Vowel duration and aspiration effects in Icelandic

Stefano Coretta

August 22, 2016

Contents

Chapter 1

Literature review

1.1 Laryngeal

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 contrasts. 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], [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 1.1 shows the mean duration of vowels before the four alveolar stops as reported by Maddieson and Gandour (1976, 47).

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

consonant	vowel duration (msec)
/t/	160
/d/	184.5
/t^h/	184.75
/d^h/	196

Durvasula and Luo (2012) performed a more controlled task and found clear evidence that aspiration lengthens the preceding vowel. He also noted that

1.3 Icelandic

Icelandic [GLOTTO: `icel1247`[1]] is nowadays 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`[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 **??** (Árnason, 2011, 60).

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

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,

[1]Entry for "Icelandic" at Glottolog: `http://glottolog.org/resource/languoid/id/icel1247`.

[2]Entry for "Faroese" at Glottolog: `http://glottolog.org/resource/languoid/id/faro1244`.

Table 1.2: Major consonantal allophones of Icelandic.

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, nasals 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. This Nordic language has voiced and voiceless nasals, laterals and trills.

Ladefoged and Maddieson (1996) reviews 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 voiceleness in laterals can be distinguished according to a variety of acoustic parameters. Given that in approximants the vocal tract is open while in fricative 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). For

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]

Silverman (1997, 61) states that in sonorants the optimal phasing for contrastive

[3]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).

gestures envisages the truncation of laryngeal feature in respect to supralaryngeal ones and sequencing of non-modal phonation and modal phonation. Thus, in nasals, optimal phasing is obtained with laryngeal abduction truncated before the supralaryngeal occlusion is released. If abduction is sequence before voicing, voicelessness in the first part of the nasal is followed by modal voice. If, instead, abduction follows modal voicing, breathiness is found in the last part of the nasal. Laterals show two possibilities for optimal phasing of laryngeal abduction: in the first, abduction is simultaneous with the lateral gestures, leading to voiceless lateral fricative; in the second, laryngeal abduction is truncated, as in nasals. This difference between the possible realizations of optimal phasing in nasals and laterals is, according to Silverman, due to a combination of two factors: all laterals are phonetically similar and laterals 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.

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 voincing.

In Icelandic, the contexts where a contrast involving aspiration can be found posvocalically 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), 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

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 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 their birthplace. The last column, "abroad", states if the subjects spent more that 6 months outside Iceland. Participant JR had to be excluded from the analysis since he misunderstood the task, while part of participant SHG's task was lost due to a technical fault in the recording equipment.

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 Appendix A. 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

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

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 /lC/ 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).

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. To 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’

