Aitken’s Law Revised

The Scottish Vowel Length Rule in 21st century Scottish Standard English

by

Andreas J. Weilinghoff

from Heek in Westphalia

First supervisor: Prof. Dr. Sarah Buschfeld

Second supervisor: Prof. Dr. Edgar W. Schneider

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Dr. phil.)

at the Department of Cultural Studies

of TU Dortmund University

Dortmund, 21st December 2022

“Ideally, a full account – a task for a future PhD at least – would survey all the available data (...) and display the results of copious studies of tape-recordings and instrumental tests of vowel durations.”

(Adam) Jack Aitken in 1981

Table of Contents

List of Abbreviations

C	Consonant segment
CS	Compensatory shortening
EC	East central
F1	First formant
F2	Second formant
GA	General American
HHE	Highland and Hebridean English
HTK	Hidden Markov Model Toolkit
IN	Insular
LAS	Linguistic Atlas of Scotland
LLL	Low-level lengthening
MAUS	Munich Automatic Segmentation System
MFA	Montreal Forced Aligner
ms	milliseconds
NO	Northeast
RP	Received Pronunciation
SAMPA	Speech Assessment Methods Phonetic Alphabet
SCOTS	Scottish Corpus of Texts and Speech
SO	Southern
SSBE	Southern Standard British English
SSE	Scottish Standard English
SVLR	Scottish Vowel Length Rule
V	Vowel segment
VE	Voicing Effect
VOT	Vowel onset time
WC	West central

List of other symbols

< … >	spelling/letters/orthography
/ … /	phonemes
[…]	phonetic variants
#	morpheme boundary
italics	specific terms and variables
UPPERCASE	indication of STRESS or indication of lexical set (e.g. FLEECE)

List of Figures

List of Tables

Acknowledgments

This project would not have been possible without the help of many people. First, I have to sincerely thank my supervisor, Sarah Buschfeld, for her kindness, enthusiasm and unwavering support. She pushed me to go further and always offered orientation and advice when I was stuck or unsure. Sarah is a great example that, apart from hard work and dedication, academic success and fulfillment depend on passion, honesty, openness, cooperation and goodwill. There could not have been a better supervisor, mentor, boss and friend along this path. Second, I have to thank Edgar Schneider for his kindness to become my second supervisor despite his many ongoing obligations around the world. I also have to thank the Chairs of English Linguistics at TU Dortmund University and all affiliated and former staff for the supportive and unbelievably friendly working environment, in particular Patricia Ronan, Claus Weihs (especially for his help with statistical matters), Emily Weidle, Brian Hess, Theresa Neumaier, Lisa Westermayer, Julian Roth, Manuela Vida-Mannl, Barbara Hamblock, Stefanie Schröter, Niklas Nitsch, Guyanne Wilson, Maggie Glass, Kieran Harrington and my office buddies Sheena Shah, Wiebke Ahlers and Lotta Christiansen.

Another great influence was Jane Stuart-Smith who invited me to visit the University of Glasgow as an affiliated researcher in the spring of 2020. Jane's invaluable knowledge on Scottish phonetics and phonology, her enthusiasm, friendliness and academic rigour pushed me and my project further in a way that I would not have expected. The research stay in Scotland is an experience that I will not forget and it was definitely one of the most important events in my PhD journey. In this context, I also have to thank the whole GULP team and the staff at the University of Glasgow, in particular Rachel Macdonald, Jennifer Smith, Edward Marshall, Joe Pierce and also the then-affiliated PhD student Florent Chevalier. I would also like to deeply thank Madeleine, Andrew, Iona and Finlay Craig, Judith Gosling with Thea and James as well as Ann and Maurice Walker for accommodating the wee German laddie in the UK.

I also feel a deep sense of gratitude for Ulrike Gut and Ole Schützler. The seminar "English in Scotland" by Ulrike Gut and my involvement in the ICE Scotland project has been the reason why I got interested in Scottish English and in the research fields of phonetics and phonology. ICE Scotland has been one of the cornerstones which made this project possible.

Naturally, I am most grateful for all participants in my study and I have to thank Donna MacDonald, Elaine and Alan Barton in particular for their friendliness and support. I also need to thank Lauren and Stephen from Edinburgh University Library as well as all the staff at the local libraries of Hawick, Oban, Girvan, Hillhead, East Kilbride, Greenhills and Edinburgh Central Library. Special thanks also to BBC Radio Shetland, BBC Radio Orkney, Radio Borders and to the Scottish Parliament.

Such a project would not be possible without the backing of family and friends. I have to thank my parents Josef and Hildegard, my late grandparents, especially Änne, and my brothers Thomas and Markus for the uncomplicated support over the years. I know it is hard to understand what I am doing in my PhD project, but the work ethic and dedication you taught me have definitely helped in this endeavor. Thanks also to all my friends on whom I can always rely. A special cheers to Hendrik

Schmülling for keeping me focused on my project and making me keep my (self-imposed) deadlines while also distracting me on some weekends. Another special thanks to Christopher Lösbrock for introducing me to the great world of Python.

As for the errors and weaknesses that may still linger in this thesis, only I am to blame.

Dortmund, in the winter of 2022

1. Introduction

The duration of vocalic intervals varies and is influenced by many different factors. In linguistic research on Scottish English vowel duration, the Scottish Vowel Length Rule (henceforth: SVLR) has clearly received the most attention (Warren, 2018, p. 45). An early definition of the SVLR states that "all vowels and diphthongs are long in stressed open syllables, before voiced fricatives and /r/, and before morpheme boundaries; and short elsewhere; with the two exceptions of /ɪ/ and /ʌ/ which are invariably short" (McClure, 1977, p. 10). Hence, in accordance with the SVLR, words such as <bee>, <beer>, <breeze> and <brewed> would be realized with long allophones in Scottish English. The SVLR partially contradicts with the standard quantity alternation pattern present in most varieties of English and many other languages, namely the Voicing Effect (henceforth: VE) (Chen, 1970). Whereas the VE generally triggers longer vowel realizations before all voiced consonants, the SVLR conditions a shortening of vowel allophones before following laterals, nasals and voiced oral stops. Thus, whereas the VE conditions long allophones in words such as <bead>, <bean> and <feel>, the same words would be realized with shorter vowels in Scottish English due to the timing effects of the SVLR. The distinction between short and long SVLR environments leads to quasi-phonemic contrasts between words such as <crude> and <crewed> in Scottish English (Scobbie & Stuart-Smith, 2008). Whereas the vocalic nucleus in <crude> is followed by a tautomorphemic /d/ representing a short SVLR context, the same vowel in <crewed> is followed by heteromorphemic /#d/, which conditions a longer allophone.

The Scottish lexicographer Adam Jack Aitken first formulated the SVLR in 1962, which is the reason why it has also come to be known as *Aitken's Law* (Aitken, 1981, p. 156). Aitken claimed that "all varieties of Scottish speech, from the fullest vernacular to Educated Scottish Standard English, operate in some measure the Scottish Vowel-Length Rule (SVLR) (…)" (Aitken, 1984b, p. 94). Yet, he also acknowledged that his most comprehensive account of the rule published in 1981 is not fully complete and bears uncertainties (Aitken, 1981, p. 131). His description of the SVLR is largely based on impressionistic observations and he could only partially support it with empirical evidence. Aitken (1981, 135) acknowledged that, back at the time, there was a lack of instrumental measurements of Scottish vowel duration.

As a result, many subsequent studies were conducted on the SVLR investigating vowel duration patterns in different dialect regions of Scotland and with different methodological approaches. Whereas some studies could generally corroborate Aitken's description, others challenged the far-reaching validity of the rule and one study even suggested that SVLR timing effects might not be a specific Scottish phenomenon at all (Agutter, 1988a). Most studies conclude that Aitken's Law is only detectable in the high vowels /i/ and /u/ as well as in the diphthong /aɪ/ and therefore propose a much more limited SVLR vowel set than Aitken. There is also contradictory evidence on the VE in Scottish English: some studies found evidence that Aitken's Law operates alongside the VE, others could, however, not find significant VE-related lengthening effects in their analyses. There were also some studies that reported age- and gender-related variation in the application of the SVLR, which was, however, not significantly measurable in most other investigations. Previous investigations also disagree whether Aitken's Law is influenced by internal or external factors. Many researchers suggested that SVLR timing effects are weakened by contact or exposure to Southern Standard British English (henceforth: SSBE). This weakening of Aitken's Law among speakers with high SSBE contact is then often accompanied with a shift towards VE vowel timing patterns. Yet, more recent investigations in Glasgow (Chevalier, 2019; Rathcke & Stuart-Smith, 2016) conclude that this change is not promoted by dialect contact but mainly by language-internal, prosodic factors.

As for methodology, almost all earlier accounts on Scottish vowel durational patterns are based on impressionistic reports including Aitken's most detailed description of the rule published in 1981. It was only at the end of the 1970s that the first empirical studies were carried out on Scottish vowel duration. Yet, most of them had unbalanced samples in terms of the speakers' age and gender and analyzed SVLR patterns in carefully controlled experimental settings using word list or carrier sentence readings. Only the most recent studies by Chevalier (2019), Rathcke and Stuart-Smith (2016), Stuart-Smith et al. (2019) and Warren (2018) investigated Scottish vowel length patterns in naturally occurring language and implemented inferential statistics in their analyses. Yet, almost all of their data is based on Scots and does therefore not reflect the situation of Aitken's Law in contemporary Scottish Standard English (henceforth: SSE). Another shortcoming of most previous studies is their narrow geographical and phonological scope since they investigated only a limited number of vowels and environments in a particular region or city. Most studies were conducted in the Central Belt of Scotland and a few other studies focused on the Northern Isles as well as on the Northeast of Scotland and England. The vowel timing effects in other regions of Scotland, for instance in the Highlands and in the South of Scotland, were not investigated. Even the only cross-dialectal study by Stuart-Smith et al. (2019) is largely based on Central and Northeastern spoken Scots data.

Whereas Aitken (1981) proposed that almost all vowels are affected by the timing patterns of the SVLR, the more recent studies have shown a more complicated picture and proposed a much smaller vowel set. In addition, the influence of regional as well as age- and gender-related variation in Aitken's Law remains largely unresolved, especially in spoken SSE. There is also uncertainty about the operation of the VE in SSE. The present research project addresses these desiderata and sets out a new approach towards analyzing vowel duration in SSE. Unlike most previous studies, I aim to achieve representativeness for the whole of the country in terms of regional background, age and gender of the analyzed speakers and find answers for the following three research questions:

(1) Which vowels are affected by Aitken's Law / the VE in 21st century spoken SSE?
(2) What is the effect of regional, age- and gender-related variation on Aitken's Law / the VE in 21st century spoken SSE?
(3) Which prosodic factors have an influence on Aitken's Law / the VE in 21st century spoken SSE?

To answer these research questions, I will use a large and balanced dataset incorporating speech from male and female speakers of all age groups and from all regions of Scotland. I will also exclusively use speech from the 21st century to ensure the up-to-datedness of the dataset. In addition, the present investigation aims to analyze the different vowels of the Basic Scottish Vowel System (Abercrombie, 1979) in all possible phonetic contexts, an undertaking which has not yet been carried out before. Furthermore, I intend to analyze SVLR patterns in naturally occurring language and account for idiolectal variation as well as all for suprasegmental and segmental factors that are known to influence vowel duration in Scottish English, namely *tempo* (see subsection 3.1.9), *constituent-final lengthening* (see subsection 3.1.8), *prosodic* and *lexical stress* (see subsections 3.1.6 and 3.1.7), *word frequency* (see subsection 3.1.5), the number of syllables in a word (see subsection 3.1.4), the number of segments in a syllable (see subsection 3.1.3) and *intrinsic vowel duration* (see subsection 3.1.1). While there are some recent studies on Aitken's Law in spontaneously spoken Scots (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019), the state of vocalic durational patterns remains largely unclear in spoken SSE. There are, of course, studies which investigated Aitken's Law in formal Scottish English speech (Agutter, 1988b; McClure, 1977; McKenna, 1988), yet all of them elicited speech with word list or carrier sentence readings. Thus, this study will be, to the best of my knowledge, the first investigation of Aitken's Law in naturally spoken SSE on a countrywide scale. While the focus is on naturally spoken SSE, I will also distinguish between more scripted speech forms, such as found in public broadcasts and speeches and more unscripted language as found in discussions and podcasts. Whereas most previous empirical studies on vowel timing in Scottish English used descriptive statistics in their analyses, the present investigation will also implement inferential analytics as part of its methodological approach.

In chapter 2, I will briefly outline the current language situation of Scotland and describe the history and current status of Scottish Gaelic (section 2.1), Scots (section 2.2) and SSE (section 2.3). While this study analyzes SSE data, it is nonetheless important to get a broad overview of the other languages, since they can have an influence on the standard variety. Especially Scots and SSE share an intricate relationship and this connection will be discussed further in section 2.4.

Chapter 3 provides a detailed overview of previous research on vowel duration patterns. Section 3.1 summarizes the main findings of earlier studies on the factors that generally influence vowel duration in English. Section 3.2 subsequently deals with the main phenomenon under investigation, Aitken's Law. I will not only discuss the rule but also outline the history of SVLR research in the last centuries. That is, subsection 3.2.1 provides an overview of all earlier impressionistic accounts on vowel duration patterns in Scottish English. Subsection 3.2.2 then summarizes all the empirical studies which investigated Scottish vowel duration in controlled speech. The latest SVLR studies in uncontrolled speech are summarized in subsection 3.2.3.

The datasets and methodology used for answering the research questions are introduced in chapter 4. I will first give a detailed overview of the data selection and transcription criteria which guided the data collection procedure (section 4.1). I will then describe the datasets that are used in the present investigation (section 4.2). The sample includes corpus data from ICE Scotland (Schützler et al., 2017) (see subsection 4.2.1), but also self-collected SSE speech data which was retrieved in Scotland (see subsection 4.2.2) and from different online sources (see subsection 4.2.3). A major undertaking of this project was to transcribe and transform the speech data into a format that is suitable for the subsequent analysis. The data preparation procedure will therefore be explained in detail in section 4.3. As the sample includes both pre-existing speech data (ICE Scotland) and self-collected datasets, the preparation process differs accordingly: the data preparation procedure for ICE Scotland will be discussed in subsection 4.3.1 and the procedure for the self-collected data will be outlined in subsection 4.3.2. Section 4.4 will subsequently describe the data analysis, which includes information about the vowel and variable selection (see subsections 4.4.1 and 4.4.2) as well as an outline of the statistical analysis (see subsection 4.4.3).

The results of the present investigation are outlined in chapter 5. I will provide a first overview of the overall vocalic durations in section 5.1 and then discuss the short monophthongs (section 5.2), long monophthongs (section 5.3) and diphthongs of contemporary SSE individually.

The findings will then be discussed against the background of the three research questions in chapter 6. Section 6.1 deals with the first research question and evaluates which vowels are influenced by Aitken's Law or the VE in 21st century SSE. The discussion of the influence of sociolinguistic variation on SSE vowel duration patterns follows in section 6.2 and the influence of prosodic factors will be discussed in section 6.3.

A general summary of and a conclusion for the present investigation will be given in chapter 7. I will also briefly outline the strengths and weaknesses of this project as well as directions for further research.

2. The languages of Scotland

In the context of the present investigation, it is vitally important to get a brief overview of the language situation of Scotland with its three indigenous languages Scottish Gaelic, Scots and English. The history of Scotland is intertwined with each language and also the minority languages Scottish Gaelic and Scots still have an influence in Scotland today. Section 2.1 will provide a brief overview of the history and the current status of Scottish Gaelic in Scotland. In addition, I will describe one feature of Scottish Gaelic that may be relevant for the present investigation, namely that vowel length is phonemic in Scottish Gaelic. Section 2.2 will provide a brief overview of the history and current status of Scots. Section 2.3 then introduces the variety under investigation in the present study: SSE. I will discuss the linguistic structure of this standard variety in more detail and I will lay a special focus on the Basic Scottish Vowel System (Abercrombie, 1979) which will serve as the main reference phoneme inventory of the present investigation. Section 2.4 then discusses the complex relationship between Scots and SSE. While Norn was also once spoken in Shetland, Orkney and Caithness, it became extinct in the late 18th century (Price, 1984, p. 203; van Leyden, 2002, p. 1). Norn had an influence on Scots in the Northern Isles, but the language will not be discussed here as the present study investigates SSE and not Scots.

2.1 Scottish Gaelic

Scottish Gaelic descended from the Gaelic branch of the Celtic languages. According to the traditional historical view, was introduced to the Southwest of Scotland at around 500 AD (Ó Baoill, 2011, pp. 1–3). While there is debate about whether this has been a result of Irish invasion or not (Campbell, 2001), it is relatively clear that Scottish Gaelic spread from Argyle to the North and East of modern-day Scotland and became the predominant language of the country. The vast number and spread of Gaelic place names bear witness to this (Cox, 2011). Scottish Gaelic place names can be found from Aberdeen in the Northeast to the island of St. Kilda in the far West, from Caithness in the North to the Scottish Borders in the South. The only exceptions are the Southeastern area of the Borders and the most northernly part of Caithness as well as the islands of Orkney and Shetland (Cox, 2011, p. 46). Whereas the Northern Isles, Northern Caithness and the Western seaboard were long under Scandinavian influence, the Southeast of Scotland was invaded by the Anglians in the seventh century AD (Stuart-Smith, 2008, p. 49). The role of Scottish Gaelic declined in particular after the 11th century when the language of the Angles, ‘Inglis’, the ancestor of Modern Scots (Jones, 2002, p. 95), spread northwards along the east coast due to major sociopolitical changes (Ó Baoill, 2011, pp. 10–11). Scottish Gaelic was largely superseded by the Germanic language in the Lowlands and ceded to the Northwest of the country. Up until the 18th century, Scotland’s linguistic map could roughly be divided by the Highland line (see Figure 1) into a predominantly Scottish Gaelic speaking population in the Highlands (and some western parts of Galloway) and a predominantly Scots speaking population in the Scottish Lowlands (Ó Baoill, 2011, p. 16). This has also been called the “Highland/Lowland divide” representing both a linguistic and geographic separation (Ó Baoill, 1997, p. 559). Yet, the status of Scottish Gaelic further declined after the 18th century and it was generally replaced by English, not by Scots, in the Highlands (Ó Baoill, 2011, p. 16). This replacement also gave rise to the dialect classification of Highland and Hebridean English (henceforth: HHE) for the areas West of the Highland line (Ó Baoill, 1997, p. 566). While Scottish Gaelic has been replaced by English in the Highlands, previous accounts nevertheless note that Aitken’s Law is also a feature of HHE (Maguire, 2012, p. 57) and this is why this region will also be investigated in the present study. The present study will therefore implement the traditional Scottish Gaelic speaking regions as one dialect area in its analysis (see subsection 4.1.1).

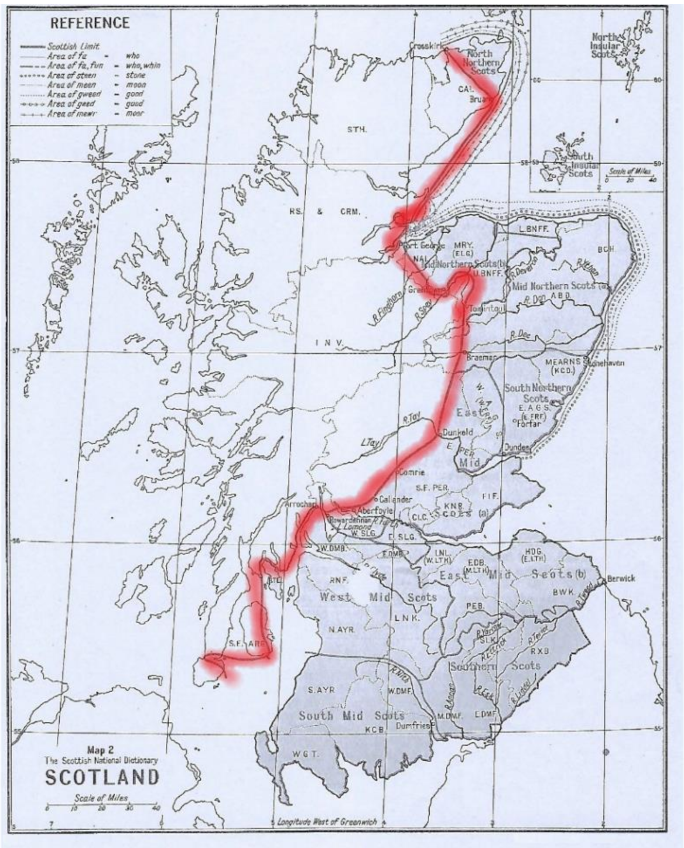


Figure 1. Map of the Scottish National Dictionary (Grant, 1931) with the Highland line highlighted in red.

Today, 87.100 people aged 3 and over in Scotland have Gaelic language skills. This equals 1.7 percent of the total population (National Records of Scotland, 2015, p. 6). The largest numbers of Scottish Gaelic speakers can be found West of the historic Highland line, more precisely in the council areas of Eilean Siar (61 percent), Highland (7 percent) and Argyll & Bute (6 percent) and generally, these speakers are bilingual in Gaelic and English (Jones, 2002, p. 2). Even though the number of Scottish Gaelic speakers continuously declined over the last decades, recent years have seen a slight increase in the number of speakers among the younger population (National Records of Scotland, 2015,

p. 11). This new tendency is a result of the political promotion of the language. One of the main political cornerstones for the protection of Scottish Gaelic was the Gaelic Language (Scotland) Act, which was unanimously passed by the Scottish Parliament in 2005. This led to the implementation of a Gaelic language board as well as to the Gaelic Language Plan, the Scottish Government's official strategy paper on how to protect and foster the Scottish Gaelic language. Recent years have also seen the opening of the first secondary schools with Scottish Gaelic as the sole medium of education as well as the launch of the Gaelic television channel BBC Alba. Thus, Scottish Gaelic still has an influence in Scotland even if its speaker numbers are relatively low. Furthermore, after centuries of decline, Scottish Gaelic seems to gain importance again due to the extensive revitalization efforts.

An important aspect, especially in the context of the present study, is that vowel length is phonemic in Scottish Gaelic (Nance, 2011). There is a triple length distinction between short, long and over-long vocalic intervals as in [tuɫ] <to go>, [uːɫ] <apple> and [suːːɫ] <eye> (Laver, 1994, p. 442). This stands in contrast to English where the distinction between long and short monophthongs is characterized by a combination of quality and quantity changes. For example, the vocalic nucleus /i:/ in <feet> is not only longer than the vowel /ɪ/ in the word <fit>, but it also differs in terms of quality due to a higher and fronter realization. The short monophthong /ɪ/ is much more centralized than /i:/ in English. Both words <feet> and <fit> do therefore constitute a minimal pair. In contrast to this, the difference between long and short monophthongs in Scottish English can be expressed solely by quantity and is not accompanied by significant quality changes. This phonemic vowel length contrast in Scottish Gaelic will be taken into consideration, especially as the present study explicitly investigates the duration of vocalic intervals.

2.2 Scots

Scots derived from the Northumbrian dialect of Old English and was introduced to the Southeast of Scotland by the Angles of Bernicia in the seventh century AD (Aitken, 1984a, p. 517; McClure, 1994, p. 23). The influence of the Germanic language grew especially in the aftermath of the Norman conquest in Britain (McClure, 1994, p. 27). Charles Jones (2002, p. 94) notes two driving political factors in this development: "(a) the policy of David I to establish burghs across Scotland peopled largely by Scots speakers and (b) the introduction of English-speaking tenants by the Norman nobility". As a result, Scots gradually superseded Scottish Gaelic in the Lowlands and became the language of administration, government and written communication and was eventually adopted by the Scottish parliament (Jones, 2002, p. 94). Scots spread from Lothian to the Southeast and the Western Central Belt and later also to the Northeast of Scotland (Aitken, 1984a, p. 518). In the seventeenth and eighteenth centuries, Scots also superseded Norse in Caithness, Orkney and Shetland and it was also brought to Northern Ireland due to the plantation of Ulster. Due to the long period of Scottish independence from the Kingdom of England, Scots is often seen as "the only Germanic variety in the British Isles besides Standard English ever to have functioned as a full language within an independent state (the Kingdom of Scotland)"

(Johnston, 2007, p. 105). The language developed different dialects and also “underwent the early stages of standardization about the same time as English did (…)” (Johnston, 2007, p. 105). Scots also developed a great corpus of literature which is comparable in its diversity to other Western European national languages (Johnston, 2007, p. 105). Yet, the role of literary Scots gradually declined after the eighteenth century and gave way to Standard Southern English (Stuart-Smith, 2008, p. 49). The eighteenth century also saw the emergence of SSE among the higher social classes. Spoken Scots, however, remained dominant in rural areas and working class contexts (Stuart-Smith, 2008, p. 49).

Today, 1.5 million people in Scotland can speak Scots and another 267.000 people claim that they could understand but not read, write or speak the language (Scotland's Census, 2011). Together, this represents around 33 percent of the total Scottish population. The largest proportions of Scots speakers are found in Aberdeenshire, Moray as well as on Orkney and Shetland in the Northeast of Scotland. There are also larger proportions of Urban Scots spoken in the Greater Glasgow area. Similar to Scottish Gaelic, the Scottish Government also supports the use of the Scots language in all possible contexts aiming to “enhance the status of Scots in Scottish public and community life” and promoting its “acquisition, use and development (…) in education, media, publishing and the arts” (The Scottish Government, 2015, p. 4).

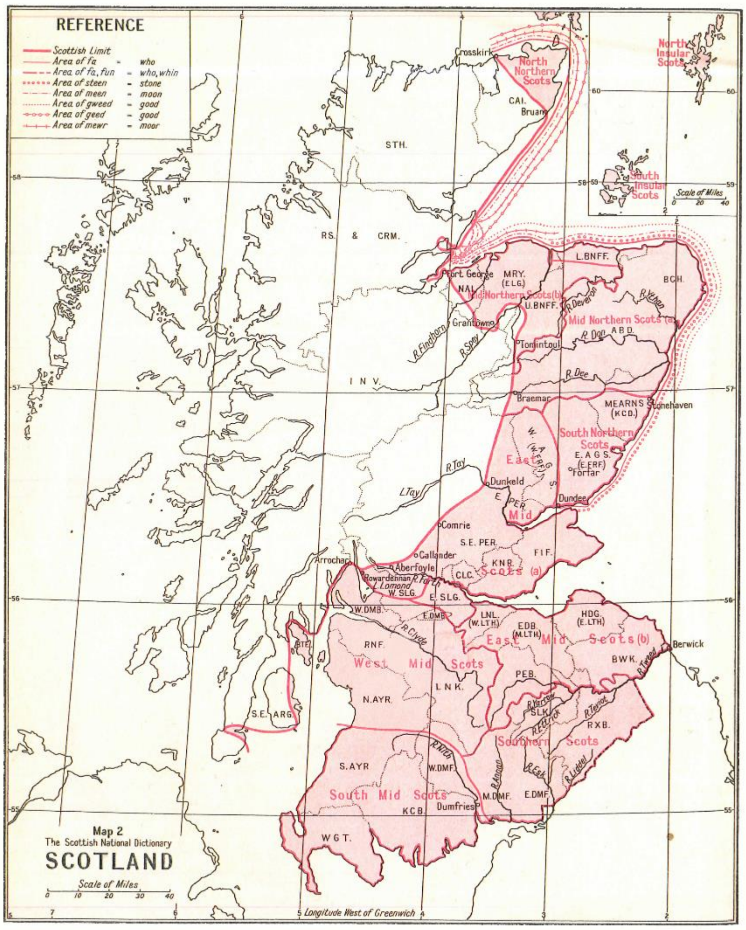


Figure 2. Dialect classification by Grant (1931) in the Scottish National Dictionary.

As for regional variation, Grant’s (1931) widely accepted and still relevant classification (Johnston, 1997, p. 437) differentiates between the overall dialect groups of Insular Scots, Northern Scots, Mid Scots and Southern Scots (see Figure 2). These dialect groups can be further subdivided into several subgroups. Insular Scots comprises the originally Norn-speaking islands of Shetland and Orkney and Southern Scots is traditionally associated with the Western and Southern Borders region (Johnston, 1997, p. 437). Northern Scots comprises the dialects of the Scottish mainland North of the Tay valley as well as Caithness (North Northern Scots) and the Black Isle. Mid-Scots covers the areas of the Central Belt where most of the Scottish population lives. Here, one can make a further threefold subdivision between West-Mid, East-Mid and South-Mid Scots. West-Mid Scots includes the areas of Glasgow and the Clydesdale and East-Mid includes Edinburgh and Lothian as well as Fife and the Eastern Borders. South-Mid is associated with the historic counties of Kirkcudbrightshire and Wigtownshire in the Southwest of Scotland. As Scots stands in a close relationship with SSE (see section 2.4), I will use Grant’s (1931) description of regional Scots variation as the main reference for the dialect area classification of this study. More information on the different dialect areas in this investigation can be found in subsection 4.1.1.

2.3 Scottish Standard English

The English language became more important in Scotland after the unification of the crowns and parliaments merging the realms of Scotland and England into the United Kingdom of Great Britain. At first, the anglicization of Scots merely concerned the written language and did not affect the spoken tongue of the masses of the population in Scotland (McClure, 1994, p. 37). However, with the departure of James VI to the English court in London, the Scottish aristocrats increasingly adapted their language towards that of England as they were spending more time in the Southern capital and many of them intermarried with the English aristocracy (McClure, 1994, p. 38). This can be interpreted as the start of SSE and over the centuries, the language of the higher social classes in Scotland became increasingly anglicized. As a result, SSE became the variety used in formal contexts and across higher social classes.

Today, almost all inhabitants of Scotland are potential speakers of SSE (Stuart-Smith, 2008, p. 49). Scotland's Census (2011) reports that 98.6 % of people in Scotland aged 3 and over speak English. English is also by far the most widely spoken language at home equaling 92.6% of the Scottish population (Scotland's Census, 2011).

SSE is generally defined as “the variety of Standard English spoken in Scotland, [which] has few lexical and syntactic characteristics that set it apart from the Standard English used in England” (Giegerich, 1992, pp. 45–46). Thus, it is assumed that, with regard to its lexis and grammar, SSE is roughly equal to SSBE. In terms of phonology, however, SSE is very similar to Scots (Johnston, 2007, pp. 112–113). In contrast to Received Pronunciation (henceforth: RP) and in accordance with Scots, SSE incorporates the voiceless velar fricative /x/ in its consonant inventory (Abercrombie, 1979, p. 71) but it only occurs in coda position due to phonotactic constraints (Scobbie & Stuart-Smith, 2008, p. 94). It can be found in words such as <loch> which would be pronounced /lɒx/ in SSE but mostly /lɒk/ in RP. Another specific consonant is the voiceless labio-velar fricative /ʍ/ which can still be found among SSE speakers leading to minimal pairs such as <which> and <witch> (Scobbie & Stuart-Smith, 2008, pp. 95–96). Another major feature of SSE is its variable rhoticity. Whereas earlier accounts have described SSE as being generally rhotic (Stuart-Smith, 2008, p. 64), newer corpus-based studies on SSE found that only half of all non-linking coda /r/ are realized (Meer et al., 2021). There are, however, many more systemic differences between the vowel inventories of SSE and RP as specified in Abercrombie's (1979) widely cited Basic Scottish Vowel System (see Table 1).

Table 1. The Basic Scottish Vowel System and the RP equivalents on the basis of Abercrombie (1979, p. 72) with reference to the lexical sets by Wells (1982).

SSE	RP	Example word by Abercrombie (1979)	Lexical set by Wells (1982)
/i/	/i/	<bead>	FLEECE
/ɪ/	/ɪ/	<bid>	KIT
/e/	/eɪ/	<bay>	FACE
/ɛ/	/ɛ/	<bed>	DRESS
/ɛ̈/		<never>	DRESS
/a/	/æ/	<bad>	TRAP
	/ɑ/	<balm>	BATH
/ɔ/	/ɒ/	<not>	LOT
	/ɔ/	<nought>	THOUGHT
/o/	/əʊ/	<no>	GOAT
/u/	/ʊ/	<pull>	FOOT
	/u/	<pool>	GOOSE
/ʌ/	/ʌ/	<bud>	STRUT
/ʌɪ/	/aɪ/	<side>	PRICE
/ae/		<sighed>	PRICE
/ʌʊ/	/aʊ/	<now>	MOUTH
/ɔe/	/ɔɪ/	<boy>	CHOICE

According to Abercrombie (1979), the phonemes /i/, /ɪ/ and /ʌ/ can be found in both SSE and RP. Apart from that, SSE lacks many tense/lax pairs which are common in RP and many other varieties of English (Scobbie & Stuart-Smith, 2008, p. 97). For example, whereas there is a distinction between /æ/ and /ɑ/ in words such as <bad> and <balm> in RP, SSE includes the uniform vowel /a/ in both contexts. This is generally accepted in the literature even though there may be social and allophonic variation (Wells, 1982, p. 403), especially among many Edinburgh SSE speakers (Abercrombie, 1979, p. 75). Likewise, whereas RP distinguishes between the shorter and open vowel /ɒ/ and the often longer open-mid vowel /ɔ/, SSE incorporates only the latter vowel in its phoneme inventory. Another, though less common difference is found among the back vowels (Abercrombie, 1979, p. 76): Whereas RP distinguishes between the shorter and more centralized phoneme /ʊ/ in <pull> and the longer phoneme /u/ in <pool>, SSE lacks the former short monophthong. It must be added, however, that /u/ is generally fronted in Scottish English and therefore often transcribed with the phoneme symbol /ʉ/ (Wells, 1982, p. 402). In contrast to these mergers, SSE includes a split between /ɛ/ and /ɛ̈/ in the lexical set DRESS where RP has the uniform vowel /ɛ/. The more centralized vowel /ɛ̈/ can be found among some speakers in the West of Scotland, the Borders as well as in Perthshire and Edinburgh (Abercrombie, 1979, p. 75). Another example is the RP diphthong /aɪ/ of the lexical set PRICE where SSE includes the two diphthongs /ʌɪ/ and /ae/. These diphthongs are perceptibly distinct and their realization depends on the morphological composition of the word (Wells, 1982, p. 405): Whereas monomorphemic words, such as <tide> and <side>, are often pronounced with the shorter diphthong /ʌɪ/, heteromorphemic words, such as <tied> and <sighed>, are pronounced with the longer diphthong /ae/. This length variation directly corresponds to the morphological conditioning of Aitken's Law and constitutes a "quasi-phonemic contrast" in words such as <side> and <sighed> (Scobbie & Stuart-Smith, 2008). These words are therefore often considered minimal pairs in Scottish English (Wells, 1982, p. 405). Apart from that, the lexical sets FACE and GOAT are monophthongal in SSE. Hence, words such as <place> or <boat> are usually pronounced /ples/ and /bot/ in SSE and not with the diphthongs /eɪ/ or /əʊ/ as in RP. The other SSE diphthongs are also different in quality when compared to the equivalent RP diphthongs. In the lexical set MOUTH, SSE uses a slightly raised onset /ʌʊ/ as opposed to RP /aʊ/ or the equivalent Scots vowel /u/ (Wells, 1982, p. 406). The lexical set CHOICE is also realized in different ways in SSE (Wells, 1982, pp. 406–407). Abercrombie (1979, p. 72) notes a more centralized offset in this diphthong so that the word <boy> is realized as /bɔe/ in SSE.

Abercrombie (1979, p. 73) notes that the Basic Scottish Vowel System is arguably the most common vowel inventory among SSE speakers. There is, of course, a lot of variation in Scottish English vowels overall; however, this variation can be best described against the background of the basic system. This means that the Basic Scottish Vowel System functions as a general reference system and other SSE vowel systems "are best described in terms of departures from it (…)" (Abercrombie, 1979, p. 74). Abercrombie's (1979) Basic Scottish Vowel System will therefore be used as a phonemic reference structure in the present study.

2.4 The Scots-English language continuum

The terms *Scots English* (MacArthur, 1979, p. 51) or *Scottish English* are often used as cover terms encompassing the varieties of Scots and SSE (Schützler, 2015, p. 1). As Scots and SSE are closely related and share a common origin, the relationship between the two is very complex. In this context, it is also a matter of debate whether Scots can be regarded as a language in its own right or whether it is just a northern dialect of English (see MacArthur (1992) for an overview). For the conceptualization of the complex relationship between Scots and SSE, Aitken (1979) proposed a very influential bipolar model with more Scots-related linguistic features on the one side and more English-related features on the other (see Table 2). The model sets out the "range of speech options actually in use among all the different groups of Lowland Scots speakers (…) and the range of alternatives theoretically available to individual speakers" (Aitken, 1979, p. 85).

Table 2. Abbreviated version of Aitken's (1979) model of Scottish Speech.

Scots			**English**	
1	2	3	4	5
<bairn>	<mair>	<before>	<more>	<child>
<kirk>	<hame>	<name>	<home>	<church>
<ken>	<hoose>	<tide>	<house>	<know>
-na		most of the inflectional system, word order and grammar		-n't

Columns 1 and 2 on the left side in Table 2 list features and lexical items that historically derive from Early Scots. On the right side, columns 4 and 5 include the equivalent "later importations from southern English" (Aitken, 1979, p. 85). Scottish English speakers can therefore choose between, for instance, the Scots-derived words <bairn>, <kirk>, or <ken> (column 1) or their English equivalents <child>, <church> and <know> (column 5). They can implement the Scots negative particle <-na> as in the sentence < I dinna ken> or <n't> in the English equivalent <I don't know>. They can also choose between Scots and English words that are structurally very similar but different in pronunciation and spelling, such as Scots <mair>, <hame> and <hoose> (column 2) or English <more>, <home> and <house> (column 4). Yet, there are also many items that are common in Scots and English as specified in column 3. This includes, for instance, lexical items such as <name> and <tide> but also most of the inflectional system, word order and grammar. Aitken (1979) notes that whenever speakers choose to include features from columns 1 to 3, they are said to be speaking Scots. When speakers exclusively use features from columns 3 to 5, they are regarded to be speaking English, or, given the terminology introduced before, SSE. Yet, many Scottish speakers can switch between the Scots and English poles in different registers, speaking more Scots in informal situations and more SSE in formal circumstances (Aitken, 1979, pp. 85–86). Aitken also notes that there are others who drift between different styles in a less predictable and more fluctuating way. There is, thus, a distinction between "dialect switchers" and "style drifters" (Aitken, 1979, p. 86). Dialect switchers consciously use either a Scots repertoire or an English repertoire according to the social situation they are engaged in. Style drifters "cannot or do not choose to control their styles in this way" (Aitken, 1979, p. 86) and drift between Scots and English-related features in a more fluctuating way. Scots is often associated with the working classes and SSE with educated middle class speakers (Stuart-Smith, 2008, p. 48). Stuart-Smith (2008, p. 48) further notes that style switching is more common in rural varieties whereas style drifting is more characteristic of urban dialects. Overall, Aitken's model of Scottish speech demonstrates a bipolar linguistic continuum between Scots and SSE. The relationship between the two is therefore not clear-cut but rather overlapping and fuzzy (Stuart-Smith, 2008, p. 48). While the bipolar model is often cited in Scottish linguistic literature and provides an accessible overview of the complex linguistic situation in Scotland, its two-dimensional setup excludes a great deal of further sociolinguistic variation in Scottish English (Aitken, 1984a, p. 519). This naturally includes regional and social variation that is not reflected by the model's two dimensions of more Scots and less standard or more standard English and less Scots. In the words of Maguire (2012, p. 55), "[a]ny particular variant might be assigned all sorts of meanings, for example: Scots, SSE, Scotland-but-not-England, working-class, educated, local, Glasgow-and-not-Edinburgh, cool, different, old-fashioned, Catholic". While the bipolar continuum serves as a good overview of the fuzzy relationship between Scots and SSE, the use of more Scots- or more English-related variables can better be understood as an indexical field (Eckert, 2008). The use of a Scots variant may not only index meaning on a scale from informal to formal or from non-standard to standard, but it can also be associated with many other forms of social meaning and belonging. As a result, it is often difficult to draw a clear dividing line between Scots and SSE. Nevertheless, as SSE is still usually spoken in formal situations and across educated middle- or upper-class Scottish speakers, the present study will focus on these social groups and on language spoken in formal situations.

2.5 Summary

The linguistic situation of Scotland is complex. The once predominant language Scottish Gaelic still has an influence in the country even though its relatively small speaker numbers are mostly clustered in the Hebrides and western Highlands. There are, however, much higher speaker numbers for Scots with large proportions in the Northern Isles, the Northeast and in the Greater Glasgow area. Scots is often used among the working classes as well as in private and informal settings. The most dominant language in Scotland is English and almost all people in Scotland are potential speakers of SSE. This standard variety is generally defined as being very close to SSBE in its grammar and vocabulary, only its phoneme inventory is similar to Scots. Both Scots and SSE have the same linguistic origin and stand in a close and fuzzy relationship with each other. This relationship is often expressed as a bipolar linguistic continuum and speakers can switch and drift between more Scots-related or more SSE-related variants in their everyday speech. There are, however, also other factors that influence linguistic variation in Scottish English outside the Scots-English bipolar continuum, such as the regional background, the age, social class affiliation and many other sociolinguistic factors.

3. Vowel Duration in Scottish English

Vowel duration is influenced by various segmental and suprasegmental factors and is also subject to considerable idiolectal variation. An important distinction that must be made in the context of the present study is that between vowel duration and vowel length. Whereas vowel duration constitutes the measurable amount of time taken up by a vocalic segment, vowel length constitutes the perceived duration which can serve as a phonemic factor. Thus, vowel duration is a phonetic feature that is usually measured in milliseconds (henceforth: ms) and vowel length is a phonological feature which can be indicated by lengthening marks (ː = long, ˑ = half-long) in transcription (Laver, 1994, p. 151). The terms quantity and timing are often used as cover terms for both the phonetic duration and the phonological length of a segment (Brown & Miller, 2013, p. 368). In English phonology, there is a general distinction between long and short vowel phonemes (see Table 3).

Table 3. Phonemically long and short vowels in Received Pronunciation (RP) and General American (GA). Retrieved from Gut (2009a, p. 64).

RP		GA	
long vowel	short vowel	long vowel	short vowel
/iː/	/ɪ/	/iː/	/ɪ/
/uː/	/ʊ/	/uː/	/ʊ/
/ɔː/	/ʌ/	/ɔː/	/ʌ/
/ɑː/	/æ/	/ɑː/	/æ/
/ɜː/	/ə/	/ɜː/	/ə/
	/e/		
	/ɒ/		

Hence, the vowel in the word <heat> would be classified as the long monophthong phoneme /iː/ and the vowel in the word <hit> would be classified as the short monophthong phoneme /ɪ/ which implies differences between the quality and quantity of the vowels. This means that the vowel /i:/ is not only longer than the vowel /ɪ/ (quantity), but also that the acoustic properties of both vowels are inherently different (quality). The long monophthong /i:/ is characterized by a higher and fronter realization than the more centralized short monophthong /ɪ/. These vocalic differences then also constitute a difference in meaning as the words <hit> and <heat> are a minimal pair. Diphthongs are generally considered to be long in English (Roach, 2010, p. 17). Yet, this classification of ‘long’ and ‘short’ vowels is phonological and may not always correlate with phonetic reality (Gut, 2009a, p. 64). This means that the actual phonetic duration of the vocalic nucleus in the word <heat> may not always be longer than the vocalic duration in the word <hit> even if it is generally the case. In the context of the present study, I will keep a strict distinction between the terms *duration* and *length* to refer to either a phonetic or a phonological description. As the present study analyses recorded speech data, it will mainly take the phonetic perspective and the focus will therefore be on vowel duration.

The present chapter provides an overview of different factors that influence the duration of vocalic intervals. Section 3.1 deals with the general segmental and suprasegmental features that influence vowel duration in English. This also includes the VE which is explained in detail in subsection 3.1.2. Section

3.2 then introduces the main timing effect that is of interest in the present study, the SVLR, and the subsections 3.2.1, 3.2.2 and 3.2.3 summarize the findings of previous studies carried out on Aitken's Law. I decided to have a clear distinction between the general factors (section 3.1) and the SVLR (section 3.2) to differentiate between those timing effects which generally influence vowel durations in spoken English and those effects which are attributed to the SVLR. Section 3.3 provides a summary of all features that influence the duration of vocalic intervals in SSE. Similar overviews of influences on vowel timing in Scottish English can be found in Warren (2018) and Weilinghoff (2019).

3.1 General factors influencing vowel duration

The present section aims to provide an overview of the most important factors which generally influence the duration of vocalic intervals in English, namely *intrinsic vowel duration* (subsection 3.1.1), the VE (subsection 3.1.2), *intrasyllabic compression* (subsection 3.1.3), *polysyllabic shortening* (subsection 3.1.4), *lexical category and frequency* (subsection 3.1.5), *lexical stress* (subsection 3.1.6), *prosodic stress* (subsection 3.1.7), *constituent-final lengthening* (subsection 3.1.8) and *tempo* (subsection 3.1.9). The present investigation will take all of these factors into account in the analysis (see section 4.4).

3.1.1 Intrinsic vowel duration

Vowel duration is influenced by tongue height (Lehiste, 1970, p. 18). There is a general tendency that low vowels display longer durations than high vowels in speech (see Figure 3).

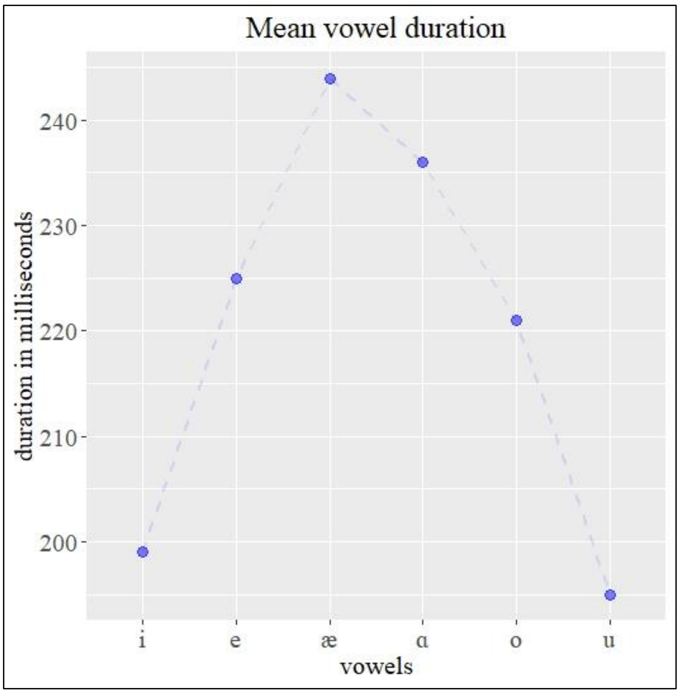


Figure 3. Plot of mean vowel durations from House and Fairbanks (1953, p. 111).

