

Vowel duration and aspiration effects in Icelandic

Stefano Coretta

August 11, 2016

Contents

1 Literature review	1
1.1 Laryngeal	1
1.2 The effect of aspiration on vowel duration	1
1.3 Icelandic	2
2 Methodology	4
2.1 Participants	4
2.2 Materials	5
2.3 Procedure	5
2.4 Annotation	6
2.5 Measurements	8
2.6 Statistical analysis	9
3 Results	10
3.1 Word duration	10
3.2 Monosyllabic words	10
3.3 Disyllabic words	12
A Word list	14
B PRAAT script	16
Bibliography	17

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], [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, 47).

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

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

Table 1.2:
Major
con-
so-
nan-
tal
al-
lo-
phones
of
Ice-
landic.

1.3 Icelandic

Icelandic [GLOTTO: icel1247¹] 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: faro1244²], 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

¹Entry for “Icelandic” at Glottolog: <http://glottolog.org/resource/languoid/id/icel1247>.

²Entry for “Faroese” at Glottolog: <http://glottolog.org/resource/languoid/id/faro1244>.

in Icelandic. 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.

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 than 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.

Table 2.1: Information on participants

participant	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

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 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’ control over their speech.

The task was presented through the software PsychoPy (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

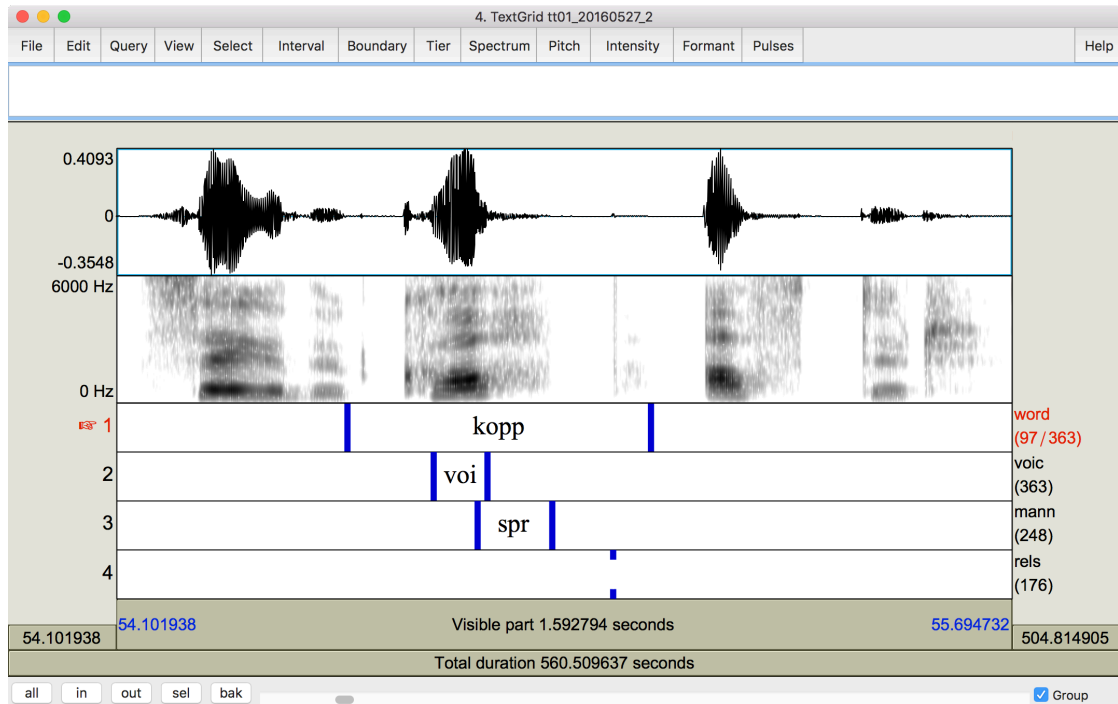


Figure 2.1: Example of the tier structure of the annotation files (PRAAT TextGrids).

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 Tier 1: words

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 Tier 2: voicing

The second tier was reserved for the portions of vocal folds vibration (voicing). The boundaries of the intervals in this tier were placed at the on-set and off-set of voicing around the target vowel. In words starting with one or more voiced continuant consonants, the portion of voicing 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.