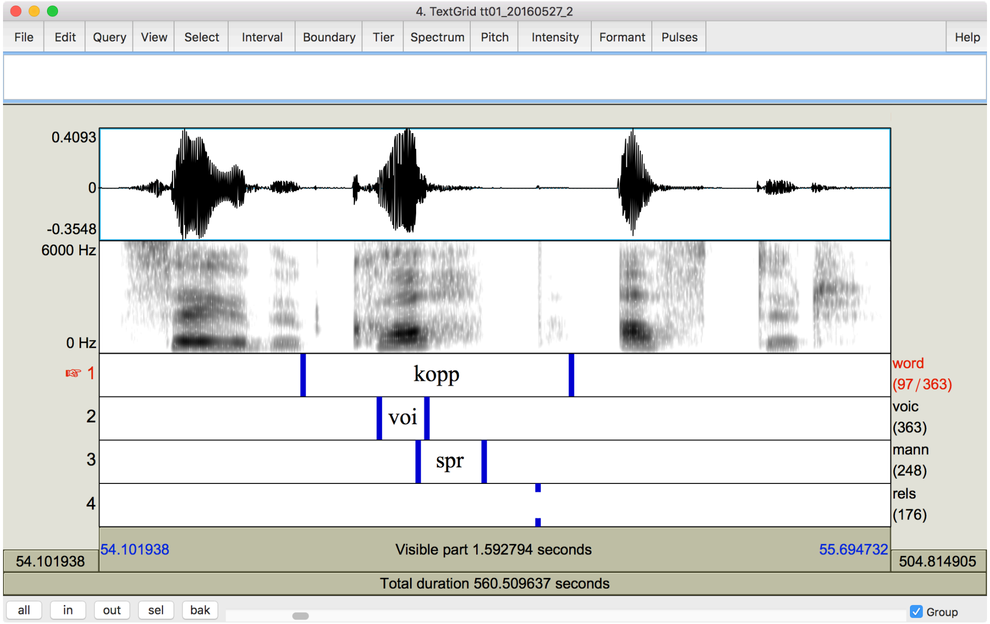

Figure 2.1: Example of the tier structure of the annotation files (PRAAT TextGrids). Tiers: 1. `word`, words; 2. `voic`, voicing; 3. `mann`, manner; 4. `rels`, release.

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 a meeting room at a travel agency, while one was recorded at the University of York and the last 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 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.

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 (*aftur*); 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.

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 Consonant release

The last tier was dedicated to marking the consonant release of the stop following the target vowel (with or without and intervening sonorant). The release of the stop consonant was marked at the onset of the burst. This is normally visible on the waveform as

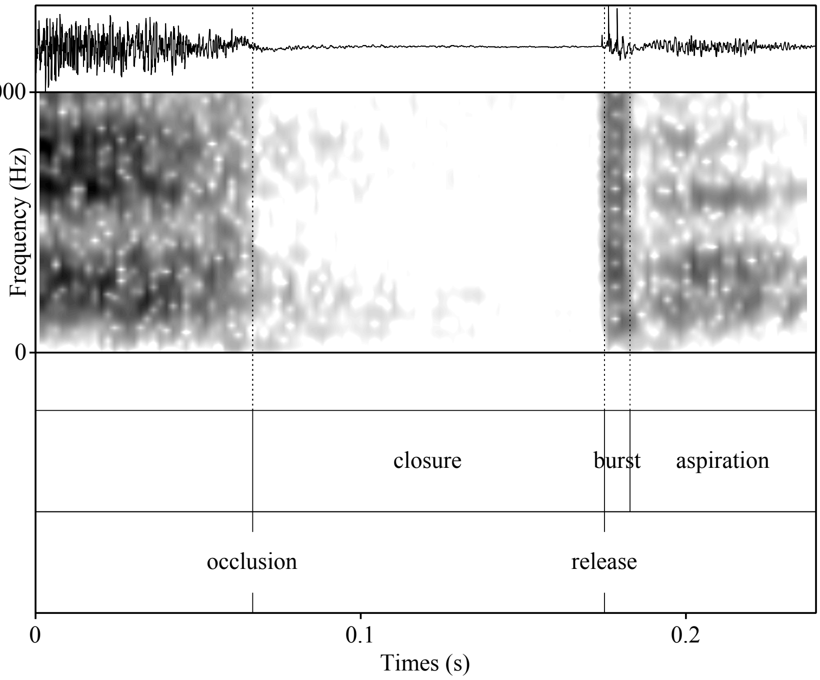

Figure 2.2: Release of [^{h}p] in the word *kopp*.

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 in Appendix B. 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.

To control for differences in speech rate, I opted to normalise the measurement. 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^{h}C, VRC, where "R" is any sonorant).[1]

2.6 Statistical analysis

I performed the statistical analysis using the R programming language (R Core Team, 2015) in RStudio (RStudio Team, 2015).

[1] Pre-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 488.55 msec (SD = 60.93), while VCCV are on average 424.99 msec long (SD = 53.42).