Thus, the low vowels /æ/ and /ɑ/ have longer average durations than the high vowels /i/ and /u/. This correlation has not only been reported for English (House & Fairbanks, 1953), but also for other languages such as German (Maack, 1949), Hindi (Ohala & Ohala, 1992), Italian (Esposito, 2002) and Swedish (Elert, 1964; Lindblom, 1982), among others.

Tauberer and Evanini (2009, p. 2211) note that there are physiological and phonological explanations for this phenomenon. Lehiste (1970, pp. 18–19), for instance, argues that open or low vowels generally tend to be longer in all languages due to the greater extent of articulatory movement involved in their production. As the jaw and tongue are lowered in the production of open vowels, the articulators need to travel a greater distance which results in longer vocalic intervals. This physiological explanation is frequently quoted in the literature (Warren, 2018, p. 67). Lindblom (1967) proposed a more complex interplay of lip and jaw movement which accounts for intrinsic vowel durations. Lisker (1974) as well as Tauberer and Evanini (2009) challenge the physiological explanations and propose that there are phonological reasons for intrinsic vowel duration.

Despite the different explanations for the durational differences between high and low vowels, the tendency that the latter display longer durations in English remains largely undisputed (House & Fairbanks, 1953, p. 111; Solé & Ohala, 2010, p. 646; Tauberer & Evanini, 2009, p. 2213).

3.1.2 The Voicing Effect

The voicing of a postvocalic consonant is a well-known influence on vowel duration in most English varieties. Henry Sweet already noted in 1877 that vowels are longer before voiced consonants than before voiceless consonants (Sweet, 1877, p. 59). Heffner (1937, p. 130) also observed that his American English vowel pronunciations are longer when they are followed by the voiced stop /d/ than when they are succeeded by the voiceless stop /t/. Thus, the high vowel in the word <bead> generally tends to be longer than the same vowel in the word <beat>. House and Fairbanks (1953, p. 108) found out that vowels preceding voiced consonants are on average 79 milliseconds longer than vowels followed by voiceless consonants which represents a ratio of 1.45 to 1.

Furthermore, also the manner of the postvocalic consonantal articulation has an influence on vowel duration. In their study, House and Fairbanks (1953) recorded 10 male students who produced 72 nonmeaningful stimulus syllables in which six vowels (/i/, /e/, /ae/, /a/, /o/, /u/) were preceded and succeeded by 12 different consonants. The recordings were made in a controlled laboratory setting. Figure 4 provides an overview of the average vowel durations in different consonantal contexts from the study by House and Fairbanks (1953). It is directly noticeable that the voiceless plosives and fricatives (highlighted in blue) are generally shorter than their voiced equivalents (highlighted in green). The vowel durations in the nasal contexts (highlighted in red) are also longer than the vocalic intervals in the voiceless plosive and fricative contexts, but shorter than the vowel durations in voiced fricative contexts. The influence of the manner of consonant production is also noticeable in Figure 4: the fricative contexts (circles) generally trigger longer vowel durations than the plosive contexts (squares). Thus, the average vowel durations in the voiceless fricative environments (/f/: 188 ms; /s/: 197 ms) are longer than those in the voiceless plosive contexts (/p/: 159 ms; /t/: 168 ms; /k/: 157 ms). Likewise, the average vocalic intervals in the voiced fricative contexts (/v/: 279 ms; /z/: 291 ms) are longer than the ones in the voiced plosive environments (/b/: 237 ms; /d/: 258 ms; /g/: 239 ms). The average vocalic intervals in the nasal contexts (triangles) (/m/: 219 ms; /n/: 245 ms) are, however, similar to the ones in the voiced plosive environments.

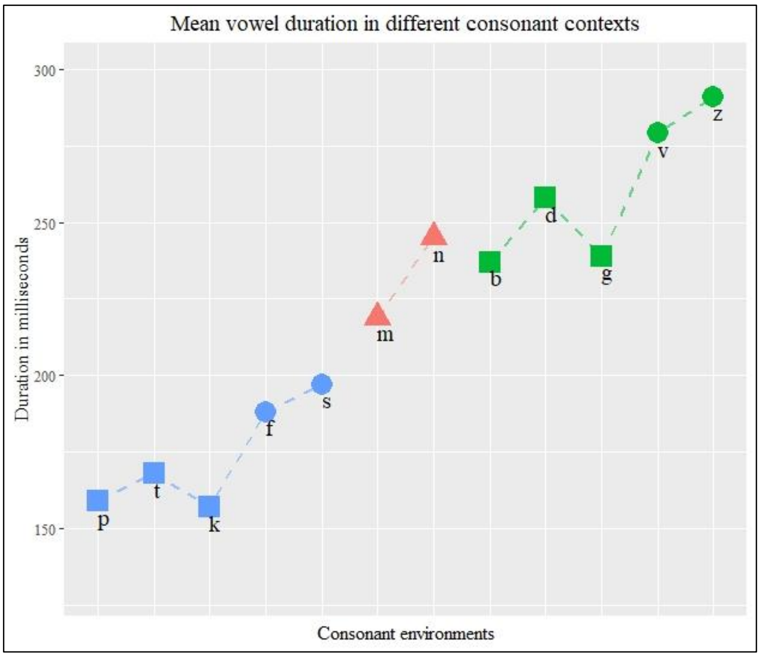


Figure 4. Mean duration of vowels in different consonant environments from House and Fairbanks (1953, p. 108). The voiced plosives and fricatives are highlighted in green and the voiceless plosives and fricatives are highlighted in blue. The nasal contexts are highlighted in red. In addition, plosive contexts are represented by squares, fricative contexts by circles and nasal contexts by triangles.

Peterson and Lehiste (1960) found an even stronger effect of postvocalic consonant voicing on vowel duration. The average duration of a syllable nucleus before voiced consonants was 297 ms and the equivalent mean duration before voiceless consonants was 197 ms (Peterson & Lehiste, 1960, p. 700) which accounts for a ratio of 1.5 (vowels before postvocalic voiced consonants) to 1 (vowels before postvocalic voiceless consonants). Furthermore, they also found that postvocalic fricatives have a stronger lengthening effect on vowel duration when compared to the effect of postvocalic plosives. In their study, Peterson and Lehiste (1960) had nine comparable CVC minimal pair sets (C = consonant segment; V = vowel segment) with following voiced and voiceless plosives as well as following voiced and voiceless fricatives. Example words for these contexts are <right> (postvocalic voiceless plosive), <ride> (postvocalic voiced plosive), <rice> (postvocalic voiceless fricative) and <rise> (postvocalic voiced fricative). Peterson and Lehiste (1960) found that the average vowel durations are the longest in postvocalic voiced fricative contexts (376 ms) followed by voiced plosive environments (280 ms), voiceless fricative contexts (228 ms) and the shortest voiceless plosive environments with an average vowel duration of 184 ms (see Figure 5).

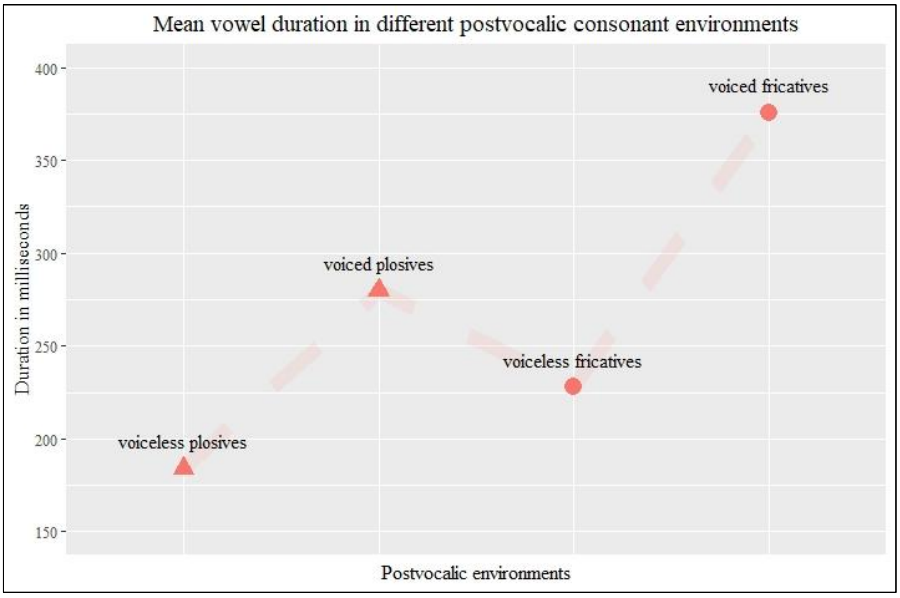


Figure 5. Mean vowel duration in different postvocalic consonant environments differentiated for the manner of articulation. The triangles represent plosive contexts and the circles represent fricative contexts. Data taken from Peterson and Lehiste (1960, p. 700).

Peterson and Lehiste (1960, pp. 700–701) also reported that the influence of prevocalic consonants on the duration of syllable nuclei is negligible. This means that, in contrast to postvocalic consonants, preceding consonants do not have a significant influence on vowel duration.

Chen (1970) tested the influence of the VE on vowel duration across English, French, Russian and Korean and he found out that it applies to all languages with different degrees. For English, he found that vowels followed by voiced consonants (average duration: 238 ms) are 61% longer than vowels followed by voiceless consonants (average duration: 146 ms) (Chen, 1970, p. 138). Hence, the ratio found by Chen (1970) even exceeds the previous ratios by Peterson and Lehiste (1960) and by House and Fairbanks (1953). The VE effects for French, Russian and Korean are, however, less pronounced but still observable. Chen (1970, p. 139) therefore concludes that the VE is presumably a language-universal phenomenon, but the extent of VE strength is determined by the language-specific phonological structure.

In contrast to this proposal of the VE being a language-universal phenomenon, Keating (1985, pp. 121–122) could not find any evidence of the VE in Polish or Czech. Furthermore, more recent studies have shown that the effect sizes of postvocalic consonant voicing in English are much smaller in connected speech than in word list or carrier sentence readings and that there are also differences between dialects (T. H. Crystal & House, 1988b; Tanner et al., 2019; Tanner et al., 2020; Tauberer & Evanini, 2009).

Tauberer and Evanini (2009, p. 2213), for instance, found that the strength of the VE varies in different North American English dialects. Using force-aligned interview speech data from the ANAE corpus (Labov et al., 2006), they found an overall VE effect size of 1.33 for Boston but no significant VE in Maine. Yet, the findings by Tauberer and Evanini (2009) could also be influenced by low token numbers. The overall effect size ratio of postvocalic consonant voicing was roughly 1.2 to 1 which represents a durational increase of 20% for vocalic intervals in voiced consonant contexts. This confirms that the VE generally operates in connected speech in most US cities. Yet, there are regional differences and the overall moderate ratio of 1.2 to 1 stands in stark contrast to those ratios found in earlier laboratory investigations (House and Fairbanks (1953): 1.45:1, Peterson and Lehiste (1960): 1.50:1, Chen (1970): 1.61:1).

One of the most recent studies by Tanner et al. (2020) also investigated the strength of the VE across a range of English varieties and found variable effect sizes. Using a multitude of speech corpora, the researchers found that the average VE size lies between 1.09 and 1.2 for different varieties (Tanner et al., 2020, p. 8). This underlines that the VE is smaller in spontaneous than in laboratory speech. As for regional variability, most Scots dialects show effectively no significant VE (Tanner et al., 2020, pp. 10–11). The VE is generally stronger in varieties such as SSBE, Irish English and, especially, in North American English dialects. African American English varieties have the largest VE in the sample (Tanner et al., 2020, pp. 10–11).

3.1.3 Intrasyllabic compression

Apart from *intrinsic vowel duration* (see subsection 3.1.1) and the postvocalic consonantal *VE* (see subsection 3.1.2), previous studies have further shown that vocalic intervals tend to be shorter in syllables with more segments than in syllables with fewer segments (Katz, 2012; Maddieson, 1985; Munhall et al., 1992). There is an overall inverse relationship between the number of segments in a syllable and the duration of the corresponding vocalic nucleus. In other words, the more consonants in a syllable, the shorter the vowel. The vowel in the word <street> tends to be shorter than the same vowel in the word <see>. This process is also often referred to as *compensatory shortening* (henceforth: CS) and concerns both the onset and coda of a syllable. Thus, one can distinguish between onset CS and coda CS. There is a further differentiation between simplex CS and incremental CS. Simplex CS is "observed in the comparison of syllables that contain one (consonantal) segment at the relevant periphery of the syllable (onset or coda) to syllables that contain no segments at the relevant periphery" (Katz, 2012, p. 391). Incremental CS is "observed in the comparison of syllables that contain one (consonantal) segment at the periphery to syllables that contain consonant clusters" (Katz, 2012, p. 391). For English, Katz (2012, pp. 395–396) reports simplex onset and simplex coda CS which means that vocalic intervals are generally longer in CV and VC words than in CVC words. Thus, words such as <eat> (VC) and <sea> (CV) tend to have longer vocalic intervals than the word <seat> (CVC). Katz

(2012, pp. 396–397) further reports a general effect of incremental CS which, however, differs by consonant manner and between onset and coda for some consonants.

3.1.4 Polysyllabic shortening

Similar to the relationship between vowel duration and the number of segments in a syllable, polysyllabic shortening “denotes the alleged property of syllable or vowel duration to be inversely related to the number of syllables in some larger prosodic unit” (Windmann et al., 2015, p. 36). This means the more syllables in a word, the shorter the vowel durations. Hence, the vowel /iː/ in the word <sleep> tends to be longer than the /iː/ in the word <sleepy> and the /iː/ in the first syllable of <sleepiness> again tends to be even shorter (Lehiste, 1972, p. 2019). This effect has generally been observed in English (Barnwell, 1971; Klatt, 1973; Lehiste, 1972; Turk & Shattuck-Hufnagel, 2000) but also in other languages (Lindblom & Rapp-Holmgren, 1971, p. 21). Rathcke and Stuart-Smith (2016, p. 415) also found evidence for polysyllabic shortening in the vowel /a/ in Glaswegian English. However, they did not find a significant effect of polysyllabic shortening in the high vowels /i/ and /ʉ/.

3.1.5 Lexical category and frequency

The lexical category of a word can have an effect on vowel duration as function words are frequently reduced and therefore shortened in spontaneous speech (A. Bell et al., 2009; Ernestus & Warner, 2011; Umeda, 1975). The determiner <the>, for instance, is usually unstressed and thus pronounced with a short schwa /ðə/ in connected speech. At the same time, there is also a relationship between the lexical category of a word and its frequency and predictability. As function words are a closed class of words that primarily fulfil grammatical roles, they are not only more frequent, but they also tend to be more predictable than content words. While it is also possible to find reduced pronunciations of content words in spontaneous speech, many previous studies report an overall inverse relationship between lexical frequency and predictability on the one hand and word duration on the other. That is, the higher the lexical frequency and predictability of a word, the shorter its duration (A. Bell et al., 2009; Gahl, 2008). Likewise, low frequency words with low predictability tend to display longer durations. This tendency coincides with the relative informativeness of a word (Priva, 2017). “More predictable units are inversely informative (less frequent words are more informative), which in turn results in more predictable words being more likely to be reduced, resulting in more centralized formants and shorter durations” (Tanner et al., 2019, p. 2). This means that the vowel in the function word <the> tends to be shorter than the vowel in the content word <glee>. Despite some conflicting evidence found in a study by Cohn et al. (2005), there is widespread agreement in the literature that word frequency and predictability can have an inverse influence on duration (Aylett & Turk, 2006; A. Bell et al., 2009; J. Bybee, 2002; Ernestus & Warner, 2011; Gahl, 2008; Jurafsky et al., 2001).

3.1.6 Lexical stress

Lexical stress refers to stress on the word level and one usually distinguishes between three different levels: primary stress, secondary stress and unstressed syllables (Laver, 1994, p. 156). A simple example would be the word <mother> which consists of two syllables, the first bearing primary word stress and the second being unstressed. Stressed syllables show greater levels of prominence and are indicated with a primary stress mark in transcription /ˈmʌðə/. Whenever a multisyllabic word comprises more than just one stressed syllable, there is a further distinction between primary and secondary word stress. For example, the word <entertaining> comprises four syllables with the third syllable carrying primary stress (indicated by the upper stress mark) and the first syllable carrying secondary stress (indicated by the lower stress mark: /ˌentəˈteɪnɪŋ/. The greater level of prominence is usually characterized by an increase in loudness, duration and often a higher pitch level (D. Crystal, 2015, p. 454). Hence, lexically stressed syllables tend to be longer in duration than lexically unstressed syllables. In this context, the phonological concept of syllable weight is often used to distinguish between light and heavy syllables: Heavy syllables comprise either a long vowel, a diphthong, or a vowel followed by a long consonant or at least two consonants (Laver, 1994, p. 156); light syllables comprise "a short vowel nucleus alone or followed by a coda of no more than one short consonant" (D. Crystal, 2015, p. 520).

3.1.7 Prosodic stress

Apart from lexical stress, prominence can also be established in higher prosodic structures and leads to differences in vowel duration. There is usually one syllable in an intonation phrase which receives the strongest accent and this is often referred to as the nucleus of the intonation phrase (Gut, 2009a, p. 111). Similar names include *prosodic stress*, *syllabic stress* (T. H. Crystal & House, 1988a), *sentence stress* (Lehiste, 1970, pp. 36–38), or *phrasal stress* (Turk & White, 1999). Despite the multitude of different conceptualizations and models for English prosody and intonation (Beckman & Edwards, 1994), there is overall agreement that prosodic stress has a strong influence on vowel duration (Chevalier, 2019; T. H. Crystal & House, 1990; Rathcke & Stuart-Smith, 2016; Turk & White, 1999; Warren, 2018). For instance, already in 1935, Parmenter and Treviño (1935, p. 130) measured that the average duration of stressed vowels is almost 75 percent longer than the duration of unstressed vowels. The heavy influence of prosodic stress on vowel duration has also been found in empirical studies on spontaneously spoken Scots (Chevalier, 2019; Rathcke & Stuart-Smith, 2016) (see subsection 3.2.3).

3.1.8 Constituent-final lengthening

Another very influential prosodic factor on vocalic duration is the position of vowels in words and utterances. Oller (1973) conducted the first empirical study on the effect of position on segmental duration in English and he found that both stressed and unstressed vowels are significantly lengthened in utterance-final syllables. Vocalic intervals that precede a pause are on average 100 ms longer than their counterparts in the middle of an utterance (Oller, 1973, p. 1236) and this lengthening effect occurs consistently in imperative, declarative and interrogative intonation patterns (Oller, 1973, p. 1238). At the same time, vowel durations are also lengthened in word-final positions (Oller, 1973, pp. 1239–1240). This means that vowels tend to be longer when they are "directly adjacent to a word boundary and become shorter with added intervening syllables" (Windmann et al., 2015, p. 36). Umeda (1975, p. 442) also found that vowels are longer in word-final positions than in non-word-final positions. A highly significant effect of constituent position has also been observed in recent studies investigating vowel durational patterns in connected Scottish English speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016, p. 415; Stuart-Smith et al., 2019): vowels tend to be longer in constituent-final position than in non-final positions.

3.1.9 Tempo

The overall tempo of speech is another factor which strongly influences the duration of vocalic intervals: vowel durations tend to become shorter the faster a speaker talks (T. H. Crystal & House, 1988a, 1990). In this context, there is a distinction between articulation rate and speaking rate which are usually measured in syllables per second (Laver, 1994, pp. 539–540). Articulation rate refers to the tempo of articulating an utterance, excluding any silent pauses. Speaking rate refers to the overall tempo of a speaker in producing all the utterances in a speaking turn, including silent pauses. Thus, whereas speaking rate refers to the overall speech tempo of individual speakers on a macro level, articulation rate reveals how fast or slow particular utterances are spoken by individual speakers. Hence, a slow speaker can be identified by a low speaking rate but this speaker can still produce some utterances with a high articulation rate. Goldman-Eisler (1968, p. 24) noted that the average articulation rate among English speakers ranges between 4.4 and 5.9 syllables per second in spontaneous speech, but there is often substantial variation even within a single utterance of a single speaker (Miller et al., 1984). Many studies showed that speech rate decreases among older speakers (Ramig, 1983; Schötz, 2007) and T. H. Crystal and House (1990) have found that articulation rate is not only influenced by interspeaker differences but also by the intonational properties of an utterance. That is, the average syllable duration depends on the number of phones per syllable as well as the proportion of stressed syllables or stressed phones in the utterance (T. H. Crystal & House, 1990, p. 110). The more recent study by Tanner et al. (2020) on the VE also included tempo in their analysis. They first calculated a mean speaker rate for each individual speaker. In the next step, they calculated a local speech rate by subtracting the articulation rate of an utterance from the corresponding speaker's mean speaking rate. This local speech rate could then be interpreted "as how fast or slow that speaker produced the vowel within that particular phrase *relative* to their average speech rate" (Tanner et al., 2020, p. 5). The results showed that both mean speaker rate and local speech rate had a significant influence on vowel duration patterns.

3.2 The Scottish Vowel Length Rule

This section summarizes and discusses all relevant research that has been conducted on the SVLR in a largely chronological order. This means that I will provide a summary of each study and then briefly discuss the study's methodology and findings in another paragraph. The section is subdivided into three different subsections based on the varying methodological approaches of the previous investigations. Subsection 3.2.1 summarizes the early impressionistic accounts on vowel duration patterns in Scottish English. This includes the publications by Murray (1873), Grant and Dixon (1921), G. Watson (1923), Dieth (1932), Zai (1942), Wettstein (1942), Wölck (1965), Lass (1974), Lodge (1984), Mather and Speitel (1986) as well as Aitken's most comprehensive description of the rule published in 1981. The second subsection includes all experimental studies which investigated Scottish vowel timing in word list and carrier sentence readings, namely McClure (1977), Agutter (1988a,b), McKenna (1988), McMahon (1991), Milroy (1995), Scobbie, Hewlett, and Turk (1999), Hewlett et al. (1999), Scobbie, Turk, and Hewlett (1999), Watt and Ingham (2000), van Leyden (2002), Scobbie (2005), Pukli (2006), Watt and Yurkova (2007) as well as Llamas et al. (2011). Subsection 3.2.3 includes all studies which investigated Scottish vowel durational patterns in naturally occurring language, namely Rathcke and Stuart-Smith (2016), Warren (2018), Chevalier (2019) and Stuart-Smith et al. (2019). While there are also many articles which discuss Aitken's Law against the background of historical linguistics and phonological theory (Anderson, 1993; Carr, 1992; Ewen, 1977; Harris, 1985; Kamińska, 1995; Kiełtyka, 2003; Lass, 1974; Taylor, 1974), they are not of interest for the present study and will therefore not be discussed. Smith and Rathcke (2016) published the first perception study by investigating the influence of SVLR patterns on word segmentation and identification tasks among speakers from Glasgow and Leeds. Yet, as this investigation is concerned with how Glaswegian speakers perceive vowels, I will not go into detail here either.

3.2.1 Impressionistic accounts on the SVLR

Despite the fact that Aitken first formulated the SVLR in 1962, he was not the first linguist who observed peculiar vowel duration patterns in Scottish English. Rather, he summarized earlier, mostly impressionistic observations made by different linguists out of which he then generated rules.

The first indication of a special Scottish durational vocalic pattern was stated by Sylvester Douglas in his work *A Treatise on the Provincial Dialect of Scotland* in 1775 (Douglas, 1991 [1775], p. 151;

Kohler, 1966, p. 33). Douglas described that the vowels in the words <pride> and <deny'd> have the same diphthongal quality in a Scottish accent, but the words do not rhyme because the vowels differ in quantity: the vowel in <pride> is shortened and the vowel in <deny'd> is protracted in Scottish English (Douglas, 1991 [1775], p. 151). This durational difference corresponds to the SVLR-related lengthening of vocalic intervals when followed by morpheme boundaries: The vowel in <deny'd> is longer because it is directly followed by a morpheme boundary. Likewise, the SVLR conditions a shortening effect in the diphthong in <pride> due to the postvocalic voiced plosive /d/ which represents an SVLR short environment. According to Aitken (1981, 151), the observation by Douglas (1991 [1775] is the earliest source which bears witness to a specific vowel timing effect in Scottish English. However, Douglas only noticed the disagreement in vowel quantity in the rhyme <pride> and <deny'd>, but he did not analyze or discuss this phenomenon further.

A much more important reference for Aitken was Murray's (1873) report on the dialect of the Southern counties of Scotland. Murray provides an in-depth but impressionistic overview of the grammar and pronunciation of his home dialect region. As for pronunciation, he generally notices that the vowel systems of Lowland Scots and English are entirely different in terms of quality and quantity (1873, pp. 93–94). Furthermore, he also describes that all vowels in Scots dialects are neutralized in unstressed final positions and that long vowels can be stopped in closed syllables in Scots without changing their vowel quality (1873, p. 96). One example would be the word <reekie> whose long high vowel in the first syllable would be shortened in Scots. Murray then states four detailed rules on Scottish long and short vowel contexts. His first rule specifies that a vowel at the end of a monosyllabic word or in the position of an accented final syllable is long. Thus, words such as <wee> or <day> would retain a long vocalic duration (1873, p. 97). Exceptions include the articles <the> and <a> which are usually short and also "possessives and prepositions like *maa*, my; *tui*, to; *wui*, with; *frœ*, from; *î*, in; which have a long sound only when emphatic, but otherwise are brief (…)" (Murray, 1873, p. 97). This claim is consistent with the observation that function words are frequently shortened in connected speech (see subsection 3.1.5). Murray's (1873, p. 97) first rule also states that vowels are long when a monosyllable is followed by a noun- or verb inflection as in "*faa, faa's, day, days, preae, preaed, preaes*". This, in turn, is in line with Sylvester Douglas' example of <pride> and <deny'd> mentioned above: vowels are longer when followed by morpheme boundaries. Murray's second rule states that vowels are also long "before the sounds of *r, z, v,* and *th* vocal (dh), however these may be written (…) or, when *s* or *d* are added in inflection, as (…) *bleez'd, leeves, leeved* (…)" (1873, p. 97). The exception is "when these consonants are followed by another consonant in a root word, as *păirt, hœ̆rt, puŏrt, cuŏrn, feărce*; contast *cāyr, cāyr'd* = cared, with *caird* = card (k*ee*r, k*ee*rd, k*e*rd)" (Murray, 1873, p. 97). Thus, the vowel in <card> would be shorter than the vowels in <care> and <cared>. The third rule then specifies that vowels followed by all other consonants are short in Scottish monosyllabic words. In words with two or more syllables, vowels are generally short before all consonants. Thus, the vowel in the word <heat> and the first vowel in <father> are short in Scots even though they are long in English. Murray's fourth rule indicates that vowel length is retained in polysyllabic words which derive from a monosyllabic stem. Thus, the first vowel in the word <rosy> would still be long whereas the same vowel in <cosy> would be shortened (Murray, 1873, p. 97). Murray further elaborates that Scottish long and short vowels tend to be longer than their English long and short equivalents. Thus, <cheap> carries a short vowel in Scots which is nevertheless a bit longer than the English short vowel /ɪ/ in <chip>, but still shorter than the English long vowel /iː/ in <cheap>. At the same time, the long vowel /iː/ in Scottish <sees> is much longer than the long /iː/ in English <cheap> (Murray, 1873, pp. 97–98). Furthermore, he also notes that the vowel length distinction in Scottish long and short contexts is more distinctively preserved in the high than the low vowels (1873, p. 98). Murray then goes on and gives a detailed description of the individual vowels of the Southern counties of Scotland using the Visible Speech Alphabet by Alexander Melville Bell (1867).

Murray was arguably the first linguist who described Scottish vowel quantity in detail. While the IPA had not yet been introduced at the time and while he had no empirical evidence for his observations, his descriptions are nonetheless very precise. Many of the observations are still accurate today and Aitken's formulation of the SVLR is primarily based on the rules set out by Murray.

A similar description of Scottish vowel duration patterns can also be found in the *Manual of Modern Scots* by Grant and Dixon (1921, p. 60): "The tense vowels **i**, **e**, **o**, **u**, **ǫ**, **ø** and the vowel **ɑ** may all be heard fully long in final accented position and before voiced fricatives and **r**." This corresponds to Murray's (1873) first and second rule. The same definition can also be found in G. Watson's *Roxburghshire Word-Book* (1923, p. 24) and in the first edition of the *Scottish National Dictionary* (Grant, 1931, pp. xliv–xlv). Grant and Dixon (1921, p. 60) further specify that the aforementioned vowels are shortened "before all voiced plosives and **l**, **m**, **n**, **ŋ**". This shortening is much more marked in Scots than it is in Southern English.

The descriptions by Grant and Dixon (1921), G. Watson (1923) and Grant (1931) correspond with the rules set out by Murray (1873). The publications provide a general overview on Scottish vowel length but they do not go into the same detail as Murray (1873). While their transcriptions follow the rules of the early IPA, the descriptions by Grant and Dixon (1921), G. Watson (1923) and Grant (1931) also lack empirical evidence.

Another source that Aitken drew on was the study by Eugen Dieth (1932) on the Buchan dialect in the Northeast of Scotland. Dieth not only gives a detailed account of the regional dialect, but he also provides a few kymograph measurements of the speech of a local man from Byth. These kymograph tracings are, to the best of my knowledge, the first durational measurements of Scottish vowels in the context of SVLR research. Dieth (1932, p. 59) notes that vowel quantity is generally relative and depends, for instance, on the stress patterns in words or on the consonants following the vowel. Moreover, he also describes that vowel duration is influenced by the tongue position, thus indicating different intrinsic durational properties for long vowels (see subsection 3.1.1): "(…) those of high tongue position, like iː, uː being shorter than ɑː, oː (…)" (1932, pp. 59–60). This claim is also corroborated by the kymographic tracings of the words <malt>, <meat> and <meet> which reveal that the vocalic interval of the high vowel is shorter than that of the low vowel: the open vowel in <malt> is 210 ms long, the mid vowel /e/ 170 ms and the high vowel /i/ 140 ms. Similar to Murray's (1873) first and second rule, Dieth (1932, p. 60) describes that full vowel length is preserved in final open syllables, before /r/ and before the voiced fricatives /v/, /ð/, /ʒ/ and /z/. Hence, the words <knee>, <dare> and <booze> would carry long vocalic nuclei in the Northeast of Scotland. The same goes for vowels followed by inflectional endings and "before enclitic pronouns that sacrifice their vowels (…)" (Dieth, 1932, p. 63). An example for the latter would be <give it to me> pronounced as [ˈgiːt tə ˈmiː] where full vowel length would be preserved in <give> while the following pronoun is reduced (1932, p. 63). Dieth (1932) further specifies that full vowel length is also preserved in some words with postvocalic /l/, /m/, or /n/, for example <whale> and <dwalm> which was not mentioned by Murray (1873). Yet, Dieth (1932, p. 61) also indicates that long vowels are reduced in front of stops and voiceless fricatives and that this shortening process is much more consistent across the high vowels /iː/ and /uː/ which further corroborates Murray's observation (1873, p. 98). Dieth (1932, p. 61) elaborates that the shortening process before stops and voiceless fricatives is common in Germanic languages, but much stronger in Scots than in Standard English or German. He also specifies that the short vowels /ɪ/, /ɨ/ and /ʌ/ are always short in all contexts (1932, p. 67).

Dieth (1932) provides a detailed description on vowel length in the Northeast of Scotland and he is the first to implement acoustic measurements of Scottish vowel duration. Nevertheless, his account is still generally impressionistic; only his observation on intrinsic vowel duration is backed up by kymograph measurements. Dieth was also the first to explicitly state that there are vowels which are invariably short (/ɪ/, /ɨ/, /ʌ/) and this was later also taken up by Aitken. In addition, Dieth (1932) also had an influence on the subsequent PhD studies by Zai (1942) and Wettstein (1942) because he was their supervising professor.

Rudolph Zai (1942) conducted a study on a particular dialect in the South of Scotland, namely the dialect of the village Morebattle in the Scottish Borders. The investigation area is therefore comparable to the one of Murray's study (1873). Zai (1942, p. 15) notes that vowel length is influenced by various factors, including intrinsic vocalic properties, the following consonant, the number of syllables following the vowel and, to a lesser degree, the origin of the vowel. As for the Morebattle dialect, he distinguishes between three main groups of vowel phonemes which are summarized in Table 4. The first group comprises the vowel sounds /ɛ̈/, /ʌ/ and, in unstressed positions, /ɪ/ and /ə/ which are always short and of the same quality in all contexts. One also has to take into consideration that the vowel /ɛ̈/ represents the lowered and centralized short high vowel in words such as <fish> and <big> (Zai, 1942, p. 12). Thus, /ɛ̈/ can be considered as a typical Scottish pronunciation of the vowel /ɪ/. Zai (1942, p. 15) labels the first group as the "phonemes of invariable quantity and quality". The second group comprises the vowel phonemes /i/, /e/, /æ/, /ɑ/, /o/ and /u/ which do not change their quality but can have different lengths. They are therefore categorized as “phonemes with quantitative variants” (Zai, 1942, p. 15). Only the phoneme in the third group opts out of this pattern as it is always long (Zai, 1942, p. 11).

Table 4. Phoneme groups divided for different quantitative and qualitative properties in the Morebattle dialect from Zai (1942, pp. 15–16).

Phonemes of invariable quantity and quality	Phonemes with quantitative variants	Phoneme with quantitative and qualitative variant
/ɛ̈/ /ʌ/ /ɪ/ /ə/	/i/ /e/ /æ/ /ɑ/ /o/ /u/	[œː]

Similar to previous descriptions, Zai (1942, p. 16) also notes that the vowels of variable quantity are fully long in final positions and before word-final voiced fricatives and /r/. They are, however, shorter when followed by other consonants. In contrast to Murray (1873), Zai (1942) further subdivides the short contexts into short and half-long environments and he also differentiates between the historically long vowels /i/, /e/, /u/ on the one hand and /æ/, /ɑ/, /o/ on the other.

Despite the greater level of detail, Zai’s (1942) account generally corresponds with the previous descriptions of long and short vowel contexts. Zai (1942) is, however, one of the first researchers to explicitly categorize Scottish English vowels into different groups based on their quality and quantity changes. The grouping of vowels was later also taken up by Aitken (see below).

Paul Wettstein (1942) provides a similar account on Scottish vowel length in his study on the Berwickshire dialect in the Southeast of Scotland. Similar to Zai (1942), he distinguishes between vowels of invariable quantity and vowels of variable quantity in monosyllables and end-stressed syllables (Wettstein, 1942, p. 6). The first group comprises the vowels /ʌ/ and /ɛ/ which are short in all contexts. Wettstein does not include the symbol /ɪ/ in his vowel system, so, for instance, <fish> would be represented as /fɛʃ/ in the Berwickshire dialect (1942, pp. 1–3). As for the second group, Wettstein (1942) states that its vowels (/ɛ̈/, /u/, /i/, /e/, /o/, /a/ and /ɒ/) are generally long in stressed open syllables or before the final voiced fricatives /z/, /v/, /ð/ and /r/. This corresponds to the previous descriptions of Scottish English long contexts. In addition, the vowels /e/, /o/, /a/ and /ɒ/ tend to be more variable in terms of their quantity. The vowel /e/, for instance, is also fully long before other consonants but half-long when followed by consonant clusters. Similarly, the vowel /a/ tends to be longer before /r/, voiced plosives and voiced consonant clusters than before /l/, voiceless fricatives, or voiceless consonant clusters. The vowel /o/ is generally unstable in terms of its quantity (1942, p. 7). Only the high vowels /i/ and /u/ tend to be more stable in displaying fully long realization in long contexts and short variants in short environments. Wettstein (1942, p. 6), who also conducted some kymograph measurements in his conversational interviews, provided the following average durations: long vowels and diphthongs were 200 ms long, half-long vowels 140 ms, short vowels 80 ms and unaccented vowels 50 ms. Unfortunately, Wettstein (1942) does not specify his measurements any further. Nevertheless, he is also aware that vocalic durations are influenced by different segmental and suprasegmental factors in conversational speech. Apart from the influence of the following consonant, he notes that vowel intrinsic properties, following morpheme boundaries, lexical and phrasal stress as well as historical quantity can strongly influence vowel durations (Wettstein, 1942, p. 6). The short vowels /ʌ/ and /ɛ/, for example, can be lengthened when carrying phrasal stress and/or when occurring in pre-pausal positions (Wettstein, 1942, p. 9). Thus, phrasal stress and utterance-final positions generally lengthen vocalic durations.

Overall, Wettstein's (1942) approach is very similar to Zai (1942) because he also sorts the vowels into groups based on quantity and quality changes. Similar to his PhD supervisor Dieth (1932), Wettstein also provides kymograph measurements but these measurements only represent general averages. His descriptions on Scottish vowel length are therefore also impressionistic and they are generally in line with earlier accounts.

Another purely impressionistic study on the Buchan dialect was conducted by Wölck (1965). In his study, the researcher distinguishes between the phoneme system of the Buchan dialect and the phoneme system of SSE spoken by the local population. Similar to Zai (1942) and Wettstein (1942), Wölck (1965, 21–23) also differentiates between several different degrees of quantity (very long, long, half-long, short, very short). He specifies that the vowels /i/, /e/, /a/, /o/ and /u/ are very long in stressed open syllables before a pause and generally long when followed by word-final voiced fricatives, /r/, /l/, /m/ and /n/ (Wölck, 1965, pp. 21–22). Hence, the vocalic nuclei in the words <fee>, <grieve>, <wear>, <fool>, <cream> and <moon> are long in Buchan which fully corresponds with Dieth's (1932) previous description of the same dialect. The vowels are, however, half long when followed by voiced stops, voiceless fricatives as well as consonant clusters with stops and /r/ as in <read>, <roof> and <hard>. Likewise, half long vowel realizations also occur in polysyllabic words when the vowel is followed by /r/, /l/, /m/ or /n/ in the onset of the subsequent syllable as in <carry>, <fellow>, <hammer> or <dinner>. The vowels are short when followed by voiceless stops and /ŋ/ as in <feet> and <king> and when followed by voiced stops and fricatives in word-medial positions of polysyllabic words (<brother> and <reader>). Vowels are even shorter when followed by voiceless stops in word-medial position of polysyllabic words (<later>).

While Wölck (1965) provides precise categories of Scottish vowel length and while his observations are generally in line with those of previous investigations, he also lacks empirical evidence to back up his account.

Another very important source for Aitken were the collections of the Linguistic Atlas of Scotland (henceforth: LAS) by Mather and Speitel (1986). In the third volume, the survey investigated the phonology of Scots in different locations of Scotland which are located South and East of the Highland Line. Furthermore, the survey also investigated localities in Orkney, Shetland, Ulster as well as in the borough of Berwick-upon-Tweed in Northern England (Mather & Speitel, 1986, p. xi). The LAS did not create an overall vowel system on a regional or national scale but applied a polysystemic approach. That is, the fieldworkers investigated one local informant (mostly male) whose speech and vowel system would be representative for the respective town or city. As a result, the LAS generated 187 vowel systems representative for 187 locations in Scotland (Mather & Speitel, 1986, pp. 397–398). The informants had to read out a systematically compiled list of monosyllabic words in citation form with a falling intonation to avoid the influence of suprasegmental factors. The researchers grouped those words and the respective vowels into so-called polyphonemes, groups of vowels which "share certain phonetic features, show similar distribution[s] from place to place and/or are historically related" (Mather & Speitel, 1986, p. xiii). The words were further grouped into 11 different sections representative of different postvocalic contexts. For instance, section 1 includes all words with the respective polyphoneme before word-final /t/ in words such as <meet>. The fieldworkers then categorized the quality and quantity of the individual vowel pronunciations on an impressionistic basis. As for quantity, the LAS distinguishes between long and short variants. While an extensive account of the 187 LAS vowel inventories would clearly exceed the scope of the present summary, there is a general trend that SVLR long environments condition long vowel pronunciations.

Even though the LAS is an extensive and detailed data collection, the impressionistic accounts of vowel duration elicited from one informant and the use of several fieldworkers limit the reliability of its descriptions. There may be transcriptional differences between the fieldworkers and it can also be challenged whether the idiolect of one informant can be representative for a whole town. Warren (2018, p. 92), for instance, notes that the LAS reports long vowel realizations almost universally for all polyphonemes from Ayrshire. This, however, stands in direct contradiction to the later findings by McClure (1977) (see subsection 3.2.2). Warren (2018, p. 92) therefore considers the LAS findings to be not very reliable.

Aitken's formulation of the SVLR is based on the previous studies and developed over time. He first summarized the previous observations on Scottish vowel timing in an unpublished handout in 1962 called "Statement of the Phenomenon's Essentials". This was later replaced by another more comprehensive handout named "The Scottish Vowel-length Rule" (Aitken, 1975). He also mentioned the fundamental principles of the SVLR in his essay on "How to Pronounce Older Scots" (Aitken, 1977) and in his article "Scottish Speech: a historical view with special reference to the Standard English of Scotland" (Aitken, 1979). His most detailed exposition, however, was published in 1981. In this article, he summarized the findings of previous studies and also incorporated the first results of the LAS phonological survey which were not yet published at the time (Aitken, 1981, 131). Central to his description of Scottish vowel timing is his vowel chart which portrays the development of all Scots vowels and diphthongs from the Early Scots period to Modern Scots (Aitken, 1981, 132–133). As Aitken's historical outline of the Scots vowels covers several centuries, he used vowel numbers to make it easier to refer to a phoneme irrespective of its particular realizations in different periods of time (Aitken, 1981, 131). The vowel chart can also be found in his publication from 1977 and a simplified overview is given in Table 5.

Table 5. Historical outline of the vowel system of Scots by Aitken (1981, 132–133).

	Early Scots	**Modern Scots**	Examples	SVLR status
Vowel number	Long monophthongs			
1	/iː/	**/əi/**	<bite>	yes
		/aˑe/	<size>	yes
2	/eː/	**/i/**	<meet>, <see>	yes
3	/ɛː/	**-**	-	-
4	/aː/	**/e/**	<bate>, <care>	yes
5	/oː/	**/o:/**	<throat>	invariably long in some dialects
6	/uː/	**/u/**	<mouth>	yes
7	/øː/	**/ø/**	<good>	yes
		/i/	<sure>	yes
		/e/	<do>	yes
	Diphthongs ending in /-i/			
8	/ai/	**/e:/**	<bait>	invariably long in some dialects
		/e:ə/	<pair>	invariably long in some dialects
8a	/ai/	**/əi/**	<day>	unknown
9	/oi/	**/oi/**	<boy>	unknown
10	/ui/	**/əi/**	<point>	yes
11	/ei#/	**/i/**	<eye>	unknown
	Diphthongs ending in /-u/			
12	/au/	**/a:/**	<fault>	invariably long in some dialects
		/ɔ:/	<snow>	invariably long in some dialects
13	/ou/	**/ʌu/**	<loud>	yes
14	/iu/	**/iu/**	<duty>	unclear / invariably short
		/ju/	<dew>	yes
	Short monophthongs			
15	/ɪ/	**/ɪ/**	<bit>	invariably short
16	/ɛ/	**/ɛ/**	<bed>	yes
17	/a/	**/a/**	<man>	yes
18	/o/	**/o/**	<cot>	yes
19	/u/	**/ʌ/**	<buzz>	invariably short

With respect to the SVLR, Aitken identifies three principal groups of vowels in Modern Scots (1981, 134). The first group consists of /ɪ/ (vowel 15) and /ʌ/ (vowel 19), which are realized short in all environments and in all dialects. These vowels are therefore not subject to the SVLR. Aitken further notes that vowel 15 is often more centralized and lowered in Scots than in RP which reflects previous accounts on the quality of that vowel (Wettstein, 1942, pp. 1–3). In addition, vowel 14 has a rather unclear status. Aitken specifies that the original Early Scots diphthong /iu/ would be invariably short and thereby resembling vowel 15 and vowel 19. Yet, the modern realization of vowel 14 with the glide /ju/ is subject to the SVLR.

The second group consists of the vowels 8 and 12 which are realized long in both SVLR short and long environments in many areas outside the Central Scots dialect region. These vowels are therefore not affected by the SVLR in many, but not all Scots dialects. The Modern Scots representatives are the vowels /eː/ and /eːə/ as in <bait> and <pair> for vowel 8 as well as /aː/ and /ɔː/ for vowel 12. In addition, Aitken notes that vowel 5 (/oː/) shows long realizations in SVLR short contexts in some dialects outside the Central Scots and South Scots area (1981, 152). The status of the diphthong /oi/ (vowel 9) is not entirely clear.

The third group comprises the remaining vowels and most, if not all of them are subject to the SVLR. That is, these vowels occupy an allophonic range of realizations of relatively short duration in the SVLR short environments and an allophonic range of markedly longer durations in SVLR long environments when occurring in end-stressed syllables and being unaffected by the influence of phrasal position and prosodic stress (Aitken, 1981, 134–135). Aitken then summarizes the long contexts which were already noted by previous researchers, namely following voiced fricatives, /r/ and morpheme boundaries. He also notes that "nearly all Scots dialects (and Scottish Standard English) agree in displaying fully long realizations of the affected vowels in these environments" (Aitken, 1981, 135). Despite a few exceptions in some dialects, all other environments of end-stressed syllables, namely following voiced and voiceless stops, nasals, laterals, voiceless fricatives and the voiceless affricate /tʃ/, condition short vowel realizations. An overview with different examples for the vowel /i/ can be found in Table 6.

Table 6. Examples of SVLR short and long environments by Aitken (1981, 135).

SVLR Short			**SVLR Long**		
Word	Environment	Realization	Word	Environment	Realization
<leaf>	following voiceless fricative	[lif]	<leave>	following voiced fricative	[liːv]
<bead>	following voiced plosive	[bid]	<deer>	following /r/	[diːr]
<feel>	following lateral	[fil]	<sees>	following morpheme boundary	[siːz]

While Aitken (1981, 135–136) acknowledges a lack of instrumental measurements of Scottish vowel durations, he backs up his account with the little empirical evidence that Dieth (1932), Wettstein (1942) and, most importantly, McClure (1977) (see subsection 3.2.3) provided. He concludes that "Scottish vowels have their own peculiar ways of phonetic behavior which, in respect of duration at least, differ from those of all other kinds of English" (Aitken, 1981, 136). In comparison to RP, he notes that the Scots shortest durations are shorter and the longest durations are longer which corresponds with the earlier description by Murray (1873, pp. 97–98). Yet, it is especially the difference between allophonic short and long vocalic durations that sets Scots and SSE apart from other dialects of English "which appear to display vowel durations descending according to environment along a single gradual continuum" (Aitken, 1981, 136). This continuum prescribes that RP vowel length is continuously decreasing in the following contexts: open syllables > following voiced plosives > following voiceless plosive > phonologically short vowel phonemes (Gimson, 1973, p. 94). Examples of the former would be the words <bee>, <bead>, <beat> and <bit>. In Scottish English, however, there is a two-fold phonemic and phonetic division between SVLR short and long contexts. The allophonic durational ranges of SVLR short and long contexts usually do not overlap even though there is, of course, internal variability. As an example, Aitken states that while the duration of the vowel /i:/ slightly decreases in descending order in the following words <sea-voyage>, <agrees>, <agreed>, <dear>, <please>, <leave>, <leaving>, these vocalic durations are nevertheless between 1.5 or 2 times longer than those of <peace>, <greed>, <mean>, <feel> or <meet> (1981, 136). While Aitken acknowledged that word- or morpheme boundaries, following voiced fricatives and following /r/ generally yield long vocalic realizations in English, the interesting feature of the SVLR is the shortening of vowels in contexts which favor long realization in other English dialects (1981, 137). Thus, whereas the vowel in <breeze> is long in both RP and Scottish English, vowels followed by voiced plosives, nasals and laterals, as in <bead>, <bean> and <feel>, would be long in RP but short in Scots and SSE. This shortening of originally long vowels is the key distinguishing factor of the SVLR and Aitken (1981, 137) assumes that this shortening process began in the 15th century. While Aitken acknowledges some variability in particular regions and vowels, the "dual arrangement of vowel realizations specified by the SVLR does certainly operate in whole or in part in all dialects (…)" (1981, 140).

