) 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).³

³Frication 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).

1.4 Research hypothesis

The studies 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 should allow more voicing.

In Icelandic, the contexts where a contrast involving aspiration can be found post-vocally is with geminate 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), these 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 unaspirated consonants.

The corresponding null hypothesis is:

(2) **Null hypothesis (H_0)**

Vowels followed by aspirated consonants are of the same duration as vowels followed by unaspirated consonants.

Chapter 2

Methodology

In this chapter I will describe the methodology employed for this study. Section 2.1 contains information about the speakers that participated in the study. In Section 2.2, I detail the materials used in the task, which is outlined in Section 2.3. Section 2.4 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 2.5 defines the measurements that I extracted from the annotated files and describes the procedure used for data normalisation. Finally, Section 2.6 briefly mentions the features of the statistical analysis applied to the measurements and the software I used to implement it.

2.1 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 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 a compensation for their time in the form of Amazon vouchers or chocolates for the value of £5 pounds.

The information on each participant is given in Table 2.1. 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

Table 2.1: Information on participants.

id	sex	age	birthplace	languages	abroad
TT	female	24	Reykjavik	English, Danish, German	Yes
BRS	female	25	Hofn	Danish, English, Spanish	Yes
BTE	female	27	Reykjavik	English, Danish	Yes
JJ	female	46	Reykjavik (Kopavogur)	English, Danish	Yes
SHG	male	25	Selfoss	English	No
JR	male	66	Reykjavik (York)	English	Yes

their birthplace. The last column, “abroad”, states if the subjects spent more than 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 a 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.

2.2 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 (bisyllabic). The list of target words is given in Table 2.2. The target words were selected so as to control for as many of the following 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.

The wordlist contained a total of 58 inflected Icelandic words (only real word forms were used). These were a mixture of nouns (26), verbs (22), adjectives (8) and adverbs (2). The 58 words were equally divided in monosyllabic (29) and bisyllabic (29) words. Of the monosyllabic words, 20 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 /IC/ cluster (one voiced, one voiceless); one word ended with a geminate nasal. Of the bisyllabic 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 2.2: List of target words.

word	IPA	pos	gloss
stops			
kokk	k ^h o ^h k	noun	cook
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 yp	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	itt	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

word	IPA	pos	gloss
detta	te ^h ta	verb	fall
gedda	cetta	noun	pike
nasals			
kamp	k ^h amp	noun	moustache
kamb	k ^h amp	noun	comb
punt	p ^h ɣnt	noun	decoration
pund	p ^h ɣnt	noun	pound
kembt	c ^h em̩t	verb	comb
kembdi	c ^h em̩tɪ	verb	comb
kampa	k ^h ampa	noun	moustache
kamba	k ^h ampa	noun	comb
kempa	c ^h empa	noun	hero
kemba	c ^h empa	verb	comb
punta	p ^h ɣnta	noun	decoration
punda	p ^h ɣnta	noun	pound
vanta	va̩nta	verb	want
vanda	vanta	verb	do st carefully
fínn	fi̩nt̪	adjective	smart
kinn	k ^h im̩n	noun	cheek
laterals			
duld	tylt	adjective	complex
dult	tylt̪	adjective	reticent
gelta	ce̩lta	verb	bark
gelda	celta	verb	castrate
mjólka	mjou̩lka	verb	milk
ólga	oulka	verb	foam
trills			
orka	o̩rka	noun	energy
orga	orka	noun	scream

2.3 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 will of keeping the task as simple as possible. 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 familiarised 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 recorder 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.

2.4 Annotation

The analysis of the audio file 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, but I did not apply any filter. 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 2.1 shows an example of the TextGrid set-up.

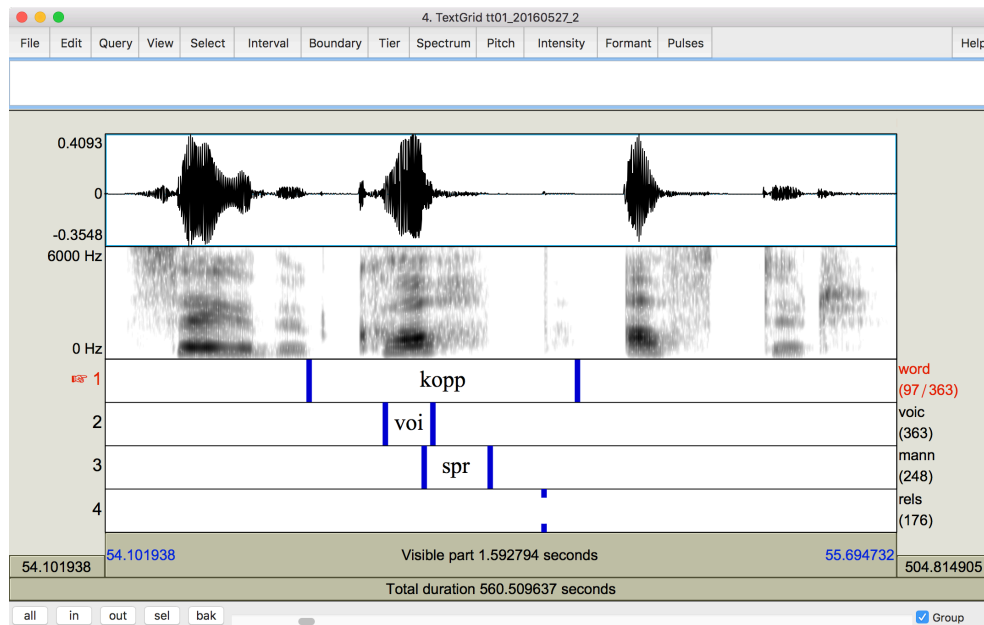


Figure 2.1: Example of the tier structure of the annotation files (PRAAT TextGrids). Tiers: 1. word, words; 2. voic, voicing; 3. mann, manner; 4. rels, release.