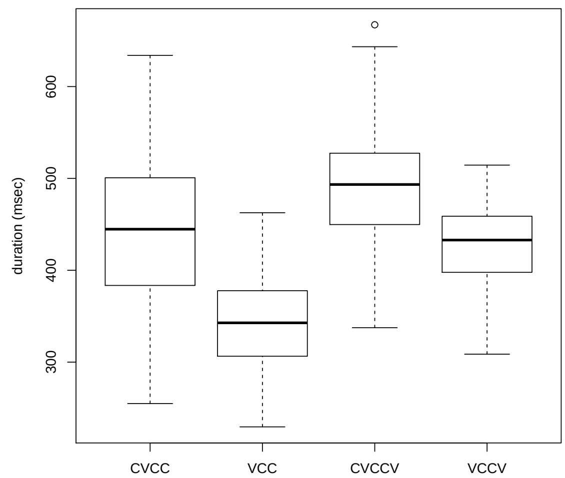

Figure 3.1: Duration (in msec) of CVCC, VCC, CVCCV, and VCCV words.

The duration of the voiced vocalic portion was not affected by the the number of syllables of the word. The boxplot in **??** shows the duration of the vocalic portion in milliseconds. According to a two-sample Wilcoxon test, the mean duration of the voiced portion in monosyllabic words () does not different significantly from the mean duration in disyllabic words () [V =]. However, the voice on-set to release was shorter in disyllabic words () than in monosyllabic words () []. Since duration normalisation was based on the voice on-set to release duration, in the next sections I will discuss monosyllabic and disyllabic words separately.

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. Figure 3.4 shows the ratio duration of the vowel in the two conditions. The difference in the mean of the vowel ratio between the non-aspirated and the aspirated class of words was significant [].

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 a independent two-sample *t*-test, the difference in ratio is significant.

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.