Overall, Aitken (1981) provides the most detailed impressionistic account on the SVLR. He is aware of the previous references and tries to back up his observations with empirical evidence.

The last impressionistic report by Lodge (1984), however, casts doubts on the far-reaching validity of the SVLR. Lodge (1984) transcribed and investigated the colloquial speech of speakers from Stockport, Shepherd's Bush, Peasmarsh, Coventry, Norwich and Edinburgh on a detailed, yet impressionistic basis. As for Edinburgh, he analyzed speech of two male speakers from Edinburgh (50 years old) and Penicuik (19 years old) and found that /a/ is often lengthened, especially before nasals and other SVLR short environments (Lodge, 1984, p. 93). Likewise, the older informant also seems to lengthen /e/ in many SVLR short contexts, such as <less> and <egg>. In addition, vowel length is strongly influenced by stress and the loss of coda /r/ also leads to increased vocalic durations (Lodge, 1984, pp. 93–94). While SVLR long contexts do also condition long realizations, the overall transcriptions show that the strict opposition of SVLR long and short contexts does often not apply in the colloquial speech of the two informants.

The results by Lodge (1984) clearly contradict the rules set out by Aitken and further underline the influence of suprasegmental factors on vocalic durations in colloquial speech. It could be the case that the effects of Aitken's Law are superseded by stronger suprasegmental factors in spontaneous speech. There is, however, no empirical evidence to back up this point.

3.2.2 Empirical studies on the SVLR in controlled speech

While the impressionistic accounts on the SVLR provide detailed descriptions of the vowels and contexts of Aitken's Law, they lack empirical evidence. Before the end of the 1970s, there were virtually no studies which provided precise durational measurements of Scottish vowels in SVLR short and long environments. This development, however, changed with Derrick McClure's study in 1977. The researcher wanted to find out whether the influences of SVLR long and short contexts on vowel duration can be confirmed experimentally: Vocalic durations should be longer in stressed open syllables and when followed by voiced fricatives, /r/ and morpheme boundaries. All the other contexts should condition shorter durations, only the vowels /ɪ/ and /ʌ/ should be short in all contexts. To test these hypotheses, McClure (1977) compiled word lists with the vowels /i/, /ɪ/, /e/, /ɛ/, /ʌ/, /a/, /ɔ/, /o/, /u/ and the diphthongs /ae/ and /ʌu/ in stressed open syllables, before word-final /t/, /s/, /d/, /z/, /r/ as well as before inflectional /#d/ and inflectional /#t/. The diphthong /ae/ represents two realizations here, namely [ʌɪ] pronunciations in SVLR short contexts and [ae] realizations in SVLR long contexts. Derrick McClure himself, who is from Ayrshire, read out the words three times; twice in isolation and once embedded in the carrier sentence "I say (…) sometimes" (1977, p. 10). He used a four-channel electric kymograph for the recordings and averaged the two measurements of the word list readings. The detailed measurements are listed in Table 7 and plotted in Figure 6.

Table 7. Vowel durations by McClure (1977, pp. 12–13) for 11 vowels and 8 contexts measured in centiseconds. (wl = word list reading; cs = carrier sentence reading)

Context	**Long monophthongs**											
	/i/		**/e/**		**/a/**		**/ɔ/**		**/o/**		**/u/**	
	wl	cs	wl	cs	wl	cs	wl	cs	wl	cs	wl	cs
open syllable	31.5	27	36	25	39	35	41.5	33	39	35	37.5	29
t	12	11	20	19	20.5	19	18.5	15	18	16	13.5	10
s	14.5	12	20.5	18	24.5	21	22	19			15.5	11
d	13	10	21.5	20	26	22	23.5	21	23	20	13	9
z	25.5	21	29	23	31.5	29	30	24	27.5	24	28	23
r	28.5	22	31	25	32	28	29	26	31	26	29	22
#d	28	22	31.5	26	35	32	33	28	31.5	27	34	25
#z	30	26	35	28	39.5	31	36	32	37	33	34.5	32

	Short monophthongs						**Diphthongs**			
	/ɪ/		**/ɛ/**		**/ʌ/**		**/ae/**		**/ʌu/**	
	wl	cs	wl	cs	wl	cs	wl	cs	wl	cs
open syllable							44.5	40	40	34
t	8.5	6	19	16	11.5	11	23	22	23.5	19
s	10.5	9	21	18	15	11	24	19	23	17
d	9.5	8	22	19	12	10	32	26		
z	8.5	8	28	25	13.5	10	40	35	31.5	26
r	10.5	9	31	27	15.5	12	40.5	32	32.5	31
#d							43	34	36	31
#z							44	38	38.5	32

Table 7 reveals a general trend that vowels are always longer in isolation than when pronounced in carrier sentences. This is true for all measurements. It is also noticeable that diphthongs and low vowels tend to be longer than high vowels which corresponds with previous accounts of vowel intrinsic durational properties (see subsection 3.1.1). The findings further reveal that SVLR long contexts condition longer durations than SVLR short contexts across all vowels. However, the durational differences between SVLR long and short contexts vary. The greatest vocalic durations are found in open syllable position and before morpheme boundaries. Voiced fricatives and /r/ also condition longer vocalic intervals than postvocalic voiceless fricatives and stops, but the differences are much more pronounced in the high vowels /i/ and /u/ than in the vowels /e/, /ɛ/, /a/, /ɔ/, /o/. The vowels "/i/ and /u/ differ from the remaining 'lengthenable' vowels in being consistently shorter in the short environments and showing a more decisive break between the figures for the 'long ' and for the 'short' group" (McClure, 1977, p. 16). The only vowels which are short in all contexts are /ɪ/ and /ʌ/ which corroborates previous observations of their invariable quantity (Aitken, 1981; Wettstein, 1942; Zai, 1942). Apart from these findings, McClure (1977, p. 14) also reports a variable VE in the vowels /ɛ/, /a/, /ɔ/, /o/, /ae/ and /ʌu/ with increased vowel duration before word-final /d/ than before word-final /t/. The VE is strongest in the [ʌɪ] pronunciation of /ae/ as the vowel in <tide> is 90 ms longer than the equivalent vowel in <tight> when read in isolation. The VE is, however, non-existent in the other vowels /i/, /e/, /u/, /ɪ/ and /ʌ/. Sometimes, there are even shorter durations before word-final /d/ than before word-final /t/ (see Table 7).

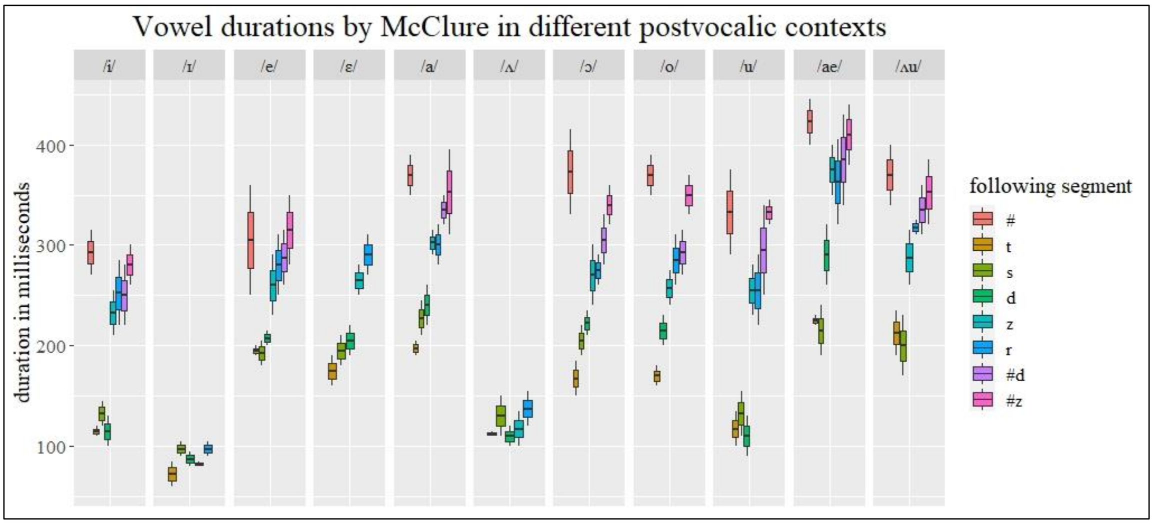


Figure 6. Averaged vowel durations by McClure's (1977) word list and carrier sentence readings.

McClure's (1977) study could provide the first systematic empirical evidence that SVLR long contexts condition longer vocalic durations than SVLR short contexts. However, the durational differences between SVLR long and short contexts are most consistent in the high vowels /i/ and /u/ which validates previous impressionistic observations by Murray (1873, p. 98) and Dieth (1932, p. 61).

Overall, the study by McClure (1977) marks a new approach in SVLR research: most subsequent studies investigated Scottish vowel duration on an empirical basis in an experimental setting. The study was, however, also criticized on grounds of objectivity because the researcher himself, who knew the purpose of the investigation, was also the one who read out the word lists (Agutter, 1988a, p. 12).

In contrast to McClure's findings, Alex Agutter challenged the proposed uniqueness of SVLR durational patterns among the varieties of English in her two articles (1988a, 1988b). She set out an experimental study investigating the durations of five vowel phonemes (/ɪ/, /ɔ/, /i/, /aɪ/ and /au/) in six SVLR long (following morpheme boundary /#/, inflectional /-d/, /ð/, /r/, /v/, /z/) as well as seven SVLR short environments (following /d/, /b/, /n/, /t/, /p/, /s/, /f/) of monosyllabic words. The words were read in a carrier sentence by four SSE and two RP students of both sexes in a soundproofed recording studio. Her later study also included another male speaker from Belfast (Agutter, 1988a), but his vocalic durational patterns will not be discussed here as the focus of the present study is on vowel duration in Scotland. All SSE students were from Edinburgh and the RP students from different parts of England. For the sake of comparability, Agutter (1988b, p. 123) implemented the same carrier sentence as McClure (1977, p. 10) and used a digital sonagraph for the durational measurements. Based on the earlier account by Aitken (1981), the researcher hypothesized that while /i/, /aɪ/ and /au/ should show variable durations in SVLR short and long contexts, the phoneme /ɪ/ should be invariably short and /ɔ/ generally long in all contexts among the SSE speakers. Agutter measured each individual vowel articulation and also calculated average vowel durations for each individual informant (1988b, p. 123). Due to high idiolectal variation, the researcher weighted the data by calculating context dependent length

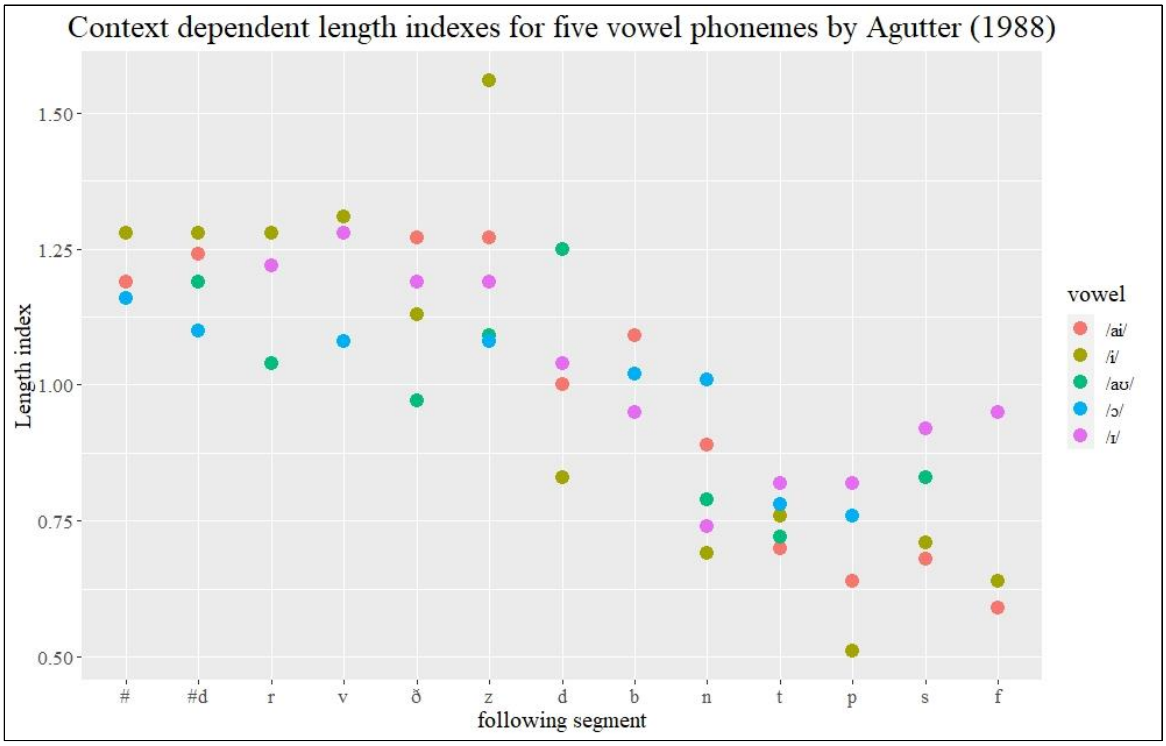


Figure 7. Context dependent length indexes for the five vowel phonemes /ɪ/, /ɔ/, /i/, /aɪ/ and /au/ in different postvocalic contexts by Agutter (1988b).

indexes and context independent vowel durational measurements. As for the context dependent length indexes, she first calculated average durations of each individual vowel in the different environments for the respective speaker groups (*v-con values*). This means that, for example for the diphthong /aɪ/ in a postvocalic /-t/ context, she averaged the measurements of the word <sight> (postvocalic context: /t/) for the two RP and the four SSE speakers respectively. She then averaged all v-con values for each of the five vowels and this provided the so-called v-tot values. In other words, the v-tot value for the diphthong /aɪ/ is the average of the 13 speaker-based v-con values of the diphthong /aɪ/. The individual v-con values were then divided by the v-tot values and, again, averaged for the different consonantal environments. As a result, a length index higher than 1.0 characterizes a comparatively long consonantal environment (v-con value is higher than v-tot value) and a length index below 1.0 represents a comparatively short consonantal environment (v-tot value is higher than v-con value). A visualization of the length indexes can be found in Figure 7. The length indexes reveal that the SVLR long environments generally condition longer vocalic durations across all phonemes investigated (see Figure 7): the vowels are longer when followed by a pause, a morpheme boundary, /r/ or voiced fricatives. In contrast, the SVLR short contexts /n/, /t/, /p/, /s/ and /f/ condition shorter durations with index values below 1.00. The following voiced plosives /d/ and /b/ show index values above and below 1.00 and are therefore intermediate in duration (Agutter, 1988b, p. 127).

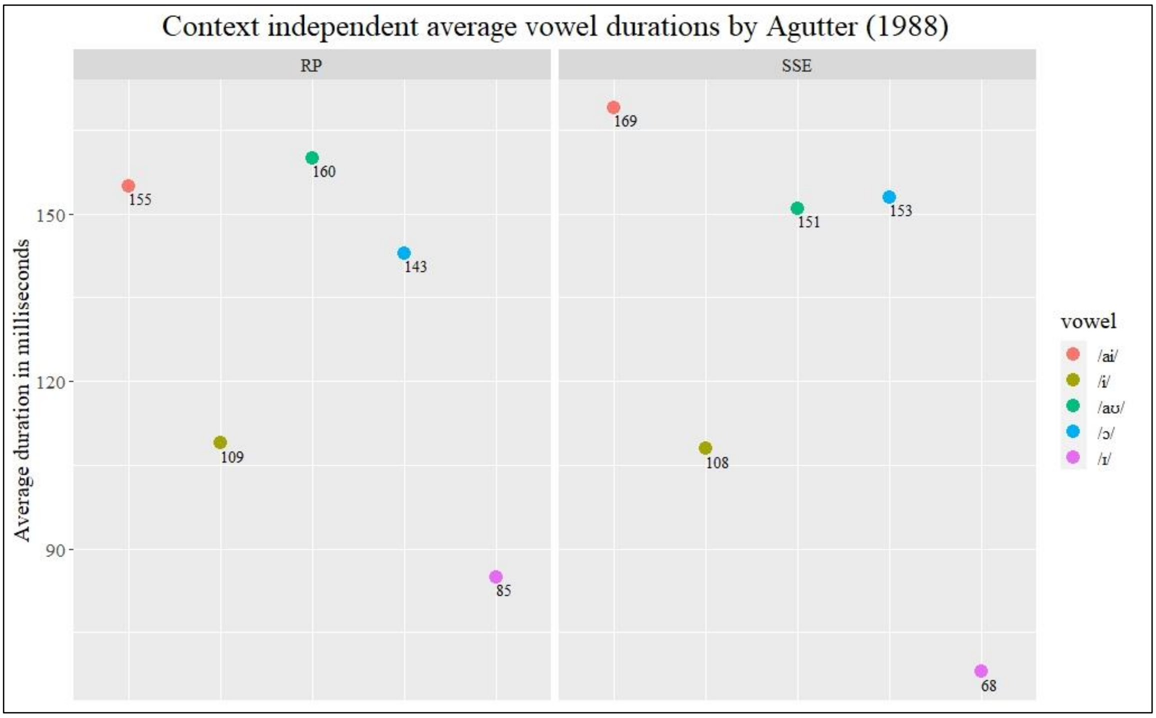


Figure 8. Context independent average vowel durations for the RP and SSE speakers by Agutter (1988b).

As for the context independent vowel durational measurements, Agutter (1988b) first calculated the average length of a particular vowel for a particular informant irrespective of the environment. The values were then multiplied by the ratio of 13 divided by the average vowel duration of the informant and these weighted mean values were then again averaged for the respective groups SSE and RP

(Agutter, 1988b, p. 126). An overview of the context independent average vowel durations can be found in Figure 8. Figure 8 shows that the vowels /ɪ/ and /ɔ/, which were reported as being unaffected by the SVLR, are nevertheless influenced by the following segment. Hence, although it is true that /ɪ/ is the shortest and /ɔ/ the longest monophthong for SSE speakers (see Figure 8), this does not mean that they are invariable in terms of their duration (Agutter, 1988b, p. 129). Figure 8 also reveals that the durational difference between /ɪ/ and /ɔ/ is wider among SSE speakers which corresponds to Aitken's (1981, 149–150) previous observation. In addition, /aɪ/ is slightly longer among SSE speakers than among RP informants which, however, may be due to quality differences in Scottish English (Agutter, 1988a, p. 9). Agutter also compares her results with the findings by McClure (1977) and applies the normalization procedure of context independent average vowel durations on his data. On the whole, McClure's normalized durations are comparable to those of the SSE speakers in Agutter's study (1988b, p. 127). McClure's overall greater range of vowel durations, however, seems to be exaggerated (Agutter, 1988a, p. 11). In this context, Agutter (1988a, p. 12) questions the objectivity of McClure's vowel articulations as the researcher knew the purpose of his own investigation. Apart from that, Agutter (1988b, p. 127) reports that the length indices show no striking differences when separated for SSE and RP speakers. Hence, while the measurements do reveal context dependent SVLR shortening and lengthening effects across different vowels, there is not a significant difference in vowel articulation between the SSE and RP groups (see Figure 8). The results therefore show the operation of Aitken's Law, but do not support the view that the vowel durational patterns are distinctively Scottish (Agutter, 1988b, p. 129). Due to the similar durational patterns in the different vowels among both SSE and RP informants, Agutter (1988b) criticizes the geographical scope of the SVLR as well as its range of vowels proposed by Aitken (1981). She suggests that all vowels may be subject to context dependent length variation and that there is a further interaction with intrinsic vowel duration (Agutter, 1988a, p. 16, 1988b, p. 129) (see subsection 3.1.1). In addition, due to the inconclusive findings by Lodge (1984), Agutter further assumes that there could be other factors in connected speech which may supersede the influence of Aitken's Law. Due to the fact that almost all previous accounts on Aitken's Law are based on perceptual judgement of Scottish speech, the researcher expresses the need for a large-scale study investigating vowel durational patterns with precise machine measurements (Agutter, 1988b, p. 131).

The study by Agutter was the first experimental investigation in SVLR research which compared two varieties and weighted its durational measurements. While the weighting procedure facilitates a comparison between groups (SSE and RP), this normalization technique was later criticized because it assumes that variation is normally distributed (McMahon, 1991, p. 40).

The subsequent investigation by Gordon McKenna (1988) comes close to Agutter's call of a large-scale empirical study on Scottish vowel duration. After McClure (1977) and Agutter (1988b), McKenna (1988) carried out the third and arguably most sophisticated experimental study on the SVLR and his findings do also not fully correspond with Aitken's (1981) proposed set of SVLR affected vowel phonemes. In the first stage of his analysis, the researcher used a list of monosyllabic CVC words containing each of the nine monophthongs of the Basic Scottish Vowel System (/i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/, /ɪ/, /ʌ/) followed by each of the four tautosyllabic alveolar consonants /t/, /d/ /s/ and /z/ (McKenna, 1988, p. 88). In the second stage of analysis, he used another word list with the six monophthongs /i/, /e/, /ɛ/, /ɔ/, /o/ and /u/ followed by either tautosyllabic /d/ and /z/ or heterosyllabic /#d/ and /#z/ after a morpheme boundary. His informants were two female and two male university students from East Central Scotland and he successfully tested their vowel inventories for the Basic Scottish Vowel System before the analysis (McKenna, 1988, pp. 90–91). The SSE informants were seated in a soundproof recording studio and read the individual words in the frame sentence “Say _______ again” two times amounting to 94 utterances per speaker (McKenna, 1988, p. 91). McKenna (1988) controlled for most factors which can influence vowel duration (i.e. tempo, stress, phrasal position, intrinsic vowel duration, voicing and manner of the following consonant, idiolectal variation) and he applied a multivariate analysis of variance. He also specified his approach toward phone segmentation (McKenna, 1988, pp. 93–95). His findings show that Aitken’s Law operates only among the high vowels /i/ and /u/. These are the only vowels which are significantly longer in the tested SVLR long contexts, namely when followed by the voiced fricative /z/ or a morpheme boundary, but short before /t/, /d/ and /s/ (McKenna, 1988, p. 145). In addition, the vowels /ɪ/ and /ʌ/ “do not show any significant differences in duration according to the phonological context (…)” (1988, p. 145) which corresponds to Aitken’s (1981) assumption that they are not subject to the SVLR. However, there is a difference between the short vowels with /ʌ/ being significantly longer than /ɪ/. This corresponds to another important influence on vowel duration detected in McKenna’s (1988, p. 143) experiment, namely vowel height. There was a consistent pattern that open vowels are longer than closed vowels which corresponds to the effect of intrinsic vowel duration (see subsection 3.1.1). In addition, the open and mid vowels /a/, /e/, /ɛ/, /ɔ/ and /o/ are consistently longer before the voiced consonants /d/ and /z/ than before their voiceless counterparts /t/ and /s/ (McKenna, 1988, p. 144). While this fully corresponds with the VE, it partly contradicts the SVLR because environments with following voiced stops should, according to Aitken (1981), condition short durations. Furthermore, morpheme boundaries do not lead to increased duration among the mid and open vowels either. McKenna’s study (1988) therefore shows that while Aitken’s Law operates in the high vowels /i/ and /u/ in SSE, which further corroborates the findings by McClure (1977) and the observations by Murray (1873, p. 98) and Dieth (1932, p. 61), a significant durational difference between SVLR long and short contexts is not detectable in the vowels /a/, /e/, /ɛ/, /ɔ/ or /o/. This, again, bears evidence for a more limited SVLR vowel set than originally proposed by Aitken (1981) and further indicates that the VE operates among the mid and open vowels of the four investigated SSE speakers.

While the outline of McKenna’s (1988) study is similar to the study designs by McClure (1977) and Agutter (1988b), his approach to data preparation and his statistical analysis is more sophisticated. McKenna (1988) is the first SVLR researcher who specifies his segmentation procedure and he is the first to apply a multivariate analysis of variance on Scottish vowel duration. He is fully aware of the different segmental and suprasegmental factors that can influence vowel duration and he is also the first who tests the vowel inventories of his informants. The findings by McKenna (1988) are therefore very reliable.

Whereas April McMahon (1991) largely discusses Aitken’s Law against the background of lexical phonology, she also performs a reanalysis of Agutter’s data which has had important implications for SVLR research. She first introduces the SVLR with reference to Wells (1982) which generally corresponds with Aitken’s (1981) description. That is, Aitken’s Law applies to the monophthongs /i/ /u/ /e/ and /o/, to /a/ and /ɔ/ in some dialects whereas they are generally long in other varieties and to the first element of the diphthong /ʌɪ/ of the PRICE lexical set (McMahon, 1991, p. 33). The effect of Aitken’s Law on the diphthongs /au/ and /ɔɪ/ is unclear. The monophthongs /ɪ/, /ʌ/ and /ɛ/ are described as non-lengthening lax vowels even though the status of /ɛ/ remains contradictory (McMahon, 1991, p. 34). McMahon’s main concern about previous SVLR studies is that Aitken’s Law is seen as an isolated phenomenon. Other factors which influence vowel duration (see section 3.1) are often not accounted for. Whereas other scholars have already noted the influence of the postvocalic consonantal context on vowel duration in other varieties of English (Agutter, 1988b; Aitken, 1981, 137; McKenna, 1988, p. 24), she is the first scholar who points out that Aitken’s Law is linked to another vowel lengthening process, namely the VE (see subsection 3.1.2), or, as McMahon (1991, p. 38) calls it, Low-Level Lengthening (henceforth: LLL). She proposes that both SVLR and LLL operate in partially overlapping contexts in Scots and SSE: while LLL operates before all voiced consonants, Aitken’s Law only applies before voiced fricatives, /r/ as well as morpheme boundaries. She therefore reconsiders Agutter’s study with the assumption that LLL alone applies in RP, but that the two overlapping processes of Aitken’s Law and the LLL operate in SSE simultaneously. In addition, McMahon (1991, p. 40) did not apply the weighting procedure by Agutter since it assumes that any variation found will be normally distributed. However, a skewed distribution would be more reasonable given that Aitken’s Law is an accent specific process affecting only particular vowels in specific contexts: therefore, Agutter’s normalization might have hidden the exact variation that would have been of interest (McMahon, 1991, p. 40). Instead, McMahon (1991) calculated simple means and standard deviations as well as a ratio between short and long environments. Based on the assumption that the SVLR interacts with the LLL effect, the study differentiated between three consonantal context groups: short environments (following /f/, /t/, /s/, /p/), LLL long environments (following /b/, /d/, /n/) and SVLR long environments (following morpheme boundary or/and following /ð/, /r/, /v/, /z/). Due to Aitken’s (1981) vowel classification, McMahon (1991, p. 41) sorted /ʌɪ/ and /i/ into one group (SVLR affected phonemes) and /ɔ/ and /ɪ/ into another (SVLR unaffected phonemes). The diphthong /au/ was considered individually as its status is unclear. McMahon's (1991) findings are visualized in Figure 9.

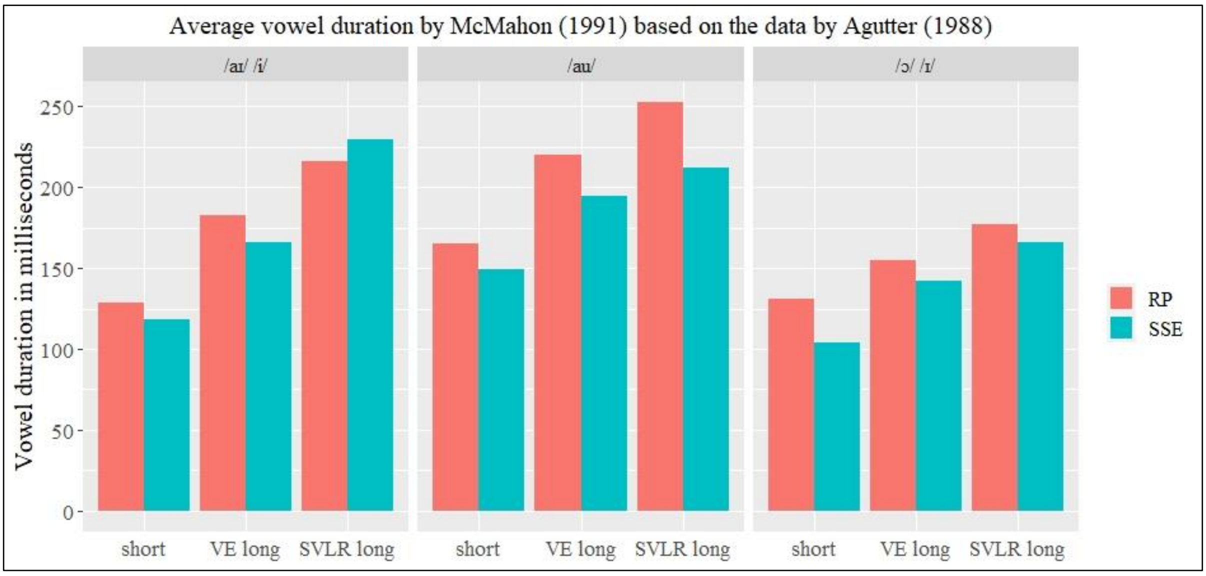


Figure 9. Average vowel durations for different vowel classes and contexts by McMahon (1991) based on the data by Agutter.

As illustrated in Figure 9, the vocalic durations of RP speakers are generally longer than those of the SSE informants. The only exception are the SVLR affected vowels /i/ and /ʌɪ/ where this relationship is reversed (McMahon, 1991, p. 41). This tendency of SVLR lengthening in /i/ and /ʌɪ/ becomes even clearer when vowel duration is recalculated as a percentage of the duration in the short contexts (McMahon, 1991, p. 44). The researcher interprets this additional lengthening as a sign that both LLL and Aitken's Law operate among SSE speakers: LLL produces the lengthening shared by RP and SSE whereas the "SVLR accounts for the peculiarly Scottish ADDITIONAL lengthening which affects /ʌi i/ (and also /u e o/ and perhaps /a ɔi/, although these were not tested by Agutter) in the traditional SVLR environments" (McMahon, 1991, p. 44).

The reanalysis by McMahon (1991) provides a new interpretation of the relationship between Aitken's Law and the VE. The study also shows that the methodological approach can have a profound effect on the results in a study. Although she used the exact same dataset as Agutter (1988b), she grouped the vowels differently and she did not weight the data which led to completely different results. The vowel groups by McMahon's (1991) were, however, also criticized in subsequent studies (Scobbie, Hewlett, & Turk, 1999, 236 f.).

In his small study on Northumbrian English, a variety outside Scotland, Milroy (1995) could find that the SVLR applies most strongly in the diphthong /aɪ/. His results also indicated that SVLR patterns are less consistent among younger speakers, thus suggesting age-related variation. The study by Milroy demonstrated that SVLR patterns are not confined to Scotland alone but do also occur in the Northeast of England.

A first meta study on Aitken's Law was conducted by Scobbie, Hewlett, and Turk (1999). The researchers took into account the previous instrumental studies on Aitken's Law by Agutter (1988b), McClure (1977), McKenna (1988) as well as McMahon (1991) and they also conducted two instrumental studies on vowel duration patterns in SSE themselves (Hewlett et al., 1999; Scobbie, Turk, & Hewlett, 1999). The researchers made a clear distinction between consonantal and morphological SVLR conditioning. Whereas consonantal SVLR conditioning occurs in the word <breeze> due to the postvocalic voiced fricative /z/, morphological SVLR conditioning occurs in the word <brewed> due to the following morpheme boundary. Similar to McMahon (1991), Scobbie, Hewlett, and Turk (1999, p. 236) also criticize the study by Agutter for its limited dataset and inadequate normalization procedure. They do, however, also challenge the reanalysis by McMahon (1991) on conceptual and methodological grounds. On the conceptual level, the researchers do not share the assumption that LLL operates in both RP and SSE as this would cross-cut the distinction between consonantal and morphological SVLR environments (Scobbie, Hewlett, & Turk, 1999, p. 236). Instead, they propose that "Scottish varieties and RP have their own language-specific, partially phonetically motivated systems sitting on top of a more universally natural phonetic base" (Scobbie, Hewlett, & Turk, 1999, p. 236). As for methodology, they criticize McMahon's (1991) inadequate pooling of the diphthong /aɪ/ with the monophthong /i/ as well as the combination of /ɔ/ and /ɪ/. As for the first pair, the diphthong /aɪ/ is much longer and less flexible in its duration than /i/. In addition, whereas /i/ only varies in terms of its quantity in Scottish English, the diphthong /aɪ/ "has quality *and* quantity differences in its allophones (short [ʌɪ] and long [ae]] (…)" (Scobbie, Hewlett, & Turk, 1999, p. 236). The pooling of /ɪ/ and /ɔ/ is also misleading as the former phoneme is generally short and does only occur in closed syllables. The phoneme /ɔ/, however, is long and can occur in closed and open syllables, such as <lord> or <law>, respectively. The substantial durational differences between the two pooled vowel sets is also detectable in Agutter's (1988b) raw data (Scobbie, Hewlett, & Turk, 1999, p. 237). When the researchers reassessed the individual vowel measurements of Agutter without any pooling and with the clear distinction of consonantal and morphological conditioning, they also find a clear durational contrast between SVLR short and long environments in /i/ and /aɪ/. The vowels /ɪ/, /ɔ/ and the diphthong /au/, however, did not show any SVLR effect which, again, partially contradicts Aitken's (1981) description of the rule (see Table 5). Scobbie, Hewlett, and Turk (1999, p. 237) therefore conclude that "Agutter's study provides an indication that Aitken's SVLR does not transfer in a simple fashion to SSE". Apart from Agutter's data, the researchers also summarize the studies by McClure (1977) and McKenna (1988), the latter of which they regard as the most detailed and informative. The researchers conclude what has been summarized before: the investigations by McClure (1977) and McKenna (1988) both indicate that only the phonemes /i/, /u/ and /aɪ/ are subject to Aitken's Law in both consonantal and morphological contexts (Scobbie, Hewlett, & Turk, 1999, p. 239).

The related instrumental investigation by Hewlett et al. (1999) found similar results. The researchers investigated SVLR and VE patterns in four minimal and near-minimal pairs with the vowels /i/ and /u/

among seven middle-class children living in Corstorphine, Edinburgh. An overview of the example words can be found in Table 8.

Table 8. Experimental items of the study by Hewlett et al. (1999).

VE environment		**SVLR environment**	
voiceless plosives	voiced plosives	voiceless fricatives	voiced fricatives
<seat>	<seed>	<fleece>	<please>
<foot>	<food>	<loose>	<lose>

All children were aged six to nine years and two of them had two Scottish English speaking parents. Two other children had one Scottish English speaking parent and one parent who spoke another English accent and the remaining three infants had parents neither of whom spoke Scottish English (Hewlett et al., 1999, p. 2158). Applying an acoustic analysis on each token, the researchers determined the start of the vowel by the onset of the first formant and the end of a vowel by the offset of the second formant (Hewlett et al., 1999, p. 2158). The results demonstrated that the SVLR is firmly established in the speech of children who have a Scottish English family background (Hewlett et al., 1999, p. 2160). The findings also showed that the VE applies to only a minimal degree in the analyzed contexts, thus contradicting both Agutter (1988b) and McMahon (1991): "contra Agutter, differential lengthening in SVLR and VE environments and, contra both, only very modest lengthening in the VE environment" (Hewlett et al., 1999, p. 2160).

The study concludes that an exposure to a non-Scottish accent, for example RP, weakens the durational patterns of Aitken's Law. Therefore, it is one of the first studies which concludes that language-external factors can influence the SVLR.

The other related study (Scobbie, Turk, & Hewlett, 1999) could provide further evidence for SVLR patterns in the vowels /i/, /u/ and /aɪ/. In this investigation, the researchers focused on the morphological conditioning of the SVLR among Glaswegian male and female speakers from two different age groups and two social classes (working class and middle class). The researchers used a word list with near-minimal pairs with the vowel nuclei /i/, /u/, /aɪ/ /o/ and /ɔ/ in contexts with following tautomorphemic /d/ and inflectional /#d/ as in <greed> and <agreed>. The results fully confirm the findings by McKenna (1988) that SVLR are found in the vowels /i/ and /u/ but not in /o/ and /ɔ/ (Scobbie, Turk, & Hewlett, 1999, p. 1619). Furthermore, the results are consistent across the two different sexes, age groups and social class backgrounds. In other words, sociolinguistic variation in SVLR durational patterns could not be found in the speaker sample from Glasgow. Aitken's Law appears to be stable across the Glaswegian community investigated in this study.

The findings by Scobbie, Hewlett, and Turk (1999) therefore contradict Milroy's (1995) results. Whereas Aitken's Law seems to be weakening among the younger generation in Tyneside, it is firmly established in Glasgow across different age groups.

In contrast to this, Watt and Ingham (2000) could also detect age- and gender related variation in SVLR patterns in Berwick-upon Tweed, the northernmost town of England. As Berwick English shares many Scottish phonological features, the researchers' interest was whether Aitken's Law also operates in that particular rural variety. For that purpose, they recorded eight speakers from Berwick, four males and four females in two distinct age groups, who were reading out a word list with 66 items. Due to the findings by Hewlett et al. (1999), the researchers took the parental background of the informants into account. The word list comprised the nine monophthongs /i/, /e/, /ɛ/, /a/, /ɑ/, /ɔ/, /o/, /u/ as well as the diphthong /aɪ/ in voiced and voiceless fricative and stop contexts testing for the influence of the VE and the SVLR. The researchers carried out an acoustic analysis applying the segmentation scheme by Hewlett et al. (1999). The durational measurements were then transferred into voicing ratios in order to maximize comparability between speakers. That is, the vocalic durations in voiced stop contexts were divided by their equivalents in voiceless stop contexts to test the strength of the VE. Likewise, the durations in voiced fricative contexts were divided by their equivalents in voiceless fricative environments to account for the influence of Aitken's Law. The division of the SVLR ratio by the VE ratio then indicated whether SVLR contexts account for an increased lengthening or not. The results for the four older speakers confirmed that the VE as well as the SVLR operate independently in Berwick English. Furthermore, the findings demonstrated that the effects of Aitken's Law are much stronger in the vowels /i/, /u/ and in the diphthong /aɪ/. The effects of SVLR vowel lengthening are less pronounced especially among the open vowels which corroborates earlier findings. The results for the younger speakers are, however, less consistent. While the young male speakers show a similar but weaker pattern of SVLR lengthening, the young females show a tendency in which the effects of the SVLR are neutralized relative to the effects of the VE (Watt & Ingham, 2000, p. 222). In particular, the diphthong /aɪ/ does not show SVLR-related lengthening among the young female speakers.

The study by Watt and Ingham (2000) therefore suggests that while Aitken's Law can be detected in Berwick English, its effects are weakened across younger and especially younger female speakers. This trend implies a convergence of the Berwick English vowel system towards an "VE-only model typical of southern forms of [SSBE]" (Watt & Ingham, 2000, p. 222).

van Leyden (2002) comes to a similar conclusion in her study on the relationship between vowel and consonant duration in Shetland, Orkney and Edinburgh. She carried out three identical production experiments with 13 natives from Shetland, 12 natives from Orkney and 12 native SSE speakers from Edinburgh. The methodology was highly similar to the one by McClure (1977) and she used the same carrier sentence as McKenna (1988). As for her experiment on the SVLR, the researcher included monosyllabic words with the short monophthong /ɪ/ and the long high vowel /i/ in six SVLR long and fifteen SVLR short contexts. Although aspiration was considered part of the vowel, the effect of vowel onset time (henceforth: VOT) on vowel duration appeared to be negligible (van Leyden, 2002, p. 5). The findings show that the SVLR still applies strongly in the Shetland dialect and conditions relatively short vowels in SVLR short contexts and significantly longer vocalic intervals in SVLR long environments. The results for Orkney and Edinburgh are less consistent as there is also increased vowel duration before following tautomorphemic /d/. The researcher suggests that this levelling of vocalic durations towards the VE in Orkney and Edinburgh might be induced by the influence of SSBE (van Leyden, 2002, p. 14).

Hence, van Leyden (2002) also suggests that weakening of SVLR patterns in Orkney and Edinburgh might result from an increased influence of SSBE. This is in line with the findings by Hewlett et al. (1999), Milroy (1995) as well as Watt and Ingham (2000).

Scobbie (2005) published another study which investigated the effects of the SVLR and the VE among 12 young adult speakers from Shetland. Similar to Hewlett et al. (1999), Scobbie took the parental background of his subjects into account: four of the informants' parents were both from Shetland, four other adolescents had one parent from Shetland and one parent from mainland Scotland and the remaining four speakers were brought up by a parent from Shetland and a parent from England. Thus, the study analyzed the speech of monolingual subjects who were brought up in the same community, but who were exposed to different accents of English because of their parents (Scobbie, 2005, p. 2). The sample was balanced in terms of gender with six male and six female informants. Similar to previous experimental studies, Scobbie (2005) used word list readings with the vowels /i/, /e/, /a/, /ɔ/, /o/, /u/ followed by tautosyllabic /t/ and /d/ to test the influence of the VE and following heterosyllabic /#d/ to test for the morphological conditioning of the SVLR. The results demonstrated that Aitken's Law is indeed present among the young informants in the high vowels /i/ and /u/, especially among the speakers whose parents are from Shetland and mainland Scotland. This is also largely the case for the male informants who have one English parent. However, the female participants with one English parent showed a pattern in which the increase in vowel duration is far greater in VE-long contexts than in a morphological SVLR environment. Scobbie (2005, p. 7) therefore concludes that the female speakers with one parent from England adopted an SSBE vowel system despite being brought up in Shetland.

These findings further support the notion that the parental accent has a profound influence on a speaker's vowel system. As a result, also the study by Scobbie (2005) suggests that a weakening of SVLR patterns and a shift towards the VE is likely to be influenced by dialect contact and an exposure to SSBE. Furthermore, the study could corroborate previous accounts that SVLR patterns are present in the high vowels /i/ and /u/.

Monika Pukli (2006) could also detect Aitken's Law among the high vowels in her PhD thesis. She investigated the speech of 12 informants from the town of Ayr in the Southwest of Scotland. The sample includes 7 female and 5 male speakers of different ages (Pukli, 2006, pp. 141–142). Similar to previous investigations, the informants had to read a list of mostly monosyllabic words in a carrier sentence to minimize the influence of suprasegmental conditioning factors on vowel duration (Pukli, 2006, p. 133). In the first part of the analysis, she investigated the vowel articulations of all 12 informants for /i/, /u/ and /aɪ/ followed by tautosyllabic /t/ and /d/, heterosyllabic /#d/ as well as open syllable boundaries to account for the influences of the VE and the SVLR. In the second stage, she investigated all vowels of the Basic Scottish Vowel System in the same environments across six speakers. She then also conducted another analysis on SVLR patterns in other consonantal contexts with just one speaker (Pukli, 2006, p. 146). The spectrographic analyses revealed that the morphological conditioning of the SVLR affects only /i/, /u/ and /aɪ/. The other vowels do not show a significant increase in duration before heterosyllabic /#d/ which contradicts McClure's (1977) findings for his Ayrshire accent. Pukli (2006) could also detect a modest VE operating across most vowels, with the exception of the high vowels /i/ and /u/ in which the VE is not detectable.

The study by Pukli (2006) therefore provides further evidence that the SVLR operates across /i/, /u/ and /aɪ/ in the SSE of Ayrshire. In contrast to McClure (1977), the findings by Pukli (2006) are more reliable because she interviewed several informants with a structured procedure.

In contrast to the widespread evidence for SVLR patterns in /i/, /u/, and /aɪ/, Watt and Yurkova (2007) found less conclusive results in their instrumental investigation of Aberdeen English. The researchers recorded the speech of nine adult speakers, five males and four females, with an age range from 21 to 62 years. The informants, who were all from Aberdeen, read aloud a list of monosyllabic words including minimal and near-minimal pairs with the vowels /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/ and the diphthong /aɪ/ positioned in open syllables or followed by tautomorphemic /t/, /d/, /s/, /z/ or heteromorphemic /#d/. Similar to Watt and Ingham (2000), the researchers used ratios to normalize the duration of the measured vocalic intervals and to account for the effects of the SVLR and VE. To be precise, they calculated ratios for the respective durational increases between voiceless and voiced stop contexts (VE) and voiceless and voiced fricative contexts (SVLR) assuming additional lengthening in the latter environments due to Aitken's Law (McMahon, 1991). They also calculated ratios comparing the vocalic durations in tautomorphemic /d/ contexts (e.g. <brood>) and heteromorphemic /#d/ environments (e.g. <brewed>) to account for the morphological conditioning of the SVLR. The findings show that there is VE-related lengthening in almost all vowels and across all speakers in tautomorphemic /d/ contexts. The results for the Aitken's Law are, however, less clear. As for the phonological conditioning of the SVLR, there was only one middle-aged female speaker who always had a higher fricative ratio than plosive ratio. Most other speakers often showed a higher plosive ratio contradicting the assumed additional lengthening of Aitken's Law. As for the vowels, Watt and Yurkova (2007, p. 1523) found no phonological SVLR vowel lengthening in /i/, /u/ or /aɪ/. Instead, the ratios suggest that the vowel /e/ conforms most consistently to both the VE and to the SVLR whereas the diphthong /aɪ/ conforms least to that pattern. This trend is, however, reversed for the morphological SVLR contexts: the diphthong /aɪ/ as well as the monophthongs /a/ and /ɔ/ are always longer when followed by heteromorphemic /#d/. A similar trend can be observed for the high vowel /u/ for almost all speakers. In contrast to this, the vowel /e/ is often shorter when followed by a morpheme boundary. Overall, the ratios exhibit a great deal of intra- and interspeaker variability and the lengthening of SVLR long contexts is often very limited (Watt & Yurkova, 2007, p. 1524).