2.4.1 Word boundaries

The first tier was segmented by target words. The left boundary of the word was considered to be the off-set 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 boundary was placed at the off-set 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 (*afur*); if a clear glottal stop separated the target word from the following, the boundary coincided with the on-set 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.

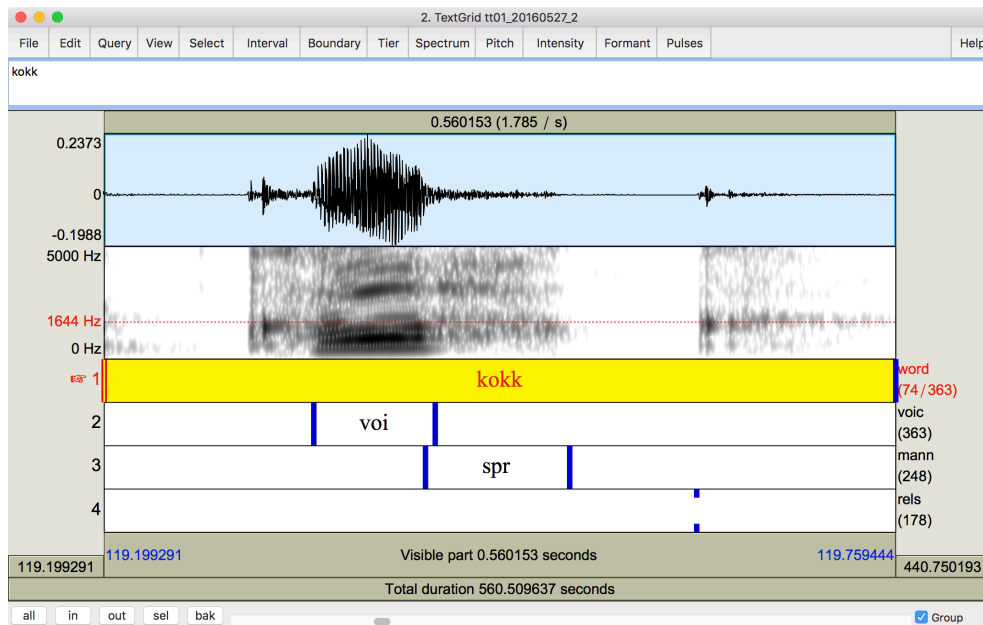


Figure 2.2: The word *kokk* ‘cook’ as an example of the annotation of glottal spread.

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

2.4.3 Glottal spreading

A single tier was used to annotate glottal spread, nasal airflow, laterality and rhoticity. 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; the vocal folds start moving apart from each other in an abduction gesture while they still vibrate (breathy voice);

then, vocal fold vibration stops while voiceless friction remains (at the glottis or at the oral cavity, depending on the place of articulation vowel). 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 2006 (cited in Nance and Stuart-Smith 2013, 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.

2.4.4 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). 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.

2.4.5 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. If the burst was not identifiable from the waveform nor from the spectrogram, no release was marked.

2.5 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

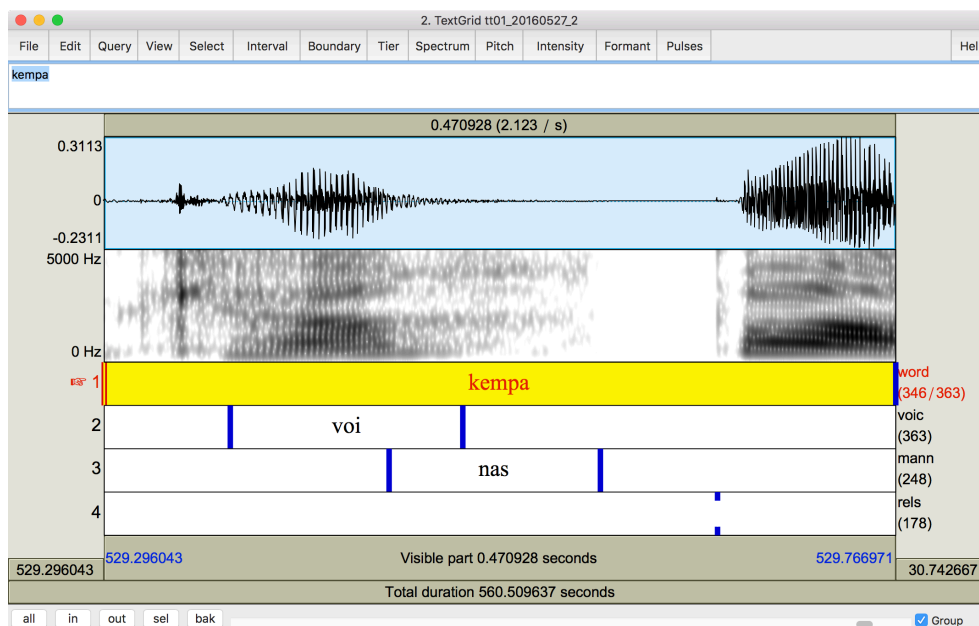


Figure 2.3: The word *kempa* ‘hero’ as an example of the annotation in the manner tier.