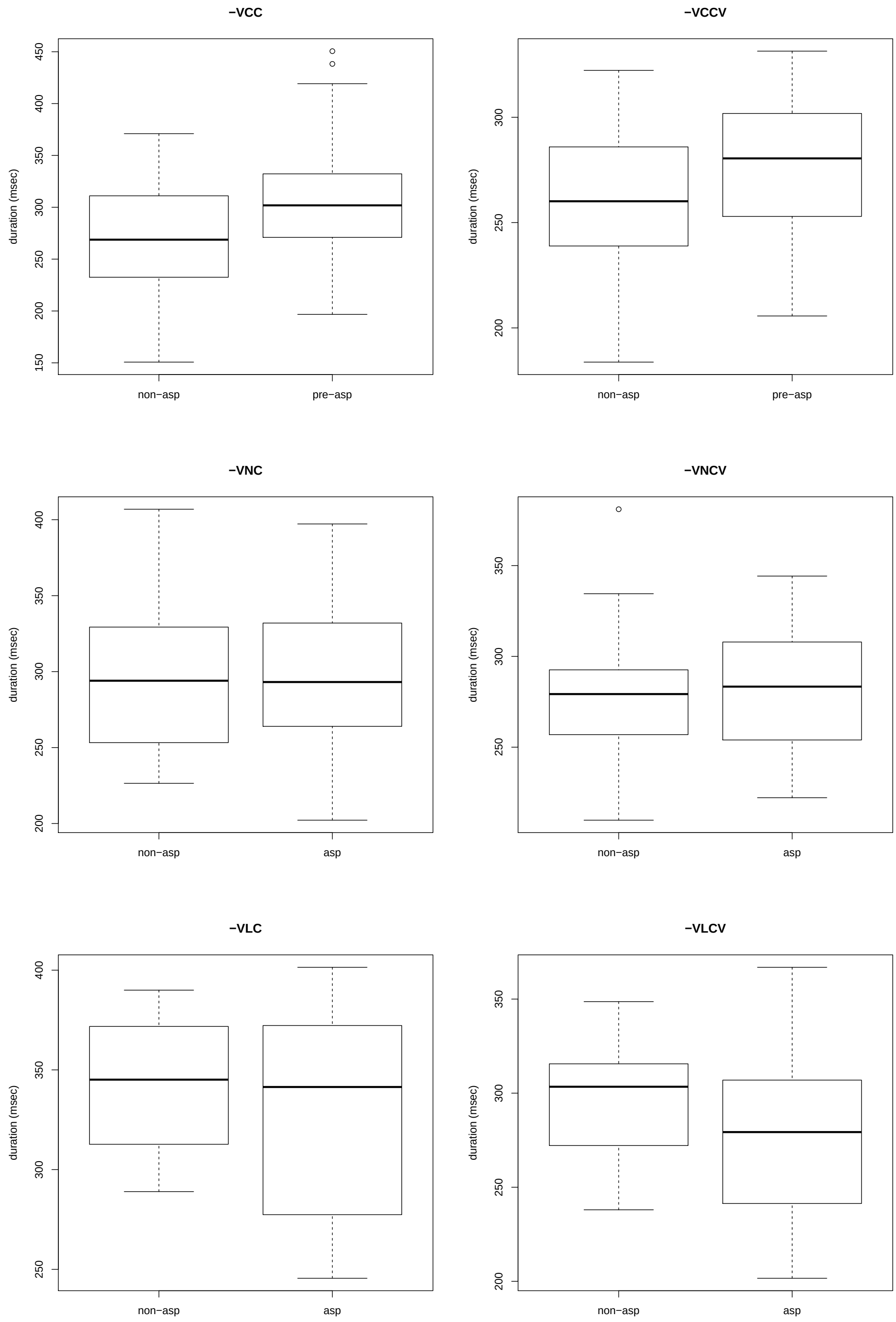

Figure 3.2: Duration (in msec) of the Voice Onset to Release (VOR).

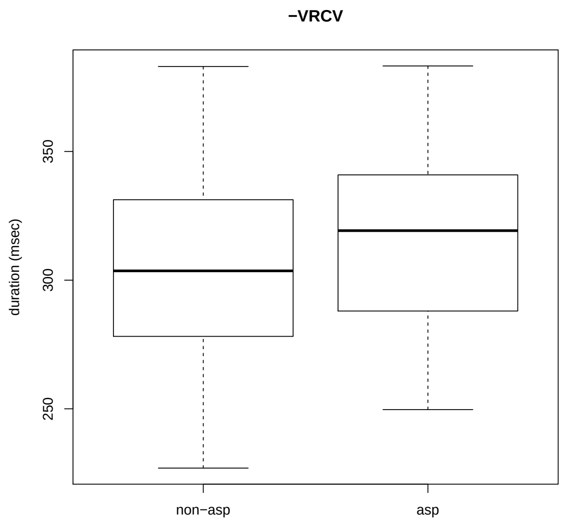

Figure 3.3: Duration (in msec) of the Voice Onset to Release (VOR) of CVRC and CVRCV words.

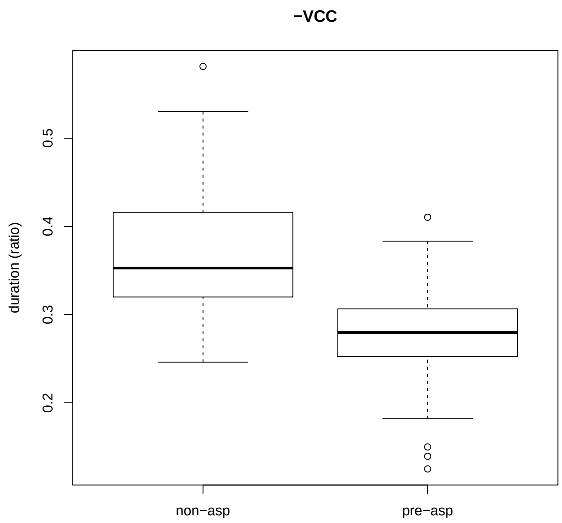

Figure 3.4: Ratio vowel duration in -VCC words.

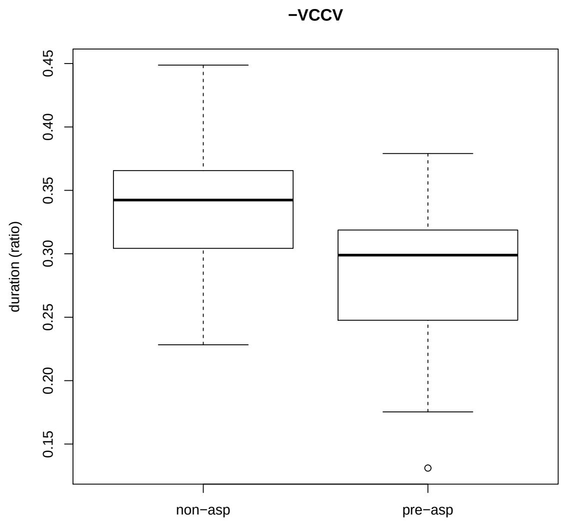

Figure 3.5: Ratio vowel duration in -VCCV words.