The patterns found by Watt and Yurkova (2007) largely contradict previous findings on Aitken's Law in wordlist readings. A possible conclusion would be that Aitken's Law operates differently in Aberdeen English thus signifying regional variation in the application of the SVLR. Yet, as Watt and Yurkova's (2007) acknowledge, their study is based on a limited dataset and more data is needed to make it more representative.

In terms of its methodology, a more recent study by Llamas et al. (2011) is comparable with Watt and Yurkova (2007). Llamas et al. (2011), however, focused on Tyneside English, a variety in the Northeast of England which was also investigated by James Milroy sixteen years before. In accordance with the findings of most previous studies, the researchers chose to "focus only on the close vowels /iː uː/ and the diphthong /aɪ/, on the assumption that if such a rule [SVLR] is present it is most likely to affect these three vowels" (Llamas et al., 2011, p. 1283). Similar to Watt and Yurkova (2007), the researchers prepared a word list with monosyllabic words followed by voiced and voiceless plosives as well as voiced and voiceless fricatives to test for the influence of the VE and Aitken's Law, respectively. They also included words with following heterosyllabic /#d/ and /#z/ to test for the morphological conditioning of the SVLR. Eight men from Newcastle and Gateshead aged between 25 and 68 years read out the word lists twice, with the words arranged in different orders. Llamas et al. (2011) recorded the speakers and averaged the durational measurements of the two readings. They then calculated VE and SVLR ratios following the approaches of Watt and Ingham (2000) and Watt and Yurkova (2007). The VE ratios are all positive indicating a moderate and sometimes strong VE across all speakers and vowels. The SVLR ratios show a similar picture, but there is great interspeaker variation, especially in the close vowels /i/ and /u/. There are speakers whose vowels are twice as long in voiced fricative contexts than in voiceless fricative environments, but there are also two speakers with negative ratios for /i/ and /u/. The SVLR ratios are much more consistent and higher in the diphthong /aɪ/ with a mean ratio of 1.94 across all speakers tested (Llamas et al., 2011, p. 1284). This indicates that the vowel /aɪ/ is almost twice as long when followed by tautomorphemic voiced fricatives than by tautomorphemic voiceless fricatives. A one-tailed t-test also revealed that the SVLR-related durational increase in /aɪ/ is the only significant one of the three vowels investigated (Llamas et al., 2011, p. 1284). As for the comparison of SVLR and VE ratios, the researchers note that there is "no agreed threshold at which the difference between VE- and SVLR-context conditioned lengthening becomes sufficient to establish whether it is valid to talk of an SVLR effect distinct from the VE (…)" (Llamas et al., 2011, p. 1284). Nevertheless, they propose a SVLR-to-VE difference of 20 percent to be a reasonable estimate. Based on this assumption, the comparison of SVLR and VE ratios reveals that Aitken's Law applies most consistently in the diphthong /aɪ/. The effects on /i/ and /u/ are less pronounced and the researchers report no age-related variation. As for the morphological conditioning of Aitken's Law, Llamas et al. (2011) only found a significant lengthening effect in /u/ and /aɪ/ when followed by heteromorphemic /#z/ in words such as <frees> and <lies>. Contrary to most previous studies, the duration of /i/ is not lengthened in postvocalic /#d/ or /#z/ environments and there is also no significant durational increase for /aɪ/ when followed by heteromorphemic /#d/.

Overall, the study by Llamas et al. (2011) corroborates the findings by Milroy (1995) in that the SVLR operates most consistently in the diphthong /aɪ/. However, in contrast to Milroy (1995) and Watt and Ingham (2000), the researchers could not detect age-related variation in the application of Aitken's Law in the Northeast of England.

3.2.3 Empirical studies on the SVLR in uncontrolled speech

Up until 2016, the empirical studies on Aitken's Law analyzed controlled speech elicited in word list or carrier sentence readings. While controlled experimental settings reduce the influence of confounding factors (e.g. phrase-final lengthening) and thus provide a clearer picture of the variable under investigation (e.g. SVLR) (Scobbie, Hewlett, & Turk, 1999, p. 235), it is unclear whether the effects of Aitken's Law also apply in naturally-occurring language. In other words, it could be that the SVLR is only detectable in word list or carrier sentence readings but not in spontaneous speech. The effects of Aitken's Law could be superseded by other segmental or suprasegmental factors (see section 3.1) in uncontrolled spoken language.

The first study which investigated Aitken's Law in spontaneous speech was conducted by Rathcke and Stuart-Smith in 2016 and they focused on the Glaswegian vernacular. They were also the first who carried out a longitudinal study with data from the 1970s and 2000s extracted from the force-aligned *Sounds of the City* corpus (more information on forced alignment can be found in section 4.3). Rathcke and Stuart-Smith (2016) wanted to find out whether the SVLR has changed in Glasgow over time and if so, they also wanted to test whether this change has been promoted by external influences due to dialect contact with SSBE or by internal prosody-related factors. In other words, the researchers wanted to test whether a change in SVLR durational patterns is in line with the *Dialect Contact Hypothesis* or with the *Prosodic Timing Hypothesis*. Whereas the former hypothesis predicts a weakening of SVLR patterns and a shift towards the VE among high-contact speakers, the latter hypothesis predicts a weakening of Aitken's Law in accentuated or phrase-final positions (Rathcke & Stuart-Smith, 2016, pp. 410–411). The researchers analyzed a total of 1520 vowel articulations of 8 male speakers from the 1970s and 8 male speakers from the 2000s. The respective speaker groups were further subdivided into 4 teenage and 4 adult speakers for each period and the study therefore incorporates both an apparent time and real time approach comparing the vowel durations of two different age groups at two different points of time. Based on previous findings, they restricted their investigation to the SVLR-affected high vowels /i/ and /u/ and the SVLR-unaffected vowel /a/ carrying out two separate analyses: one for the high vowels and another for the open vowel. They did not investigate the diphthong /aɪ/ due to its quality changes in different contexts in Scottish English (Wells, 1982, p. 405). To assess the influence of external factors, the amount of every informant's contact with SSBE was described as either high or low

(Rathcke & Stuart-Smith, 2016, p. 411). As for the internal prosody-related factors, the researchers distinguished between three levels of prominence and two phrasal positions. As for the latter, they differentiated between phrase-medial and phrase-final syllables to account for the effect of *constituent-final lengthening* (see subsection 3.1.8). As for prominence, they coded stressed syllables for metrical stress, nuclear stress, or a pitch accent. Here, "metrical stress was labelled in primary stressed syllables which showed metrical prominence (…) but did not carry a pitch accent" (Rathcke & Stuart-Smith, 2016, p. 412). The difference between accented and nuclear syllables was made based on the relative importance of the syllable within the intonational phrase. That is, nuclear accents were observed when the syllable had a high status in the information structure of the utterance (Rathcke & Stuart-Smith, 2016, p. 412). In addition, they also coded for *word frequency* (see subsection 3.1.5), the number of syllables within a word to account for *polysyllabic shortening* (see subsection 3.1.4) and the number of segments within a syllable to account for *intrasyllabic compression* (see subsection 3.1.3) (Rathcke & Stuart-Smith, 2016, p. 413). The latter three factors were included as covariates in the analyses. More importantly, the researchers differentiated between different three SVLR and VE contexts (see Table 9).

Table 9. VE and SVLR context specification by Rathcke and Stuart-Smith (2016, p. 413) with example words.

VE			**SVLR**		
short	long	unspecified	short	long (phonemic)	long (morphemic)
<look> <last> <sweeping>	<good> <fool> <reads>	<new> <see> <two>	<good> <senior> <keeping>	<believe> <excuse> <used>	<wee> <knew> <doing>

The VE contexts include only vowels followed by tautomorphemic consonants and the voicing decision was based on the phonemic voicing status. This means that they followed the phonemic categorization of the consonants, but they did not investigate the acoustic properties of the consonants to see whether the sounds are actually voiced or not. The researchers assumed that an intervening morpheme boundary does not constitute a VE-related environment and thus sorted all respective tokens into the class *unspecified* (Rathcke & Stuart-Smith, 2016, 412). *VE short* tokens are followed by voiceless plosives and voiceless fricatives and the *VE long* tokens are followed by voiced consonants. As for Aitken's Law, *SVLR short* contexts include vowels with following voiced and voiceless stops, voiceless fricatives as well as nasals and this category therefore partially overlaps with the *VE long* category. The class *SVLR short* does also include all /a/-tokens as this vowel is, according to previous studies, not affected by Aitken's Law. The *SVLR long* contexts were further divided into phonemically and morphologically long environments even though this division was collapsed in the final statistical analysis (Rathcke & Stuart-Smith, 2016, p. 415). Phonemically SVLR-long tokens are followed by tautomorphemic voiced fricatives. The class of morphemically SVLR long contexts is identical with the VE class *unspecified*. This includes tokens such as /u/ in the bimorphemic word <doing> but also vowels in constituent-final open syllable positions as in the words <see> or <knew>. As for the statistical analysis and in contrast to all previous studies, Rathcke and Stuart-Smith (2016) applied linear mixed effects models and reported estimates instead of durational measurements. Whereas previous studies used different normalization procedures for their speech data which was elicited in controlled settings (see subsection 3.2.2), Rathcke and Stuart-Smith chose linear mixed effects models as they can cope well with unbalanced datasets typical of spontaneous speech (2016, p. 414). They applied a backward fitting procedure and the best model fit was established through likelihood ratio tests and the Akaike information criterion (Rathcke & Stuart-Smith, 2016, p. 415). The whole procedure was conducted in R (R Core Team, 2021) with the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. The overall results show that there is no clear timing distinction between the VE short and long contexts among the high contact speakers in the SVLR-unaffected vowel /a/ providing weak support for the *Dialect Contact Hypothesis*. While there was a slight increase in the duration of /i/ and /u/ in VE-long contexts as opposed to VE-short environment, this durational difference did not reach statistical significance. There was only a significant lengthening of /i/ in VE-unspecified contexts, but as stated above, this category is identical to the long contexts of the morphological SVLR conditioning (see Table 9). Dialect contact or an exposure to SSBE does therefore not lead to an erosion of SVLR durational patterns in Glasgow which contradicts many earlier investigations in other regions (Hewlett et al., 1999; Scobbie, 2005; van Leyden, 2002; Watt & Ingham, 2000). In contrast, the findings do provide evidence for the *Prosodic Timing Hypothesis* as there is a significant interaction between Aitken's Law and stress as well as phrase-final lengthening. The synchronic perspective shows that, whereas SVLR short vowels are unaffected by prosodic stress, SVLR long vowels show a considerable degree of lengthening under prosodic stress, which, in turn, increases the durational contrast between the short and long environments (Rathcke & Stuart-Smith, 2016, p. 421). The difference between vocalic durations in SVLR long and short contexts is also larger in phrase-final positions overall, but this effect was absent in phrase-medial positions. This means that the difference between SVLR long and short contexts is greater in final positions than in non-final positions. The vowels are also generally affected by *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4), but the effect of *lexical frequency* (see subsection 3.1.5) turned out to be insignificant. When comparing the SVLR durational patterns across the different speaker groups from a diachronic perspective, the durational differences between SVLR short and long contexts weaken over time. Especially the oldest speaker group, namely the middle-aged men in the 1970s, had a stronger SVLR-related lengthening effect in phrase-final positions than the subsequent generations. This weakening of the SVLR in later speaker groups therefore fully corresponds with the *Prosodic Timing Hypothesis* as the segments in phrase-final, or, to a lesser extent, stressed syllables are subject to change. Rathcke and Stuart-Smith (2016, p. 422) further argue that this diachronic change might also be reinforced by the process of urban regeneration which happened in Glasgow between the 1950s and mid-1970s. This process led to drastic changes in the city landscape as inner-city housing was demolished and inhabitants were moved to new, largely suburban homes. This had a profound influence on the city's communities and, consequently, their linguistic systems. As a weakening of the SVLR durational patterns can be observed among those speaker groups who grew up during the time of urban regeneration, it seems that the shifts in network structure may also have led to a weakening of the traditional SVLR vowel timing system in the communities (Rathcke & Stuart-Smith, 2016, p. 422). Overall, the researchers conclude that the SVLR remains operative in the Glaswegian vernacular which is in line with the previous findings by Scobbie, Turk, and Hewlett (1999). While there are also slight signs of SVLR weakening, this process, however, is "not accompanied by a straightforward shift towards the constraints of the consonantal Voicing Effect" (Rathcke & Stuart-Smith, 2016, p. 423).

The study by Rathcke and Stuart-Smith (2016) marks a change in SVLR research. Whereas previous studies investigated Aitken's Law in highly controlled experimental settings, Rathcke and Stuart-Smith (2016) were the first who took on the difficult task of analyzing SVLR patterns in spontaneous speech. Their statistical analysis is very sophisticated as they applied linear mixed effects models and they also provide a reasonable discussion of their findings. Nevertheless, the diachronic focus of the study is very broad when compared to the limited number of tokens investigated. For example, there are 425 tokens for the vowel /i/ but there are also four speaker groups (70Y, 70M, 00Y, 00M) with two levels of SSBE contact (high, low), three SVLR and three VE context categories, two phrasal positions, three levels of prominence as well as three covariates (word frequency counts, number of syllables in a word, number of segments in a syllable). As the linear mixed effects models include all these predictor variables, the actual number of tokens in the individual categories can become relatively small. The representativeness of the sample might therefore be limited, also since it only includes male speakers from Glasgow. Furthermore, the study generally analyses Scots data, so it remains unclear how Aitken's Law operates in Glaswegian SSE.

David Warren (2018) conducted another investigation into the VE and SVLR in the Northeast of Scotland as part of his unpublished PhD thesis. This region has received little attention in vowel timing research, with the exception of Watt and Yurkova's (2007) conference paper on Aberdeen English summarized above. Warren (2018), however, focused on rural varieties in the Northeast and split his study into two stages. In the first stage, he analyzed the vowel pronunciations of six middle-aged female speakers with data elicited in a map task and word list readings. He controlled for the speakers' birthplace, their age as well as for the parental accent of the speakers. The researcher elicited monosyllabic words to avoid the influence of polysyllabic shortening and included the monophthongs /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/ and the diphthong /aɪ/ as vowel nuclei followed by tautomorphemic /t/, /d/, /s/ and /z/. Similar to Rathcke and Stuart-Smith (2016), Warren (2018) implemented linear mixed effects models in his statistical analysis in R (R Core Team, 2021) fitting two independent models for the word list and map task data. As the map task elicited semi-spontaneous speech, he also controlled for the suprasegmental factors *prosodic stress* and *phrase position* distinguishing between nuclear and non-nuclear accents as well as final and non-final positions. In the second stage of his analysis, he investigated the same vowel set with the additional context of following heteromorphemic /#d/ to test the morphological conditioning of Aitken’s Law. He also used another speaker sample with two male and four female speakers in a slightly larger age range (36-57 years) and restricted his analysis to speech elicited in controlled settings. That is, he avoided the map task and used word list readings with the lexical items read in isolation and in the frame sentences “I say [word] now” and “I say [word] again” (Warren, 2018, pp. 182–183). The overall results contradict most previous studies on Scottish vowel duration in that he found a strong and consistent VE across all vowels. Similar to the findings by Watt and Yurkova (2007), the vocalic intervals were significantly lengthened before following tautomorphemic /d/. Interestingly, the often-cited SVLR-affected vowels /i/, /u/ and /aɪ/ were subject to the greatest durational increases in this VE long context (Warren, 2018, p. 250). In contrast to this, heteromorphemic /#d/ conditioned only a small increase in vowel duration and the researcher could not find a significant statistical durational difference between tokens in tautomorphemic /d/ and heteromorphemic /#d/ environments. Warren (2018, p. 230) therefore concludes that the morphological conditioning of Aitken’s Law does not apply in the Northeast of Scotland although he acknowledges that this claim needs to be corroborated by further statistical evidence. As for the phonological conditioning of the SVLR, the analyses found that all vowels are significantly longer before tautomorphemic /z/, especially the monophthongs /i/, /u/, /o/ and the diphthong /aɪ/. Indeed, the longest average vocalic durations are found in voiced fricative contexts. However, due to the fact that the findings show a strong VE and a minimal morphological SVLR, it is unclear whether vowel lengthening in these environments can be attributed to the effects of Aitken’s Law as both the VE and the SVLR condition long allophones in voiced fricative contexts. Apart from this, the analyses also reveal an influence of intrasyllabic compression (Warren, 2018, p. 203) as well as interactions between prosodic stress and phrasal position with vowel duration in different coda contexts. That is, vowels followed by /d/ and /z/ are generally shorter in phrase-medial and non-nuclear contexts which corresponds to the observations by Rathcke and Stuart-Smith (2016). While the vowel duration patterns were relatively stable across the social and demographic groups (Warren, 2018, p. 247), the researcher could find significant age-related variation in his second analysis. Here, the younger speakers have “an overall reduced distinction between long and short vowels phrase-finally compared to the older speakers” (Warren, 2018, p. 250). The researcher believes that this indicates ongoing system-internal reconfigurations of the vowel timing system in the Northeast.

The study by Warren (2018) shows that the vowel timing patterns in the Northeast are different from those in other regions of Scotland. Hence, his findings corroborate the earlier observations by Watt and Yurkova (2007) on Aberdeen English. However, this claim needs further evidence due to the limited age range and small speaker sample investigated. While the study by Warren (2018) is very detailed, the overall sample includes only 12 speakers and a large part of his analysis focused on controlled speech.

Based on the approach of Rathcke and Stuart-Smith (2016), Florent Chevalier (2019) conducted another study on the effects of the SVLR and the VE in the Glaswegian vernacular. He also retrieved data from the force-aligned *Sounds of the City corpus* and conducted his statistical analysis in R (R Core Team, 2021) applying linear mixed effects modeling with the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. The researcher also differentiated between tokens under nuclear or non-nuclear prosodic stress and in final or non-final positions and the speakers' degree of contact with SSBE was also described as being either high or low. He also implemented a real- and apparent time approach with young and middle-aged speakers from the 1970s and 2000s and he used the same SVLR and VE context classification as Rathcke and Stuart-Smith (2016). However, in contrast the previous study on the Glaswegian vernacular, Chevalier (2019) investigated only 12 female speakers to test for possible gender-related variation in the application of Aitken's Law. In addition, he only focused on the high vowels /i/ and /u/ and excluded most function words, all proper nouns as well as all tokens which were followed by /r/. The researcher also ran two separate models to test the effects of the VE and the SVLR independently. The findings largely correspond with the results by Rathcke and Stuart-Smith (2016): SVLR long contexts condition longer vocalic intervals than SVLR short contexts and the durations are also longer in phase-final positions and in syllables under nuclear prosodic stress. There is also a significant interaction between suprasegmental factors and Aitken's Law as the durational contrast between SVLR short and long vowels is increased in nuclear and phrase-final contexts. The best model for the SVLR context classification also showed significant interactions between Aitken's Law and the speaker group. Similar to the findings by Rathcke and Stuart-Smith (2016), the durational difference between SVLR short and long contexts weakens over time with a shortening of SVLR long vowels, especially in phrase-final positions. When comparing this trend with the male speakers of Rathcke and Stuart-Smith (2016), Chevalier (2019) notes that the weakening of Aitken's Law is stronger among the female speakers in Glasgow. The difference between VE short and long contexts is small with an average durational difference of only 8.5 ms. The linear mixed effects model with the VE context classification returned no significant durational differences between VE long and short environments. However, there was a significant interaction between the VE and two speakers with high contact to SSBE. The two speakers belonged to the middle-aged 1970s group and their vowel pronunciations showed a smaller difference between SVLR short and long contexts. Apart from that, the data also suggests effects of polysyllabic shortening and intrasyllabic compression in the vowel pronunciations of the speakers. Overall, the study could find similar patterns as Rathcke and Stuart-Smith (2016): the SVLR operates among female speakers in Glasgow even though signs of weakening are also detectable and slightly stronger among the women than among their men. Except for two speakers with high SSBE contact, this weakening is not associated with a shift toward the VE but interacts with prosodic factors, especially the phrasal position.

The study by Chevalier (2019) is very similar to the previous investigation by Rathcke and Stuart-Smith (2016). Both the study design and the findings are comparable, the main difference is that Chevalier (2019) investigated female Glaswegian speakers from the *Sounds of the City* corpus. Together, the studies provide a thorough insight into SVLR patterns in Glaswegian Scots over time. Overall, SVLR patterns are stable in Glasgow and the SVLR also interacts with segmental and suprasegmental factors. The studies do not, however, provide any results for SSE because their samples mainly include Glaswegian Scots.

The most recent study on the VE and SVLR in Scotland was carried out by Stuart-Smith et al. (2019) in the context of the SPADE project. The researchers applied a large-scale approach towards analyzing vowel timing patterns in Scottish English implementing files from different corpora in their analysis and distinguishing between five Scottish dialect regions: *Northern*, *Glasgow*, *Edinburgh*, *South* as well as *Highlands, Islands and Insular*. An overview can be found in Table 10.

Table 10. Corpora and data structure by Stuart-Smith et al. (2019) separated for the five dialect regions with the number of speakers and tokens.

Region	**Glasgow**	**Edinburgh**	**Northern**	**South**	**Highland, Islands and Insular**
Corpora	SCOTS, Sounds of the City Corpus, Brains in Dialogue Corpus	SCOTS, Edinburgh (Arthur the Rat), Doubletalk Corpus	SCOTS, 1Speaker-2Dialects Corpus	SCOTS	SCOTS
Number of Speakers	177 (51.6%)	85 (24.78%)	49 (14.29%)	17 (4.96%)	15 (4.37%)
Number of Tokens	152364 (47.74%)	41418 (12.98%)	105692 (33.11%)	13860 (4.34%)	5842 (1.83%)

As seen in Table 10, the dialect regions *Glasgow*, *Edinburgh* and *Northern* represent the vast majority of the data in terms of speaker and token numbers: 47.74% of the dataset is represented by the dialect region *Glasgow*, 33.11% by *Northern* and 12.98% by *Edinburgh*. In contrast to this, the dialect regions *South* as well as *Highlands, Islands and Insular* only make up a very small proportion of the dataset. In terms of corpora, most of the speech data was taken from the spoken component of the *Scottish Corpus of Text and Speech* (henceforth: SCOTS). Files of this corpus were used for all dialect regions. SCOTS includes a variety of recordings from different genres and registers of Scottish English. Most of the speech data, however, can be categorized as Scots; only a few files represent SSE speech. Apart from that, the researchers also used files from the *Sounds of the City corpus* and the *Brains in Dialogue corpus* for *Glasgow*. The dialect region *Edinburgh* is further represented by files from the *Edinburgh (Arthur the Rat)* recordings and the *DoubleTalk* corpus and it is the only dialect area which is largely represented by SSE speakers. The *Arthur the Rat* files are recordings of read speech from different time periods. The *Northern* dialect region is also represented by recordings from the *1Speaker2Dialects* corpus. Overall, the dataset comprises speech from 343 speakers (171 females) who were born in the decades from 1890 to 1990. In this dataset, Stuart-Smith et al. (2019) investigated the timing effects of Aitken's Law and the VE across monosyllabic words with the vowel nuclei /i/, /ɪ/, /e/, /ɛ/, /a/, /ɔ/, /ʌ/, /o/ and /ʉ/ over time. Based on previous research, the researchers expected that the SVLR operates in /i/ and /ʉ/ and that it might also affect /e/ and /o/. However, the vowels /a/, /ɔ/ are unlikely to show SVLR patterns in most dialects and Aitken's Law should generally not operate in the short monophthongs /ɪ/, /ɛ/ or /ʌ/. The data analysis was conducted via ISCAN (McAuliffe et al., 2019) and the researchers applied linear mixed effects modeling on the log-transformed vowel durations in R (R Core Team, 2021). As the data represents read and spontaneous speech, the researchers also controlled for the speech rate, the phrasal position and the lexical frequency of each individual token. As for the tempo, Stuart-Smith et al. (2019) calculated a logarithmic speech rate deviation by subtracting the articulation rate of an utterance from the corresponding speaker's mean speaking rate. Logarithmic word frequency counts were obtained on the basis of the SUBTLEX-UK dataset (van Heuven et al., 2014). The phrase position, logarithmic speech rate, logarithmic word frequency as well as the vowel, the following context, the gender, the birth decade and all possible interactions were included as fixed factors in the model building. Word and speaker were treated as random intercepts. The overall results show that there are no SVLR nor VE patterns in the short monophthongs /ɪ/, /ɛ/, or /ʌ/ which corresponds to earlier observations. Aitken's Law can, however, be detected in /ɔ/ and /e/ in the dialect area *Northern* and its patterns are generally found in the high vowels /i/ and /ʉ/ across all dialect regions. Interestingly, the analysis could also find an "anti-Voicing Effect" (Stuart-Smith et al., 2019) in the vowels /i/, /ʉ/, /e/ and /o/ across most dialects. This means that there is an overall tendency that voiceless stops condition longer vocalic intervals than voiced stops which completely contradicts the VE. Exceptions can be found in the vowel /ʉ/ in the dialect region *Northern* and in /i/ across the speakers from the Highlands and Islands where a VE could be detected. The analysis could not find any clear effect of gender on vowel timing and there was also no difference between the predominantly SSE speakers from Edinburgh and the largely Scots speakers from the other dialect regions. Apart from this, the vocalic durations are generally shorter at faster speech rates and in words with a high lexical frequency. Phrase-finality leads to increases in vowel duration overall and the contrast between SVLR short and long vowels is also widened in phrase-final positions which corresponds to earlier findings on Aitken's Law in spontaneous speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Warren, 2018). The study concludes that Aitken's Law remains operative across male and female speakers of Scots and SSE in different Scottish dialect regions in the twentieth century.

The latest study by Stuart-Smith et al. (2019) sets a new benchmark in terms of token numbers for SVLR research. The study has by far the largest sample and it includes different dialect regions of Scotland. However, the dataset is not balanced in terms of the regional background of the speakers and most of the data is Scots. Furthermore, the sample includes several datasets with completely different speaking styles. Whereas the dialect region *South* includes 17 SCOTS files which mainly represent spontaneous speech, the dialect region *Edinburgh* also includes *Arthur the Rat* recordings which represent controlled read speech. The speaking styles are therefore not balanced for each dialect region. Furthermore, as the sample includes speakers who were born between 1890 and 1990, the study provides an overview of SVLR patterns over a large time span. However, much of the data is therefore not representative for the 21st century. The statistical analysis also does not account for the influence of prosodic stress.

3.3 Summary

There are various factors which influence the duration of vocalic intervals. Apart from idiolectal variation and the phonological classification of long and short vowel phonemes, there are some general tendencies that can be observed in English. On the segmental level, open vowels tend to be longer than close vowels. Thus, the word <bat> tends to have a longer vocalic nucleus than the words <beat> or <boot>. The syllable and word structure also affect vowel duration due to compensatory lengthening and shortening: the more segments in a syllable and the more syllables in a word, the shorter the vocalic intervals tend to be. For instance, the monosyllabic word <streets> tends to have a shorter vowel with its CCCVCC structure than the word <seat> with its CVC segments. Likewise, the vowel /iː/ in the monosyllabic word <speed> tends to be longer than the same vowel in the trisyllabic word <speediness>. There is also an influence of the lexical category and frequency on vowel duration. As function words are often reduced in speech, their vowel durations tend to be shorter than those of content words. High frequency content words also tend to be pronounced quicker than words with a low frequency and the vowel durations in the former therefore tend to be shorter. Another fundamental influence on vowel duration is stress. On the word level, lexically stressed syllables, such as the first syllable in <payment>, tend to be louder and longer in speech than lexically unstressed syllables. On the larger prosodic level, vowels in syllables carrying a nuclear accent also tend to be lengthened. Vowels are also lengthened in word-final and especially utterance-final positions preceding a pause. Another very important suprasegmental factor is tempo: the faster a speaker talks, the shorter the vocalic intervals become. A high articulation rate therefore leads to shorter vowel articulations. Apart from that, there is also the influence of the postvocalic context on vowel duration. In most varieties of English, the VE conditions longer vowels before voiced consonants than before voiceless consonants.

The situation is, however, different for Scottish English with Aitken’s Law which conditions short vowel allophones before voiced plosives, nasals and laterals. Whereas the earlier impressionistic accounts on the SVLR, including Aitken’s most elaborate description in 1981, assumed that most Scottish English vowels are subject to Aitken’s Law, the more recent empirical studies suggest a limited SVLR vowel set. Most studies conclude that the timing effects of Aitken’s Law are only detectable in the high front vowel /i/ (Chevalier, 2019; Hewlett et al., 1999; McClure, 1977; McKenna, 1988; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019; van Leyden, 2002; Watt & Ingham, 2000), the high back vowel /u/ (Chevalier, 2019; Hewlett et al., 1999; McClure, 1977; McKenna, 1988; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019; Watt & Ingham, 2000) as well as in the diphthong /aɪ/ (Llamas et al., 2011; McMahon, 1991; Pukli, 2006; Scobbie, Hewlett, & Turk, 1999; Scobbie, Turk, & Hewlett, 1999). Apart from that, most empirical studies agree with earlier SVLR descriptions that the monophthong /ɪ/ (McClure, 1977; McKenna, 1988; Scobbie, Hewlett, & Turk, 1999; Stuart-Smith et al., 2019; van Leyden, 2002) and the monophthong /ʌ/ (McClure, 1977; McKenna, 1988; Stuart-Smith et al., 2019) are generally short and therefore unaffected by Aitken’s Law. The effects of Aitken’s Law on the other vowels are less consistent.

From a geographical perspective, most previous empirical studies on Aitken’s Law investigated speakers from the Central Belt of Scotland, in particular, Edinburgh and Glasgow (see Figure 10). There were two studies that focused on Ayrshire (McClure, 1977; Pukli, 2006), two that investigated the Northeast (Warren, 2018; Watt & Yurkova, 2007), two conducted in the Northern Isles (Scobbie, 2005; van Leyden, 2002) and three other investigations which focused on vowel timing patterns in the Northeast of England (Llamas et al., 2011; Milroy, 1995; Watt & Ingham, 2000). There are, however, no empirical findings for the Highlands and Hebrides, nor for the Borders region or the Southwest of Scotland. While Stuart-Smith et al. (2019) include these regions in their study, the proportionally low token numbers for these regions can only provide a very limited overview.

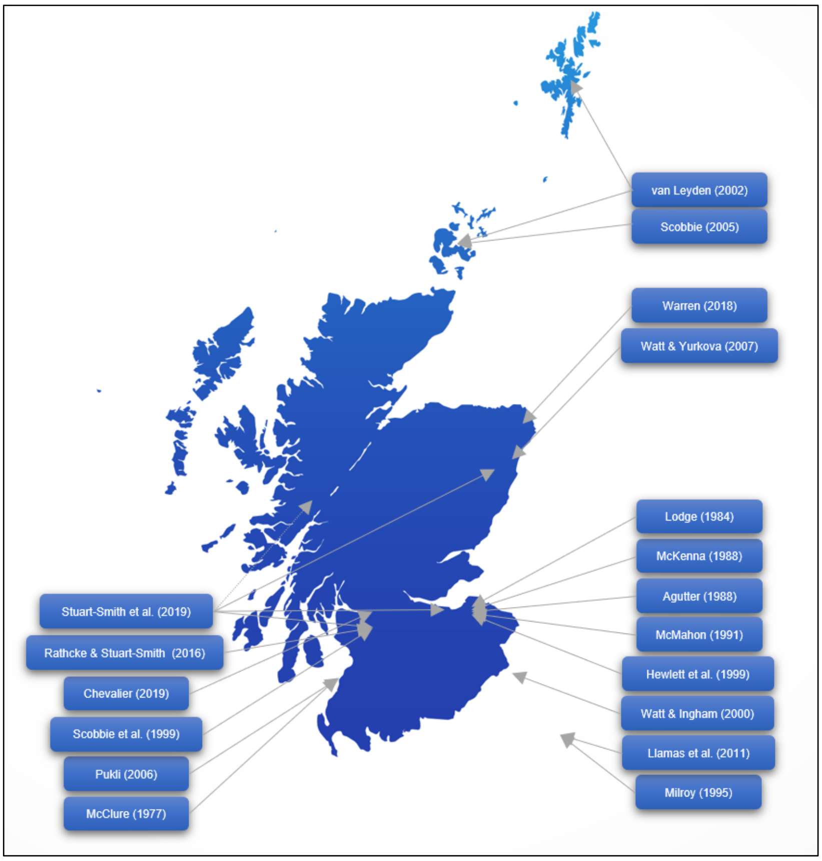


Figure 10. Geographical locations of previous empirical studies on Aitken’s Law.

Overall, the SVLR seems to be relatively stable in Glasgow as previous studies did not find significant variation with regard to social class or gender (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019). While the diachronic studies found a weakening of SVLR patterns over time (Chevalier, 2019; Rathcke & Stuart-Smith, 2016), Aitken’s Law is still in operation in Glasgow and there is no shift towards VE timing effects. A different situation can be found in the Northeast of Scotland. The related studies by Watt and Yurkova (2007) and Warren (2018) found a strong and consistent VE across all vowels investigated. The timing effects of the SVLR are, however, less consistent. A similar picture can be found for the Northeast of England with the VE consistently operating in Tyneside and Berwick-upon-Tweed (Llamas et al., 2011; Milroy, 1995; Watt & Ingham, 2000). The studies also suggest that Aitken’s Law partly operates alongside the VE and its timing effects are most consistent in the diphthong /aɪ/ in Northern England. Aitken’s Law can also be detected /i/ and /u/ in Edinburgh, Orkney and Shetland (Hewlett et al., 1999; McKenna, 1988; Scobbie, 2005; van Leyden, 2002). However, McKenna (1988) reports that the vowels /a/, /e/, /ɛ/, /ɔ/ and /o/ are also affected by the VE among Edinburgh speakers. The timing effects of Aitken’s Law are also detectable in the high vowels /i/ and /u/ in Ayrshire, but there is also a variable VE influence in the other vowels (McClure, 1977; Pukli, 2006).

Apart from regional variability, some studies also report variation in SVLR vowel timing with regard to the speakers’ age, gender as well as exposure to SSBE. Milroy (1995), Watt and Ingham (2000), Rathcke and Stuart-Smith (2016) as well as Chevalier (2019) report age-related variation in the application of Aitken’s Law as SVLR patterns are weakened among younger, often female speakers. However, Scobbie, Turk, and Hewlett (1999) and Llamas et al. (2011) found no evidence of age-related variation among their speaker samples from Glasgow and Tyneside. There are also some studies which suggest that SVLR patterns are weakened due to high SSBE contact. Children from Edinburgh and Shetland with non-Scottish parents are frequently reported to adopt VE-related vowel timing patterns (Hewlett et al., 1999; Scobbie, 2005). However, this trend could not be observed in the spoken Glaswegian vernacular (Chevalier, 2019; Rathcke & Stuart-Smith, 2016).

The effects of the SVLR and the VE also vary with regard to the speaking style. McClure (1977) already noted that the timing effects of Aitken’s Law are weaker when words are spoken in carrier sentences than when read in isolated word list readings. The most recent studies on the SVLR and VE report even weaker effect sizes in naturally spoken Scottish English (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Tanner et al., 2020). Furthermore, there are also significant interactions between Aitken’s Law and the VE with prosodic factors: SVLR and VE long and short contrasts are less pronounced in phrase-medial and non-nuclear syllables.

4. Data and Method

This chapter provides an overview of the datasets and methodological approaches used in the present study. A special feature of the present investigation is that it includes precise data selection criteria which will be explained in detail in section 4.1. This means that I will first determine what an ideal dataset should look like in the context of the present investigation. This not only includes sociodemographic criteria of the sample (subsection 4.1.1) but also information on speaking style (subsection 4.1.2) and the transcription format (subsection 4.1.3). The actual sample of this study is then described in section 4.2. Here, the corpus *ICE Scotland* was first used as a base dataset and I will discuss the structure of the corpus as well as its advantages and disadvantages in subsection 4.2.1. Yet, as this corpus cannot fully meet the data selection criteria set out in section 4.1, I collected further speech data to complement the ICE Scotland files chosen for analysis. Hence, the sample of this study incorporates both pre-existing data (ICE Scotland) but also self-collected datasets. The self-collected speech data include interview data which was collected in Scotland and other spoken data which was retrieved from the internet. I will provide an overview of the speakers and I will also discuss the advantages and disadvantages of the self-collected datasets in the subsections 4.2.2 and 4.2.3. An essential task was to harmonize the different transcription formats of the ICE Scotland files and the self-collected data and these harmonization procedures will be explained in detail in section 4.3. In this section, I will also provide an overview of the speech analysis software Praat (Boersma & Weenink, 2019), the process of forced alignment and the phone segmentation approach used in the present study. Subsection 4.3.1 then describes the data preparation procedures of the ICE Scotland files and subsection 4.3.2 does the same for the self-collected datasets. The data analysis will be explained in section 4.4. I will provide an overview of the vowel selection (subsection 4.4.1) and the predictor variables included in the analyses (subsection 4.4.2). After that, I will describe the statistical analysis (subsection 4.4.3) which incorporates means of descriptive statistics but also inferential statistical procedures. A summary of the dataset and method follows in section 4.5.

4.1 Data selection and transcription criteria

The data selection process was based on multiple criteria to create a balanced and adequate corpus suitable for the subsequent analyses. As many previous studies on the SVLR are clearly outdated (see subsections 3.2.1 and 3.2.2), an essential criterion for the present investigation was that its speech data is up-to-date. This means that the sample should only contain speech from the 21st century. Other selection criteria are the sociodemographic background of speakers (subsection 4.1.1) and the speaking style (subsection 4.1.2) and these criteria will be explained in the following subsections. Apart from that, I will also describe what an ideal transcription format should look like in subsection 4.1.3.

4.1.1 Sociodemographic background of the speakers

In contrast to previous studies on Aitken's Law (see section 3.2), the dataset of the present investigation should be balanced in terms of the age, gender and the regional background of the speakers. The sample should therefore be representative for the whole of Scotland and only include Scottish speakers whose age, gender and regional upbringing are known.

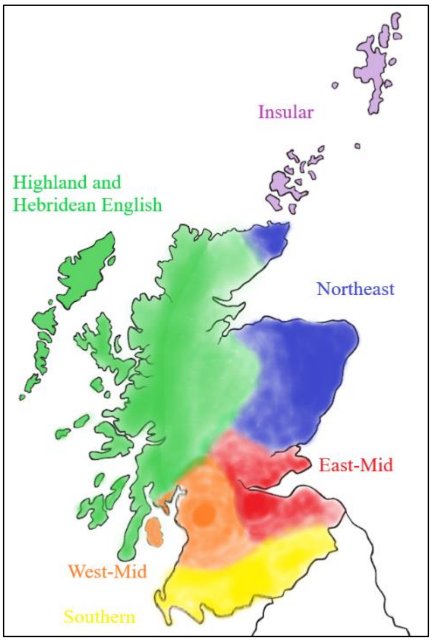


Figure 11. Dialect differentiation applied in the present investigation

Based on the linguistic situation of Scotland (see chapter 2) and Grant's (1931) dialect classification, the present study will differentiate between six dialect regions to investigate potential regional variation in Aitken's Law: *Insular*, *Northeast*, *East central*, *West central*, *Southern* and *Highland and Hebridean English* (henceforth: *HHE*) (see Figure 11). *Insular* comprises the islands of Shetland and Orkney and *HHE* covers the originally Scottish Gaelic speaking areas North and West of the old Highland Line (cf. Figure 1). The dialect region *Northeast* incorporates the traditional Northern Scots areas in Caithness, Moray, Aberdeenshire and Angus. In accordance with Grant's (1931) classification, *West central* represents Glasgow and the Western Central Belt and *East central* represents Edinburgh and the Eastern Central Belt. *Southern* incorporates the Borders as well as Galloway in the Southeast of Scotland. Each dialect area should be represented by a comparable number of speakers and tokens.

As for gender, roughly half of the speakers in the sample should be male and the other half should be female. A roughly equal distribution should also be true for the six dialect regions: each region should be represented by a comparable number of male and female speakers. A comparable distribution of the genders makes it possible to investigate gender-related variation in Aitken's Law.

Likewise, the sample should also be balanced in terms of age: there should be comparable numbers of young, middle-aged and old speakers of both genders for each dialect region. While age groups are fundamentally arbitrary, I will use the categories of *under 30 years* (young), *30-60 years* (middle) and *over 60 years* (old) for data collection purposes to ensure that each dialect region has a comparable amount of younger, middle-aged and older speakers. This ultimately makes it possible to investigate age-related variation in SVLR durational patterns. Furthermore, if there is a roughly balanced distribution of speakers in terms of *age*, *gender* and *regional background*, it is also possible to check for interactions. It could be the case, for instance, that young male speakers from a certain dialect region produce different vowel durational patterns than the rest of the sample. In this case, a balanced dataset would facilitate the identification of such an interaction.

4.1.2 Speaking style

Whereas most previous studies investigated the SVLR in controlled speech (see subsection 3.2.2) or in predominantly spontaneous spoken Scots (see subsection 3.2.3), the present study is the first which exclusively investigates vowel duration in naturally spoken SSE. Hence, this study will be the first to investigate Aitken's Law in SSE on a countrywide scale. The spoken data should therefore preferably derive from formal and/or public contexts. The formality of the speech situation ensures that speakers use SSE and avoid drifting toward Scots (Schützler, 2015, pp. 47–48). Sound files with many Scots expressions will therefore be excluded from the dataset.

Furthermore, I decided to include both scripted and unscripted speech forms in the sample. Scripted speech can be found in, for example, public TV broadcasts, radio broadcasts, or in political speeches. Political speeches, for instance, are often read or at least rehearsed so that the speakers already know how to formulate their phrases and sentences. The same is true for news broadcasts on the TV or radio. Unscripted formal speech can be found in, for instance, broadcasted discussions and interviews. Here, the language is more spontaneous as the speakers have little time to plan their conversational contributions. The selection of both scripted and unscripted language makes it possible to investigate SVLR vowel timing patterns in more controlled and more spontaneous SSE. An ideal sample should therefore have a comparable amount of scripted and unscripted speech data. This means that the dataset should not only be balanced in terms of the sociodemographic criteria specified above (subsection 4.1.1) but that it should also have comparable shares of scripted and unscripted speech for each dialect region. As the effect sizes of Aitken's Law and the VE are generally stronger in controlled than in spontaneous speech (Tanner et al., 2020, pp. 2–3), I expect that the strength of Aitken's Law is more consistent in scripted than in unscripted speech forms.

4.1.3 Transcription format

As this study investigates vowel duration, an essential part is that the transcription of sound files is time-aligned. This means that the transcription format should specify when a vowel starts and when it ends to calculate its duration. Additionally, the transcription format should also include a *syllable*, *word* and *utterance* level as well as the corresponding durations of these levels. This not only reflects the most important elements of the prosodic hierarchy, but the inclusion of these levels also makes it easier to account for the general factors influencing vowel duration (see section 3.1). For example, the *syllable* level should provide information about the number of phones in the syllable to account for the effects of *intrasyllabic compression* (see subsection 3.1.3). Likewise, the *word* level should provide information about the number of syllables within the word to account for *polysyllabic shortening* (see subsection 3.1.4). It would also be helpful to transfer the syllable transcriptions into vowel (V) and consonant (C) symbols to get an overview of the overall syllabic structure. The syllable <mem> in the word <members> could, for example, be represented by CVC and this would allow for comparisons with other syllables of the same type. The *word* level can also be used to account for the *lexical category* and *frequency* (see subsection 3.1.5). This is important because function words and frequent content words tend to be shorter than low frequency content words. In addition, the *word* level can also be used to account for *lexical stress* (see subsection 3.1.6) and the number of syllables in an utterance makes it possible to calculate the local articulation rate (see subsection 3.1.9). Dividing the total duration of the utterance by the corresponding number of syllables within that utterance makes it possible to obtain the local *average syllable duration*. For example, if an utterance has a total duration of two seconds and consists of five syllables, the *average syllable duration* would be 0.4 seconds (2/5 = 0.4). This means that the lower the *average syllable duration*, the higher the articulation rate. Similarly, dividing the number of syllables by the total duration of the utterance makes it possible to obtain the *average number of syllables per second* in that utterance. In the previously stated example, this would equal 2.5 syllables per second (5 syllables divided by 2 seconds = 2.5 syllables per second). Hence, the local articulation rate is faster if the *number of syllables per second* is higher. Apart from that, an *utterance* and *syllable* level would also make it possible to distinguish between, for example, utterance initial, medial and final syllables. Syllables before a pause can be categorized as *final*, syllables following a pause as *initial* and syllables within an utterance as *medial*. This would make it possible to control for *constituent-final lengthening* (see subsection 3.1.8). The *syllable* and *utterance* level further facilitate the identification of *prosodic stress* (see subsection 3.1.6): if each utterance and syllable is transcribed, it will be easier to identify which syllable received prosodic stress in an utterance. The final data format should further include information about the following phone of each vowel to identify the postvocalic context of a vowel. This will facilitate the identification of VE (see subsection 3.1.2) and SVLR environments (see section 3.2). For example, when a vowel is followed by a voiced plosive in a stressed syllable, this environment can be categorized as a long VE context but also as a short SVLR context. More information about the SVLR and VE categorization schemes can be found in subsection 4.4.2.

4.2 Datasets

In the data selection process, the corpus *ICE Scotland* was first used as the base dataset. However, due to a lack of unscripted speech and a shortage of speakers from specific dialect regions and age groups in *ICE Scotland*, I collected further speech data to counter these imbalances. As a result, the final dataset includes speech data from *ICE Scotland* (see subsection 4.2.1) but also interview data which was collected in Scotland (see subsection 4.2.2) and data which was retrieved from different sources of the internet (see subsection 4.2.3). For a better overview, the speech data which was collected in Scotland will be named *Alba Recordings* ("Alba" is the Scottish Gaelic name for "Scotland") and the speech data from the internet will be called *Home Recordings* (speech data retrieved at home) in the following. As a result, the final dataset of this study includes speech data from three sources: *ICE Scotland* as well as the two self-collected datasets of spoken SSE. In the following subsections, I will provide an overview of the data sources and discuss their advantages and disadvantages against the background of the data selection criteria (see section 4.1). A detailed overview of the whole dataset can be found in the summary of this chapter (see section 4.5).

4.2.1 ICE Scotland

One of the most important datasets for the present analysis is the spoken component of the ICE Scotland corpus. This corpus was compiled by Ulrike Gut and Ole Schützler at the universities of Münster and Bamberg and is the first one-million-word corpus of SSE. The spoken component comprises 600000 words from monologues (e.g. broadcast talks, news reports) and dialogues (e.g. private conversations, broadcast discussions), most of them recorded after 2013 (Schützler et al., 2017). Following the ICE guidelines, all speakers of the corpus were at least 18 years old, they were born and raised in Scotland and they have a high level of education (Schützler et al., 2017, p. 290). It includes male and female speakers from different age groups and different dialect regions of Scotland. The corpus is therefore in line with the data selection criteria specified in the subsections 4.1.1 and 4.1.2: it includes scripted and unscripted SSE speech and there is information on the *age*, *gender* and *regional background* of many speakers. Yet, the sociodemographic background information is not given for all speakers of the corpus and some regions and age groups are better represented than others. Overall, many speakers grew up in the Central Belt of Scotland, but there are very few from the South of Scotland and none from the Northern Isles. Furthermore, young speakers are less well represented than other age groups.