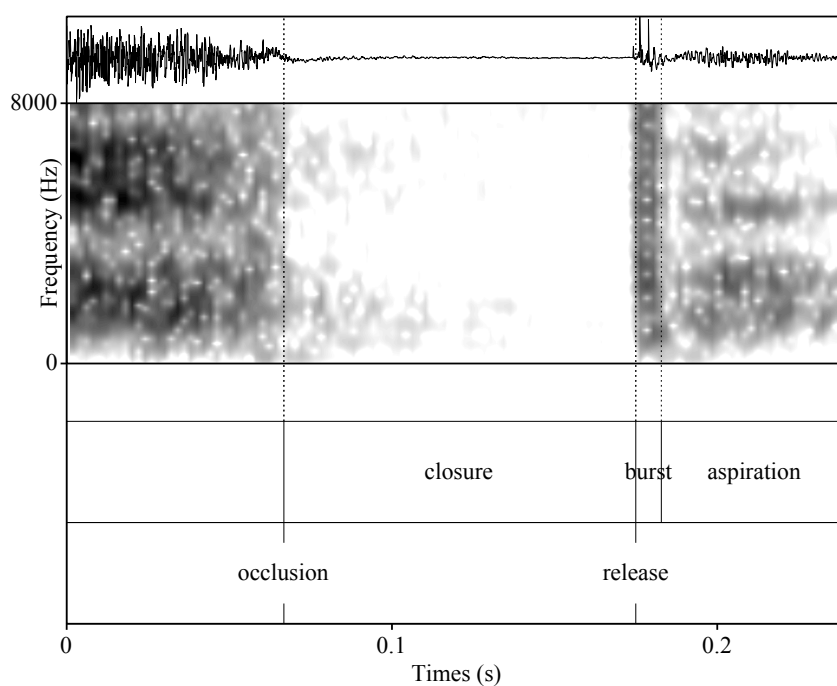


Figure 2.4: Release of [h]p in the word *kopp*.

in ???. The output of the extraction procedure is a .csv file which contains the relevant measurements. The following measures were taken in milliseconds:

- *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 absolute voicing 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 off-set of voicing, glottal spread or consonant manner, and the release.
- *Closure duration*: the duration of the interval between the on-set 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 spread*. 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 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 (VCC, V^hC, VRC, where “R” is any sonorant).¹

2.6 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 Wilcoxon 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 [...].

¹Pre-aspirated stops are phonologically geminates.

Chapter 3

Results

3.1 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 (SD = 78.24), while for VCC is 338.71 msec (SD = 54.86). CVCCV words have a mean duration of 487.25 msec (SD = 59.94), while VCCV are on average 424.99 msec long (SD = 53.42).

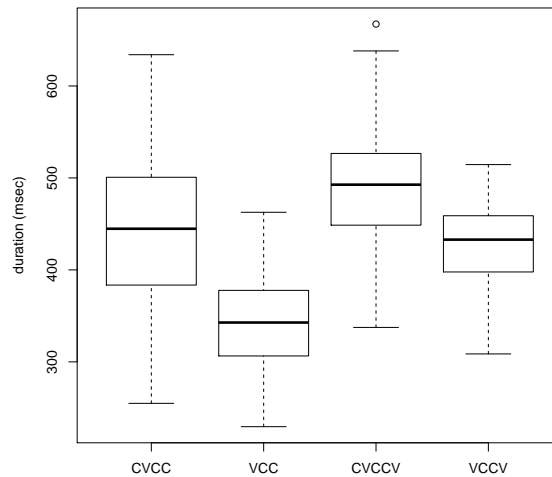


Figure 3.1: Duration (in msec) of CVCC, VCC, CVCCV, and VCCV words.