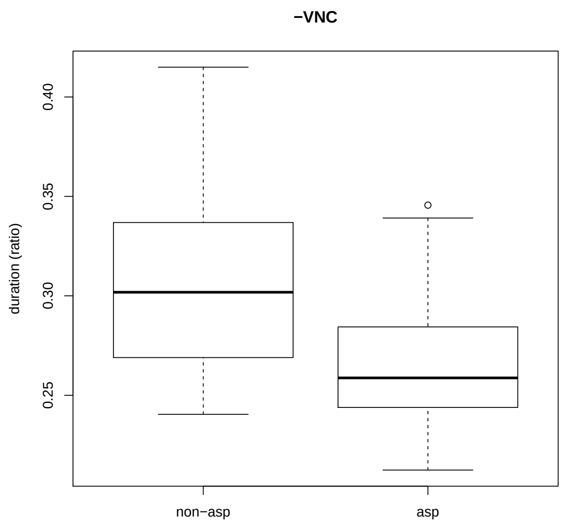

Figure 3.6: Ratio vowel duration in -VNC words.

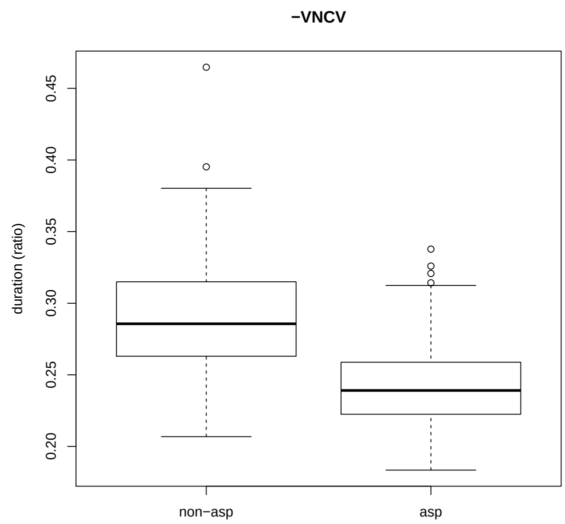

Figure 3.7: Ratio vowel duration in -VNCV words.

CVNCV words have a mean voiced vocalic portion duration of 81.72 msec if the nasal is voiced and 67.83 msec if it is voiceless. As percentages, the vocalic portion is 29.47% of the VOR if followed by a voiced nasal, and 24.22% when the nasal is voiceless. A *t*-test showed that this difference is significant.

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 Wilcoxon test shows that the difference is significant [].

The voiced vocalic portion in CVLCV 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 36.2% of the VOR if followed by a voiced lateral, and 23.26% when the lateral is voiceless.

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.

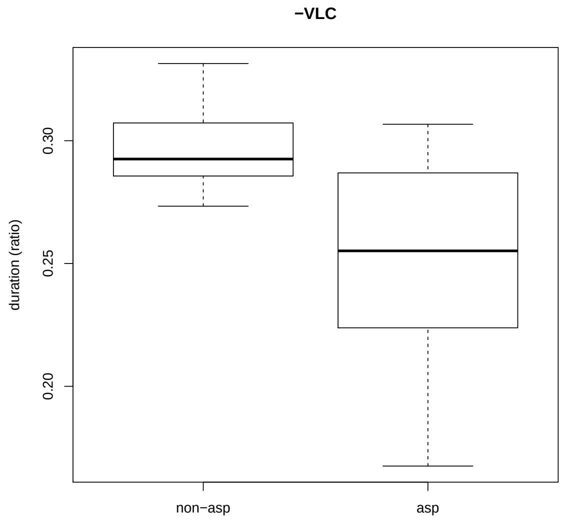

Figure 3.8: Ratio vowel duration in -VLC words.