A very valuable feature of ICE Scotland is that its spoken component includes force-aligned and manually corrected transcriptions on the word and segment level (more information on the force-alignment process of ICE Scotland can be found in section 4.3). This means that the ICE Scotland transcriptions do not only specify word and phoneme labels but also the start and end times of the words and phonemes. This corresponds to the basic requirement of the transcription format specified in subsection 4.1.3: the transcriptions are time-aligned and it is therefore easy to retrieve their duration. However, the ICE Scotland transcriptions do not provide an *utterance* or *syllable* level.

Out of the spoken component of ICE Scotland, a total of 91 files were selected for the present study. These files were chosen because they included all relevant sociodemographic background information of the speakers (see subsection 4.1.1). All files also had a high audio quality and they were all recorded after the year 2013. After examining the sound files, I further categorized the speech as being either *scripted* or *unscripted* based on the selection criteria specified in subsection 4.1.2. In the category *scripted*, speakers read aloud a written speech in a public context, mostly in the Scottish Parliament or in a news broadcast. This means that the speakers had time to prepare their utterances and had few, if none, false starts or hesitations in their speech. In the category *unscripted*, speakers spoke freely and subsequently, their spoken contributions are less structured and include more hesitations, false starts and repetitions. An overview of the selected files from ICE Scotland can be found in Table 11. Due to data protection issues, Table 11 does not list the names of the speakers.

Table 11. Files of the ICE Scotland corpus used in the analysis. The parentheses behind a file name indicate that the file has more than one speaker. The addition "(s3)", for instance, indicates that this is the third speaker of the sound file.

File	**Duration (minutes)**	**Category**	**Word count**	**Regional upbringing**	**Gender**	**Age** (at time of recording)	**Style**
bdis_01 (s1)	8	Broadcast discussions	529	West central	female	44	unscripted
bdis_09 (s2)	9	Broadcast discussions	586	West central	male	38	unscripted
bdis_09 (s4)	9	Broadcast discussions	525	West central	female	57	unscripted
bdis_10 (s5)	6	Broadcast discussions	281	West central	female	57	unscripted
bdis_13 (s1)	5	Broadcast discussions	376	West central	female	44	unscripted
bdis_13 (s3)	5	Broadcast discussions	640	West central	female	69	unscripted
bdis_14 (s1)	6	Broadcast discussions	370	West central	male	49	unscripted
bdis_14 (s2)	6	Broadcast discussions	755	East central	female	45	unscripted
bdis_14 (s3)	6	Broadcast discussions	406	West central	male	38	unscripted
bdis_16 (s1)	7	Broadcast discussions	350	West central	female	44	unscripted
bdis_20 (s1)	6	Broadcast discussions	298	West central	female	44	unscripted
bdis_20 (s3)	6	Broadcast discussions	796	Northeast	male	58	unscripted
bdis_21 (s1)	10	Broadcast discussions	542	West central	female	44	unscripted
bdis_21 (s4)	10	Broadcast discussions	744	West central	male	45	unscripted
bint_07 (s1)	10	Broadcast interviews	646	East central	male	55	unscripted
bint_09 (s1)	2	Broadcast interviews	466	HHE	female	34	unscripted
bint_09 (s2)	2	Broadcast interviews	195	HHE	male	34	unscripted
bint_10 (s1)	2	Broadcast interviews	593	East central	male	50	unscripted
bint_11 (s2)	4	Broadcast interviews	528	East central	female	24	unscripted
bnew_06 (s1)	1	Broadcast news	241	West central	male	30	scripted
bnew_08 (s1)	2	Broadcast news	448	HHE	male	50	scripted
bnew_13 (s1)	2	Broadcast news	294	West central	female	44	scripted
bnew_16	3	Broadcast news	443	Northeast	male	50	scripted
bnew_19 (s1)	22	Broadcast news	1711	HHE	male	45	scripted
bnew_19 (s5)	22	Broadcast news	175	HHE	male	45	unscripted
bnew_20 (s1)	5	Broadcast news	540	HHE	male	45	scripted
bnew_21 (s1)	18	Broadcast news	2048	Northeast	female	41	scripted
bnew_29 (s1)	25	Broadcast news	1787	Northeast	female	41	scripted
bnew_30 (s1)	26	Broadcast news	2807	West central	male	49	scripted
bnew_31 (s1)	22	Broadcast news	1936	West central	male	49	scripted
bnew_35 (s1)	3	Broadcast news	391	Northeast	female	37	scripted
bnew_36	2	Broadcast news	227	East central	female	32	scripted
btal_01	2	Broadcast talks	415	West central	female	18	scripted
btal_02	2	Broadcast talks	455	West central	male	25	scripted
btal_03	3	Broadcast talks	473	East central	female	60	scripted
btal_04	2	Broadcast talks	381	East central	male	67	scripted
btal_05	7	Broadcast talks	933	West central	male	52	scripted
btal_07	2	Broadcast talks	486	HHE	female	30	scripted
btal_08	1	Broadcast talks	215	Northeast	female	52	scripted
btal_09	4	Broadcast talks	424	West central	female	61	scripted
btal_10	3	Broadcast talks	432	Northeast	female	57	scripted

btal_11	2	Broadcast talks	437	West central	male	45	scripted
btal_12	2	Broadcast talks	393	East central	female	50	scripted
btal_13	3	Broadcast talks	456	East central	female	45	scripted
btal_14	3	Broadcast talks	443	West central	female	60	scripted
btal_15	2	Broadcast talks	402	Northeast	female	52	scripted
btal_16	3	Broadcast talks	513	HHE	female	17	scripted
btal_17	6	Broadcast talks	1017	West central	female	44	scripted
btal_18	8	Broadcast talks	1439	East central	female	48	scripted
btal_19	4	Broadcast talks	706	West central	female	50	scripted
btal_20	4	Broadcast talks	708	HHE	female	45	scripted
btal_21	9	Broadcast talks	1609	West central	male	63	scripted
btal_22	12	Broadcast talks	2053	East central	male	50	scripted
btal_23	8	Broadcast talks	1427	East central	male	52	scripted
btal_25	3	Broadcast talks	505	Southern	male	61	scripted
btal_26	5	Broadcast talks	802	East central	female	66	scripted
btal_27	7	Broadcast talks	1310	Northeast	female	73	scripted
btal_28	4	Broadcast talks	726	West central	male	45	scripted
btal_29	17	Broadcast talks	2576	East central	female	55	scripted
btal_30	3	Broadcast talks	414	West central	female	45	scripted
btal_31	2	Broadcast talks	454	East central	male	45	scripted
btal_32	2	Broadcast talks	395	East central	female	55	scripted
btal_33	3	Broadcast talks	435	East central	male	50	scripted
btal_34	7	Broadcast talks	993	West central	female	63	scripted
btal_35	7	Broadcast talks	1095	East central	male	60	scripted
btal_36	9	Broadcast talks	1291	East central	male	47	scripted
btal_37	16	Broadcast talks	2069	West central	male	78	scripted
btal_38	4	Broadcast talks	597	West central	female	57	scripted
btal_39	2	Broadcast talks	370	East central	male	51	scripted
btal_40	10	Broadcast talks	1596	West central	male	60	scripted
btal_41	5	Broadcast talks	844	West central	female	44	scripted
btal_42	4	Broadcast talks	647	West central	female	52	scripted
btal_44	4	Broadcast talks	459	HHE	male	65	scripted
btal_49	7	Broadcast talks	1117	HHE	male	54	scripted
leg_13	6	Legal cross-examinations	992	East central	male	63	scripted
leg_14	4	Legal cross-examinations	641	East central	male	63	scripted
leg_15	3	Legal cross-examinations	599	East central	male	59	scripted
leg_16	6	Legal cross-examinations	925	East central	male	59	scripted
leg_18	0	Legal cross-examinations	101	West central	male	65	scripted
nbtal_06	30	Non-broadcast talks	3887	West central	male	59	scripted
nbtal_07	2	Non-broadcast talks	357	HHE	female	35	scripted
nbtal_08	2	Non-broadcast talks	473	HHE	female	21	scripted
nbtal_11	4	Non-broadcast talks	892	West central	male	21	scripted
parl_05 (s3)	36	Parliamentary debates	1400	HHE	female	50	unscripted
parl_05 (s7)	36	Parliamentary debates	1700	Northeast	male	56	unscripted
parl_06 (s3)	15	Parliamentary debates	700	Northeast	male	56	unscripted
unsp_02	2	Unscripted speeches	444	West central	female	44	unscripted
unsp_06	4	Unscripted speeches	680	Northeast	male	50	unscripted
unsp_07	3	Unscripted speeches	619	West central	female	35	unscripted
unsp_08	4	Unscripted speeches	648	East central	female	30	unscripted
unsp_13	2	Unscripted speeches	433	West central	male	65	unscripted

The selected ICE Scotland files include 72677 words in total. The total duration of the files amounts to 10 hours and 40 minutes and the average duration of a file is seven minutes. There is, however, great durational variation between the files and it must be added that for some files, only some specific speakers with, for example, a certain dialect background were analyzed. This is why a longer file duration does not always coincide with a higher word count in the present study.

As seen in Table 11, the files were retrieved from the ICE categories *broadcast discussions* (14 files), *broadcast interviews* (5 files), *broadcast news* (13 files), *broadcast talks* (42 files), *legal cross-examinations* (5 files), *non-broadcast talks* (4 files), *parliamentary debates* (3 files) and *unscripted speeches* (5 files). The categories *broadcast discussions* and *broadcast interviews* include mostly televised discussion groups and interviews and the category *broadcast news* includes news broadcasts on Scottish radio and television channels. The files from *broadcast talks* are mostly prepared speeches read by public guests in the Scottish Parliament and the *legal cross-examinations* are recordings from speeches made by Scottish judges in court. The category *non-broadcast talks* comprises prepared speeches by Scottish university lecturers and politicians which were not broadcasted live. The *parliamentary debates* and most of the *unscripted speeches* were also held in the main chamber of the Scottish Parliament. Thus, all speeches and interviews were held in a very formal and public context and all selected speakers used SSE.

Table 11 also shows that most of the files belong to the category *scripted* and that the vast majority of speakers grew up in the Central Belt of Scotland. The dialect region *West central* is represented by 39 files and *East central* by 25 files. 14 files represent the region *HHE* and 12 other files the *Northeast* of Scotland. There is only one male speaker from the region *Southern* and the dialect region *Insular* is not represented at all.

Table 12. Word counts of the ICE Scotland files separated for style and dialect region with total numbers.

Style	East central	HHE	Northeast	Southern	West central	Total
scripted	17424	6812	7028	505	24483	56252
unscripted	3170	2236	3876	-	7143	16425
Total	20594	9048	10904	505	31626	72677

These distributional differences are even clearer in the word count (see Table 12). A total amount of 56252 words (77.4%) derive from *scripted* registers and only 16425 words (22.6%) from *unscripted* speech forms. As for the dialect regions, *West central* is represented by 31626 words (43.52%) and *East central* by 20594 words (28.34%). There are 10904 words (15 %) for the *Northeast* and 9084 words (12.45%) for *HHE*. The dialect region *Southern* is represented by only 505 words (0.69%) and all of them represent scripted speech. As for *gender*, 41599 of the words were spoken by male speakers (57.24%) signifying a small imbalance. A much greater imbalance is noticeable for *age* as 78% of all words were produced by speakers which are between 30 and 60 years old. The middle-aged speakers are therefore clearly overrepresented. To conclude, there is a clear lack of speakers from the regions *Insular* and *Southern* in the selected ICE Scotland files. The dataset lacks unscripted speech in general and it includes relatively few young and old speakers. These imbalances are the reason why the present study incorporated further datasets to counterbalance these distributions.

4.2.2 Alba Recordings

In addition to the data from ICE Scotland, I collected further spoken data in the form of short unstructured interviews. The main goal was to record samples of spontaneous SSE from different speakers to counterbalance the lack of unscripted speech data in ICE Scotland (see subsection 4.2.1). To obtain spontaneous speech and mitigate the observer's paradox (Labov, 1972, p. 209), I used a free structure in the interviews to make participants feel at ease. Following the approach of "let the informant talk" (Tagliamonte, 2006, p. 46), I did not interrupt the participants but tried to establish an unconstrained conversation. Nevertheless, my presence as an outsider, a young male researcher from Germany with a non-Scottish accent might have influenced the language behavior of most participants. This, however, had the positive effect that most speakers used SSE throughout the interviews. Hence, the avoidance of Scots features might be a result of accommodation towards the interviewer (Di Paolo & Yaeger-Dror, 2011, p. 11; Schützler, 2015, pp. 47–48).

The interviews were carried out in February and early March 2020 before the COVID-19 pandemic in different locations in Scotland, namely East Kilbride, Glasgow, Hawick, Edinburgh, Girvan, Stranraer and Oban. I explicitly chose Hawick, Girvan and Stranraer in the hope to interview speakers from the South of Scotland and I chose Oban to record speakers from the Highlands and Hebrides. Most of the interviews were conducted in local libraries due to their relatively quiet environment. After getting official permission from the libraries and the corresponding councils, I randomly asked library visitors for a short interview of about 10 minutes. Instead of conducting long interviews (Labov, 1984, p. 8), the goal was to interview as many people as possible in a relatively short timeframe. I explicitly conducted short interviews so that the speech data is comparable to the length of the unscripted *broadcast discussions* of ICE Scotland (see Table 11). The interviewees were first briefed that the recordings will be anonymized and only used for academic purposes. I then collected written consent from each of the interviewees and asked for other sociodemographic data which can have an influence on the language use of the speakers (age, gender, city or region of upbringing, city of residence, highest level of education, profession, longer periods stayed outside Scotland, ethnicity, regional background of parents, other languages than English). The consent form can be found in the appendix. The speakers were informed after the interview about the purpose of the investigation.

The recordings were made with a Zoom H5 recorder which was placed on a table between the interviewer and the interviewee. The microphone was directed towards the interviewee and the device recorded stereo files in a .wav format at a sampling rate of 44100 Hz with a 16-bit resolution. Recordings with significant and constant background noise were excluded from the analysis.

Overall, a total number of 22 interviews with a total duration of approximately five hours were used for the analysis. The total number of words amounts to 33875. An overview of the dataset can be found in Table 13.

Table 13. Overview of the interview data of Recordings Alba used for analysis.

File	Duration (minutes)	Word count	Regional upbringing	Gender	Age (at time of recording)	Style
03_Alba	9	948	HHE	female	54	unscripted
06_Alba	11	1518	Northeast	female	23	unscripted
12_Alba	25	2928	HHE	female	42	unscripted
16_Alba	16	1812	East central	female	49	unscripted
18_Alba	15	1475	HHE	female	31	unscripted
22_Alba	13	1303	East central	female	64	unscripted
23_Alba	10	1412	Southern	female	41	unscripted
25_Alba	21	2343	Southern	female	51	unscripted
27_Alba	10	1222	Southern	female	60	unscripted
28_Alba	9	1309	Southern	male	56	unscripted
31_Alba	17	2380	HHE	male	44	unscripted
39_Alba	14	1472	East central	male	52	unscripted
42_Alba	7	864	East central	female	48	unscripted
43_Alba	11	1408	Northeast	male	50	unscripted
48_Alba	9	940	Northeast	male	55	unscripted
52_Alba	11	1233	Southern	male	87	unscripted
54_Alba	9	961	Southern	female	42	unscripted
56_Alba	12	1009	Southern	male	56	unscripted
60_Alba	19	2148	HHE	male	52	unscripted
62_Alba	12	1502	HHE	female	53	unscripted
63_Alba	11	1538	HHE	male	58	unscripted
64_Alba	23	2150	HHE	female	68	unscripted

Due to the unstructured format of the interviews, there were some participants who wanted to finish early (minimum duration: 9 minutes) and others who were happy to have a longer conversation (maximum duration: 25 minutes). The average duration of an interview is 13.3 minutes. The interviews elicited spontaneous SSE and this is why all files were categorized as being *unscripted.*

Table 14: Word counts of the Alba recordings files separated for gender and dialect region with total numbers.

Gender	East central	HHE	Northeast	Southern	Total
female	3979	9003	1518	5938	20438
male	1472	6066	2348	3551	13437
Total	5451	15069	3866	9489	33875

The female speakers produced 20438 words in total (60.33%) (see Table 14). In contrast to the *ICE Scotland* dataset (see subsection 4.2.1), *Alba Recordings* includes more female (N=13) than male speakers (N=9). As for the dialect regions, *East central* is represented by 5451 words (16.09%) and *HHE* by 15069 words (44.48%). There are 3866 words (11.41%) for the *Northeast* and 9489 words (28.01%) for the dialect region *Southern*. However, similar to ICE Scotland, there are no speakers from the dialect region *Insular*. While I intended to visit other dialect regions for recording interviews, the outbreak of the COVID-19 pandemic in March 2020 and subsequent lockdowns made this impossible. This is the reason why further data was retrieved via online sources.

4.2.3 Home Recordings

Another dataset was compiled from the internet during the COVID-19 pandemic in 2020 and 2021. These recordings will be named *Home Recordings* in the following and they are very similar to the sound files of ICE Scotland (see subsection 4.2.1). Like ICE Scotland, much of the spoken SSE data was retrieved from the Scottish Parliament and from local Scottish radio stations. This includes scripted political speeches and news broadcasts but also unscripted language retrieved from parliamentary debates and broadcasted interviews. I especially tried to retrieve spoken data from speakers who grew up in those dialect regions which are not well represented in *ICE Scotland* (see subsection 4.2.1) and *Alba Recordings* (see subsection 4.2.2), namely *Insular*, *Southern*, *HHE*, *Northeast* and partly *East central*. I contacted the speakers and institutions and asked for permission to use the spoken data. Whenever I could contact an individual person, I also asked them to sign the same consent form which was given to participants of *Alba Recordings*. This not only gave me personal consent, but it also provided me with further sociodemographic background information about the speakers. The recordings from Scottish Parliament TV are publicly available but I nevertheless got consent from the Scottish Parliament to use the material under the Scottish Parliament copyright license (Scottish Parliament, 2022). I used only material for which I received written consent.

Table 15. Overview of the interview data of Home Recordings used for analysis

File	**Duration (minutes)**	**Word count**	**Regional upbringing**	**Gender**	**Age** (at time of recording)	**Style**
01_Home_A	5	415	Insular	Male	28	scripted
01_Home_B	5	312	Insular	Male	28	scripted
02_Home	6	394	Insular	Male	59	scripted
03_Home_A	6	450	Insular	Male	48	scripted
03_Home_B	3	297	Insular	Male	48	scripted
04_Home	27	3562	Insular	Male	67	scripted
05_Home	25	3202	Insular	Male	33	scripted
06_Home	22	2812	Insular	Female	25	scripted
08_Home	32	961	Insular	Female	51	scripted
09_Home	28	3147	Insular	Female	65	scripted
10_Home	21	3436	Insular	Male	20	unscripted
11_Home	5	1012	Insular	Female	59	scripted
12_Home	8	1389	Southern	Male	45	unscripted
13_Home	26	2375	Insular	Male	25	scripted
14_Home	23	1683	Southern	Male	28	unscripted
16_Home_A	26	785	Southern	Female	23	unscripted
16_Home_B	26	1702	Southern	Female	25	unscripted
18_Home	6	969	Southern	Male	28	unscripted
19_Home	7	1182	Southern	Male	50	unscripted
20_Home	7	1173	Southern	Female	50	scripted
21_Home	7	1049	Southern	Male	63	scripted
22_Home	6	1022	Insular	Female	65	scripted
24_Home	6	1127	Southern	Male	51	scripted
25_Home	7	1141	Northeast	Male	67	scripted
26_Home	6	931	Northeast	Female	63	scripted
28 Home	4	766	Insular	Female	28	scripted
29_Home	5	647	Insular	Female	31	scripted
30_Home	5	509	Insular	Male	45	unscripted
31_Home	4	828	Insular	Male	66	scripted
32_Home	10	1682	Southern	Male	29	scripted
33_Home	8	1677	Southern	Male	45	unscripted
34_Home	7	1154	East central	Female	31	unscripted
35_Home	3	554	Northeast	Male	25	unscripted
36_Home	4	611	Northeast	Female	60	unscripted

The recordings were made with the audio editor Audacity (Audacity Development Team, 2019) in a 16-bit resolution and with a sampling rate of 44100 Hz. The recording quality is therefore comparable to the recording quality of the *Alba Recordings* (see subsection 4.2.2). The files were subsequently exported from Audacity in a .wav format. Overall, I retrieved 34 sound files with a total duration of 6.6 hours and a word count of 44956. An overview can be found in Table 15.

The average duration of the sound files is 11.6 minutes and the average number of words equals 1322. Yet, there is great variability with a minimum duration of 3 minutes (297 words) for the shortest file and a maximum duration of 32 minutes (961 words) for the longest file in this dataset. Nevertheless, a longer duration does not always coincide with a higher word count. For example, the file *09_Home* is four minutes shorter than the longest file (08_Home) but it has a much higher word count with 3147 words. The reason for this is that the interviewed speaker in *09_Home* talked a lot more than the interviewed speaker in *08_Home* which, ultimately, led to a different number of words.

Table 16. Word counts of the Home Recordings files separated for style and dialect region with total numbers

Style	East central	Insular	Northeast	Southern	Total
scripted	-	6143	2072	5031	**13246**
unscripted	1154	20004	1165	9387	**31710**
Total	**1154**	**26147**	**3237**	**14418**	**44956**

As seen in Table 16, the majority of the files belong to the category *unscripted* (70.5%) and many of the speakers grew up in the dialect region *Insular*. The word count for *Insular* amounts to 26147 words which equals 58.16% of all words from the dataset. *Home Recordings* is therefore the first dataset that incorporates speakers from that dialect region. The rest of the speakers hail from the dialect regions *Southern* (32.07%), *Northeast* (7.20%) and there is also one speaker from the dialect region *East central* whose speech represents 2.57% of the dataset *Home Recordings*. The dialect regions *HHE* and *West central* are not represented. As for *gender*, a total of 62.8% of all words were spoken by male speakers.

4.3 Data preparation

As this study incorporates several datasets with different formats, an essential task was to convert these datasets into a uniform format that meets the requirements set out in section 4.1. This means that the final data format should include information about the general factors that influence vowel duration (see section 3.1), the SVLR (see section 3.2) and the VE (see subsection 3.1.2), but it should also entail information about the speaker, his or her age and the dialect background to investigate possible idiosyncratic and social variability in SSE vowel durational patterns. For this task, I used different tools and prepared different algorithms to semi-automatically transform the datasets into the desired uniform format. Yet, whereas ICE Scotland already provides time-aligned transcriptions on the word and segment level, the self-collected data still needed to be transcribed. Hence, the preparation processes of the data sources differ and I will therefore discuss them separately in the following subsections. The data preparation process of the ICE Scotland files will be described in subsection 4.3.1 and the transcription process of the self-collected data, including *Alba Recordings* and *Home Recordings*, will be described in subsection 4.3.2.

While the data preparation process will be discussed separately for the datasets in the following subsections, I will first introduce software packages and transcription procedures which are used for all data sources. This includes the speech software Praat (Boersma & Weenink, 2019), the process of forced alignment as well as the segmentation approach used in the present study.

The speech software Praat (Boersma & Weenink, 2019) proved very useful for the present investigation and it was therefore used extensively in the data preparation process. Praat (Boersma & Weenink, 2019) is a widely used speech analysis tool with a visual interface and uses the ".TextGrid" format for speech segmentation and transcription. A TextGrid transcription consists of at least one tier and one can split this tier into different intervals. The TextGrid format was especially useful as it can be adapted both manually in the Praat interface and automatically with the help of different algorithms. The Praat (Boersma & Weenink, 2019) interface is visualized in Figure 12 and one can see the waveform and the corresponding broadband spectrogram at the top as well as three TextGrid transcription tiers at the bottom. While the waveform and broadband spectrogram represent the sound file in Praat, the transcription tiers at the bottom are saved in a text file in a TextGrid format. For the segmentation of speech sounds, one can manually create, delete or move the blue interval boundaries of the TextGrid transcription tiers and one can also edit the transcription annotations via the Praat visual interface

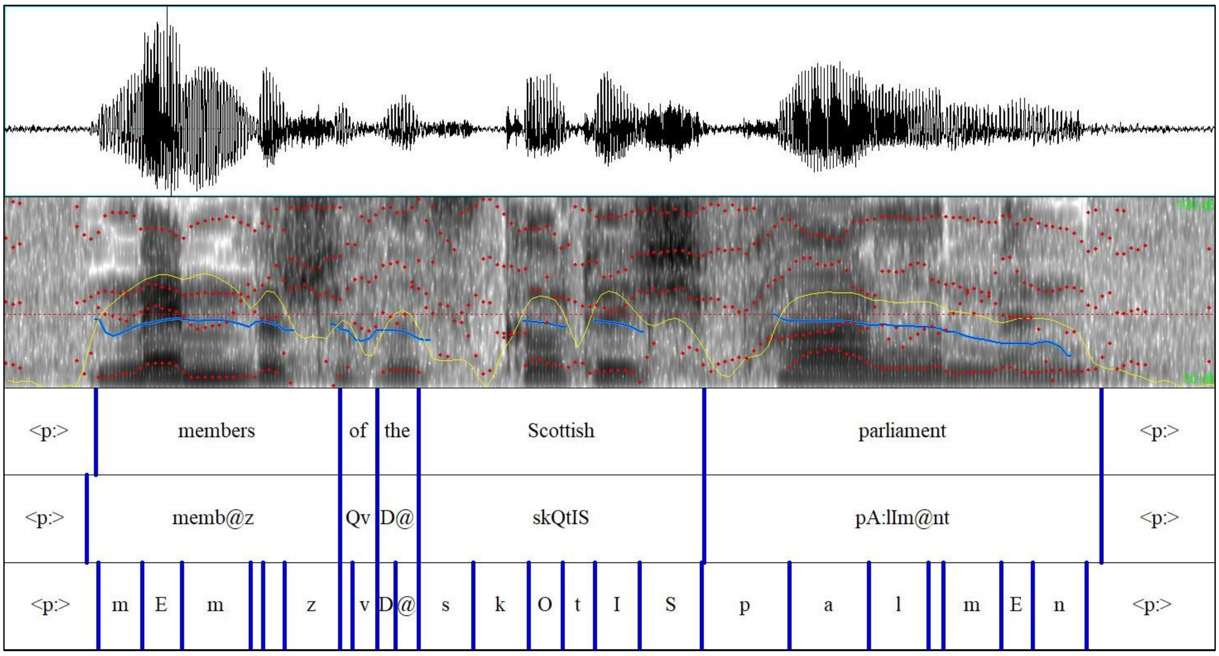
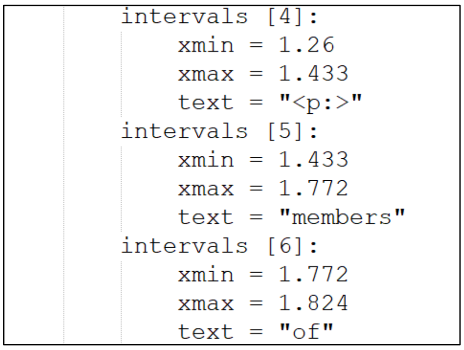


Figure 12. Extract from a forced-aligned and manually corrected ICE Scotland sound file with the corresponding transcription tiers visualized in Praat. The yellow line in the spectogram is the intensity curve, the blue lines represent the detected pitch. The red dots are the automatically traced formants by Praat.

(Boersma & Weenink, 2019). However, it is also possible to edit the tiers directly in the TextGrid file without the visual interface. The core structure of three TextGrid intervals is visualized in Figure 13.

As seen in Figure 13, a TextGrid interval always contains a start time (xmin) and end time (xmax) as well as the corresponding transcription within that given time frame (text). The end time of one interval is always identical to the start time of the next interval so that there are precise boundaries which are also indicated by the blue interval boundary lines in the Praat visual interface (see Figure 12). If one wants to retrieve the duration of a particular interval, one can simply subtract the start time (xmin) from the end time (xmax) of that interval. While Praat includes a built-in programming language (Boersma & van Heuven, 2001, p. 344), most of the automatized TextGrid preparation was carried out with self-designed algorithms in Python (Python Software Foundation, 2021) using the library *Praat-textgrids* (Nieminen, 2019)[1]. In contrast to the general-purpose Praat programming language, Python code executes much faster and its syntax allows for many more operations.



```
intervals [4]:
    xmin = 1.26
    xmax = 1.433
    text = "<p:>"
intervals [5]:
    xmin = 1.433
    xmax = 1.772
    text = "members"
intervals [6]:
    xmin = 1.772
    xmax = 1.824
    text = "of"
```

Figure 13. Core structure of three Textgrid intervals from the word tier of an ICE Scotland file. The same word intervals are visualized in the Praat interface in Figure 12.

The process of forced alignment is also essential for the data preparation procedure and it therefore needs further explanation. To provide more information on how forced alignment works, I will describe the transcription procedure of ICE Scotland which also implemented a forced alignment system. In the first step, broad utterance-based orthographical transcriptions were created manually via ELAN (Max Planck Institute for Psycholinguistics, 2022) for the ICE Scotland files by different transcribers. These broad transcriptions were subsequently force-aligned with the *Munich Automatic Segmentation System* (henceforth: MAUS) (Kisler et al., 2017). MAUS takes the audio file and its broad utterance-based orthographic transcription as input and automatically transfers the orthographic transcription into a phonological transcription with the help of a pronunciation dictionary (Schützler, 2015, p. 293). This process is called *grapheme-to-phoneme conversion* as the orthographic words are converted into the corresponding phoneme sequences. For example, the orthographic word <fish> is usually represented by the phonemic transcription /fɪʃ/. As the vowel system of SSE differs fundamentally from RP and GA (see section 2.3), the ICE Scotland researchers created a specific SSE pronunciation dictionary for

---

[1] The scripts and algorithms used for data preparation can be found in the repository which is accessible via this URL: https://tu-dortmund.sciebo.de/s/cRlncU1KwPSmh3m

MAUS written in the *Speech Assessment Methods Phonetic Alphabet* (henceforth: SAMPA). In the following step, MAUS automatically time-aligned the phonological SAMPA transcription to the audio of the sound recording with the help of a pre-trained acoustic model for English. The MAUS acoustic model for English is based on the *Hidden Markov Model Toolkit* (henceforth: HTK) (Young et al., 1999) and matches the phonemes of the SAMPA transcription with the acoustic properties of the corresponding timeframe in the audio file. Based on the acoustic transitions from one phoneme to the next, MAUS estimates the boundaries between the individual segments and subsequently between the different words (Schützler et al., 2017, p. 293). If the pronunciation dictionary includes multiple transcription options for individual words, MAUS chooses the transcription which fits best to the acoustic patterns in the recording (Schützler et al., 2017, p. 293). A classic example would be the indefinite article <a> with its stressed diphthongal realization /eɪ/ and its more frequent reduced realization /ə/: if the pronunciation dictionary contains both transcription options and the acoustic properties in the sound file indicate a diphthong, MAUS would ideally choose the diphthong /eɪ/ for transcription. Hence, MAUS always chooses the "'best guess' for a time-aligned phonemic transcription even if the acoustic data does not closely match any of the entries in the pronunciation dictionary" (Schützler et al., 2017, p. 293). The force-aligned transcriptions were then manually corrected by different research assistants in Praat (Boersma & Weenink, 2019).

The use of a forced aligner accelerates a task that is crucial for the present investigation: the segmentation of speech sounds. This process is very important because the beginning and end of a vocalic interval specify its duration and the duration subsequently forms the basis for the analysis of the

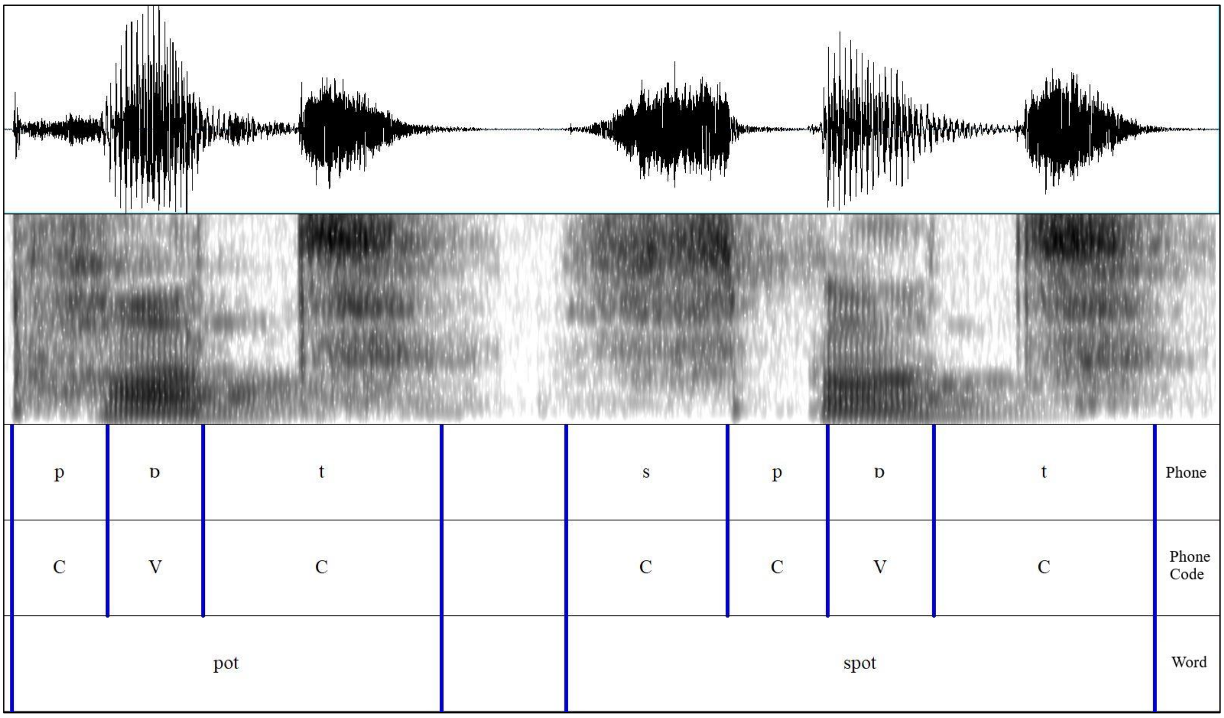


Figure 14. Visualization of the segmentation approach used in the present study applied on the words <pot> and <spot> said in isolation.

present study. While oral constriction criteria can be beneficial for vowel segmentation in carefully designed acoustic experiments (Turk et al., 2012), I follow the approach of previous SVLR research (Hewlett et al., 1999; McKenna, 1988; Rathcke & Stuart-Smith, 2016; Warren, 2018) and use voicing criteria for the segmentation of vocalic intervals. This means that the start of a vowel is determined by the "onset of periodic striations in the region of the first formant" (McKenna, 1988, p. 160) and the end of a vowel is determined "from the offset of the second formant" (Hewlett et al., 1999, p. 2158). This segmentation approach therefore only considers stable vocal fold vibration as part of the vocalic interval and does not take into account any other acoustic features such as VOT (see Figure 14). As speech segmentation is essential for measuring vowel duration, the segmentation approach will be discussed further in the following paragraphs.

As seen in Figure 14, the vocalic intervals in the words <pot> and <spot> are defined by the stable vibration pattern in the broadband spectrogram (second row) and by the high amplitude in the waveform. The aspiration which follows the word-initial plosive [p^h] in <pot> is considered part of the consonant and does not belong to the vowel. While there is no aspiration after the plosive in <spot> due to the preceding sibilant /s/, the short closure stage is considered part of the plosive /p/ and the vowel only starts at the onset of voicing. As for postvocalic contexts in <pot> and <spot>, the drop in the amplitude and the cessation of all formants indicates the end of the vowel and everything after this boundary is considered part of the /t/ in coda position. While this segmentation approach can lead to great variability in terms of consonant duration, it provides reliable measurements for vowel duration as vocalic intervals are outlined by stable patterns of the first formant (henceforth: F1) and the second formant (henceforth: F2). Furthermore, this segmentation approach is also in line with previous studies on Aitken's Law (Hewlett et al., 1999; McKenna, 1988; Rathcke & Stuart-Smith, 2016; Warren, 2018) and it is consistent

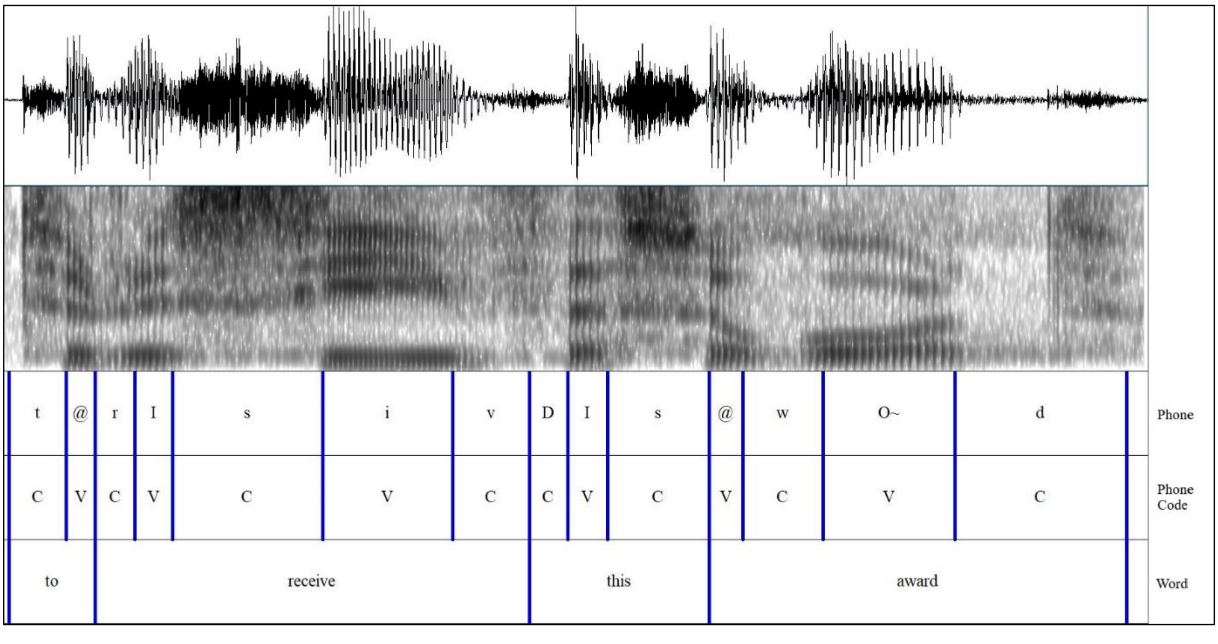


Figure 15. Sound file from ICE Scotland with time-aligned and manually corrected segmentations on the word and phone level, a corresponding SAMPA transcription and phoneme codes.

with the phonemic segmentation procedure of ICE Scotland (see Figure 12). However, the segmentability of vowel phones depends on the phonetic context in which they are embedded. As seen in Figure 14, it is relatively straightforward to detect the start and end boundaries of a vowel when it is preceded and succeeded by oral stops in a CVC or CCVC context (Turk et al., 2012, p. 5). It is also relatively easy to segment vowels when they are preceded or followed by nasals, affricates or fricatives (Turk et al., 2012, p. 5). As seen in the second syllable of the word <receive> in Figure 15, the boundary between the sibilant /s/ and the high vowel /i/ is placed at the offset of frication and the onset of voicing. Likewise, the end boundary of the same vowel is positioned at the end of the stable F2 pattern and at the start of the noise of the voiced fricative /v/. A similar pattern can also be seen in the word <this> in Figure 15: the formant pulses in the spectrogram and the high amplitude in the waveform indicate the vocalic interval which is surrounded by the noise of the interdental fricative /ð/ (SAMPA: /D/) and the sibilant /s/. It is, however, relatively difficult to establish acoustic boundaries between vowels and preceding or following glides or approximants (Turk et al., 2012, p. 5). As for the lateral approximant /l/, it is usually easier to identify a segment boundary when a vowel is following the consonant such as in the word <light> due to a "clear spectral discontinuity at constriction onset and release" (Turk et al., 2012, p. 15). However, it is often much more difficult to establish an interval boundary between a vowel and the lateral approximant in a postvocalic position due to a lack of acoustic cues. Word-initial /w/ and /j/ are usually fully voiced and consonant-vowel transitions are characterized by uninterrupted formant movements from the consonant to the vocalic nucleus. This can also be seen in the word <award> in Figure 15: while the segment of the intervocalic voiced labio-velar approximant /w/ includes parts of the formant pulses of the preceding and succeeding vowels, the overall transitions are very smooth. Hence, it is difficult to decide where the schwa /ə/ (SAMPA: /@/) ends and where the glide /w/ starts and it is also difficult to decide where the glide /w/ ends and the vowel /ɔː/ (SAMPA: O~) starts. In the same word, the tilde symbol "~" in the long vowel (SAMPA: O~) indicates rhoticity in the ICE Scotland transcription. While the rhotic oral constriction is acoustically perceptible as an alveolar approximant and partly visible by the formant change in the spectrogram, it is also difficult to draw a boundary between the vowel and the postvocalic rhotic sound. The same is true for the rhotic sound /r/ in the first syllable of the word <receive>. This rhotic sound is realized also as an alveolar approximant /ɹ/ and due to the continuous formant transition from the consonant to the vowel, it is difficult to identify a precise boundary (Turk et al., 2012, p. 14).

4.3.1 ICE Scotland

The transcription format of ICE Scotland has advantages and disadvantages. A clear advantage is that ICE Scotland provides a precise and consistent phoneme segmentation. This will be very useful in the present analysis. As seen in Figure 12, ICE Scotland provides orthographic transcriptions on the word level (first tier), the corresponding word-level phonemic transcriptions in SAMPA from the pronunciation dictionary (second tier) and the manually corrected phoneme transcription (third tier). The transcription in tier 3 is also partly phonetic as most segments have been manually corrected. An example would be the word <parliament> in Figure 12. Whereas the idealized phonemic SAMPA transcription includes a word-final /t/ (tier 2), the third tier does not include the segment [t] as the speaker did not actually produce the plosive at the end of the word. Another advantage of the manually corrected transcriptions of ICE Scotland is their precise and consistent segmentation of vocalic intervals. Following standard phonetic criteria (Peterson & Lehiste, 1960, p. 694), the start of a vowel is determined by the beginning of a stable F1 and the end of a vowel is determined by the end of a stable F2 (Gut, 2009b, p. 70). This can also be seen in the spectrogram in Figure 12: While the first vowel /e/ in the word <members> is surrounded by the bilabial nasal /m/, its beginning and end can be clearly defined by the darker formants and the overall increase in the amplitude. Especially the second formant is a lot darker in the vowel than in the nasals and this is also visible in the spectrogram. These acoustic cues made it relatively easy to identify the start and end of the vowel /e/ in this context. Another example would be the vowels in the word <Scottish> in Figure 12. Here, the vowel durations are determined by the two short voicing patterns in the spectrogram. The segmentation of the intervocalic /t/, however, takes up the silence, plosion and aspiration of the plosive between the vowels. The break of the plosive can also be seen in the drop of the yellow intensity curve. Hence, the vowel segmentation in ICE Scotland is based on the onset and offset of vocal fold vibration and does not include VOT (Turk et al., 2012, p. 9). The phoneme transcriptions and segmentations of the third tier are therefore very precise and they are in line with the segmentation approach of the present study.

However, there are many important features that are not represented by the transcriptions of ICE Scotland. These features must be added to the ICE Scotland files so that the format meets the requirements specified in subsection 4.1.3. As seen in Figure 12, there is no utterance or syllable level and there is also no information about lexical or prosodic stress in the ICE Scotland transcriptions. Another problem with the ICE Scotland transcriptions arises from the manual corrections of the phoneme segments. As the start and end times of most segments have been corrected by hand, they are sometimes not identical with the start and end times of the corresponding words. For example, the end boundary of the segment /n/ is not identical with the end boundary of the corresponding word <parliament> (see Figure 12). As a result, the total duration of the word is often not identical with the sum of the durations of its segments. While these inconsistencies are usually very small, they can lead to problems in the subsequent data analysis (see section 4.4). If the start and end times of the levels do not coincide completely, it will be difficult to identify which segments belong to which syllable, which syllables belong to which word and which words belong to which utterance. It would therefore be very good if the start and end times of the different levels match exactly so that the duration of the syllable is equal to the sum of its segment durations, the word duration is equal to the sum of its syllable durations and the utterance duration is equal to the sum of its word durations. Apart from that, there are also some general inter-transcriber inconsistencies in the ICE Scotland transcriptions. While the research assistants were given the same guidelines for manual correction, there were, of course, some transcribers which were more precise than others. Some transcribers annotated glottal stops more often and some others sticked more to the idealized phonemic transcription than to the actual phonetic realization. While small inconsistencies are unavoidable in such a big transcription project with different researchers involved, it would be good to check and correct significant errors and inconsistencies, especially when they occur in the vowel segments.

I used different tools and prepared different algorithms to correct inconsistencies in the ICE Scotland transcriptions and to convert them into a format that meets the requirements specified in subsection 4.1.3. The data preparation of the ICE Scotland files was conducted in five consecutive steps.