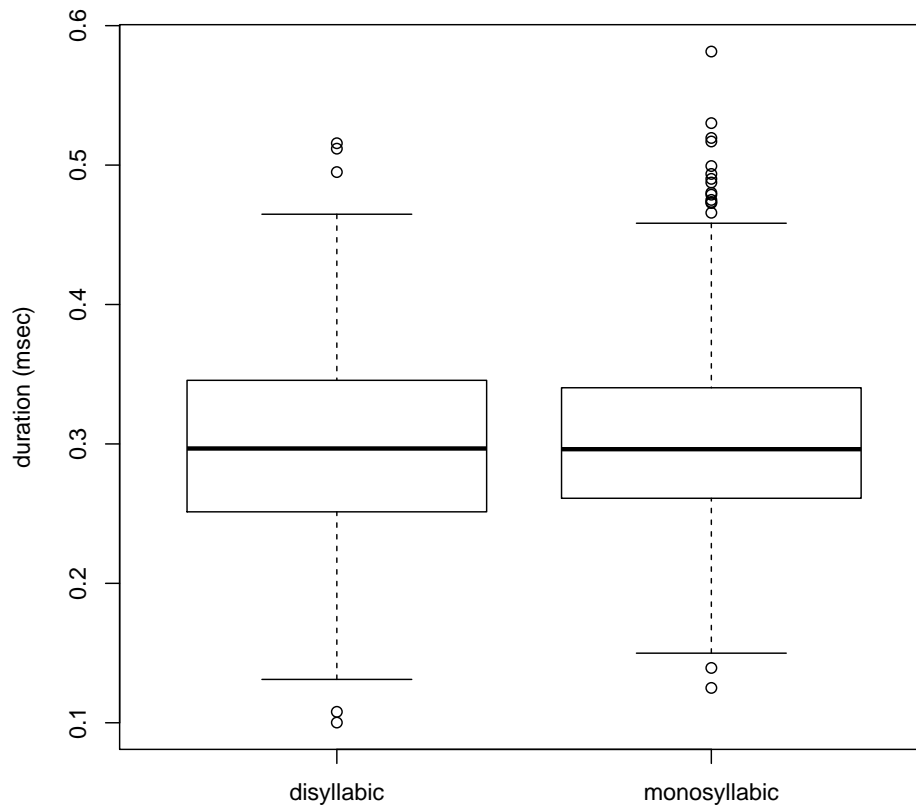


Figure 3.2: Vowel duration in monosyllabic vs. disyllabic words.

The ratio duration of the vowel was not affected by the the number of syllables of the word. The boxplot in Figure 3.2 shows the duration of the vowel in milliseconds. According to a two-sample Wilcoxon test, the mean duration of the vowel in monosyllabic words (88.67) does not differ significantly from the mean duration in disyllabic words (84.21) [$V = 35302$, $p < 0.001$].

3.2 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 -VCC words, did not have a

significant difference in mean duration in both the non-aspirated and aspirated condition.

3.3 Vowel duration

3.3.1 Geminate stops

The mean duration of vowels in milliseconds was 98 in monosyllabic words ending in a non-aspirated geminate stop, while it was 84.47 if followed by a pre-aspirated geminate stop. Section 3.3.1 shows the ratio duration of the vowel in the two conditions. The difference in the mean of the vowel ratio between the non-aspirated (0.37) and the aspirated class of words (0.28) was significant [$W = 7793$, $p\text{-value} < 0.001$].

In disyllabic words, the voiced vocalic portion is 86.97 msec long if followed by a non-aspirated geminate, and 78.04 msec long if followed by a pre-aspirated geminate. In ratio terms, the voiced part makes up the 33.61% of the VOR in word with plain geminates, and the 28.27% in words with pre-aspirated geminates. According to an independent two-sample t -test, the difference in ratio is significant [$t = 4.39$, $df = 50.46$, $p < 0.001$].

3.3.2 Nasals

Words of the form (C)VNC showed a mean voiced portion duration of 89.96 msec if the nasal is voiced and 77.73 msec if voiceless. The normalised duration was, respectively, 0.31 and 0.26. The difference in ratio duration was significant [$W = 836$, $p < 0.001$].

CVNVCV words have a mean voiced vocalic portion duration of 81.72 msec if the nasal is voiced and 67.48 msec if it is voiceless. As percentages, the vocalic portion is 29.47% of the VOR if followed by a voiced nasal, and 24.16% when the nasal is voiceless. A t -test showed that this difference is significant [$t = 7.04$, $df = 119.99$, $p < 0.001$].

3.3.3 Laterals

In the case of words ending in a /lC/ cluster, the mean duration of the vocalic portion were 102.05 and 80.7 msec when followed, respectively by a voiced and voiceless lateral. These correspond to the 29.69% and the 24.98% of the VOR. A t -test shows that the difference is significant [$t = 4.1$, $df = 20.43$, $p < 0.001$].

The voiced vocalic portion in CVLVCV words is on average 107.68 msec long when the lateral is voiced and 66.66 msec when it is voiceless. The voiced portion takes the