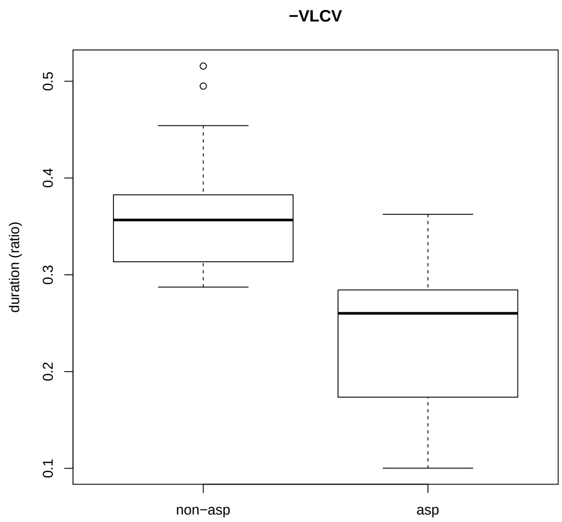

Figure 3.9: Ratio vowel duration in -VLCV words.

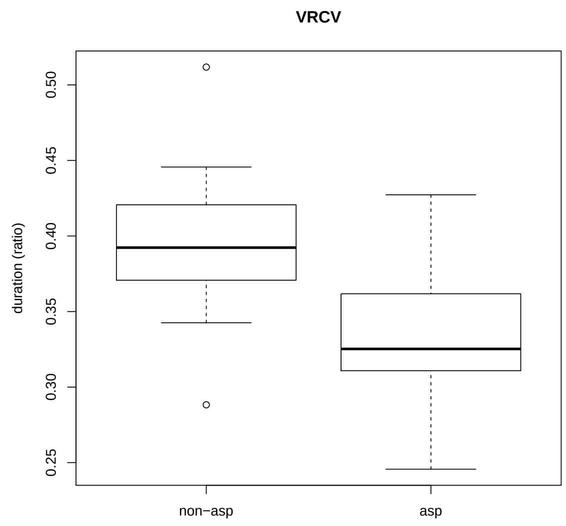

Figure 3.10: Ratio vowel duration in VRCV words.

The voiced portion takes the 39.83% of the VOR if followed by a voiced rhotic, and 33.43% when the rhotic is voiceless.

3.4 Stop closure duration

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 a /lC/ [] and /rC/ cluster []. 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.

3.5 Duration of glottal spread

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). Moreover, these results overlap with those in Jessen and Pétursson (1998), Bombien (2006) and Silverman (2012).

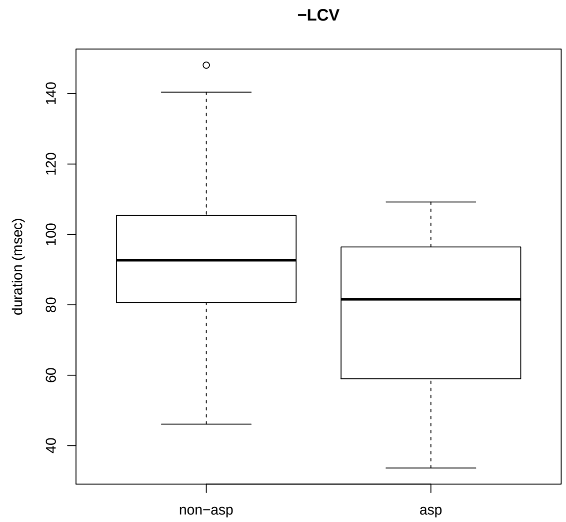

Figure 3.11: Stop closure duration (msec) in -LCV words.