In the first step, the second tier with the idealized SAMPA transcription was automatically removed from all selected ICE Scotland files. This tier does not correspond with the manually corrected segment tier and will therefore not be helpful in the analysis. An algorithm subsequently parsed through the remaining orthographic *word* tier and aligned the word boundaries with the boundaries of the corresponding first phoneme and last phoneme in the *phoneme* tier. As a result, the word start time (xmin) is always identical with the start time (xmin) of the first phoneme of the word and the word end time (xmax) is always identical with the end time of the last phoneme in the word (xmax). The word duration is therefore the exact sum of the duration of its segments. Another algorithm subsequently copied the *word* tier and converted it into an orthographic *utterance* transcription. A helpful feature of the ICE Scotland files is that pause intervals are marked with the symbol "<p:>" (see Figure 12). The algorithm thus conglomerated all the word transcriptions that are positioned in between two pause interval symbols. Another algorithm copied the *phoneme* tier and replaced the segments with either vowel (V), consonant (C), or pause symbols (<p:>). To obtain a syllable structure and information on lexical stress, I used the pipeline service from WebMAUS (Kisler et al., 2017) and activated the *syllabification* and *word stress* function. This means that I force-aligned all ICE Scotland files again; the goal, however, was not to obtain force-aligned segments but to automatically obtain two tiers with information on syllable structure and word stress. The new *syllable* and *word stress* tiers from WebMAUS (Kisler et al., 2017) were then automatically added to the corresponding TextGrid files of ICE Scotland. This process had, however, one disadvantage. As the segment intervals of ICE Scotland were manually corrected, they are much more precise than the newly added force-aligned *syllable* and *word stress* tiers from WebMAUS (Kisler et al., 2017). This means that the start and end times of the force-aligned *syllable* and *word stress* tiers did often not correspond with the start and end times of the manually corrected *phoneme* tier from ICE Scotland. Also, the idealized phonemic syllable structure from WebMAUS (Kisler et al., 2017) did often not correspond with the actual pronunciations of the speakers. For example, an idealized phonemic transcription of the word <family> contains three syllables /ˈfæ.mə.li/. In fast connected speech, however, the word-medial schwa is often deleted resulting in a bisyllabic realization [ˈfæm.li]. For these problems, another two-step algorithm first checked whether the structure of the WebMAUS *syllable* tier coincides with the structure of the manually corrected ICE Scotland *phoneme* tier. Whenever there was a disagreement between the two tiers, the algorithm printed out the corresponding start times (xmin) of the cases. Using these start times for orientation, I subsequently corrected the cases manually in the *syllable* tier via the visual interface of Praat. As soon as all errors were corrected, the algorithm then aligned the structure as well as the start and end times of the *syllable* tier with the start and end times of the *phoneme* tier and also adapted the start and end times of the *lexical stress* tier accordingly. A visualization of the resulting format can be seen in Figure 16. The TextGrid files comprised six tiers after the first step of data preparation, namely an *utterance* tier, a *word* tier, a *word stress* tier, the corrected *syllable* tier from WebMAUS (Kisler et al., 2017), a *phoneme code* tier and a *phoneme* tier (see Figure 16). In addition, the interval boundaries do all match with the next higher prosodic level.

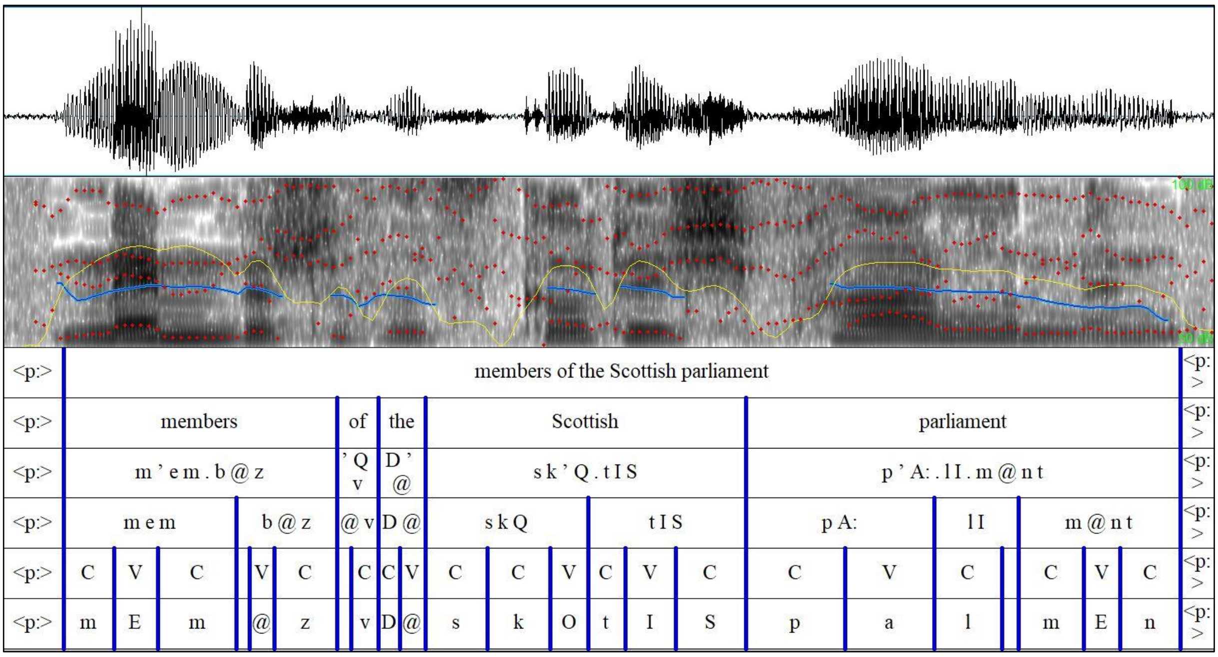


Figure 16. Adapted TextGrid structure of an ICE Scotland file after the first step of data preparation visualized in Praat.

In the second step, the *syllable* tier was first copied four times and the four copies were renamed *syllable structure* tier, *syllable position* tier, *lexical stress* tier and *prosodic stress* tier. The prosodic stress tier was left blank and will only be used in the third step of the data preparation process (see below). An algorithm then replaced the idealized SAMPA characters from WebMAUS (Kisler et al., 2017) in the *syllable* tier with the corresponding manually corrected segment transcriptions from the *phoneme* tier. This means that, for example, the idealized SAMPA transcription "m@nt" in the word <parliament> was replaced by "mEn" from the manually corrected ICE Scotland *phoneme* tier. This made the syllable transcription more precise as the word-final /t/ was not pronounced by the speaker (see Figure 12 and Figure 16). Another algorithm replaced the text in the *syllable structure* tier with the symbols from the *phoneme code* tier. As a result, the first syllable in the word <Scottish> would be represented by /skO/ in the *syllable* tier and by <CCV> in the newly created *syllable structure* tier. Hence, the *syllable structure* tier will make it possible to compare syllables of the same type.

Subsequently, another algorithm parsed through the *syllable position* tier and specified whether the syllables are either in utterance-initial, utterance-medial, or utterance-final position. Syllables preceding a pause symbol ("<p:>") were categorized as utterance-final, syllables following a pause symbol as utterance-initial and the remaining syllables as utterance-medial. A subsequent algorithm transferred the information from the *word stress* tier into the newly created *lexical stress* tier. As seen in the *word stress* tier in Figure 16, the word <Scottish> carries primary lexical stress in the first syllable. This is indicated in the SAMPA transcription by the stress symbol (single quote ‘) and the syllable boundary is also specified by the full stop: /sk’Q.tIS/. The algorithm used the full stop and quotation mark for orientation and inserted the labels <stressed>, <unstressed>, or <mono> into the respective syllables in the *lexical stress* tier. This means that primary stressed syllables of polysyllabic words were marked <stressed>, lexically unstressed syllables with <unstressed> and monosyllabic words were labeled with <mono>.

In the third step of the data preparation process, I used the Praat plugin ProsoBox (Goldman & Simon, 2020) for obtaining information about prosodic stress. ProsoBox takes the audio file as well as the *phoneme*, *syllable* and *word* tiers as input and carries out automatic prosodic analyses. The plugin can automatically identify pitch accents in the different utterances and marks these accented syllables with the letter <P> in a new TextGrid tier with the name *promauto.* I then manually checked the results of ProsoBox and transferred the <P> with the label <accented> into the *prosodic stress* tier. On the basis of the accented syllables in the *prosodic stress* tier, I subsequently identified and marked nuclear stressed syllables in all TextGrid transcriptions with the help of auditory analyses. The categorization scheme in the *prosodic stress* tier therefore distinguishes between nuclear and non-nuclear stressed syllables on the suprasegmental level. The final TextGrid structure includes 11 tiers and a visualization can be found in Figure 17.

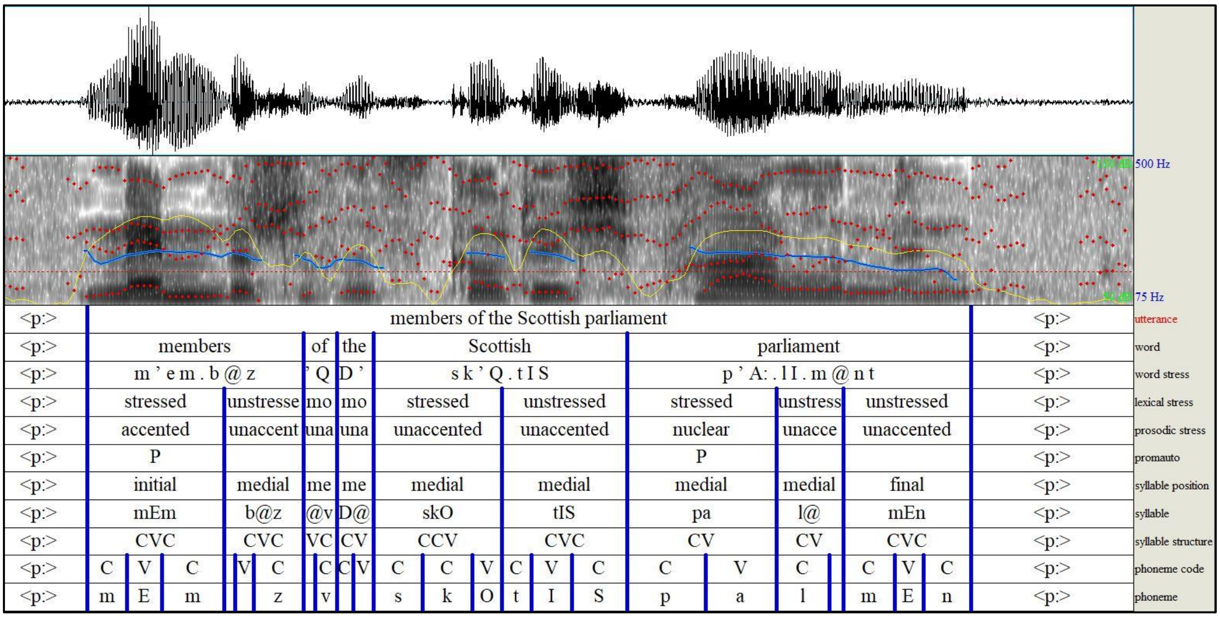


Figure 17. Final TextGrid structure of an ICE Scotland file after the third step of data preparation visualized in Praat.

In the fourth step of the data preparation process for the ICE Scotland files, all TextGrids were converted into one csv-file with the help of another Python algorithm. While the TextGrid format proved very useful for the data preparation process in Praat, the simpler .csv format is more suitable for the subsequent statistical analyses in R (see section 4.4). In the csv-file, each segment of the *phoneme* tier represents one row (see Table 17) so that each segment carries all relevant information (Hay, 2011, p. 198). While converting the format, the algorithm extracted not only the label of each segment but also its duration by subtracting the start time (xmin) from the end time (xmax) for each segment. In the same way, the algorithm also extracted the label and duration of the corresponding syllable, word and utterance interval in which the segment is positioned. For instance, in the fifth row of Table 17, the schwa (phone_label: /@/) has a total duration of 30 ms and is embedded in the CVC syllable /b@z/. The total duration of the syllable is 125 ms and represents the sum of the duration of its segments (17 ms (/b/) + 30 ms (/@/) + 78 ms (/z/) = 125 ms = 0.125 seconds). Hence, as seen in Table 17, each row contains information about the segment but also about the corresponding higher prosodic level.

Table 17. Extract from the csv structure of the ICE Scotland dataset. Each row is represented by one segment.

dataset	filename	phone_ label	phone_ xmin	phone_ xmax	phone_dur	phone_ code	syl_ structure	syl_label	syl_ duration	num_phone_ in_syl	+ other columns
ICE	btal_10	m	1.433	1.495	0.062	C	CVC	mEm	0.214	3	…
ICE	btal_10	E	1.495	1.550	0.05	V	CVC	mEm	0.214	3	…
ICE	btal_10	m	1.550	1.647	0.097	C	CVC	mEm	0.214	3	…
ICE	btal_10	b	1.647	1.664	0.017	C	CVC	b@z	0.125	3	…
ICE	btal_10	@	1.664	1.694	0.03	V	CVC	b@z	0.125	3	…
ICE	btal_10	z	1.694	1.772	0.078	C	CVC	b@z	0.125	3	…
+ other rows	…	…	…	…	…	…	…	…	…	…	…

In addition, the algorithm also extracted the number of segments in each syllable, the number of syllables in each word as well as the number of syllables in each utterance. The number of segments in each syllable makes it possible to account for *intrasyllabic compression* (see subsection 3.1.3) and the number of syllables in each word makes it possible to control for *polysyllabic shortening* (see subsection 3.1.4). Furthermore, the algorithm transferred the information of the *lexical stress* and *prosodic stress* TextGrid tiers (see Figure 17) into separate columns of the csv-file. These columns allow for investigating the influence of *lexical stress* (see subsection 3.1.6) and *prosodic stress* (see subsection 3.1.5) on vowel duration. The information of the *syllable position* tier (see Figure 17) was also added to the csv-file which makes it possible to control for the influence of *constituent-final lengthening* (see subsection 3.1.8). To facilitate the categorization of SVLR and VE environments (see section 4.3), the algorithm further transferred the label of the following segment into another column of the csv-file. For example, if the segment /i:/ is followed by the segment /z/ in the word <freeze>, the following voiced fricative indicates that this vowel is positioned in a long SVLR context (see section 3.2). The detailed SVLR and VE classifications will be discussed in subsection 4.4.2. Apart from that, the algorithm added the filename, the name of the dataset and the labels of all other tiers from the TextGrid format into subsequent columns of the csv-file (see Table 17) so that no information is lost while transforming the data structure.

In the fifth and last step of the data preparation process, the csv-file was read into R (R Core Team, 2021) and I added the relevant sociodemographic data (speaker, age, gender, regional background) via R commands to the dataset. I also created another column and identified proper nouns by searching for capitalized words. Based on the list of weak forms by Roach (2010, pp. 89–96), I created another column which specifies whether the segment is positioned in a content word or in a function word. Thus, there is a distinction between proper nouns and common nouns and another distinction between function words and content words in the dataset. The distinction between content words and function words makes it possible to account for the influence of *lexical category* (see subsection 3.1.5). As for the influence of *lexical frequency*, I used the Zipf scale values of the SUBTLEX-UK database (van Heuven et al., 2014) and added them to the respective words in the dataset. The SUBTLEX-UK database was chosen because, unlike other reference corpora for word frequency counts, it only contains contemporary British English speech data that was retrieved in mostly formal contexts (van Heuven et al., 2014, p. 1177). Hence, the structure of SUBTLEX-UK fits well with the spoken SSE data of this study. The Zipf scale is a logarithmic scale with values from 1 (very low frequency words) to 7 (high frequency words) and provides a better overview of word frequency effects than raw word frequency counts (van Heuven et al., 2014, p. 1179). The pronoun <you>, for instance, has a Zipf value of 7.31 which is much higher than the corresponding Zipf value of 1.47 for the city name <Penicuik>. The numeric Zipf value makes it possible to check whether there is a correlation between *lexical frequency* and vowel duration (see subsection 3.1.5). In addition to the numeric Zipf scale value, I also added the categorical variable *freq* which distinguishes between low frequency words (Zipf scale <3.5) and high frequency words (Zipf scale >3.5). Furthermore, I also used the English component of the CELEX database (Baayen et al., 1995) to obtain information about the morphological structure of words. This will be important because the morphological conditioning of Aitken's Law states that vowels are long when followed by a morpheme boundary as in the word <kneed>. The CELEX morphological database specifies that the word <kneed> has the morpheme structure <knee + ed> and this information was added to the dataset in another column. Based on the utterance duration and the number of syllables per utterance, I calculated the number of *syllables per second* by dividing the number of syllables by the total utterance duration and added the values into another column of the dataset. The values for the local articulation rate in terms of *syllables per second* make it possible to account for the influence of *tempo* (see subsection 3.1.9).

4.3.2 Self-collected data

The recordings of the self-collected data were orthographically transcribed on an utterance level via ELAN (Max Planck Institute for Psycholinguistics, 2022) and subsequently exported as TextGrid files. The subsequent data preparation of the files was conducted in five consecutive steps.

In the first step, the broad utterance transcriptions were force-aligned with two forced-aligners, WebMAUS (Kisler et al., 2017) and the updated second version of the Montreal Forced Aligner (henceforth: MFA) (McAuliffe et al., 2017). WebMAUS and the MFA were both used to harness the strengths of each system and compensate for their individual weaknesses. A great advantage of WebMAUS (Kisler et al., 2017) is that the system can provide transcriptions on multiple prosodic levels. It is therefore easy to obtain not only forced-aligned segments on the phoneme level but also the corresponding syllable, word and utterance segmentations and WebMAUS can also provide further information on lexical stress (see subsection 4.3.1). Furthermore, WebMAUS also includes a Scottish English based pronunciation dictionary. This dictionary was also used for the forced alignment of the ICE Scotland corpus (see section 4.3). Another advantage of WebMAUS is that it can handle unknown words because it always chooses the most likely phonemic transcription even if the word is not listed in its pronunciation dictionary (Schützler et al., 2017, p. 293). However, a great disadvantage of WebMAUS is the moderate alignment accuracy (Gonzalez et al., 2020). When compared to other common forced-aligners for English, MAUS produces the least accurate alignments and it is especially inconsistent in boundary placement between vowels and postvocalic segments (Gonzalez et al., 2020, p. 9).

In contrast to this, the MFA is arguably the strongest forced alignment system in terms of its segmentation accuracy. The error rates of the MFA in vowel onset boundary detection come close to the human segmentation standard (Gonzalez et al., 2020, p. 9). In contrast to all other forced aligners for English, MFA is not based on HTK (Young et al., 1999), but uses the more up-to-date Kaldi speech recognition toolkit (Povey et al., 2011). This means that the MFA employs more advanced speech recognition technology than MAUS which ultimately leads to a higher segmentation accuracy (Gonzalez et al., 2020, p. 9). Another advantage of the MFA is its trainability (McAuliffe et al., 2017, p. 499). Similar to MAUS, the MFA provides pre-trained acoustic models and pronunciation dictionaries for different languages which can be used for forced alignment. However, it is also possible to train the MFA on new data. This means that one can build MFA acoustic models that were trained on specific datasets. Instead of aligning SSE speech data with a pre-trained American English acoustic model, one can train a specific SSE acoustic model which fits better to the SSE dataset. The alignment accuracy can be significantly enhanced when acoustic models are used which were trained on the data itself (McAuliffe et al., 2017, p. 499). To train an acoustic model with the MFA, one must first generate an adequate pronunciation dictionary with the MFA's *g2p function*. This function takes the orthographic transcriptions of, for example, a selection of TextGrid files and matches the words with corresponding phonemic transcriptions using the transcription codes of *ARPABET* (see Table 18).

Table 18: Extract from a pronunciation model built with the MFA g2p function. The left column shows the orthographic words and the right column the corresponding phoneme strings (transcription code: ARPABET).

scotland	S K AA1 T L AH0 N D
scotland's	S K AA1 T L AH0 N D Z
scotrail	S K AA1 T R EY2 L

When training an acoustic model, the MFA then takes the audio files, the corresponding TextGrid transcriptions and the pronunciation dictionary as input. It then trains an acoustic model based on the audio files and the corresponding time-aligned TextGrid transcriptions by mapping the phoneme strings of the pronunciation dictionary to the corresponding acoustic cues in the audio files. The acoustic model can then be used for forced alignment. The trainability of the MFA is a great advantage as the forced alignment process can be adapted to a particular dataset which in turn improves segmentation accuracy. A small disadvantage is, however, that the generation of pronunciation dictionaries with the *g2p function* is based on American English. The ARPABET transcription of the word <Scotland> in Table 18, for example, represents the GA transcription /skɑːtlənd/ with the long open vowel /ɑː/. The GA phoneme system, however, partly differs from the SSE phoneme inventory (see section 2.3). Another disadvantage of the MFA is that it only produces alignments on the word and phoneme level. It does not provide segmentations for the syllable or utterance level. Furthermore, the word alignments of the MFA do not include capitalization in the transcription; all words are in lowercase (see Table 18). While this is not a serious problem, it can make the identification of proper nouns more difficult at a later stage.

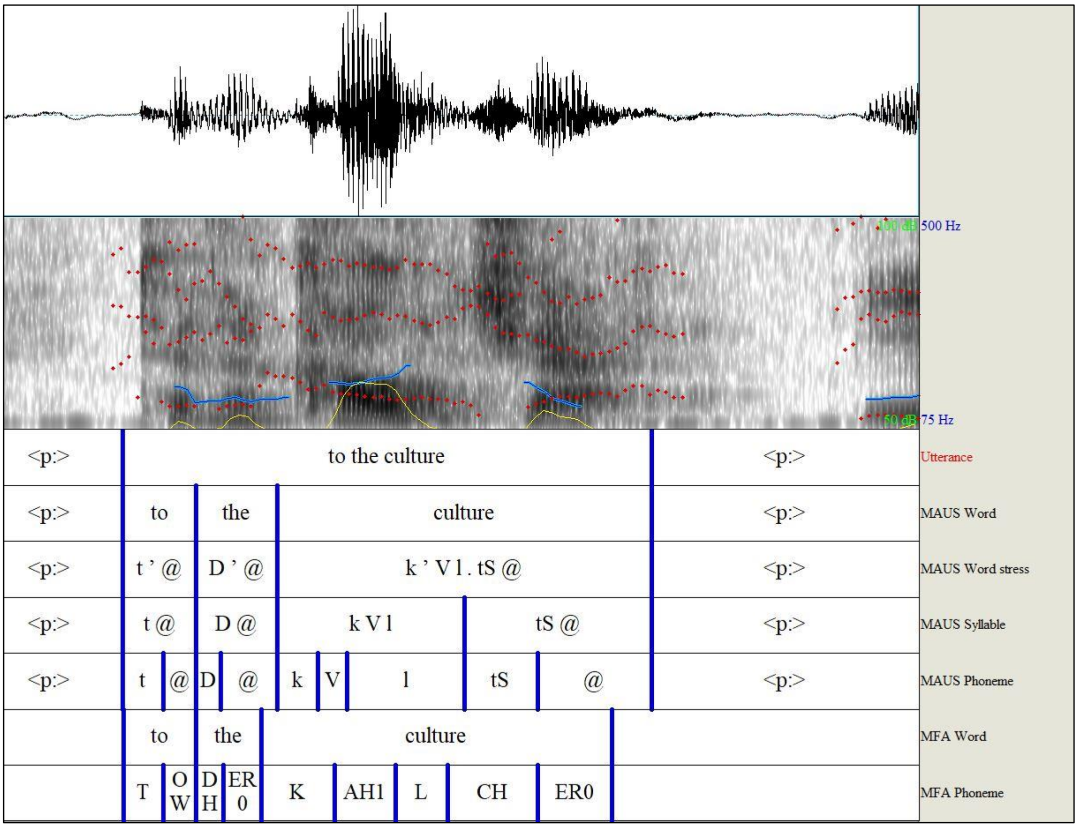


Figure 18. TextGrid structure of a file from Home Recordings after the forced alignment with WebMAUS and the MFA.

In the data preparation process of the self-collected datasets, I first used the pipeline service from WebMAUS (Kisler et al., 2017) with the activated *syllabification* and *word stress* function to obtain force-aligned transcriptions on the word, syllable and segment level with corresponding information about lexical stress and the syllabic structure. This produced four tiers, namely a *word* tier, a *word stress* tier, a *syllable* tier and a *phoneme* tier. I then used the *g2p function* of the MFA and created a pronunciation dictionary for forced alignment. I then trained an acoustic model on the basis of the whole dataset, including the already force-aligned and manually corrected transcriptions of the ICE Scotland files. This acoustic model was then used to force-align the broad utterance transcriptions of the self-collected data with the MFA. In other words, I created an acoustic model on the basis of the whole dataset of this study and this acoustic model was then used for force-aligning the broad utterance transcriptions of *Alba Recordings* and *Home Recordings*. The MFA alignment produced TextGrid files with *word* and *phoneme* tiers and these two tiers were added to the respective force-aligned files from WebMAUS. As a result, the TextGrid incorporated seven tiers in total: the original *utterance* tier, the force-aligned WebMAUS *word*, *word stress*, *syllable* and *phoneme* tiers as well as the force-aligned MFA *word* and *phoneme* tiers (see Figure 18.) While the transcriptions of WebMAUS and the MFA are relatively similar, the segmentation of the MFA is often more precise than the WebMAUS segmentation. As seen in Figure 18, the vowel /ʌ/ (ARPABET: AH1) in the MFA *phoneme* tier fits very well to the voicing pattern in the spectrogram. It starts at the beginning of a stable F1 and also ends at the end of the F1 and F2 voicing pattern. The WebMAUS segmentation, however, is not as precise: the same vowel /ʌ/ (SAMPA: /V/) starts before the vocal fold vibration and ends in the middle of the voicing pattern. The final schwa (ARPABET: ER0) in the word <culture> is also much more precise in the MFA segmentation than in the WebMAUS segmentation (SAMPA: /@/) (see Figure 18). While the start boundaries for that vowel are similar for both aligners, the MFA end boundary stops at the end of the stable F1 and F2 voicing pattern. The MFA end boundary is therefore very precise. In contrast to this, the WebMAUS segmentation for the same vowel ends at a much later point and includes noise that cannot be associated with a vocalic interval. A similar difference can be seen in the schwa of the word <the> for which the MFA also produced a more accurate segmentation than WebMAUS. To use the more precise vowel phoneme segmentation of the MFA while preserving the Scottish based SAMPA transcription as well as the *syllable* and *word stress* structure of WebMAUS, I created an algorithm which first parsed through the WebMAUS and MFA *phoneme* tiers and searched for overlaps in the vowel segmentation. For that, the algorithm transferred the ARPABET vowels codes of the MFA into the appropriate SAMPA vowel codes which were used by WebMAUS. Whenever there was an overlap in the vowels and timeframes, the start and end times of the WebMAUS segmentation were moved by the algorithm to the respective interval boundaries of the MFA segmentation. This process is visualized in Figure 19. This means that the WebMAUS phoneme segmentation was adapted to the more precise MFA phoneme segmentation while keeping the original Scottish based SAMPA transcription. The

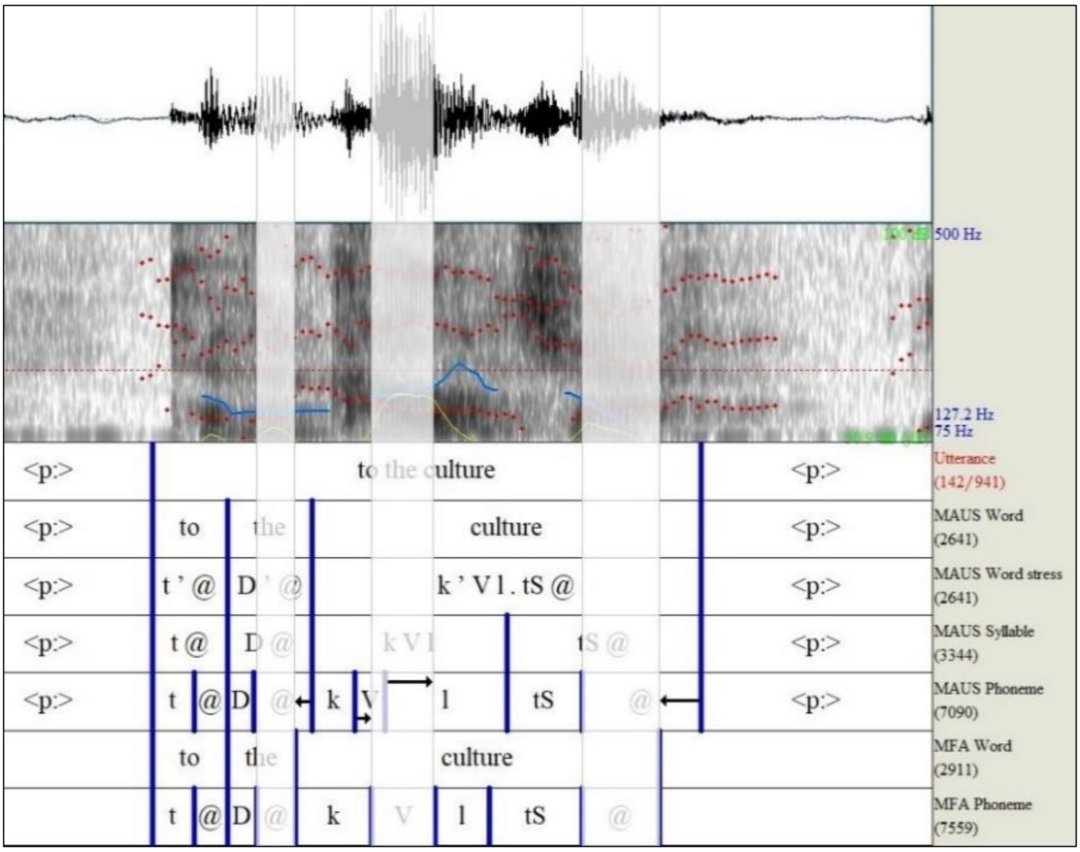


Figure 19. Visualization of the segmentation correction in the MAUS Phoneme tier. The grey bars and arrows show the changes which are performed by the algorithm.

algorithm then also adapted the intervals of the WebMAUS *syllable*, *word stress*, *word* tiers with the corrected WebMAUS *phoneme* tier so that the start and end times of the different prosodic levels match precisely. As a result, the duration of a word is the sum of the duration of its syllables and the syllable duration is the sum of the duration of its segments. After this correction procedure, the MFA *word* and *phoneme* tiers were removed from the files. Another algorithm then copied the *phoneme* tier, renamed the copy *phoneme code* tier and replaced the segments with either vowel (V), consonant (C), or pause symbols (<p:>).

The subsequent steps of data preparation are identical to the preparation procedure of the ICE Scotland files (see subsection 4.3.1). This means that the syllable tier was first copied four times and the copies were renamed *syllable structure* tier, *syllable position* tier, *lexical stress* tier and *prosodic stress* tier. The intervals of the tiers were corrected so that the duration of each prosodic element is the exact sum of the corresponding intervals on the next lower prosodic level. The syllables were categorized in terms of internal structure (e.g. CVC), phrasal position (e.g. initial, medial, final) and lexical stress and I also used ProsoBox (Goldman & Simon, 2020) to obtain information on prosodic stress. The final TextGrid structure contained 11 tiers (see Figure 20) and was transformed into a .csv format with the help of the same Python algorithm that was used for the ICE Scotland files. The csv-file was then read into R (R Core Team, 2021) and further intralinguistic data (word type, word frequency, proper noun or common noun, average syllable duration per utterance, average syllable per second for each utterance, morphological structure) as well as sociodemographic information (speaker, age, gender, regional background) was added.

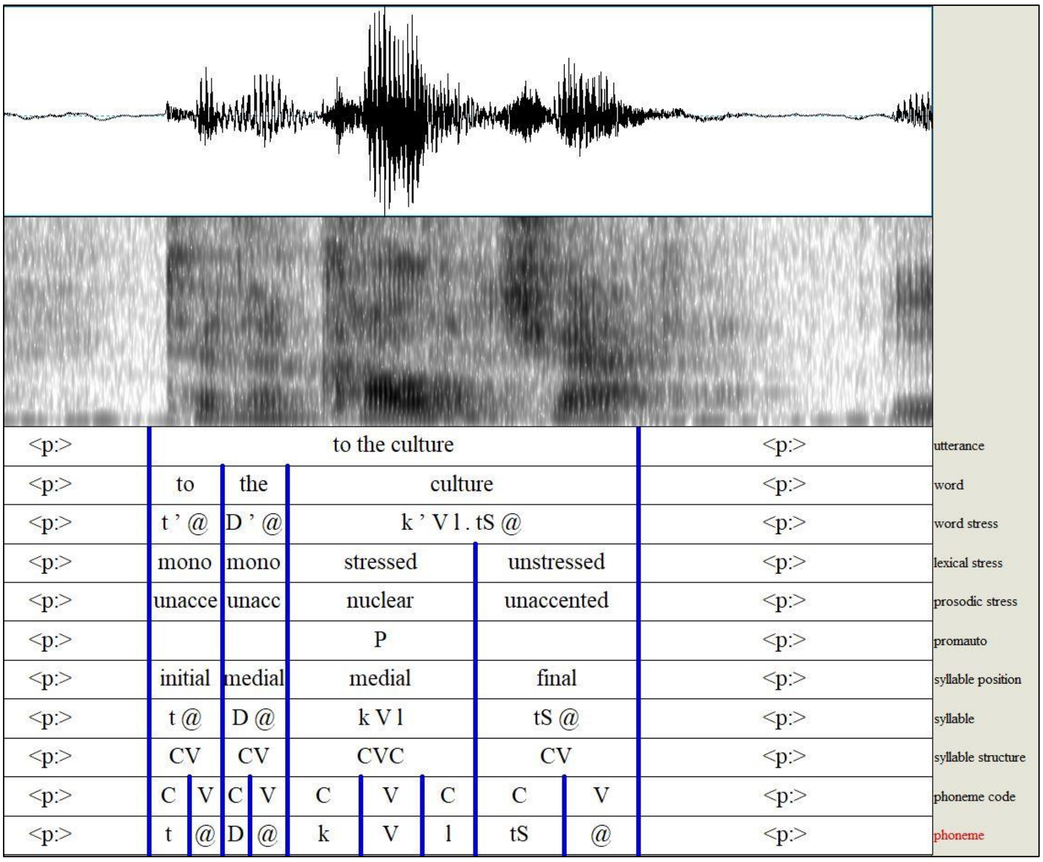


Figure 20. Final TextGrid structure of an Alba Recordings file after the third step of data preparation visualized in Praat.

4.4 Data analysis

The data analysis was conducted in R (R Core Team, 2021). In the first step, I merged the csv-files of the data sources (ICE Scotland and self-collected data) into one csv spreadsheet and excluded all rows in which the phone labels are pauses and hesitations. This study investigates vowel duration and therefore, the rows with hesitations and pauses are not of interest. In the following, I will provide an overview of the vowel selection criteria (subsection 4.4.1) and the predictor variables that will be implemented in the analysis (subsection 4.4.2). The statistical analysis will be discussed in subsection 4.4.3.

4.4.1 Vowel selection

As described in section 2.3, the vowel selection of the present study is largely based on the Basic Scottish Vowel System by Abercrombie (1979). Thus, in contrast to all previous investigations, this will be the first study which analyses the SVLR and the VE in all vowels of the Basic Scottish Vowel System. There are only two minor modifications. In the lexical set DRESS, the present study will not distinguish between the vowel /ɛ/ and the more centralized variant /ɛ̈/ as the latter can only be found among some speakers in some regions (Abercrombie, 1979, p. 75). Furthermore, the present study will investigate all tokens of the lexical set PRICE. Many previous studies note a quality change in SSE for this lexical set, but this study will not distinguish between /ʌɪ/ tokens in monomorphemic words and /ae/ tokens in heteromorphemic words. This means that overall the present analysis will investigate 12 vowels: the monophthongs /i/, /ɪ/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/ and /ʌ/ as well as the diphthongs /ʌɪ/ or /ae/, /ʌʊ/ and /ɔe/. For better referentiality, I will use the names of the lexical sets (Wells, 1982) when referring to the vowels (see Table 19).

Table 19. Vowels investigated in the present study with the corresponding lexical sets and the SVLR status according to Aitken (1981).

Vowel(s)	Lexical Set	SVLR status according to Aitken (1981)
/i/	FLEECE	yes
/ɪ/	KIT	invariably short
/e/	FACE	yes
/ɛ/	DRESS	yes
/a/	CAT	yes
/ɔ/	THOUGHT	yes
/o/	GOAT	invariably long in some dialects
/u/	GOOSE/FOOT	yes
/ʌ/	STRUT	invariably short
/ʌɪ/ /ae/	PRICE	yes
/ʌʊ/	MOUTH	yes
/ɔe/	CHOICE	unknown

The vowels and their duration will be analyzed collectively but I will also investigate each vowel independently. The collective analysis of the vowels makes it possible to compare their durations and check for the effects of *intrinsic vowel duration* (see subsection 3.1.1). However, in contrast to many previous studies, I will also analyze each vowel independently to provide a clear and precise overview of the durational patterns in each vowel. This mitigates effects of collinearity because the vowels are not grouped together. As discussed in subsection 3.2.2, the study by McMahon (1991) was criticized by Scobbie, Hewlett, and Turk (1999) because she put the diphthong /aɪ/ and the high vowel /i/ into one class for the analysis. Yet, this pooling is inadequate as the diphthong /aɪ/ is generally longer than the monophthong and this in turn can distort the results. It could, for instance, be the case that the SVLR operates in only one of the vowels, but the grouping erroneously leads to findings which suggest that both vowels are affected by the SVLR. Thus, all vowels in this study will be analyzed separately (see sections 5.2, 5.3 and 5.4).

4.4.2 Variable selection

The response variable of the analysis is *phone duration* which is measured in milliseconds following voicing criteria (see section 4.3). I will investigate the raw durations of the vowels, but I will also logarithmically transform the measurements. A log-transformation is important for the inferential statistical analysis and has also been carried out by many of the latest studies (Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Tanner et al., 2020). However, whereas the latest studies have exclusively analyzed either log-transformed (Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Tanner et al., 2020) or raw durations (Chevalier, 2019), the present analysis will include both in its analysis.

The analysis further includes multiple language-internal and language-external predictor variables. An overview of the predictors can be found in Table 20.

Table 20. Predictor variables analyzed in the present study.

Predictor variable (abbreviation)	Class	Levels/Range
age (age)	Linear	18-87
age group (age_group)	Categorical	young, middle, old
gender (gender)	Categorical	female, male
regional background (reg)	Categorical	East_central (EC), West_central (WC), Northeast (NE), Highland and Hebridean English (HHE), Southern (SO), Insular (IN)
style (style)	Categorical	scripted, unscripted
syllable phone count (num_pho_syl)	Linear	1-7
word syllable count (num_syl_word)	Linear	1-7
word frequency (lex_freq)	Linear	1.173 – 7.422
categorical word frequency (freq)	Categorical	high, low
word type (word_type)	Categorical	content_word, function_word
stress (stress)	Categorical	nuclear, primary, unstressed
phrasal position (position)	Categorical	initial, medial, final
syllables per second (syl_per_sec)	Linear	0.098 – 25.000
mean articulation rate (mean_rate)	Linear	3.771 – 6.915
F1 (F1)	Linear	69.19 – 3052.74
F2 (F2)	Linear	384.1 – 4632.2
phonological SVLR (SVLR1)	Categorical	long, short
morphological SVLR (SVLR2)	Categorical	long, short
general VE (VE1)	Categorical	long, short
VE in plosive contexts (VE2)	Categorical	long, short

To account for sociolinguistic variability in SVLR and VE patterns, I coded for the variables *age*, *age group*, *gender* and *regional background* of the speakers. Age is a linear variable describing the age of the speakers at the time of recording. As described in subsection 4.1.1, *age group* is a categorical variable and encompasses three classes: the age group *young* consists of speakers between 18 and 30 years of age, the age group *middle* includes speakers between 30 and 60 years and the age group *old* incorporates speakers who are older than 60 years. While the variable *gender* can have more categories, I included this predictor as a binary variable because all speakers in the sample stated that they were either male or female. The variable *regional background* includes six regions as specified in subsection 4.1.1: *Insular*, *Northeast*, *East central*, *West central*, *Southern* and *HHE.*

Apart from that, I also coded for several intralinguistic predictor variables. The variable *style* was included to check whether there are style-related differences in vowel durational patterns. The variable *style* distinguishes between *scripted* and *unscripted* speech (see subsection 4.1.2).

The variable *syllable phone count* was included to test the effects of *intrasyllabic compression* (see subsection 3.1.3) by specifying the number of phones in a syllable. The variable *word syllable count* was incorporated to account for effects of *polysyllabic shortening* (see subsection 3.1.4) as it specifies the number of syllables within words. For instance, the vowel /iː/ in the word <frequently> would have a *syllable phone count* of 3 /friː/ and a *word syllable count* of 3 /ˈfriː.kwənt.li/.

In accordance with the possible influence of *lexical category and frequency* on vowel duration (see subsection 3.1.5), the variable *word type* categorizes words into content words and function words and the variable *word frequency* represents the Zipf scale frequency values of the SUBTLEX-UK database (van Heuven et al., 2014). In addition, the categorical variable *categorical word frequency* distinguishes between words of low (Zipf scale value below 3.5) and high frequency (Zipf scale value above 3.5).

The variable *stress* distinguishes between lexically stressed and unstressed syllables on the word level (see subsection 3.1.6) but it also incorporates the category of nuclear *prosodic stress* (see subsection 3.1.7). Thus, the variable *stress* includes both *lexical* and *prosodic stress.*

The predictor variable *phrasal position* indicates whether a phone is positioned in the initial, medial or final syllable position of an utterance (see subsection 3.1.8). Utterances are defined as groups of words which are separated by pauses (Tanner et al., 2020, p. 5) and following the approach of Tsao and Weismer (1997), pauses need to be at least 150 ms long. For example, if the phrase <I like you> is separated by pauses of at least 150 milliseconds, <I> would be classified as initial, <like> as medial and <you> as a vowel in a phrase-final syllable.

As for *tempo* (see subsection 3.1.9), the variable *syllable per second* represents the local articulation rate in an utterance. Similar to Tanner et al. (2020), I also calculated a *mean articulation rate* for each speaker. Slower and faster speakers can therefore be identified by the *mean articulation rate* and the variable *syllable per second* provides more detailed information about the local speed of speech in a particular utterance.

I also included the vowel quality measurements *F1* and *F2*, but these variables are only used for the modeling of all vowels. More information about vowel quality measurements can be found at the beginning of section 5.1.

While these intralinguistic and extralinguistic factors are important for analyzing vowel durational patterns, the SVLR and VE context classifications are even more crucial predictor variables for the present study. The SVLR and VE categorizations are particularly important as they can strongly influence the results. Moreover, the categorizations need to be very precise due to the many overlaps between SVLR and VE short and long contexts. Previous studies have shown that postvocalic fricative contexts trigger generally longer vowel durations than postvocalic stop environments (House & Fairbanks, 1953) (see subsection 3.1.2). It could be the case, for instance, that greater vowel duration in SVLR long contexts simply results from the manner of articulation (postvocalic voiced fricatives), but this lengthening might not be related to Aitken's Law. To investigate the effects of the postvocalic environments on vowel duration closely, I will implement four different categorization schemes, two for Aitken's Law and two for the VE. As the SVLR and VE environments partly overlap, each model in the analysis will only be fitted with one of the categorization schemes. The four categorization schemes were added as additional columns to the overall dataset in R (R Core Team, 2021) with the help of the *following phone* column. A principle for all categorization schemes is that they exclude function words and proper nouns. Function words are frequently reduced in speech which has an influence on vowel quality and quantity (see subsection 3.1.5). It could be the case, for instance, that the phonemic transcription of the word <the> is /ði/, but the actual phonetic realization by the speaker is unstressed with a schwa /ðə/. Likewise, the idealized phonemic transcription of many proper nouns can deviate from the actual pronunciation by a speaker. The Scottish town Hawick, for instance, is pronounced /hɔɪk/ and not /hawɪk/. To avoid possible distortions, all categorization schemes focus exclusively on content words and do not include proper nouns.

The first SVLR categorization (henceforth: *SVLR1*) focuses on the phonological conditioning of Aitken's Law: long *SVLR1* contexts are stressed open syllables and stressed syllables with postvocalic voiced fricatives, affricates or postvocalic /r/. Hence, the vowel /iː/ in the words <bee>, <freeze>, <legion> and <beer> are positioned in *SVLR1* long contexts. As the *SVLR1* scheme focuses on the phonological conditioning of Aitken's Law, tokens with postvocalic morpheme boundaries (e.g <freed>) will be excluded from the respective datasets for model building. Vowel tokens with following glottal stops will be excluded from the datasets as well because they cannot be categorized as SVLR environments. The *SVLR1* short environments include vowels which are followed by nasals, voiceless fricatives, voiceless affricates as well as voiced and voiceless stops as in the words <bean>, <peace>, <bleach>, <feed> and <feet>. An overview can be seen in Table 21.

The second categorization scheme for the SVLR (henceforth: *SVLR2*) focuses exclusively on the morphological conditioning of Aitken's Law. This means that long contexts are vowels in lexically stressed syllables which are followed by morpheme boundaries such as in the words <agreed> and <bees>. Postvocalic voiced fricatives, nasals, affricates and /r/ in monomorphemic words are excluded from the *SVLR2* long contexts to have a clear separation between the phonological and morphological conditioning of the SVLR. In contrast to the bimorphemic word <agreed> which represents a long *SVLR2* context, the monomorphemic word <greed> is a short *SVLR2* environment. The *SVLR2* categorization therefore makes it possible to analyze whether there are quasi-phonemic contrasts (Scobbie & Stuart-Smith, 2008) between those environments in naturally spoken SSE.

In contrast to the *SVLR1* and *SVLR2* categorizations, the *VE1* long contexts include vowels which are followed by voiced fricatives, voiced plosives, /r/ and nasals and the *VE1* short contexts include postvocalic voiceless plosives and fricatives. Similar to the *SVLR1* categorization, *VE1* also excludes tokens which are succeeded by postvocalic glottal stops. Yet, the *VE1* classification also excludes vowels in stressed open syllables and vowels which are followed by morpheme boundaries. These environments have not been categorized as VE contexts in previous studies and it has already been observed that vowels are generally longer in open than in closed syllables. The *SVLR1* and *VE1* contexts partly overlap as both categorizations include voiced fricatives for the long contexts and voiceless fricatives and voiceless plosives for the short contexts (see Table 21). The main distinction between the phonological SVLR and VE conditioning therefore lies in the contrast between postvocalic voiced stop and postvocalic voiceless stop contexts (Rathcke & Stuart-Smith, 2016, p. 406). Whereas voiced stops constitute long environments for the *VE1* classification, they would be short for the *SVLR1* categorization scheme.

To have a sharper separation between the phonological SVLR and VE conditioning effects, I also included another categorization scheme (henceforth: *VE2*) which only incorporates vowels which are followed by voiced and voiceless stops. I excluded the fricative contexts in this categorization scheme because postvocalic fricatives are known to generally trigger longer vowel durations than postvocalic stop environments (House & Fairbanks, 1953). It could be the case, for instance, that greater vowel duration in *SVLR1* long contexts simply results from the manner of articulation (postvocalic voiced fricatives). To exclude these potentially confounding effects, the *VE2* classification focuses exclusively on stop environments. If the vowel durations are significantly longer before voiced stops than before voiceless stops, this indicates the operation of the VE. Yet, if there are no significant differences between the vowel durations in these contexts, this should indicate the operation of Aitken’s Law. For a similar reason, the *VE2* classification also excludes vowels in open syllables to avoid possible confounding effects of *constituent-final lengthening* (see subsection 3.1.8). An overview of the categorization schemes can be found in Table 21.

Table 21. Overview of the SVLR1, SVLR2, VE1 and VE2 categorization scheme with the three classes "long", "short" and "excluded", the corresponding postvocalic contexts and example words.