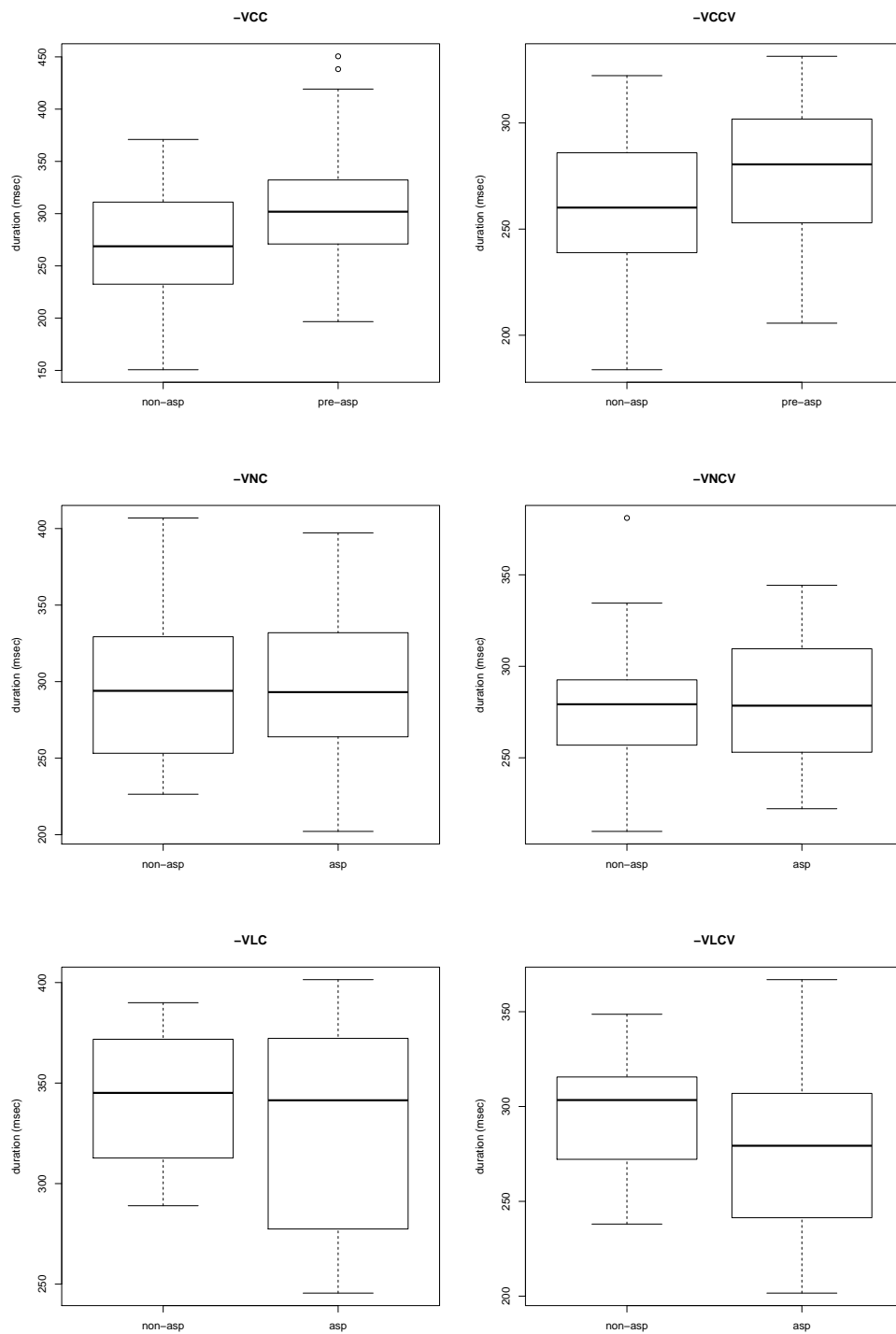


Figure 3.3: Duration (in msec) of the Voice Onset to Release (VOR).

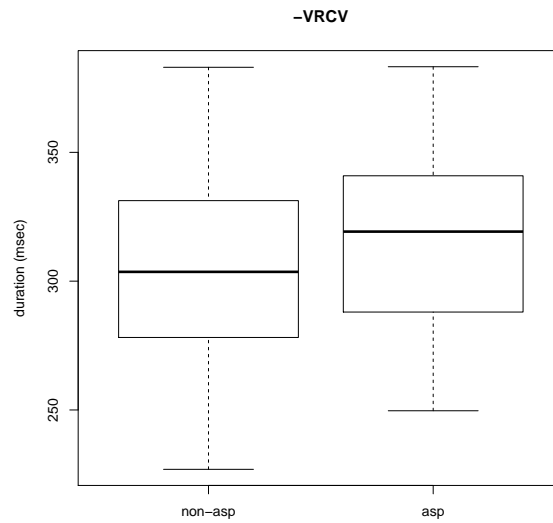


Figure 3.4: Duration (in msec) of the Voice Onset to Release (VOR) of VRCV words.

36.2% of the VOR if followed by a voiced lateral, and 23.26% when the lateral is voiceless. A t-test reveals the difference to be significant [$t = 7.47$, $df = 50.72$, $p < 0.001$].

3.3.4 Rhotics

With words containing an /rC/ cluster, the mean duration of the voiced vocalic portion is 121.07 msec long when the rhotic is voiced and 104.78 msec when it is voiceless. The voiced portion takes the 39.83% of the VOR if followed by a voiced rhotic, and 33.43% when the rhotic is voiceless. A t-test gave a significant result [$t = 3.4$, $df = 27.98$, $p < 0.001$].

3.4 Duration of stop closure

The ratio duration of stop closure in words containing a sonorant + stop cluster 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 [$t = 5.62$, $df = 27.82$, $p < 0.001$]. 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.