2.4.3 Tier 3: glottal spreading

The third 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 vocalic is accompanied by modal voicing; the glottis starts an abduction gesture while still vibrating the vocal folds (breathy voice); vocal fold vibration stops and voiceless friction remains (at the glottis or at the oral cavity, depending on the 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 spreading. However, at times 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 Tier 3: nasals, laterals and rhotics

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 rhotics. I placed the right boundary of these intervals (nasal, lateral, rhotic) 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 Tier 4: consonant release

In tier 4, the release of the stop consonant following the target vowel was marked at the onset of the burst. If the burst was not visible in the waveform, no release was marked.

2.5 Measurements

In the third phase, I extracted the durational properties of the annotated intervals through an automated routine. The routine was run with a PRAAT script, specifically written for this study. The script with its documentation can be found in Appendix B. The output is a .csv file with the relevant measurements. The following measures were taken (all measures are in seconds):

- *Word duration*: the duration of the whole word as annotated on tier 1
- *Duration of voicing*: the duration of the voicing interval on tier 2
- *Duration of manner*: the duration of the manner interval on tier 3 (this can be glottal spreading, nasality, laterality or the rhoticity)
- *Duration of absolute voicing*: this was measured as the duration of the interval between the onset of voicing and the onset of the manner interval, in words containing a pre-aspirated geminate or a sonorant (nasal or liquid); in words with non-aspirated geminates, the duration of absolute voicing is the same as the duration of voicing
- *Closure duration*: the duration of the stop closure, calculated as the duration of the interval between the off-set of voicing or, if present, manner and the release

- *Absolute closure duration*: the duration of the interval between the on-set of the manner and the release; in words with non-aspirated geminates, it is the same as the closure duration

2.6 Statistical analysis

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

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 439.31 msec (SD = 78.66), 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).

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 Monosyllabic words

3.2.1 Geminate stops

The mean duration of the voiced portion in milliseconds was 98 in words ending in a non-aspirated geminate stop, while it was 84.97 if followed by a pre-aspirated geminate stop. Figure 3.2 shows the normalised ratio of the voiced portion. The difference in the mean of the normalised voiced ratio between the non-aspirated and the aspirated class of words was significant [].

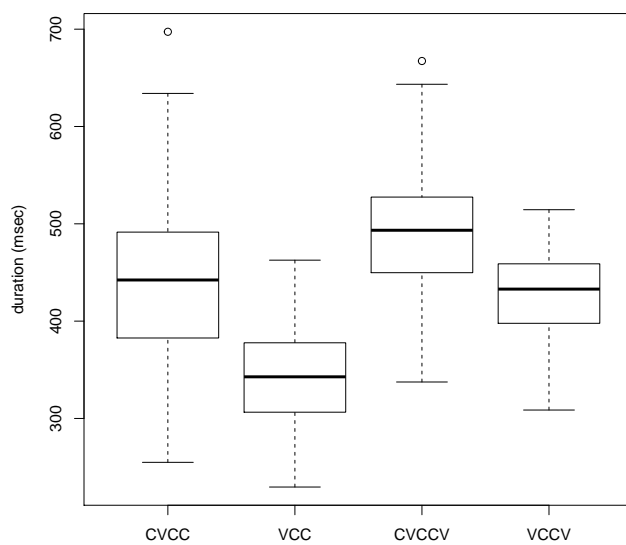


Figure 3.1: Duration (in msec) of CVCC, VCC, CVCCV, and VCCV words.