	SVLR1			**VE1**		
	long	**short**	**excluded**	**long**	**short**	**excluded**
Postvocalic context; /phoneme/; <Example word>	postvocalic /z/ <freeze> postvocalic /ʒ/ <prestige> postvocalic /v/ <achieve> postvocalic /ð/ <breathe> postvocalic /dʒ/ <liege> postvocalic /r/ <beer> stressed open syllable <bee>	postvocalic /m/ <seem> postvocalic /n/ <seen> postvocalic /ŋ/ postvocalic /s/ <peace> postvocalic /ʃ/ <leash> postvocalic /f/ <beef> postvocalic /θ/ <heath> postvocalic /p/ <leap> postvocalic /t/ <feet> postvocalic /k/ <seek> postvocalic /b/ <feeble> postvocalic /d/ <feed> postvocalic /g/ <league> postvocalic /tʃ/ <leech>	postvocalic /ʔ/ <righ't> unstressed syllables <agreed> function words <the> proper nouns <Keith> morpheme boundary <agreed>	postvocalic /m/ <seem> postvocalic /n/ <seen> postvocalic /ŋ/ postvocalic /b/ <feeble> postvocalic /d/ <feed> postvocalic /g/ <league> postvocalic /z/ <freeze> postvocalic /ʒ/ <prestige> postvocalic /v/ <achieve> postvocalic /ð/ <breathe> postvocalic /dʒ/ <liege> postvocalic /r/ <beer>	postvocalic /p/ <leap> postvocalic /t/ <feet> postvocalic /k/ <seek> postvocalic /s/ <peace> postvocalic /ʃ/ <leash> postvocalic /f/ <beef> postvocalic /θ/ <heath> postvocalic /tʃ/ <leech>	postvocalic glottal stop /ʔ/ unstressed syllables <agreed> function words <the> proper nouns <Keith> morpheme boundary <agreed> stressed open syllable <bee>
	SVLR2			**VE2**		
	long	**short**	**excluded**	**long**	**short**	**excluded**
	morpheme boundary /#d/ <agreed> morpheme boundary /#z/ <bees>	postvocalic /d/ <greed> postvocalic /s/ <peace>	prevocalic /j/ <years> prevocalic /w/ <wheeze> postvocalic /r/ <beer> postvocalic glottal stop /ʔ/ unstressed syllables <agreed> function words <the> proper nouns <Keith> postvocalic voiced fricatives postvocalic nasals postvocalic affricates stressed open syllable <bee>	postvocalic /b/ <feeble> postvocalic /d/ <feed> postvocalic /g/ <league> postvocalic /dʒ/ <liege>	postvocalic /p/ <leap> postvocalic /t/ <feet> postvocalic /k/ <seek> postvocalic /tʃ/ <leech>	prevocalic /j/ <years> prevocalic /w/ <wheeze> postvocalic /r/ <beer> postvocalic glottal stop /ʔ/ unstressed syllables <agreed> function words <the> proper nouns <Keith> morpheme boundary <agreed> stressed open syllable <bee> postvocalic voiced fricatives postvocalic nasals

4.4.3 Statistical analysis

The statistical analysis includes means of descriptive and inferential statistics and was carried out in R (R Core Team, 2021). I investigated the influence of the different intralinguistic and extralinguistic predictor variables (see subsection 4.4.2) on the duration of the vowels.

In the descriptive part of the analysis, I primarily created boxplots and jitterplots with the R package *ggplot2* (Wickham, 2016). This means that vowel duration was plotted on the y-axis and the individual vowel tokens were grouped for the respective variable categories on the x-axis. Boxplots were used because they directly show where the majority of data points of a continuous variable are located: 50% of the data is located between the upper and lower hinges of the box and the line in the middle of the box shows the median. The whiskers above and below the box are not longer than 1.5 times the interquartile range and all data points beyond the box and whiskers can be considered outliers (Levshina, 2015, p. 58). Yet, while boxplots provide a graphical summary of the data, they do not show the distribution of the exact values. Therefore, I added semi-transparent jitterplots to the corresponding boxplots. These jitterplots do not just show every individual data point but they also distribute the data points so that tokens with similar durations are not laid on top of each other. The combination of boxplots and jitterplots provides a detailed visualization of the data and they made it possible to see whether vowel duration differs for the respective categories of the variables. The most important variables are the postvocalic categorizations of *SVLR1*, *SVLR2*, *VE1* and *VE2*. Whenever these variables were significant, I created separate plots that subdivided Aitken’s Law and the VE for specific intralinguistic and extralinguistic variables. This provided a first overview to see not only the influence of SVLR and VE contexts but also whether these influences differ among different speaker groups or in specific contexts. Apart from the SVLR and VE environments, I also investigated whether the rate of *syllable per second* affects vowel duration in terms of *tempo* (see subsection 3.1.9). The variable *phrasal position* was also plotted against vowel duration to check whether there is an influence of *constituent-final lengthening* (subsection 3.1.8) and the variable *stress* was plotted to see whether vowel duration differs in terms of *lexical* and *prosodic stress* (see subsections 3.1.6 and 3.1.7). I used the *number of phones in a syllable* to check whether the durations are affected by *intrasyllabic compression* (subsection 3.1.3) and I used the *number of syllables in a word* to account for effects of *polysyllabic shortening* (subsection 3.1.4). In addition, I also plotted the data against the Zipf scale values of the SUBTLEX-UK database (van Heuven et al., 2014) to check whether vowel duration differs with regard to *lexical frequency* (subsection 3.1.5).

Apart from the plots, I further calculated the average duration and standard deviation of each vowel as well as the average vowel durations and standard deviations for the different categories of the extralinguistic and intralinguistic variables. For example, I did not only plot vowel duration against the variable *SVLR1*, but I also calculated the average duration and standard deviation of the vowels in *SVLR1* long and *SVLR1* short contexts.

While the descriptive analysis provides a good overview of the overall durational distributions, the inferential statistical analyses convey whether these differences are statistically significant. In the present study, I will fit linear mixed effects models on the dependent variable *log-transformed vowel duration.* I will fit models for the whole dataset incorporating all SSE vowels (see section 5.1), but I will also subdivide the dataset and fit models for each vowel independently to avoid effects of collinearity (see sections 5.2, 5.3 and 5.4). In the model building process, the whole dataset as well as the datasets for the individual vowels are further subdivided for the *SVLR1*, *SVLR2*, *VE1* and *VE2* categorization schemes. This means that there are four sub-datasets which were filtered for the long and short contexts of each category. The subdivision of the dataset not only reduces effects of collinearity, but it also ensured that tokens from the level "exclude" are removed. For example, the sub-datasets which will be used for modeling vowel duration with *VE2* only include tokens with postvocalic plosives (see Table 21).

Similar to the most recent studies on SVLR patterns in spontaneous speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016), the linear mixed effects models will be fitted with the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. I will apply a stepwise regression procedure with a backward selection for model building. I start with a full model including all variables and all possible interactions and I use the generic step function of the lmerTest package (Kuznetsova et al., 2017) to exclude the factors and interactions which return no significant effects. The R^2 values for the models will be generated via the *r.squareGLMM()* function from the MuMin package (Bartón, 2022) and the best model fit will be found via maximum likelihood ratio tests. For a better interpretation of the model estimates, I am going to apply the procedure by Tanner et al. (2020) and take the exponent of the model parameter's value. For, example, an estimate of $e^{0.19} = 1.2$ represents a vowel duration increase of 20% when compared to the intercept (Tanner et al., 2020, p. 8).

An advantage of linear mixed effects models is that they can handle unbalanced datasets typical of spontaneous speech (Rathcke & Stuart-Smith, 2016, p. 414). This is very important in the context of the present study: the data derives from different contexts and includes scripted and unscripted speech from many different speakers. The calculation of ratio durations (i.e. Watt & Ingham, 2000) or length indexes (i.e. Agutter, 1988b) is possible when the data was retrieved in word list readings, but these normalization procedures are useless in spontaneous speech as there are no stable points of reference (Rathcke & Stuart-Smith, 2016, p. 414). Linear mixed effects models, however, are multivariate which means that they can test the effect of several predictor variables simultaneously while controlling for the effect of other predictor variables (Schweinberger, 2022). The models can control for the influence of postvocalic SVLR lengthening *while* accounting for the effect of other factors, for instance, *tempo*, *phrasal position* and *stress.*

Another great advantage of linear mixed effects models is that they can handle both fixed and random effects. Fixed effects are, for instance, categorical or numerical predictor variables such as *stress*, *regional background*, *SVLR1*, or the *number of segments in a syllable*. Random effects are nongeneralizable, non-independent categorical cluster variables such as *speaker*, *word*, or *syllable*. For instance, if several vowels are produced by different speakers, the corresponding vowel durations are not generalizable for the whole dataset, but they depend on the speakers and can therefore be grouped for the speakers. In contrast, vowel duration can generally vary in different *phrasal positions*; the influence of this variable is generalizable across different speakers, words, or syllables. Linear mixed effects models can therefore control for the influence of both fixed factors (e.g. *SVLR1*, *VE1*) and random factors (e.g. *speaker, word*).

In contrast to simple or multiple linear regression, linear mixed effects models allow for varying intercepts and varying slopes. For example, it could be the case that for the vowel FLEECE, the duration is generally longer in *SVLR1* long contexts than in *SVLR1* short contexts, but this effect might not apply to all speakers. There could be, for instance, a few speakers whose vowels are longer in *SVLR1* short contexts than in *SVLR1* long contexts. Linear mixed effects models can fit varying intercepts and slopes for the individual speakers and the model can therefore control for speaker-based deviations from the overall lengthening pattern. In addition, it is also possible to implement interactions in linear mixed effects models. This means that the influence of a predictor variable on the response variable (e.g. vowel duration) depends on another predictor variable (Winter, 2020, p. 133). For instance, it could be the case that only female speakers from a particular *regional background* have typical SVLR patterns; the male speakers from that region, however, do not. When the variable *regional background* is implemented as a fixed factor without any interactions, it could be the case that the durational patterns of the female and male speakers balance each other out. As a result, the model could return that the influence of that *regional background* is not significant. An interaction between *regional background* and *gender* in the model could, however, show that there is a significant influence of the female speakers from that region.

A small disadvantage of linear mixed effects modeling is that their predictive power is limited to the observed data. Linear mixed effects modeling takes the whole dataset as input and tries to fit a model which can explain as much variation in the dataset as possible. The main focus lies on the model fit; how well the model fits to the observed data. This has the effect that the model predictions often only make sense in the range of the observed data, extrapolating beyond the dataset can result in unreliable estimates (Winter, 2020, p. 72). In other words, a model can fit relatively well to a dataset and it can produce reliable estimates for this dataset, but the predictive power of the trained model can be very limited if one applies it to new data. In section 5.1, I will therefore stochastically subsample the dataset via the PrinDT function (Weihs & Weilinghoff, 2023) and evaluate the predictive power of the models with training and test datasets. More information can be found at the end of section 5.1. An overview of the fixed and random effects for the model building of the present study can be found in Table 22.

Table 22. Overview of the random and fixed factors used for linear mixed effects modeling.

Random factors	**Type**	**Levels**
speaker	categorical	*130 individual speakers*
syllable label	categorical	*8739 individual syllable labels*
word label	categorical	*11262 individual word labels*
Fixed factors (abbreviations)	**Type**	**Levels**
SVLR1 (SVLR1)	categorical	*long, short*
SVLR2 (SVLR2)	categorical	*long, short*
VE1 (VE1)	categorical	*long, short*
VE2 (VE2)	categorical	*long, short*
syllable phone count (num_pho_syl)	numerical	*1-7*
word syllable count (num_syl_word)	numerical	*1-7*
first formant (F1)	numerical	*69.19 – 3052.74*
second formant (F2)	numerical	*384.1 – 4632.2*
categorical word frequency (freq)	categorical	*high, low*
stress (stress)	categorical	*nuclear stressed, primary stressed, unstressed*
phrasal position (position)	categorical	*final, medial, initial*
syllable per second (syl_per_sec)	numerical	*0.947 – 25.0*
speaking style (style)	categorical	*scripted, unscripted*
gender (gender)	categorical	*female, male*
regional_background (reg)	categorical	*Insular, Northeast, East central,, West central, Southern and HHE*
age group (age_group)	categorical	*young, middle, old*

I also included the *syllable label* as another random factor because the syllable label offers a more precise representation of a vowel's phonetic context. The short vowel /ɪ/ is, for instance, embedded in the last syllables of the words <shaving> and <waving> and the structure of the syllables is identical: /vɪŋ/. The duration of /ɪ/ could therefore depend not just on the word but also on the syllable in which the vowel is embedded.

The independent fixed variables include the categorization schemes *SVLR1*, *SVLR2*, *VE1* and *VE2*. Apart from that, I also included the *syllable phone count*, the *word syllable count*, the categorical *word frequency*, *stress*, the *phrasal position*, the local articulation rate in terms of *syllables per second*, the *style* of speech and the extralinguistic variables *gender*, *regional background* and *age group* as fixed variables in the model building process. For the model on all vowels (see section 5.1), I also included the vowel quality measurements of *F1* and *F2*.

4.5 Summary

The present study uses an up-to-date dataset that is balanced in terms of *gender*, *age*, *regional background* and speech *style* (see Table 23). While ICE Scotland (Schützler et al., 2017) served as the base dataset for this study, I also collected further spoken SSE speech in different parts of Scotland and I retrieved spoken data from the internet. In total, the dataset comprises 150995 words and the total duration of the sound files adds up to 22 hours and 16 minutes. For each file, I created time-aligned phonemic transcriptions on multiple prosodic levels with the help of different software tools and algorithms. The final transcription format includes a phoneme, syllable, word and utterance level which made it possible to account for the different variables that affect vowel duration (see section 3.1). The segmentation approach follows voicing criteria which means that the beginning of a vowel is determined by a stable F1 and the end of a vowel is marked by the offset of a stable F2. In contrast to all previous studies, this investigation will analyze all vowels of the Basic Scottish Vowel System (Abercrombie, 1979), both collectively and independently. The analysis incorporates multiple extralinguistic and intralinguistic variables and implements four different categorization schemes to account for SVLR and VE effects as precisely as possible. The statistical analysis uses means of descriptive and inferential statistics for investigating durational patterns in contemporary SSE vowels. More specifically, I will use linear mixed effects modeling and include the *speaker*, *word* and *syllable* label as random factors.

Table 23. Overview of the number of words separated for the variables regional background, gender, age group and style.

Variable	**Level**	**Number of words**
Regional background	East central	27199
	HHE	23604
	Insular	26147
	Northeast	18007
	South	24412
	West central	31626
Gender	female	67726
	male	83269
Age group	old (60+)	33165
	middle (31-60)	94683
	young (18-30)	23147
Style	scripted	85044
	unscripted	65951
Total		**150995**

5. Results

The results of the analyses are presented in this chapter. For a better overview, the findings are divided into different sections. In section 5.1, I will first provide an overview of the durations of all vowels. This means that I will investigate which predictor variables (see subsection 4.4.2) influence vowel duration in general using means of descriptive and inferential statistics. I will also provide a brief overview of vowel quality at the beginning of section 5.1.

In the subsequent sections, I will repeat the analysis for each vowel individually. I will first briefly introduce the vowel and then plot the duration of the vowel against the different predictor variables, I will calculate average values, standard deviations and I will fit linear-mixed effects models for the *SVLR1*, *SVLR2*, *VE1* and *VE2* categorization schemes. Due to space issues, I will only plot those variables which show interesting distributions and I will only provide detailed information on the best model for each vowel. Section 5.2 incorporates the findings for the short monophthongs KIT (subsection 5.2.1), STRUT (subsection 5.2.2) and DRESS (subsection 5.2.3). Section 5.3 then continues with the findings for the long monophthongs of SSE. Subsection 5.3.1 comprises the results for the lexical set GOOSE and subsection 5.3.2 the findings for FLEECE. The following subsections deal with the lexical sets THOUGHT (subsection 5.3.3), FACE (subsection 5.3.4), GOAT (5.3.5) and CAT (subsection 5.3.6). After that, I will continue with the findings for the diphthongs (section 5.4). This includes the lexical sets MOUTH (subsection 5.4.1), PRICE (subsection 5.4.2) and CHOICE (subsection 5.4.3). A summary of the findings will be provided at the end of this chapter (section 5.5).

5.1 Vowel overview

This section provides a general overview of the duration of all vowels. Yet, as vowel duration also interacts with vowel quality in English, I will first give an exemplary overview of the quality of the vowels of 21st century spoken SSE. For this purpose, I used the Python library *parselmouth* (Jadoul et al., 2018) which works as an interface to Praat (Boersma & Weenink, 2019) and I automatically extracted F1, F2 and F3 measurements at the midpoints of the vowels. I chose the midpoints for the measurements to minimize the effects of the surrounding phonetic context. I then normalized the data to factor out the physical differences in vowel production which result from the anatomical differences between the speakers (Watt et al., 2011, p. 111). I used the bark transformation (Traunmüller, 1990) to see how the vowel quality differences are perceived by listeners (Watt et al., 2011, p. 112). Following the procedure of vowel formant analyses (Di Paolo et al., 2011, pp. 88–89; Fruehwald, 2013, p. 117; Tanner et al., 2020, p. 5), I excluded vowels in function words, unstressed syllables and all measurements below 50 ms to avoid reduced realizations. I also chose only monosyllabic and bisyllabic content words to minimize the effect of *polysyllabic shortening* (see subsection 3.1.4). Moreover, I selected only vowels which are followed by voiceless plosives to reduce the influence of coarticulation (Thomas, 2011, p. 49): vowel formant tracks are heavily influenced by the surrounding phonetic context so the formant measurements for the vowel /iː/ in the words <bean> and <beat> can differ drastically. A following nasal can change the formant tracks to a great extent, so I focused on voiceless plosives to have a more neutral environment (Di Paolo et al., 2011, p. 88). Despite all these precautions, the resulting plot shows great variation in terms of vowel quality (see Figure 21). The normalized vowel measurements are spread over a very large area. This variability is, however, not surprising because the data represents naturally spoken language and not carefully controlled speech that was elicited under laboratory conditions. Due to the many factors that can influence the production of vowels (see section 3.1), vowel quality can differ widely. Nevertheless, despite all the variation, the mean formant values of the vowels largely correspond to the general structure of a vowel chart. The open vowels CAT and STRUT are relatively low and back with high F1 values and low F2 values on average. The vowel THOUGHT is in a higher position and the vowel DRESS is clearly more fronted. GOOSE is also very fronted and comes close to the quality of KIT which corresponds to the observations that this vowel is fronted in SSE: /ʉ/ (Wells, 1982, p. 402). The vowel GOAT is slightly backer but clearly higher than THOUGHT. The monophthong FACE is relatively high and front and comes close to the quality of FLEECE, which is in the highest and most fronted position. Overall, the average values of the vowels are following a diagonal line from a low back to a high front position, but there is great variability overall.

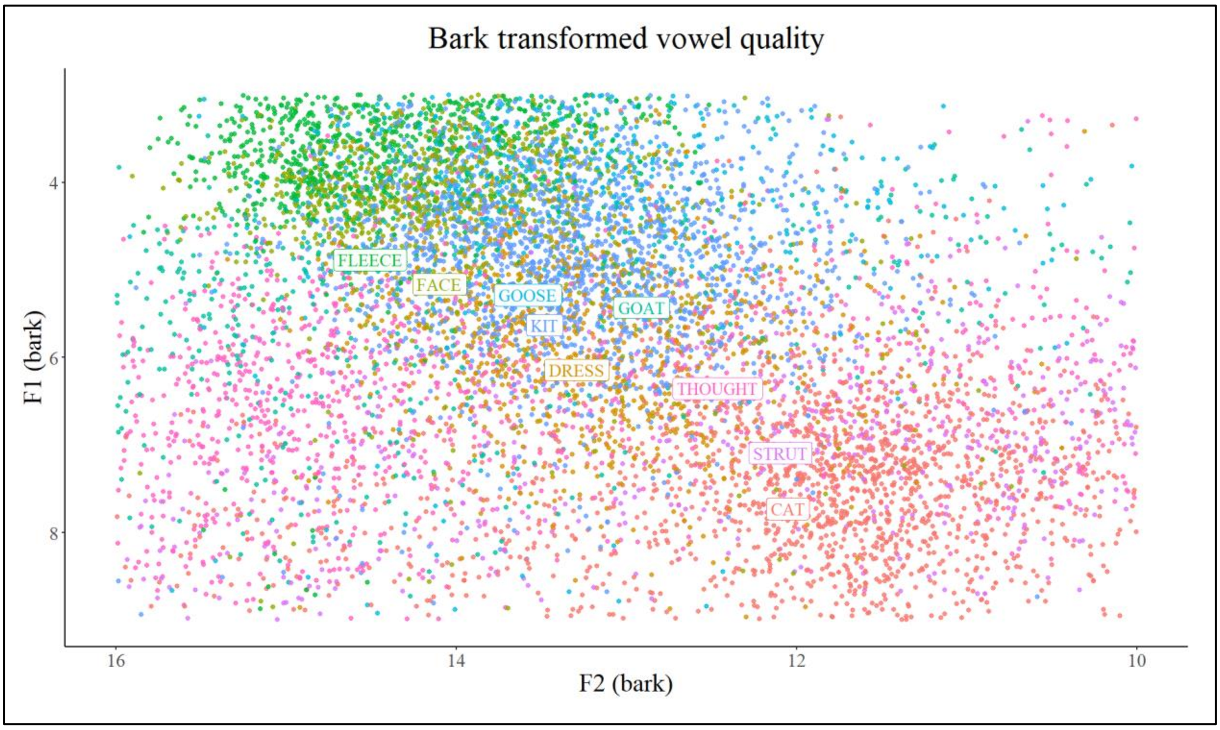


Figure 21. Bark-transformed vowel quality plots for the SSE monophthongs FLEECE, FACE, GOOSE, GOAT, KIT, DRESS, THOUGHT, STRUT and CAT with the mean values represented by the text boxes. The plots only include tokens with following voiceless plosive contexts. All tokens in function words, unstressed syllables, or tokens with a duration below 50 ms were excluded.

The mean durations and standard deviations of the respective vowels are listed in Table 24. The raw durations are also visualized in the boxplots in Figure 22. Overall, there is great durational variability across all vowels. This variability cannot only be seen in the standard deviations (GOAT has the highest standard deviation with 70.32 ms) but also in Figure 22: the boundaries of the whiskers at the bottom of the boxplots indicate that some vowels have very short realizations. At the same time, each boxplot has a multitude of outliers at the top which indicates that there are also many long realizations in the dataset.

Table 24. Mean durations and standard deviations of all vowels sorted for the mean duration in ascending order.

Vowel(s)	**Lexical set**	**Type**	**Token number**	**Mean duration (ms)**	**Standard deviation (ms)**
/ɪ/	KIT	Short monophthong	44382	64.71	41.14
/ʌ/	STRUT	Short monophthong	6467	71.72	35.70
/ɛ/	DRESS	Short monophthong	13714	91.76	47.27
/u/	GOOSE	Long monophthong	6351	92.33	55.55
/i/	FLEECE	Long monophthong	9146	100.42	53.91
/ɔ/	THOUGHT	Long monophthong	7762	112.64	55.79
/ʌʊ/	MOUTH	Diphthong	9394	114.90	56.95
/e/	FACE	Long monophthong	6708	113.00	57.34
/o/	GOAT	Long monophthong	10868	119.10	70.32
/a/	CAT	Long monophthong	2992	119.71	51.23
/ʌɪ/ /ae/	PRICE	Diphthong	9540	125.02	60.52
/ɔe/	CHOICE	Diphthong	480	140.74	59.09

Despite the great variability, there is a general tendency that the diphthongs are longer than the long monophthongs and the long monophthongs tend to be longer than the short monophthongs. The lexical set KIT has the shortest mean duration of 64.71 ms followed by STRUT (71.72 ms) and DRESS (91.76 ms). The long monophthongs have longer average durations than the short monophthongs. CAT is the monophthong with the longest mean duration (119.71 ms). Only the diphthongs PRICE (125.02 ms) and CHOICE (140.74 ms) are longer than that. The diphthong MOUTH (114.90 ms) is, however, shorter than FACE, GOAT and CAT.

Apart from the overall increase in vowel duration based on the phonemic classification of the vowels, there is also a tendency that low vowels are longer than high vowels. The short low monophthong STRUT is longer than KIT and the high vowels GOOSE and FLEECE are clearly shorter

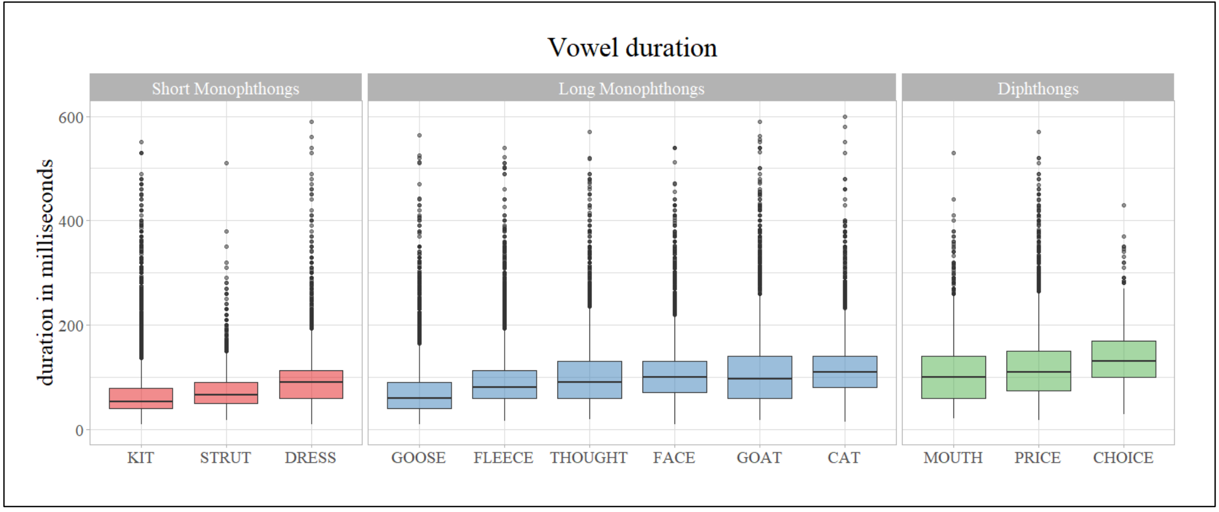


Figure 22. Boxplots of the raw durations of all vowels sorted for the short monophthongs, long monophthongs and diphthongs.

than THOUGHT, GOAT, FACE and the low vowel CAT. This tendency is consistent with the effects of *intrinsic vowel duration* (see subsection 3.1.1): open vowels are longer than close vowels.

The classifications representative for Aitken's Law and the VE are visualized in Figure 23. As for the categorization of *SVLR1*, there is a small overall difference between vowels in SVLR long and short contexts. The boxplot of the *SVLR1* long contexts is higher than the boxplot of the *SVLR1* short contexts and the average vowel duration is over 15 ms longer in the long environments (mean duration: 98.48 ms) than in the short environments (83.23 ms). This difference suggests that the phonological conditioning of Aitken's Law applies: vowels tend to be longer in stressed open syllables and in stressed syllables with postvocalic voiced fricatives, affricates, or postvocalic /r/ (see subsection 4.4.2). A similar tendency can also be seen for the *SVLR2* contexts. The *SVLR2* long contexts (mean duration 104.78 ms) are generally longer than the *SVLR2* short contexts (mean duration 83.68 ms). This suggests that also the morphological conditioning of Aitken's Law influences vowel duration. It is interesting that the average durational difference between *SVLR2* long and short contexts (21.1 ms) is greater than the average difference between *SVLR1* long and short contexts (15.25 ms). As the *SVLR2* long categorization only includes vowels with postvocalic morpheme boundaries (e.g. <agreed>, <bees>), it is clear that the increased vowel duration derives from the morphological structure of the word. Hence, the vowels tend to be longer when they are followed by morpheme boundaries.

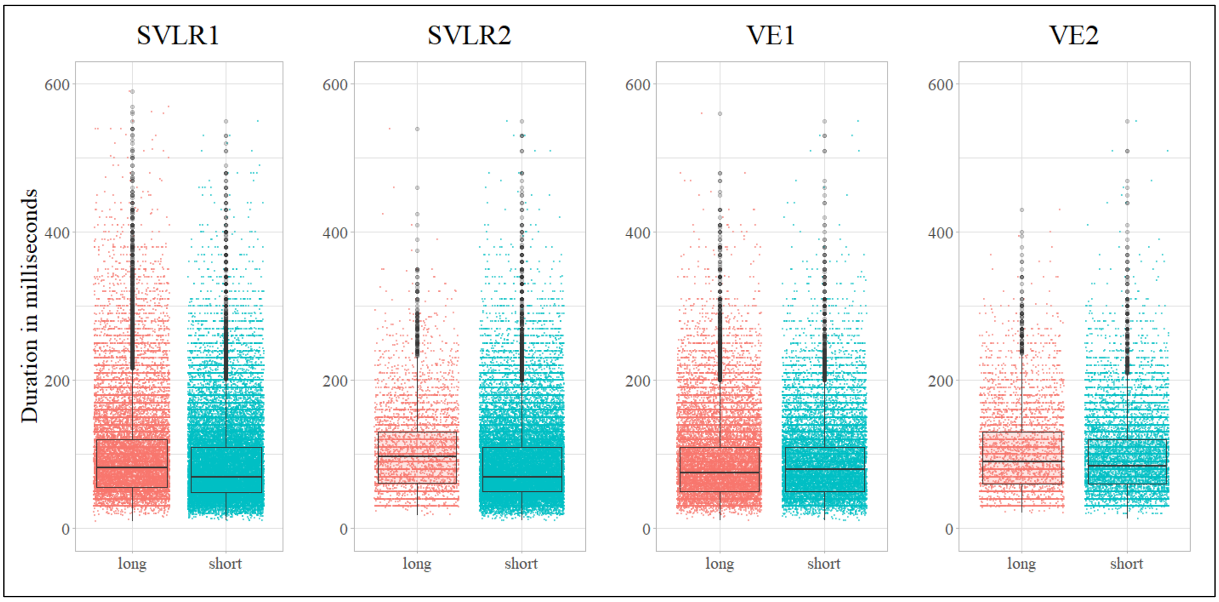


Figure 23. Boxplots and jitterplots of raw vowel duration for the categories SVLR1, SVLR2, VE1 and VE2.

The categorizations of the VE, however, show a slightly different picture. The *VE1* long contexts are not longer than the *VE1* short contexts. The boxplots of the categories are on a similar level and one can see no real difference between the two classes in Figure 23. Interestingly, the average duration of the *VE1* long contexts (86.47 ms) is even slightly shorter than the average duration of the *VE1* short contexts (87.59 ms). This means that there tends to be no overall difference in duration when vowels are followed by either voiced or voiceless consonants. This distribution does therefore contradict the VE (see subsection 3.1.2). The difference between *VE2* long and short contexts is also not very pronounced. The boxplot for the *VE2* long contexts is a bit higher than the boxplot of the *VE2* short contexts and there is a difference in average vowel duration (*VE2* long: 103.62 ms; *VE2* short: 93.81 ms). The average durational difference between *VE2* long and short contexts amounts to only 9.8 ms which is shorter when compared to the equivalent differences between the *SVLR1* and *SVLR2* long and short environments. As the *VE2* categorization only includes postvocalic voiced and voiceless plosive contexts, this further supports the notion that the effects of the VE are relatively small overall.

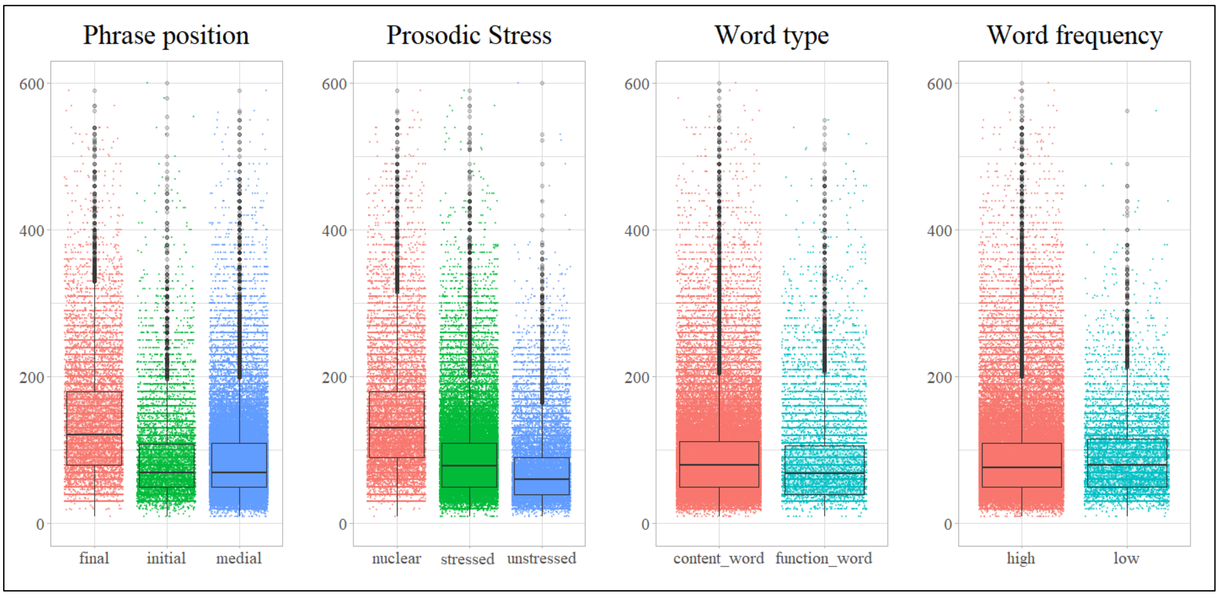


Figure 24. Boxplots and jitterplots of raw vowel duration separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

As for the other intralinguistic predictor variables, Figure 24 shows that there is increased vowel duration in phrase-final position. Whereas vowels in phrase-initial and phrase-medial syllables are comparable in terms of their duration with average durations of 84.46 and 82.80 ms respectively, phrase-final syllables have an increased vowel duration overall (mean: 136.74 ms). This shows an effect of *constituent-final lengthening* (see subsection 3.1.8): vowels are longer in pre-pausal syllables.

It is also visible that *stress* has an effect on vowel duration. Lexically unstressed syllables (mean duration: 70.31 ms) are shorter than lexically stressed syllables (mean duration: 86.12 ms). This means that, for example, the first syllable in the word <needy> is longer than the second syllable because lexical stress falls on the first syllable of the word. Hence, *lexical stress* tends to have an influence on vowel duration overall (see subsection 3.1.6). The effects of *prosodic stress* are, however, even stronger: nuclear stressed syllables have an average duration of 141.86 ms and are therefore clearly longer than lexically stressed and unstressed syllables. It must be taken into consideration that some vowel tokens under nuclear stress are also lexically stressed so that there is an overlap in the categories. Despite this overlap, nuclear stressed syllables are characterized by a drastic increase in vowel duration. Therefore, *prosodic stress* (see subsection 3.1.7) also has an influence on vowel duration.

Whereas the effects of stress and constituent-final lengthening are clearly visible in Figure 24, there seems to be no difference between content words and function words nor between high frequency words and low frequency words in terms of vowel duration. On average, vowels in content words (mean: 89.26 ms) are less than five ms longer than vowels in function words (mean: 84.52 ms). The difference is even smaller for the predictor variable *word frequency*: high frequency words have an average duration of 88.78 ms and low frequency words have an average duration of 89.29 ms. It must be noted, however, that the token numbers are also very different for the classes of *word type* and *word frequency*. The dataset comprises 126952 content words but only 12701 function words. Similarly, there are 128210 words with a high frequency but only 11444 low frequency words. This can also be seen in the jitterplots (Figure 24): there are many more data points for content words and high frequency words than for function words and low frequency words. While it must be taken into consideration that the variable *word frequency* is not necessarily binary (see subsection 3.1.5), the plots nevertheless suggest that there is not a stark difference between high frequency and low frequency words in terms of vowel duration.

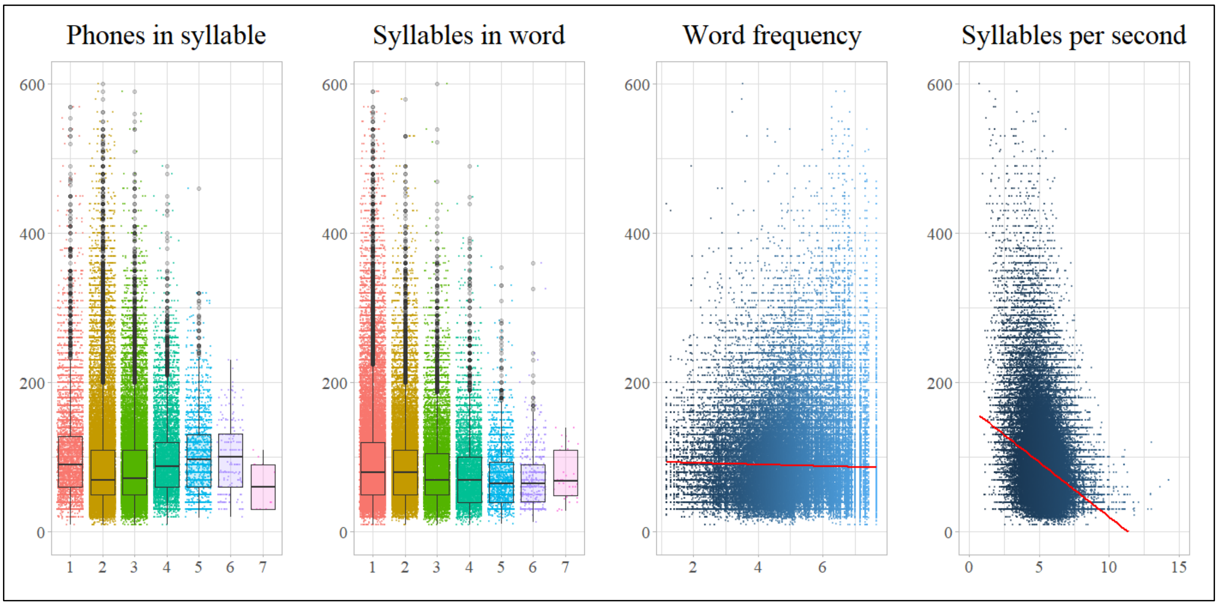


Figure 25. Plots of raw vowel duration separated for the linear variables number of phones in a syllable, number of syllables in a word, word frequency and syllables per second.

Figure 25 provides further information about the influence of several linear intralinguistic predictor variables on vowel duration, namely the *phone syllable count*, the *syllable word count*, *word frequency* and the local articulation rate in terms of *syllables per second*. As for the number of phones in a syllable (*phone syllable count*), it is clear to see that there are fewer longer vowels in syllables with a high number of phones. This means that extremely long vowels become less frequent the higher the number of phones in a syllable. However, the average duration of the vowels and does not continuously decrease with a higher phone count. While there is an overall decrease in vowel duration from syllables with one (mean: 100.26 ms), two (mean duration: 87.45 ms) and three (mean duration: 86.04 ms) phones, this trend is reversed in syllables with four (mean duration: 95.03 ms), five (mean duration: 102.92 ms) and six (mean duration: 101.29 ms) phones. The shortest average duration can be found in syllables with seven phones (mean duration: 63.33 ms). This overall trend can also be seen in the boxplots: instead of a continuous decrease in the height of the boxplots, there are different peaks for the syllable phone counts 1, 5 and 6. While it must be taken into consideration that the token numbers decrease with a higher phone syllable count from 2 phones onward, the plot does not reveal a general effect of *intrasyllabic compression* (see subsection 3.1.3). It is true that very long vowels become less frequent in syllables with high phone counts, but the average duration does not steadily decrease.

A slightly different trend can be seen for the *syllable word count*. The height of the boxplots steadily decreases if the number of syllables increases. The only exception from this trend can be seen for the words with seven syllables where there is a slight increase in duration. Yet, this finding has to be taken with caution as this category is represented by only 33 word tokens (i.e. <telecommunications>). Apart from that, there is also the trend that extremely long vowels tend to become less frequent if the *word syllable count* increases. In other words, the outliers above the boxplots become fewer and shorter in words with higher syllable counts. Furthermore, the token numbers and the average vowel duration also steadily decrease in words with many syllables. Hence, the higher *word syllable count*, the lower the token numbers and the average vowel duration. The overall distributions therefore correspond with the effect of *polysyllabic shortening* (see subsection 3.1.4): the more syllables in a word, the shorter its vowel articulations tend to be.

The plot of the linear variable *word frequency* does not show a clear tendency. Most of the data points have a word frequency between 3 and 6 on the Zipf scale and a phone duration between 30 and 180 ms. I also added a red regression line to the plot to see more clearly whether the predictor variable (*word frequency*) has an influence on the response variable (*phone duration*). The slope of the regression line is slightly negative so words with a high Zipf scale value (= high frequency) tend to be slightly shorter than words with a low Zipf scale value (= low frequency). However, the slope of the regression line comes close to zero so there seems to be no strong correlation between lexical frequency and vowel duration. This corresponds to the plot of the associated categorical variable *word frequency* (see Figure 24) in which no striking difference between high and low frequency words could be observed.

The plot for the local articulation rate indicates that the speed of speed is relatively stable in the range from 3 to 6 syllables per second. More precisely, the average speech rate is 5.3 syllables per second with the first quantile at 4.6 and the third quantile at 5.9 syllables per second. This closely corresponds to previous findings on speech rate in English (Goldman-Eisler, 1968, p. 24). While the articulation rate is clustered in this range, there is an overall tendency that higher articulation rates lead to shorter vowel pronunciations. Whereas there are many long vowel pronunciations in the range of 2 to 6 syllables per second, these long vowel pronunciations become less frequent if the speed of speech is higher. The red regression line also has a clear negative slope which means that a higher speech rate leads to shorter vowel articulations overall (see subsection 3.1.9).

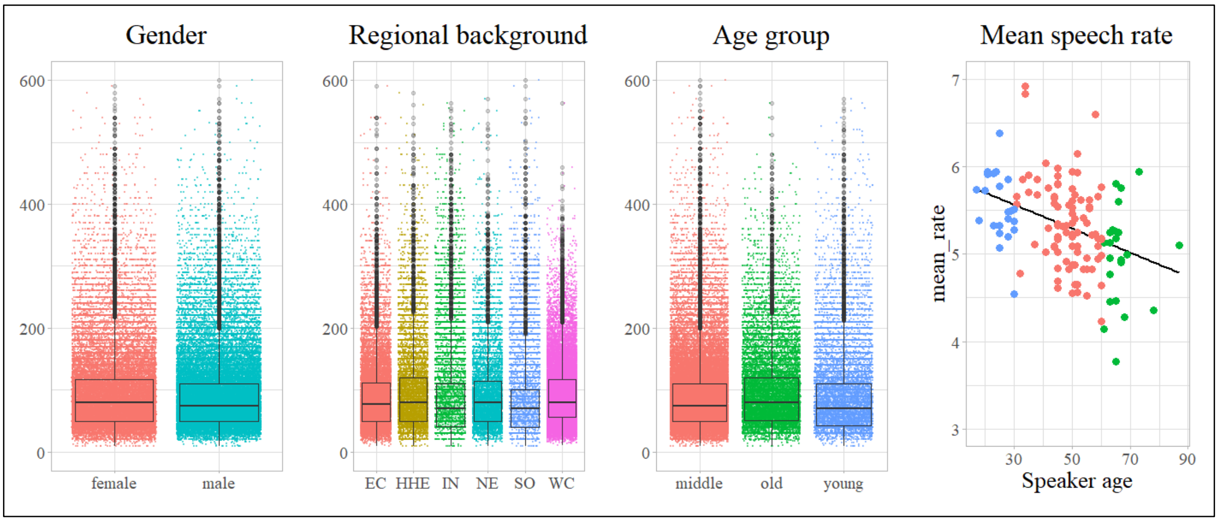


Figure 26. Plots of raw vowel duration separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean speech rate (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean speech rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

As seen in Figure 26, the *gender* of a speaker does not have a striking influence on the duration of vocalic intervals. There is great variability overall, the boxplots are on a very similar level and the average vowel duration for female and male speakers is almost equal with 89.96 ms and 87.81 ms respectively.

The average vowel duration is also highly similar for the different regions of Scotland. The jitterplots show that there is high durational variability across all regions and the boxplots indicate that the regions EC, HEE, IN, NE and WC are all on a highly similar level. Only the boxplot for SO tends to be a bit shorter than the rest. This is also reflected in the average vowel durations: whereas the average vowel durations of all other regions are in between 88.35 ms (EC) and 91.88 ms (WC), the dialect region SO has a slightly shorter average duration with 82.72 ms. While speakers from the *Southern* parts of Scotland tend to produce shorter vowels in the present dataset, these differences are not very pronounced.

Vowel duration is also relatively similar for the *age groups*. The boxplots for the age groups *young* and *middle* are on a very similar level, only the boxplot for the older speakers is a bit higher than the rest. This overall tendency can also be seen in the average vowel durations: whereas the age group *middle* has an average vowel duration of 87.34 ms and the age group *young* a mean duration of 88.12 ms, the age group *old* has a slightly higher vowel duration of 94.37 ms. Older speakers therefore tend to produce slightly longer vowel pronunciations overall.

The tendency that older speakers produce slightly longer vowels can also be seen in the plot for the mean speech rate (see Figure 26). While there is great interspeaker variability among all age groups, there is an overall trend that older speakers speak with a lower mean articulation rate. This cannot only be seen in the dotplots but it is also indicated by the black regression line which has a clear negative slope. The vowel durations in the dataset therefore corroborate the observation that the mean articulation rate decreases in higher age groups (see subsection 3.1.9).

I also plotted the vowel durations for the variable *style*, but there seems to be no big difference between scripted and unscripted speech forms. This can be seen in the average durational values and ratios in Table 25: the ratios between the scripted and unscripted speech forms are relatively comparable for the *SVLR1*, *VE1* and *VE2* long and short contexts in scripted and unscripted speech. Only the difference for the *SVLR2* categorization stands out: the difference between the long and short contexts is stronger in scripted speech (ratio: 1.30) than in unscripted speech (ratio: 1.16). Furthermore, the mean durations for the *VE1* contexts are relatively similar; the ratios are very small and the *VE1* long vowels are on average shorter than the *VE1* short vowels in unscripted speech forms.

Table 25. Average values of vowel duration in milliseconds separated for the short and long contexts of the SVLR1, SVLR2, VE1 and VE2 categories as well as for scripted and unscripted speech forms. The ratios of the long and short contexts are stated in italics for the scripted an unscripted mean values of each categorization scheme.