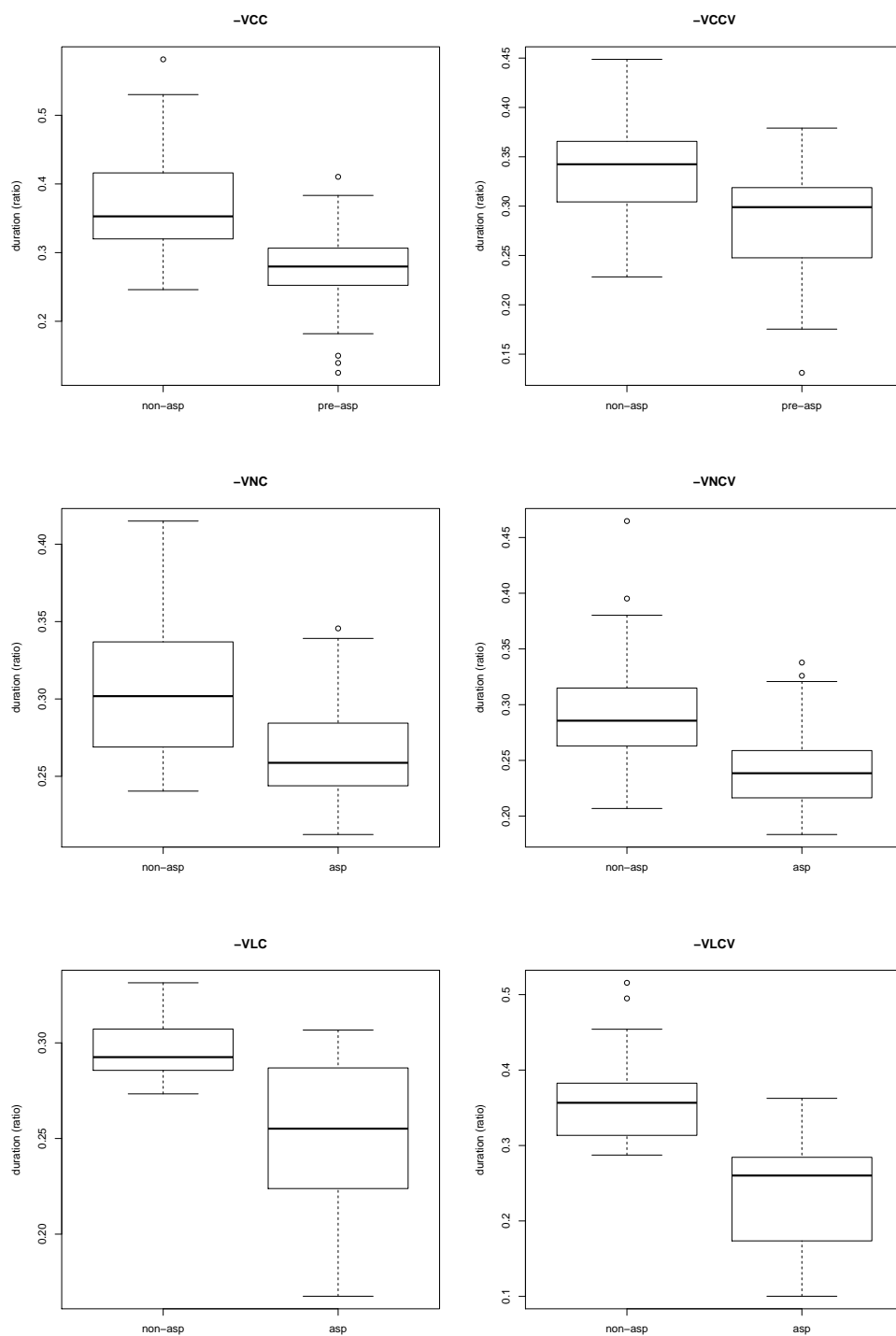


Figure 3.5: Vowel duration (ratio) depending on the presence vs. absence of pre-aspiration in the following consonant

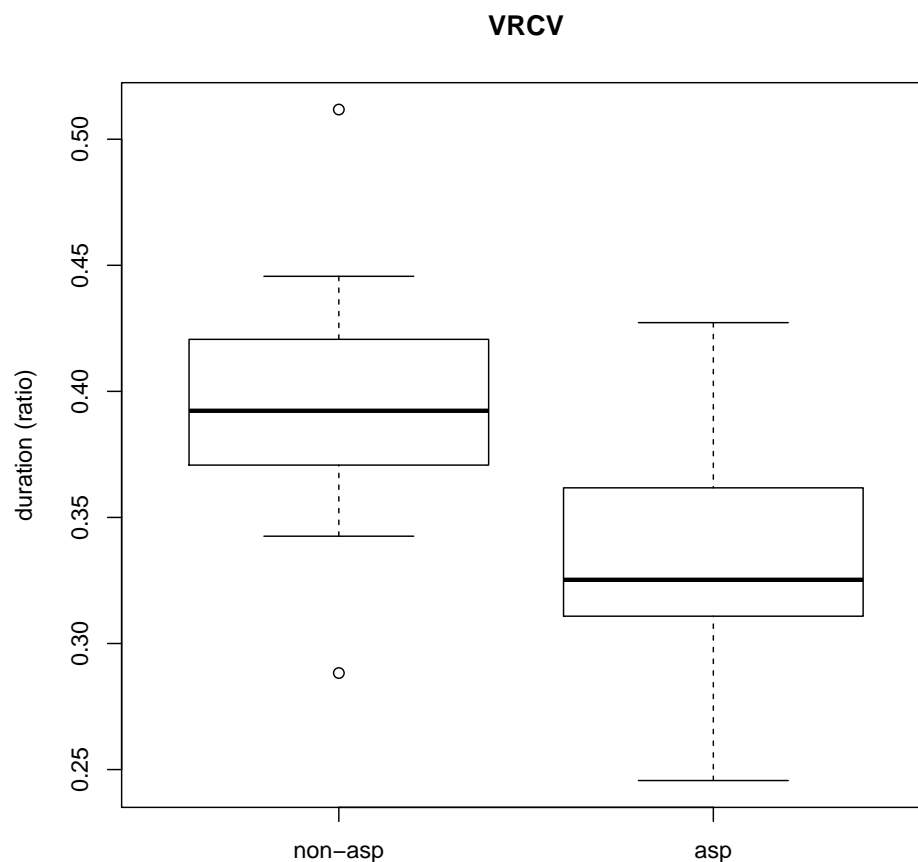


Figure 3.6: Ratio vowel duration in words with an /rC/ cluster.

3.5 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 as the level of significance []. A post-hoc Games-Howell test for unequal variances showed that, while the duration of the spreading gesture did not differ within the sonorant classes, the duration of glottal abduction in stops was significantly different from that one in lateral and nasals (but not from the one in trills).