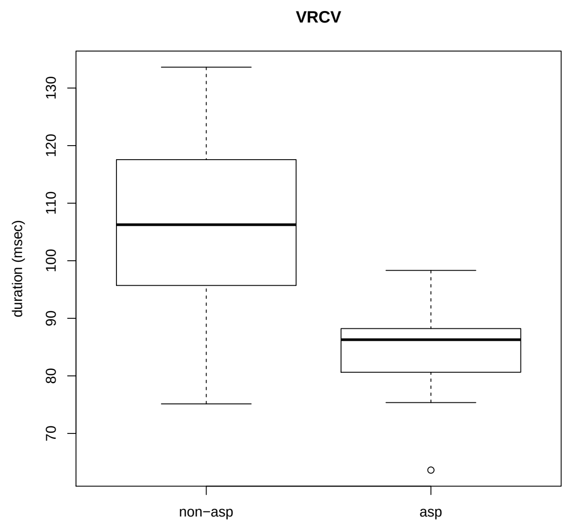

Figure 3.12: Stop closure duration (msec) in VRCV words.

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 voiced vocalic portion 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 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 reached the critical point where the abduction is enough to produce breathiness. This would imply that the gestured might start 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 word 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.

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

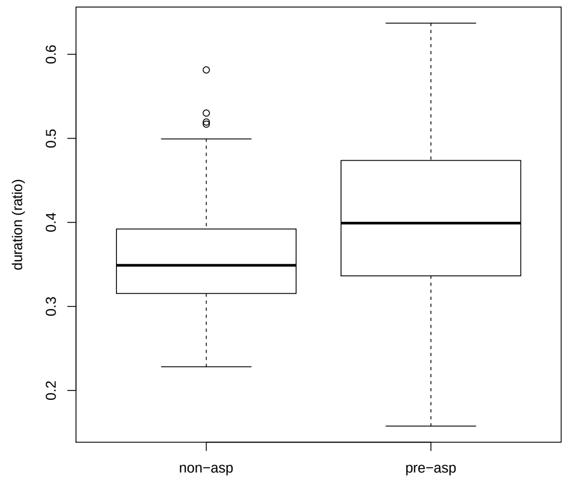

Figure 4.1: Ratio duration of the voicing interval depending on presence or absence of glottal spread.

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 /N̥C/ 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 characteristics 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 Individual differences

4.4 Diachrony and aspiration effects

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 in. 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.

Another 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 **??**, 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 () did not recognise most of the 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 spectrogram.

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 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 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 open and when it has been closed during speech.

Appendix A

Word list

Table A.1: List of target words

word	IPA	word	IPA
kokk	kʰoʰk	kembt	kem̥t
gogg	kokk	kembdi	kemtɪ
dökk	tœʰk	kampa	kʰam̥pa
dögg	tœkk	kamba	kʰampa
kopp	kʰoʰp	kempa	kʰem̥pa
kubb	kʰʏpp	kemba	kʰempa
vítt	viʰt	punta	pʰʏn̥ta
vídd	vitt	punda	pʰʏnta
þítt	θiʰt	vanta	van̥ta
þíddi	θittɪ	vanda	vanta
fætt	faiʰt	fínn	fitn̥
fæddi	faittɪ	kinn	kʰɪnn
ýtt	iʰt	duld	tʏlt
ydd	ɪtt	dult	tʏl̥t
ótt	ouʰt	gelta	kel̥ta
odd	ott	gelda	kelta
sets	sess	orka	or̥ka
sett	seʰt	orga	orka
feits	feiss	mjólka	mjoul̥ka
feitt	feiʰt	ólga	oulka
vots	voss	hefna	hepna
vott	voʰt	vopna	voʰpna
takka	tʰaʰka	nafla	napla
kagga	kʰakka	japla	jaʰpla
detta	teʰta	kafli	kaplɪ
gedda	ketta	kapli	kaʰplɪ
kamp	kʰam̥p	tefla	tepla
kamb	kʰamp	tipla	tɪʰpla
punt	pʰʏn̥t		
pund	pʰʏnt		

Appendix B

PRAAT script

Bibliography

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

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.

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

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

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.

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

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

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.

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 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.

Thráisson, H. (1978). On the phonology of Icelandic prespiration. *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.