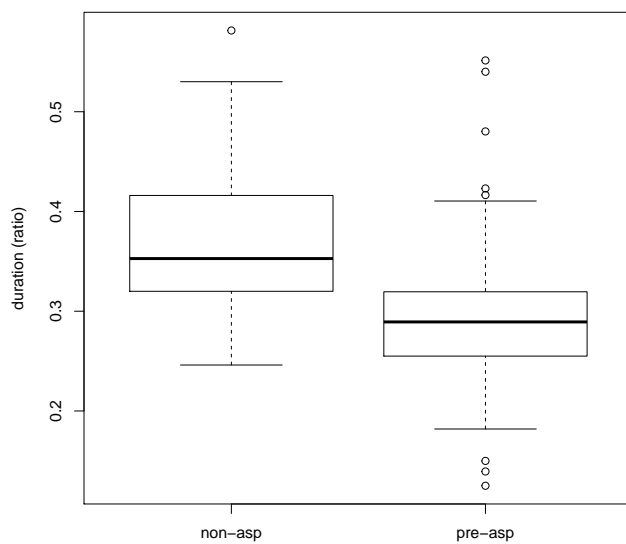


Figure 3.2: Duration (in ratios) of the voiced vocalic portion in geminate-final words.

3.2.2 Nasals

Words of the form (C)VNC showed a mean voiced portion duration of 98 msec if the nasal is voiced and 84.97 msec if voiceless. The normalised duration was, respectively, 0.37 and 0.3.

3.2.3 Laterals

In the case of words ending in a /lC/ cluster, the mean duration of the vocalic portion were 89.96 and 77.73 msec when followed, respectively by a voiced and voiceless lateral. These correspond to the 30.99% and the 26.44% of the VOR. A Wilcoxon test shows that the difference is significant [1].

3.3 Disyllabic words

3.3.1 Geminate stops

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

CVNVCV words have a mean voiced vocalic portion duration of 86.97 msec if the nasal is voiced and 78.04 msec if it is voiceless. As percentages, the vocalic portion is 33.61% of the VOR if followed by a voiced nasal, and 28.27% when the nasal is voiceless. A *t*-test showed that this difference is significant.

3.3.3 Laterals

The voiced vocalic portion in CVLVCV words is on average 86.97 msec long when the lateral is voiced and 78.04 msec when it is voiceless. The voiced portion takes the 33.61% of the VOR if followed by a voiced lateral, and 28.27% when the lateral is voiceless.

3.3.4 Rhotics

With words containing an /rC/ cluster, the mean duration of the voiced vocalic portion is 86.97 msec long when the rhotic is voiced and 78.04 msec when it is voiceless. The voiced portion takes

the 33.61% of the VOR if followed by a voiced rhotic, and 28.27% when the rhotic is voiceless.

Appendix A

Word list

Table A.1: List of target words

word	IPA	word	IPA
kokk	k ^h o ^h k	kembt	kem̩t
gogg	kokk	kembdi	kem̩tɪ
dökk	tø ^h k	kampa	k ^h am̩pa
dögg	tøkk	kamba	k ^h ampa
kopp	k ^h o ^h p	kempa	k ^h em̩pa
kubb	k ^h ypp	kemba	k ^h empa
vítt	vi ^h t	punta	p ^h ynta
vídd	vitt	punda	p ^h ynta
þítt	θi ^h t	vanta	van̩ta
þíddi	θittɪ	vanda	vanta
fætt	fai ^h t	fínn	fin̩
fæddi	faittɪ	kinn	k ^h inn
ýtt	i ^h t	duld	tylt
ydd	itt	dult	tylt
ótt	ou ^h t	gelta	kel̩ta
odd	ott	gelda	kelta
sets	sess	orka	or̩ka
sett	se ^h t	orga	orka
feits	feiss	mjólka	mjou̩lka
feitt	fei ^h t	ólga	oulka
vots	voss	hefna	hepna
vott	vo ^h t	vopna	vo ^h pna
takka	t ^h a ^h ka	nafla	napla
kagga	k ^h akka	japla	ja ^h pla
detta	te ^h ta	kafli	kaplɪ
gedda	ketta	kapli	ka ^h plɪ
kamp	k ^h am̩p	tefla	tepla
kamb	k ^h am̩p	tipla	ti ^h pla
punt	p ^h ynt		
pund	p ^h ynt		

Appendix B

PRAAT script

Bibliography

- Árnason, K. (2011). *The phonology of Icelandic and Faroese*. Oxford: Oxford University Press.
- 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].
- 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.
- 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.
- 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].
- Thráisson, H. (1978). On the phonology of Icelandic prespiration. *Nordic Journal of Linguistics* 1(01), 3–54.