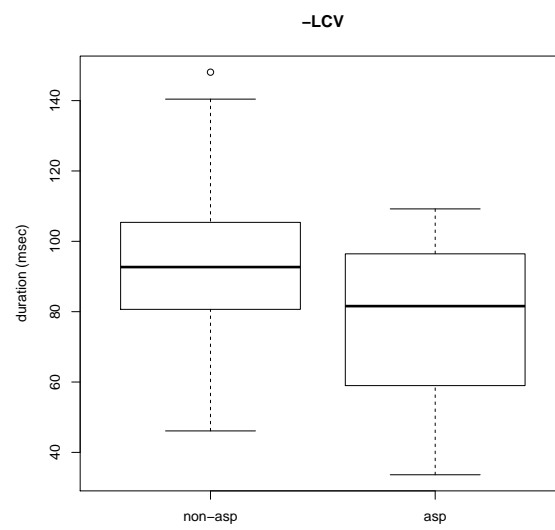


Figure 3.7: Stop closure duration (msec) in -LCV words.

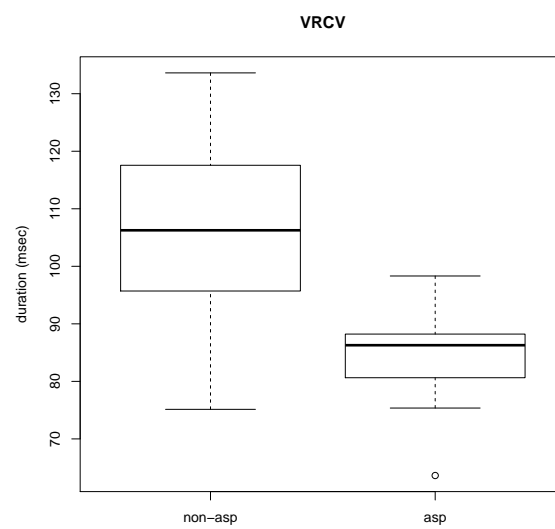


Figure 3.8: Stop closure duration (msec) in VRCV words.

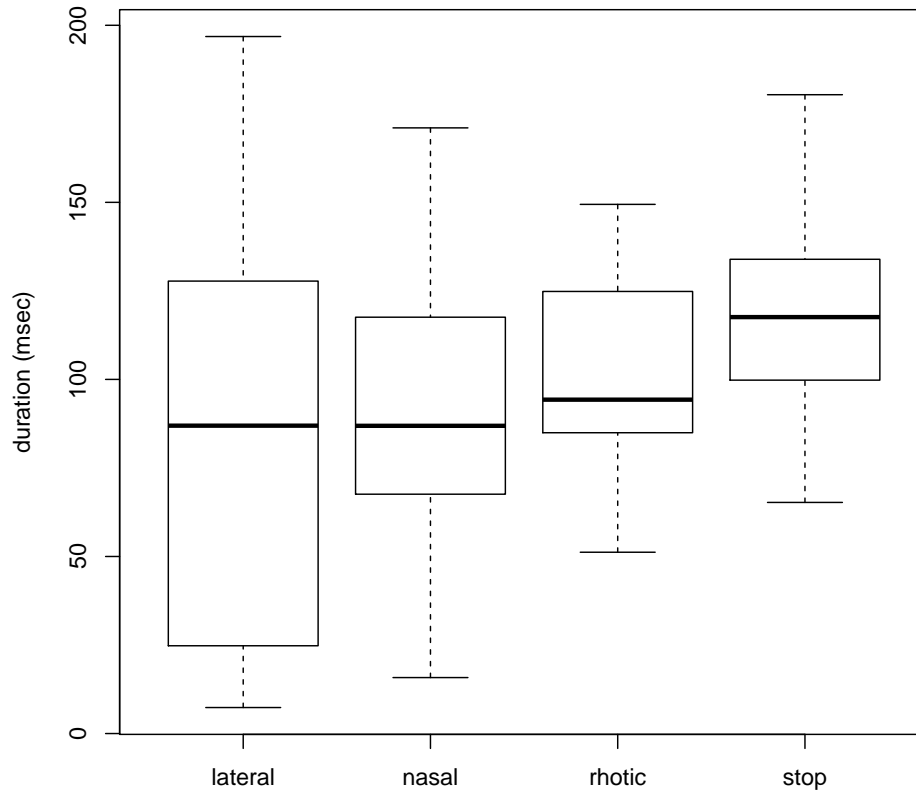


Figure 3.9: Duration (in msec) of the glottal spread gesture in sonorants and stops.

3.6 Duration of sonorant consonants

The duration of voiced sonorant consonants was significantly different from the duration of voiceless sonorants. As a trend, the voiceless sonorants are longer in duration than the voiced ones. 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 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 overlap with those in Jessen and Pétursson (1998), Bombien (2006) and Silverman (2012).

Chapter 4

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 3), I found a significant difference between the duration of the vowel in words with aspirated consonants and its duration in words with non-aspirated consonants. This finding seems to support the timing hypothesis.

4.1 Geminate stops

We have seen that the absolute duration of vowels is smaller when they are followed by a pre-aspirated geminate 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 . 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 aspiration while the absolute duration of the vowel is *smaller*. Other things being equal, we would expect 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 voiced interval in words with plain geminates. Instead, laryngeal spreading starts quite early in the word, while voicing is still active even after glottal spread is initiated.

It can be then argued that the relative earlier timing of glottal abduction reduces

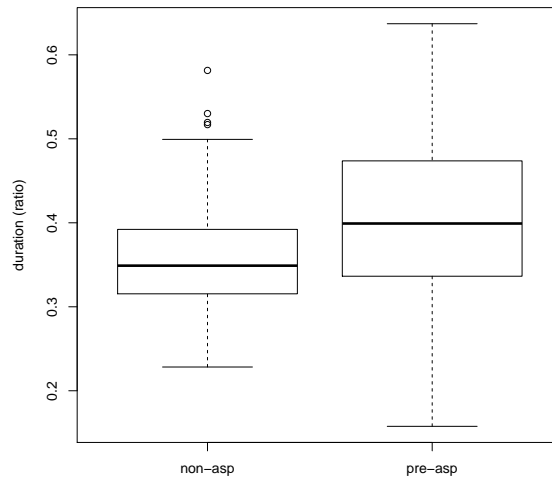


Figure 4.1: Ratio duration of the voicing interval depending on presence or absence of glottal spread.

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 (which is voicing with concomitant abducted vocal folds) 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 gestured might have started even earlier, strengthening the central idea proposed in this research.

However, the hypothesis rested on the idea that voicing would cease earlier in vowels followed by pre-aspirated stops, resulting in a shortened vowel. What I found, instead, points to a longer voicing portion in words with pre-aspirated stops. The duration of voicing is, in fact, greater in words with pre-aspirated stops than in words with plain geminates. This is not surprising, since glottal spread is not incompatible with voicing and could actually allow voicing to be maintained longer. A plain geminate is produced by keeping the closure for a longer time and voicing during closure is more difficult to maintain. Thus it is natural to expect that, other things being equal, voicing duration would be greater if glottal spread is involved. This fact does not fit well with the idea that voicing should cease earlier due to glottal spread. Even if the physiological explanation seems not to hold, the relative timing on itself can be provisionally considered to be right.

Contrary to what stated in Section 1.4, glottal abduction does not prevent voicing. On the contrary, longer voicing can be found in words with glottal spread. A possible explanation of this phenomenon is that the absence of a closure at the VC transition allows for voicing to continue. In words without glottal spread, the closure of the plain geminate—which is produced immediately after the vowel—quickly produces an increase in oral pressure, until the point is reached when voicing can no longer be sustained. At the same time, the variance of voicing in the two classes is not the same, and voicing in the aspirated condition has greater variance. In some aspirated tokens, 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, 70–71) state that the glottal width in Icelandic pre-aspirated stops is wider than in non-aspirated stops, but they argue 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. 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 preceding vowel and the release of the geminate. Since normalisation was done with the VOR as a base, the portion of VOR subtracted by the modal vocalic portion would correspond to the geminate. Thus, if the ratio duration of glottal spread which should correspond to the duration of the geminate were to be the same both in pre-aspirated and plain stops, the ratio duration of the vowel should have been the same as well. However, I found that the ratio duration of the vowel was shorter if followed by a pre-aspirated stop.

4.2 Sonorants

The class of sonorant consonants (nasals, laterals and rhotics) 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 aspirated and non-aspirated conditions, except with laterals and rhotics in disyllabic words. The stop consonants in words with /NC/ and /ŋC/ clusters had identical closure durations, so the difference in vowel duration cannot be attributed

to closure properties. The presence vs. the absence of aspiration did not affect the duration of stop closure.

On the contrary, the ratio duration of sonorant consonants (nasal, lateral and rhotic) 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.

4.3 Linguistic 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, 423) point out that “speakers and listeners employ extensive and subtle phonetic knowledge.”

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, ar-

gue that the phonetic constraints normally called for in the literature as the cause of most phonological behaviours do not have predictive power, but rather 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 classes 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” which shape the phonological systems of the languages of the world.

We have seen that pre-aspirated stops in Icelandic are preceded by shorter vowels, if we consider the vowel to be the modally voiced open 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 nasals, 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.

4.4 Diachrony and aspiration effects

In the previous section I argued that the origin of the effect could be ascribed to timing of laryngeal but that does not mean that synchronically is the case.

4.5 Limitations and challenges

Some of the limitations of this study concern the participants recruited for the reading task. The first difficulty was finding a sensible number of Icelandic speakers in York. Due to limitations 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 taken with a cautionary note.

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 2.2, 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 in the task procedure. 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. This could have been avoided, even if at the cost of less naturalistic data, by letting the speakers read the words beforehand.

A recurrent problem with studies of phonetic duration is that speech tempo is not stable and it is difficult to control. Using ratio durations can overcome some of the shortcomings of varying tempo. However, the compressibility of phonetic material is not homogeneous between classes of sounds and contexts.

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 2.4, 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.

Even so, the annotation scheme presented some 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 only 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 measurement 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. However, 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.

4.6 Further studies

Some of the limitations presented in the previous sections could be overcome with articulatory data. Laryngeal measures can be taken using photoglottography or a laryngograph. While the first tool is invasive, the second is not and its portability makes 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 hypothesis that rely on articulatory factors, such as those of ... and ...

Combined with electropalatography, ultrasound scanning can also improve the accuracy of closure detection. As mentioned above, at times closure is not clearly visible from the spectrogram, especially in cases when voicing is maintained after it. Tracking the movements of the tongue through ultrasound 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.

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

- 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.
- Jones, M. . C. L. (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.
- 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.
- 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. (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.
- 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).
- 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 Linguistics*, Vol. 15 of *Dissertation 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 prepiration. *Nordic Journal of Linguistics* 1(01), 3–54.

- 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.