	SVLR1			**SVLR2**			**VE1**			**VE2**		
	short	long	*ratio*	short	long	*ratio*	short	long	*ratio*	short	long	*ratio*
scripted	84.53	98.68	*1.17*	85.01	110.48	*1.30*	87.91	91.74	*1.04*	98.86	109.33	*1.11*
unscripted	82.08	97.74	*1.19*	82.40	95.67	*1.16*	87.29	81.66	*0.94*	90.30	98.59	*1.09*

Based on the proportions conveyed in the plots and tables, I also fitted several linear mixed effects models on log-transformed vowel duration to see whether the influence of the different predictor variables is statistically significant. I did not include any interactions because I first wanted to see what the general influence of the variables on vowel duration is. The best model was fit with the *SVLR1* categorization and the model has a marginal R^2 value of 0.34 and a conditional R^2 value of 0.50. This means that the fixed effects in the model can explain 34 percent and the combination of fixed and random effects describe 50 percent of the variance in the dataset. In other words, the model can explain 50 percent of the variation in log-transformed vowel duration by the random and fixed predictor variables. An overview of the model output is provided in Table 26.

Table 26. Best linear mixed model fit by maximum likelihood (marginal $R^2 = 0.34$; conditional $R^2 = 0.50$). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation.
Model formula: log_dur ~ lexical_set + SVLR1 + position + stress + freq + num_syl_word + num_pho_syl + syl_per_sec + reg + F1 + (1|speaker) + (1|syl_label) + (1|word_label)

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	94304.9	94605.4	-47120.5	94240.9	88389

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-4.8533	-0.6174	0.0195	0.6006	6.0210

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.01998	0.1414
syl_label	(Intercept)	0.02655	0.1630
speaker	(Intercept)	0.00604	0.0777
Residual		0.15760	0.39699

Intercept: lexicalsetCAT; SVLR1long; positionfinal; stressnuclear; freqhigh; reg East_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.951e+00	384.2706488	2.891e-02	< 2e-16	***
lexical_setCHOICE	1.644e-01	1.1786996	5.122e-02	0.00134	**
lexical_setDRESS	-1.422e-01	0.8674526	1.582e-02	< 2e-16	***
lexical_setFACE	1.577e-02	1.0158902	1.725e-02	0.36094	
lexical_setFLEECE	-2.681e-01	0.7648072	1.901e-02	< 2e-16	***
lexical_setGOAT	-1.314e-01	0.8768281	2.188e-02	2.06e-09	***
lexical_setGOOSE	-4.839e-01	0.6164054	2.145e-02	< 2e-16	***
lexical_setKIT	-5.211e-01	0.5938477	1.601e-02	< 2e-16	***
lexical_setMOUTH	-7.118e-02	0.9312908	2.905e-02	0.01431	*
lexical_setPRICE	8.014e-02	1.0834435	1.937e-02	3.58e-05	***
lexical_setSTRUT	-4.077e-01	0.6651999	2.228e-02	< 2e-16	***
lexical_setTHOUGHT	-1.244e-01	0.8830338	1.748e-02	1.33e-12	***
SVLR1short	-2.297e-02	0.9772875	7.719e-03	0.00292	**
positioninitial	-3.070e-01	0.7356589	6.584e-03	< 2e-16	***
positionmedial	-2.760e-01	0.7588496	5.248e-03	< 2e-16	***
stressprimary	-3.029e-01	0.7386976	4.623e-03	< 2e-16	***
stressunstressed	-3.964e-01	0.6727488	7.617e-03	< 2e-16	***
freqlow	2.101e-02	1.0212299	8.297e-03	0.01136	*
num_syl_word	-3.010e-02	0.9703490	3.352e-03	< 2e-16	***
num_pho_syl	-5.583e-02	0.9456979	5.504e-03	< 2e-16	***
syl_per_sec	-1.001e-01	0.9047790	1.486e-03	< 2e-16	***
regHighland	7.911e-03	1.0079426	2.363e-02	0.73830	
regInsular	7.671e-03	1.0077001	2.541e-02	0.76326	
regNorth_east	1.891e-02	1.0190869	2.450e-02	0.44168	
regSouth	-7.018e-02	0.9322232	2.461e-02	0.00504	**
regWest_central	-1.756e-02	0.9825978	2.138e-02	0.41308	
F1	-2.119e-05	0.9999788	3.734e-06	1.39e-08	***

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

Significant fixed factors include most of the *lexical sets*, the *SVLR1* environment, the *phrase position*, *stress*, *word frequency*, the *syllable phone count*, the *word syllable count*, the *F1* vowel quality measurement and the local articulation rate in terms of *syllable per second*. The variable *style*, however, had no significant influence on vowel duration. This means that the vocalic durations in scripted and unscripted speech forms are not significantly different.

There is also one significant extralinguistic variable, namely the *regional background*, specifically the dialect region *Southern*. The other regions do not show a significant effect when compared to the intercept (= dialect region *East central*). This means that the dialect regions *Highland and Hebridean English*, *Insular*, *Northeast* and *West central* do not have significantly different vowel durations than the dialect region *East central*. The speakers from the dialect region *Southern*, however, produce significantly shorter vowels than the speakers from the Eastern central belt of Scotland. As seen in the column with the exponentiated estimates, the estimate for *South* indicates that vowels tend to be 7 percent shorter when compared to the vowels of the intercept. While the difference between the dialect region *Southern* and the other regions was already visible in Figure 26, the model output provides statistical evidence for this difference. I also checked whether this durational difference might be due to variations in the local articulation rate of the speakers from the different regions. However, none of the regions reached statistical significance in a corresponding linear mixed effects model with *syllables per second* as the dependent variable. This confirms that vowel duration is generally shorter in the South of Scotland than in the Eastern Central Belt. Only the speed of speech for the older speakers was significantly slower than the articulation rate of the other speaker groups.

The other extralinguistic variables do not show a significant influence when modeled against the dependent variable of log-transformed vowel duration. This means that the *gender* or *age group* of a speaker does not have a significant influence on vowel duration overall. This confirms the distributions in the corresponding plots (see Figure 26): there were no visible differences between female and male speakers and only a very small but insignificant difference for the older age group.

As for the lexical sets, only the lexical set FACE does not show a significant difference when compared to the intercept (= lexical set CAT). This means that the vowel durations of the lexical sets FACE and CAT are very similar. The similarity between the two lexical sets was also visible in the boxplot in Figure 22. All other lexical sets are significantly different in terms of their vowel durations. The estimates reveal that the diphthongs CHOICE and PRICE are longer than the intercept. This is also detectable in Figure 22: only the boxplots of CHOICE and PRICE are higher than the boxplot of CAT. Likewise, CHOICE and PRICE are also the only lexical sets with a longer average duration (see Table 24). All other lexical sets tend to be shorter in terms of their duration. The effect size for short monophthong KIT, for instance, is the strongest of all lexical sets and specifies that it is 41 percent shorter than CAT. This also corresponds to the plots and average durations: KIT is the shortest vowel overall with the shortest mean duration (see Table 24 and Figure 22). The model thus confirms the influence of *intrinsic vowel duration* in the dataset (see subsection 3.1.1).

In addition, also the *F1* has a significant influence on vowel duration which may further corroborate the influence of tongue height on vowel duration (see subsection 3.1.1). The *F1* generally corresponds to the openness of the mouth in an inverse manner: high vowels have low *F1* measurers and low vowels have high *F1* frequencies. I also fitted further models with interactions and in these models, *F1* always showed significant interactions with the lexical sets. This means that the higher the vowel, the lower the

F1 and consequently, the shorter the duration. In other words, high vowels with low *F1* values tend to be shorter and low vowels with high *F1* values tend to be longer. This is in line with the effect of *intrinsic vowel duration* (see subsection 3.1.1).

The model further specifies that *SVLR1* short contexts are significantly shorter than *SVLR1* long environments. The effect size, however, is not very strong. The estimate suggests that vowels in *SVLR1* short contexts are overall only 3 percent shorter than vowels in *SVLR1* long contexts. Yet, it must be taken into consideration that all vowels were included in the present model. Previous studies have shown that Aitken's Law does not operate across all vowels (see section 3.2), so it could be that some non-SVLR affected vowels diminish the effect size of the *SVLR1* categorization. Nevertheless, the difference between *SVLR1* short and long vowels is statistically significant.

The durational difference between vowels in phrase-final syllables and vowels in phrase-medial or phrase-initial syllables is also significant. The estimates reveal that vowels tend to be 25 to 28 percent shorter in phrase-medial and phrase-initial positions. These durational differences were also clearly visible in the corresponding plot (see Figure 24). The model therefore confirms the influence of *constituent-final lengthening* on vowel duration (see subsection 3.1.8).

A similar distribution can be found for the variable *stress*. The model confirms that vowels in unstressed or primary stressed syllables are significantly shorter than vowels in syllables under nuclear stress. The estimates reveal that vocalic intervals in unstressed syllables are 33 percent shorter and vowels in primary stressed syllables are 27 percent shorter than vowels in nuclear stressed syllables. This tendency was also visible in the corresponding plot (see Figure 24). Hence, vowel duration is significantly affected by *stress*.

Whereas the plots indicated no major difference between vowels in high and low frequency words (see Figure 24), the durational difference between the two levels reaches statistical significance. According to the estimate, vowels in low frequency words tend to be 2 percent longer than vowels in high frequency words. This generally fits to the inverse relationship between frequency and duration (see subsection 3.1.5); however, the effect size is relatively small. It is therefore not surprising that no clear difference between the high and low frequency words was visible in the plots.

The linear variables representative for *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4), namely the *syllable phone count* and the *word syllable count*, do also reach statistical significance in the model. For each additional syllable in a word, the vowel duration tends to become 3 percent shorter. For each additional phone in a syllable, the duration of vocalic intervals tends to become roughly 6 percent shorter.

The local articulation rate also affects vowel duration. The estimates suggest that for each additional syllable produced in a timeframe of 1 second, vowel duration tends to become 10 percent shorter. Hence, speakers with a high articulation rate produce shorter vowels and speakers with a low articulation rate produce longer vowels. This corresponds to the influence of *tempo* (see subsection 3.1.9).

The models for the other categories *SVLR2*, *VE1* and *VE2* provide highly similar results despite the partially different classification schemes. There are significant effects for the variables *lexical set*, *stress*, *phrasal position*, the *word syllable count*, the *syllable phone count*, the local articulation rate (*syllables per second*) and the regional background *Southern*. Furthermore, the predictor variables *SVLR2*, *VE1* and *VE2* all reach statistical significance. Vowels in *SVLR2* short contexts are significantly shorter than vowels in *SVLR2* long contexts. This confirms the morphological effect of Aitken's Law. Vowel duration also differs in *VE1* and *VE2* long and short contexts. Yet, there is a striking difference for the *VE1* categorization: the estimates indicate that the *VE1* short contexts are, in fact, 2 percent longer than the *VE1* long contexts. While the effect size is relatively small, this distribution is clearly not in line with the VE (see subsection 3.1.2). The difference between the more restrictive categories of the *VE2* long and short contexts follows a more usual pattern: *VE2* long contexts are significantly longer than *VE2* short contexts although the effect sizes are also very small. Hence, there seems to be an influence of the VE in the plosive contexts, but it is not exactly clear which of the postvocalic contexts lead to the unusual distribution in the *VE1* categorization scheme.

I also used the PrInDT function (Weihs & Weilinghoff, 2023) to evaluate the predictive power of the best *SVLR1* models. The PrInDT function first stochastically subsamples the dataset so that only two-thirds (66%) or four-fifths (80%) of the whole sample are used for model building. Then, PrInDT implements a specified amount of the fixed factors (i.e. 50%, 75% and 100%) as predictor variables and fits the best possible models with the subsampled datasets and the varying amount of predictor variables. The model building procedure is repeated so that many different models are generated. These models are then fitted to the whole dataset and PrInDT calculates the corresponding R^2 values. In other words, PrInDT uses different proportions of the dataset and different combinations of variables to train multiple linear mixed effects models that are then reapplied for the whole dataset. The selection procedure is random so that many different combinations are tested. In the last step, PrInDT uses the *predict()* function of the lme4 package (Bates et al., 2015) to calculate the predictions for the training and test datasets and returns those models that reach the highest R^2 values when fitted to the whole sample. In this study, I used 250 repetitions and subsamples with 66% and 80% of the dataset as well as 50%, 75% and 100% of the predictor variables. This means that PrInDT created 1500 different linear mixed effects models in total. Table 27 provides an overview of the mean R^2 values for different models that were fitted to the test datasets. For instance, those models that were trained on 66 percent of the data and implemented 50 percent of the fixed factors have an average R^2 of 0.444 when applied to the test dataset (= the remaining one-third of the sample that was not used for training). Table 27 shows that the average R^2 values increase with a larger training dataset and with a higher proportion of fixed factors in the model building process. The sample with 66% of the data has an average R^2 of 0.463 when 75% of the fixed factors are used and a mean R^2 value of 0.487 when all fixed predictor variables factors are implemented. The respective R^2 values with 80% of the sample are even higher: 0.448 with 50% of the fixed factors, 0.466 with 75% of the fixed variables and 0.490 with all fixed factors. This shows that the predictive power of the models slightly increases with a larger dataset and more variables.

Table 27. Summary of mean R^2 values for the test datasets with those models that were trained on 66 and 80 percent of the dataset (subsampled) with 50, 75 and 100 percent of the fixed predictor variables included (100 repetitions).

	50 % of fixed factors	**75 % of fixed factors**	**100 % of fixed factors**
66 % of dataset	0.444	0.463	0.487
80 % of dataset	0.448	0.466	0.490

The best model of the PrInDT function reached a maximum conditional R^2 value of 0.53 with 80 % of the observations but only 50 % of the fixed predictor variables. A summary of the model output is given in Table 28.

Table 28. Best linear mixed model created with the PrInDT method (conditional $R^2 = 0.53$). The model was trained on 80 percent of the sample and 50 percent of the fixed predictor variables. T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation.

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	77702.7	77812.7	-38839.4	77678.7	70724

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-4.7629	-0.6146	0.0208	0.6014	5.7958

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.022683	0.15061
syl_label	(Intercept)	0.062902	0.25080
speaker	(Intercept)	0.007047	0.08395
Residual		0.157676	0.39708

Intercept: stressnuclear; freqhigh; positionfinal, SVLR1long

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.529e+00	251.8919	1.532e-02	< 2e-16	***
stressprimary	-3.062e-01	0.7362	5.213e-03	< 2e-16	***
stressunstressed	-4.338e-01	0.6480	8.408e-03	< 2e-16	***
syl_per_sec	-1.024e-01	0.9027	1.668e-03	< 2e-16	***
freqlow	2.444e-03	1.0025	9.114e-03	0.78860	
positioninitial	-3.014e-01	0.7398	7.413e-03	< 2e-16	***
positionmedial	-2.723e-01	0.7616	5.915e-03	< 2e-16	***
SVLR1short	-3.255e-02	0.9679	8.810e-03	0.00022	***

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

Significant fixed factors include *stress*, the *phrasal position*, the local articulation rate in terms of *syllables per second* as well as the *SVLR1* categorization. These variables were also significant in the best manually fitted model (see Table 26). The variable *lexical frequency* was also chosen by the PrInDT function, but this variable did not return significant effects on log-transformed vowel duration in the best PrInDT model (see Table 28). When comparing the PrInDT model with the manually fitted model (Table 26), one can also notice that the scaled residuals and exponentiated estimates are highly similar. For example, the estimate for vowels in initial *phrasal positions* is 0.738 in the regular model and 0.739 in the PrInDT model. Highly similar estimates can also be found for the medial *phrasal positions*, primary *stress* and the local articulation rate in terms of *syllables per second*. The estimates for the

SVLR1 categorization and unstressed syllables are also relatively similar. Overall, the estimates show the same effects that could also be found in the regular model: vowels are generally shorter in phrase-initial (-26.02%) and phrase-medial positions (-23.84%) than in final positions (=intercept). Vowel duration is also 26.38 percent shorter in primary stressed and 35.2 percent shorter in unstressed syllables than in nuclear stressed syllables (=intercept). This underlines the influence of *prosodic stress* (see subsection 3.1.7) and the different effect sizes of primary stressed syllables and unstressed syllables further indicate an effect of *lexical stress* (see subsection 3.1.6). Like in the previous model (see Table 26), the vowels are also affected by the local articulation rate and by the phonological SVLR. The vowels become roughly 10 percent shorter for every additional syllable produced in the timeframe of one second and they are also shorter in *SVLR1* short environments than in *SVLR1* long contexts.

However, the PrInDT model does not include the variables *lexical set*, *number of phones in syllable, number of syllables in word*, or *regional background* even though these variables were significant in the model that was trained with the backward selection procedure (see Table 26). This means that these variables are not necessary for training a model that fits the dataset well. The influence of the regional background might be negligible because only the region *South* had significantly different vowel durations than the intercept. All other regions had comparable durations. The variable regional background is therefore not a great help for explaining the variation in the dataset. The same goes for the *syllable phone count* and the *word syllable count*. However, it is surprising that the exclusion of the variable *lexical set* led to higher conditional R^2 values. I also ran the PrInDT function with different repetition numbers (i.e. 100, 200 repetitions), but the variable *lexical set* was always excluded from the model. A possible explanation for the exclusion of the variable *lexical set* is the influence of multicollinearity. For example, while average duration varies for the different vowels (see Table 24), other variables, such as *stress*, the *phrasal position*, the *SVLR1* categorization or the local articulation rate might have a stronger influence on the duration that the phonological classification of the vowels. This means that the influence of *intrinsic vowel duration* (see subsection 3.1.1) is not strong in SSE speech. Overall, duration varies across the different vowels (see Figure 22), but the variables *stress, phrasal position, SVLR1* and *tempo* have a much stronger influence on the duration than the classification of different vowels themselves.

The plots and models provide a great overview of what affects vowel duration in the dataset. Yet, it must be taken into consideration that they included all vowels at the same time. It could be that there are some vowels which are affected by the SVLR and VE and some which are not. It could also be that other intra- and extralinguistic predictor variables affect some vowels more than others. Therefore, the lexical sets will be analyzed individually in the following sections. Furthermore, I will also include interactions in the model building for the individual vowels in the following sections to obtain more detailed findings. As the F1 represents the height of the vowels and generally interacts with the lexical sets, I will not include this variable in the subsequent sections because I analyze each vowel individually.

5.2 Short Monophthongs

This section comprises the findings for the short SSE monophthongs KIT, STRUT and DRESS. For all vowels, tokens shorter than 10 ms were excluded from the dataset because the vowel pronunciation will inevitably be reduced if they last for such a short period of time.

5.2.1 KIT

The lexical set KIT represents the short monophthong /ɪ/ in words such as <fit>, <bit> and <miss>. A statistical overview of the vocalic durations of KIT can be found in Table 29 and a visualization of the data can be found in Figure 27.

Table 29. Statistical summary of vowel duration for the lexical set KIT.

Token number:	44382
Average duration:	64.71 ms
Standard deviation:	41.14 ms
Minimal duration:	10 ms
Maximal duration:	780 ms
1st quantile:	40 ms
Median:	53 ms
3rd Quantile:	79 ms

KIT is the lexical set which has by far the most tokens in the dataset (N= 44381). The most frequent words are the function words <in>, <the> and <is>. The most frequent content words are <think>, <really> and <Scottish>. As seen in Figure 27 and Table 29, the majority of the data points are in the range between 40 and 79 ms. The median is 53 ms and the average duration is 64.71 ms. The average duration is therefore shorter than the mean duration in McClure's (1977) carrier sentence readings (see subsection 3.2.2). However, the current dataset comprises naturally occurring language only, so it is not

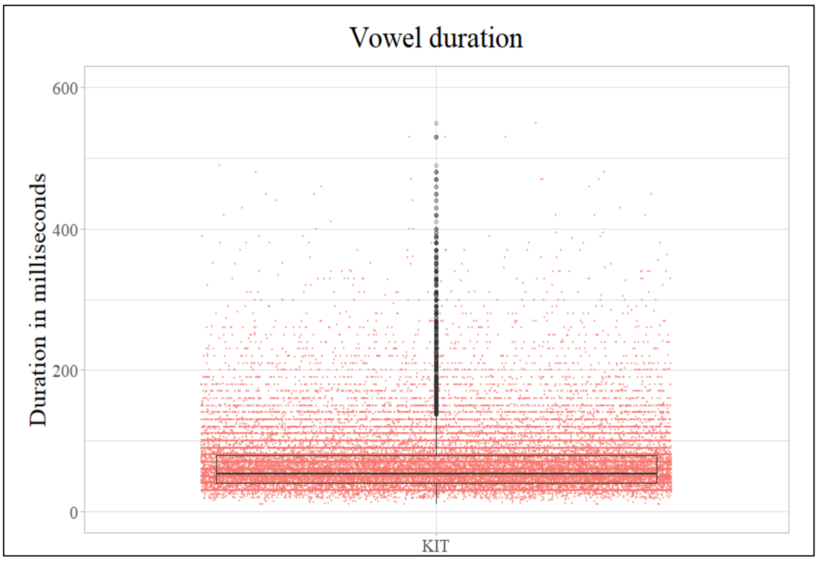


Figure 27. Vowel duration of the lexical set KIT in milliseconds.

surprising that the overall duration is shorter than in the experimental setting by McClure (1977). While most of the data is clustered in the range between 40 and 79 ms, there is still a relatively high standard deviation of 41.14 ms. The overall duration range is also very large: the shortest KIT articulations are only 10 ms long, the longest pronunciation lasts 780 ms. Thus, while it is true that KIT is generally a short vowel, it is clearly not "always short in all contexts" (Dieth, 1932, p. 67).

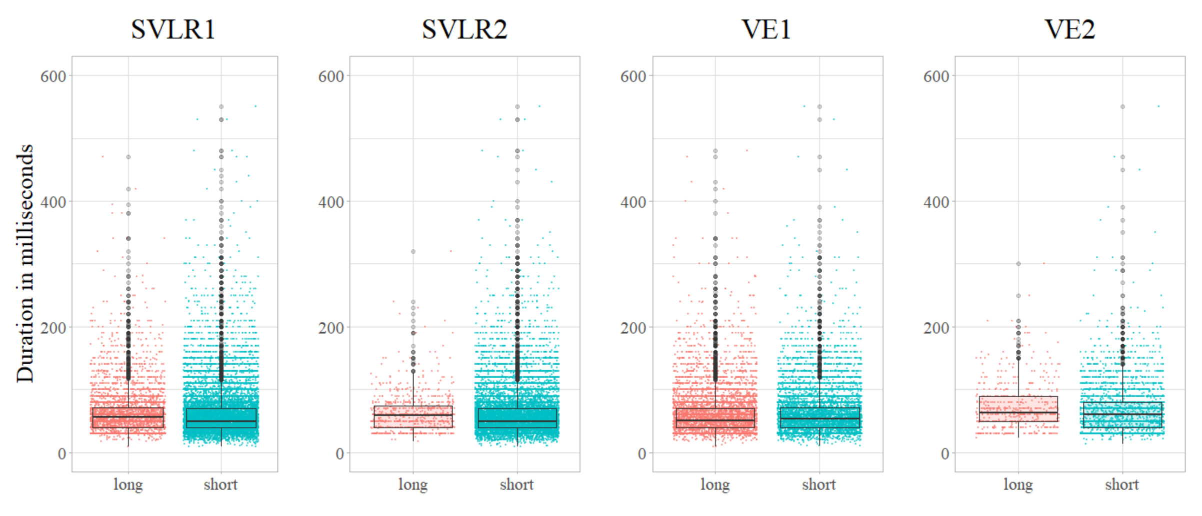


Figure 28. Boxplots and jitterplots of raw vowel duration of the lexical set KIT for the categories SVLR1, SVLR2, VE1 and VE2.

The categories representing Aitken's Law and the VE (see Figure 28) reveal that there is not a great difference between the long and short contexts. The jitterplots further show that there are fewer tokens for the long contexts than for the short contexts. As for the *SVLR1* categorization, the boxplots as well as the average durations indicate that there is not a great durational difference between the long and short environments (mean duration *SVLR1 long*: 63.97 ms; mean duration *SVLR1 short*: 60.18 ms). Hence, while KIT is slightly longer in the *SVLR1* long context overall, the average durational difference of 3.79 ms seems negligible. The difference between *SVLR2* long and short contexts is also very weak. The boxplot for the long environments is only slightly higher and the average durations are less than 3 ms apart (mean *SVLR2 long*: 62.67 ms; mean *SVLR2 short*: 59.77 ms). Hence, the morphological conditioning of Aitken's Law does not affect the duration of KIT to a great extent. As for *VE1*, there is virtually no difference between the long and short contexts. The boxplots and the average durations are almost identical (mean *VE1 long*: 62.08 ms; mean *VE1 short*: 62.09 ms). Therefore, postvocalic voiced consonants do generally not lead to increased vowel duration in the lexical set KIT. Yet, a slight difference can be detected in the *VE2* environments: the boxplot is a bit higher and the average duration is 5.87 ms longer for the long environments (mean *VE2 long*: 72.11 ms; mean *VE2 short*: 66.24 ms). This indicates that postvocalic voiced plosives lead to a slightly longer average duration than postvocalic voiceless plosives. Yet, while postvocalic voiced plosives slightly increase vowel duration overall, there is no difference for the *VE1* long and short contexts: this means that other postvocalic voiced consonants might slightly decrease vowel duration. When checking the average durations of the tokens in all postvocalic consonant environments (see Table 30), one can see that following nasals have the shortest vowel duration overall. This is in line with the categorization of Aitken’s Law as following nasals are considered short contexts in Scottish English (Aitken, 1981). At the same time, it partially contradicts the VE because nasals are voiced and usually lead to an increase in vowel duration in English (House & Fairbanks, 1953, p. 108) (see subsection 3.1.2). The postvocalic nasals might therefore be responsible for the similarity between *VE1* long and short contexts because they decrease vowel duration in KIT. Another clear tendency visible in Table 30 is that postvocalic pauses lead to a strong increase in vowel duration. This clearly shows the effect of *constituent-final lengthening* (see subsection 3.1.8).

Table 30. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set KIT. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categorizations.

Postvocalic environments	Average vowel duration
laterals	63.18 ms
nasals	54.93 ms
voiced fricatives	67.63 ms
voiceless fricatives	59.36 ms
voiced plosives	66.50 ms
voiceless plosives	63.32 ms
pauses	142.89 ms

The tendency of pre-pausal lengthening can also be seen in Figure 29: tokens in phrase-final positions are longer than in phrase-medial and phrase-initial positions. It is also clear that KIT is longer in nuclear stressed syllables than in non-nuclear stressed syllables which represents the influence of *prosodic stress* (see subsection 3.1.8). However, the influence of *lexical stress* seems negligible: primary stressed syllables and unstressed syllables have very similar durations. The boxplots of *primary* stressed syllables and *unstressed* syllables are on a very similar level and the average durations are almost identical (62.11 ms and 61.96 ms). A possible reason is that the level *primary* comprises both monosyllabic words and stressed syllables of polysyllabic words. For example, the phrase <he’s knitting it> contains three KIT vowels: /hiz ˈnɪ̱tɪ̱ŋ ɪ̱t/. The first syllable in <knitting> is lexically stressed and

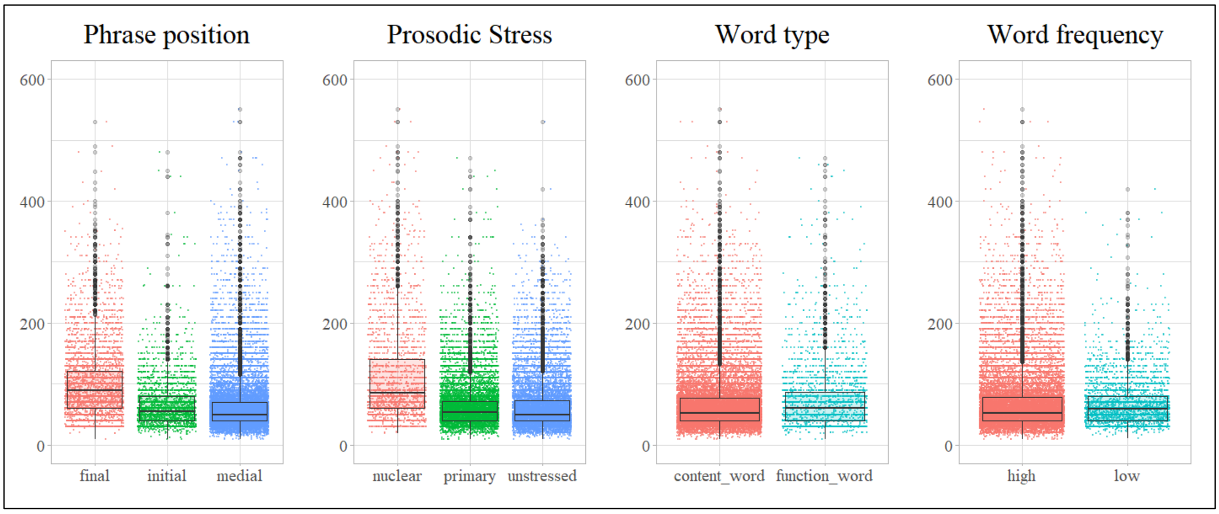


Figure 29. Boxplots and jitterplots of raw vowel duration for the lexical set KIT separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

therefore belongs to the category *primary*. The second syllable in the same word is not lexically stressed and it is therefore categorized as *unstressed*. While it is possible that the lexically stressed syllable is longer than the unstressed syllable, the monosyllabic function word <it> would also fall into the category of *primary* lexical stress. The category *primary* might therefore be as short as the category *unstressed* because it also contains words from short monosyllabic function words.

As for the word type, the jitterplot in Figure 29 shows that KIT is represented by far more content words (N= 41477) than function words (N= 2905). There is an overall tendency that function words tend to be longer than content words. The boxplot of the function words is slightly higher and the average duration is slightly longer (content words: 64.13 ms; function words: 72.98 ms). This distribution does not correspond with the effect of *lexical category* (see subsection 3.1.5), but the durational difference is also not very strong. Similar to the overall trend (see Figure 22), an even smaller difference can be found for the high and low frequency words. The boxplots are on the same level and the average durations are highly similar (high frequency words: 64.13 ms; low frequency words: 66.02 ms). The plot therefore indicates that *word frequency* has a negligible influence on the duration of KIT in connected SSE speech.

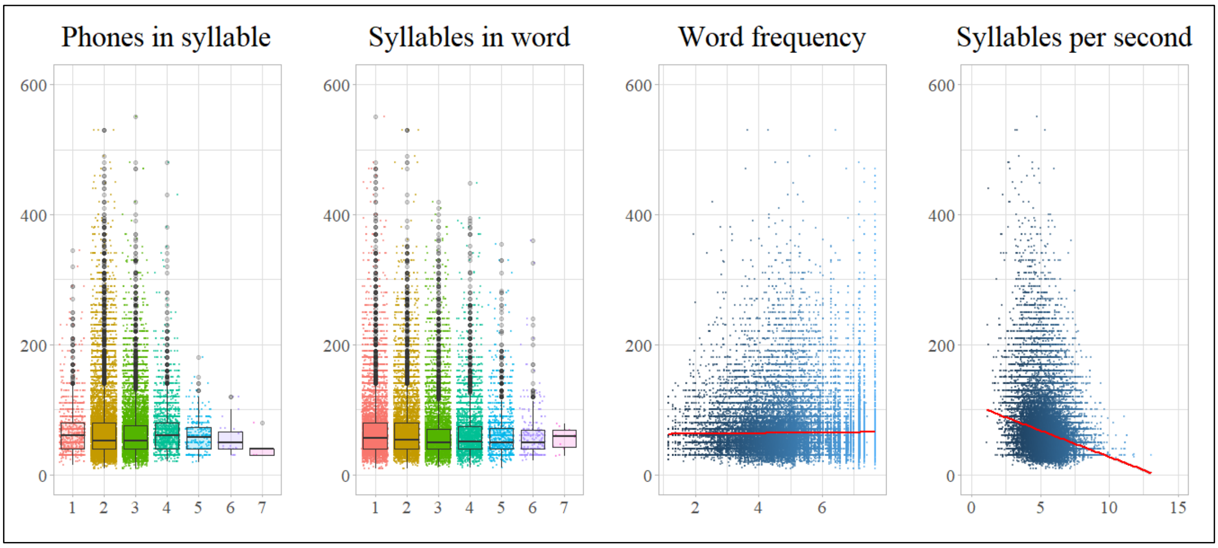


Figure 30. Plots of raw vowel duration of the lexical set KIT separated for the linear variables number of phones in a syllable, number of syllables in a word, word frequency and syllables per second.

The plots showing the linear intralinguistic variables (see Figure 30) are very similar to the equivalent plots representing all vowels (see Figure 25). Overall, a higher number of phones and syllables reduces the occurrence of very long KIT pronunciations. This means that the higher the number of phones and syllables, the fewer outliers can be found at the top of the boxplots. Furthermore, while vowel duration does not constantly decrease with a higher *phone* or *syllable count*, the average durations reveal that overall, the phone or syllable number is inversely correlated with duration. For example, vowel duration generally decreases with a higher *phone syllable count*; the only exception is that syllables with four phones (mean: 64.33 ms) are slightly longer than syllables with three phones (mean: 62.89 ms). This can also be seen in Figure 30: the boxplot with 4 syllables is the only one which is slightly higher than the neighboring boxplot on the left side. Apart from that, every additional phone in a syllable leads to a lower boxplot and to a decrease in the average vowel duration. A similar development can also be seen for the *word syllable count*: whereas vowels in monosyllabic words have the highest average duration (65.83 ms), vowels in words with seven syllables are the shortest overall (mean: 54.82 ms). This clearly indicates an influence of *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4).

The plot for *word frequency* does not show a clear effect. The slope of the red regression line comes close to zero, so it does not seem that the frequency of a word has a striking difference on the pronunciation of its vowels. This trend also corresponds with the plot representing the binary classification of high and low frequency words (see Figure 29).

The plot representing the *local articulation rate* shows that a higher speed of speech leads to shorter vowel pronunciations. There are fewer long vowels in high articulation rates and the regression line has a clear negative slope. This means the higher the articulation rate, the shorter the vowel pronunciation which is in line with the influence of *tempo* (see subsection 3.1.9).

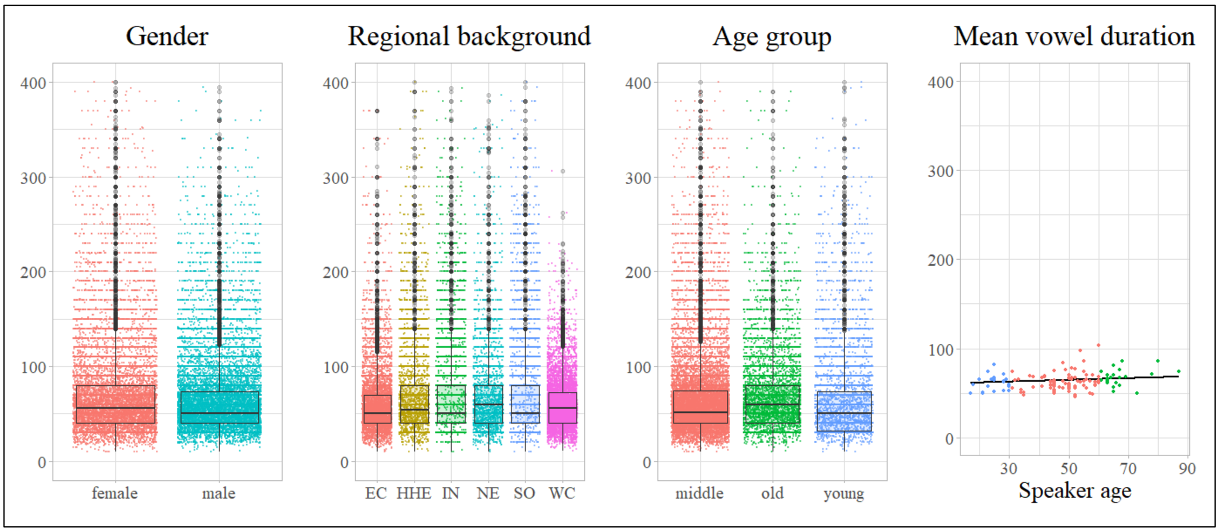


Figure 31. Plots of raw vowel duration of KIT separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean vowel duration plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

The plots for the social variables (see Figure 31) are also very similar to the corresponding plots representative for all vowels (see Figure 26). Hence, the duration of KIT is not strikingly affected by the *gender* or *regional background* of a person. There is a tendency that older speakers produce slightly longer KIT vowels. The boxplot for the *age group old* is slightly higher than the boxplots representing the other age groups and the average durations differ as well (*age group old*: 68.18 ms; *age group middle*: 63.78 ms; *age group young*: 64.18 ms). Furthermore, the black regression line in the mean vowel duration plot indicates that there is a positive relationship between the age of a speaker and the corresponding mean vowel duration. The slope of the regression line is mildly positive, so the duration of KIT increases slightly if the age of a speaker is higher. The differences might not be as striking because the durations are generally shorter for KIT. There is also great interspeaker variation, but this trend goes hand in hand with the findings concerning the mean articulation rate and the age of a speaker (see section 5.1): the older a person, the shorter the mean articulation rate. This means that older people tend to speak more slowly than younger people. Because of this, vowel duration tends to become slightly longer among older speakers.

As for inferential statistics, I fitted several linear mixed effects models in a stepwise regression procedure with a backward selection for model building. I also included all possible interactions between the fixed variables, but the implementation of the interactions did not improve the model's accuracy in terms of R^2. A general tendency is that all models for the lexical set KIT have lower R^2 values than the models for all vowels (see section 5.1). Whereas the conditional R^2 values of the model for all vowels were in the range of 0.50, the best model for KIT has a conditional R^2 value of only 0.36. This means that it is more difficult to explain the durational variation in KIT with the given fixed and random factors. Another important finding is that none of the categories *SVLR1*, *SVLR2*, *VE1* and *VE2* returned significant effects. The categories representative for Aitken's Law and the VE do therefore not have a significant influence on vowel duration in the lexical set KIT. This corresponds with the respective plots where it is not possible to detect a striking difference between the different long and short contexts (see Figure 29). The best model was subsequently fitted on the whole dataset of the lexical set KIT without any of the categories representing SVLR or the VE. Table 31 provides a summary of the model output.

Table 31. Best linear mixed model fit by maximum likelihood (marginal R^2 = 0.13; conditional R^2 = 0.36). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ position + stress + freq + num_syl_word + num_pho_syl + syl_per_sec + age_group + (1|speaker) + (1|syl_label) + (1|word_label)

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	50425.5	50556.0	-25197.7	50395.5	44367

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-5.5885	-0.6624	-0.0279	0.6067	5.6123

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.030226	0.17386
syl_label	(Intercept)	0.022748	0.15082
speaker	(Intercept)	0.006428	0.08018
Residual		0.168040	0.40993

Intercept: positionfinal; stressnuclear; freqhigh; agegroupmiddle

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. Code**
(Intercept)	5.463e+00	235.91441	4.274e-02	< 2e-16	***
positioninitial	-3.305e-01	0.7185533	1.017e-02	< 2e-16	***
positionmedial	-3.625e-01	0.6959191	7.273e-03	< 2e-16	***
stressprimary	-3.651e-01	0.6941325	1.013e-02	< 2e-16	***
stressunstressed	-4.301e-01	0.6504403	1.044e-02	< 2e-16	***
freqlow	2.352e-02	1.0237985	1.023e-02	0.021554	*
num_syl_word	-1.587e-02	0.9842540	4.653e-03	0.000653	***
num_pho_syl	-7.492e-02	0.9278142	1.052e-02	2.46e-12	***
syl_per_sec	-8.811e-02	0.9156611	2.252e-03	< 2e-16	***
age_groupold	5.143e-02	1.0527732	2.037e-02	0.012847	*
age_groupyoung	-1.679e-02	0.9833547	2.045e-02	0.413248	

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

Significant fixed factors include the *phrasal position*, *stress*, *word frequency*, the *word syllable count*, the *syllable phone count*, the local articulation rate in terms of *syllables per second* and the *age group*. All other variables were dropped in the model building process, so there is no significant influence of the variables *SVLR1*, *SVLR2*, *VE1*, *VE2*, *style*, *gender* or *regional background*. The model therefore specifies that neither Aitken's Law nor the VE has a significant effect on vowel duration in the lexical set KIT. The gender or regional background of the speakers do also not affect vowel duration and it is irrelevant whether the vowels were produced in a scripted or unscripted speech setting (= variable *style*).

The fixed factor *age group* is the only extralinguistic variable which returns significant effects: the *age group old* produces significantly longer KIT vowels than the *age group middle* (= intercept). The effect size, however, is not very strong. The estimates reveal that KIT vowels produced by older speakers are only five percent longer than the KIT vowels produced by middle-aged speakers. The *age group young* does not pronounce significantly longer vowels than the *age group middle*. While the increased vowel duration among the *age group old* was indicated in the corresponding plot (see Figure 31), the model output reveals that this difference is also statistically significant.

The intralinguistic variable *phrasal position* also returns significant effects. Vowels in phrase-initial syllables are roughly 28.15 percent shorter than vowels in phrase-final syllables (= intercept). Likewise, vowels in phrase-medial syllables are over 30 percent shorter than those in phrase-final syllables. Both the model and the corresponding plot (see Figure 29) therefore confirm the influence of *constituent-final lengthening* (see subsection 3.1.8) on vowel duration in the lexical set KIT.

The model output further confirms that vowels in unstressed or primary stressed syllables are significantly shorter than vowels in syllables under nuclear stress. Primary stressed syllables lead to 30 percent decrease in vowel duration when compared to the intercept. Vowels in unstressed syllables are almost 35 percent shorter than vowels in nuclear stressed syllables. The lexical set KIT is therefore influenced by *prosodic stress* (see subsection 3.1.7), but the influence of *lexical stress* appears to be negligible.

The binary category of high and low frequency words also reaches statistical significance in the model output even though the p-value is not very high. The estimates suggest that words with a low frequency are 2 percent longer than high frequency words. While this difference is in line with the inverse relationship between word frequency and duration (see subsection 3.1.5), the effect size is relatively small.

The model output further shows a significant influence of the *syllable phone count* and *word syllable count*. The estimates reveal that KIT vowels become roughly 2 percent shorter for every additional syllable in a word and roughly 7 percent shorter for every additional phone in a syllable. While the effect size is not very strong, the significance of both factors indicates the influence of *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4) in the lexical set

KIT. The short monophthong /ɪ/ therefore tends to be shorter in the word <street> than in the word <eat> and it also tends to be shorter in the polysyllabic word <knitting> than in the monosyllabic word <knit>.

The model also shows a significant influence of *tempo* (see subsection 3.1.9). The estimates indicate that for each additional syllable produced in a timeframe of 1 second, vowel duration tends to become roughly 9 percent shorter. In other words, the higher the local articulation rate, the shorter the KIT vowels.

The plots and models of the lexical set KIT have shown that the short monophthong is not affected by Aitken's Law nor by the VE. This corresponds to earlier impressionistic accounts (Aitken, 1981; Dieth, 1932; Wettstein, 1942; Zai, 1942) and also to the findings of previous empirical studies (McClure, 1977; McKenna, 1988; Scobbie, Hewlett, & Turk, 1999; Stuart-Smith et al., 2019). However, while KIT is a short vowel in naturally spoken SSE, it is not a "vowel of invariable quantity" (Zai, 1942, pp. 15–16). The models and plots revealed an influence of *prosodic stress*, *constituent-final lengthening*, *tempo*, *lexical frequency* as well as *intrasyllabic compression* and *polysyllabic shortening*. Furthermore, older speakers also generally produce longer KIT vowels and they also have a lower mean articulation rate.

Apart from that, the low R^2 values indicate that it is difficult to model the duration of KIT. The significant fixed factors can only explain roughly 23 percent of the variation in the dataset. The independent factors can only account for 13 percent of the variation. Roughly 64 percent of the durational variation in KIT cannot be explained by the fixed and random factors. It is therefore difficult to account for the durational variation in the short monophthong /ɪ/. In comparison to other vowels, the duration of this lexical set does not follow the overall patterns in such a clear way. The classification schemes of the SVLR and the VE do also not help in increasing the model fit.

The findings for KIT are generally in line with most previous investigations and they also correspond to Aitken's (1981) classification. Other recent studies could find no empirical evidence for SVLR- or VE-related effects in KIT (Scobbie, Hewlett, & Turk, 1999; Stuart-Smith et al., 2019) and the present investigation could find no such effects either. The short monophthong is, however, clearly affected by other sociolinguistic and prosodic factors. This contradicts many impressionistic accounts on Scottish vowel duration (see subsection 3.2.1) which claimed that KIT is invariably short.

5.2.2 STRUT

The lexical set STRUT represents the short low monophthong /ʌ/ in SSE. A statistical overview of the vocalic durations of STRUT can be found in Table 32 and a visualization of the data can be found in Figure 32.

The dataset comprises 6467 STRUT tokens and the most frequent words are <one>, <up> and <but>. STRUT has an average duration of 71.72 ms and a standard deviation of 35.70 ms. The median is 65.32 ms and most of the data points lie in the range between 50 and 90 ms. STRUT is therefore longer than KIT which corresponds to McKenna's (1988) observation that also the short monophthongs are influenced by *intrinsic vowel duration* (see subsection 3.1.1). In other words, STRUT is longer than KIT because its pronunciation is more open. Similar to KIT, however, the average duration of STRUT in this study is shorter than the mean duration of McClure's (1977) carrier sentence readings (see Table 7).

Table 32. Statistical summary of vowel duration for the lexical set STRUT.

Token number:	6467
Average duration:	71.72 ms
Standard deviation:	35.70 ms
Minimal duration:	17.00 ms
Maximal duration:	510.00 ms
1st quantile:	50.00 ms
Median:	65.32 ms
3rd Quantile:	90.00 ms

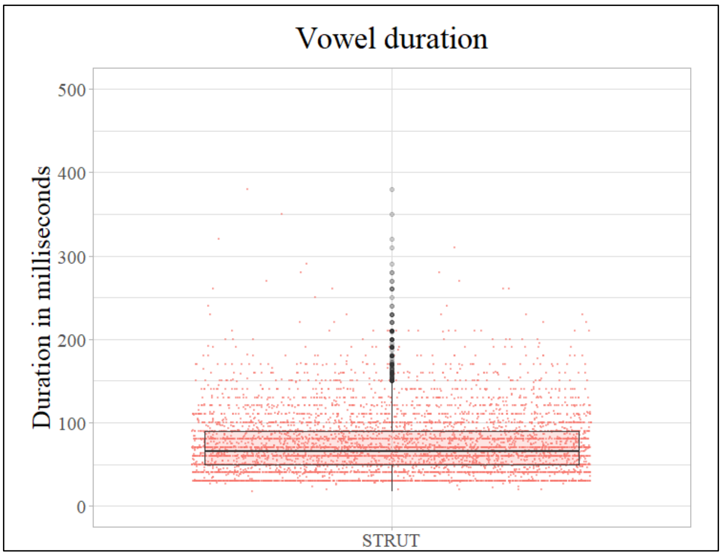


Figure 32: Vowel duration of the lexical set STRUT in milliseconds.

The boxplots representative for Aitken's Law show that there is not a great difference between the long and short contexts (see Figure 33). The average values for the *SVLR2* long and short contexts are

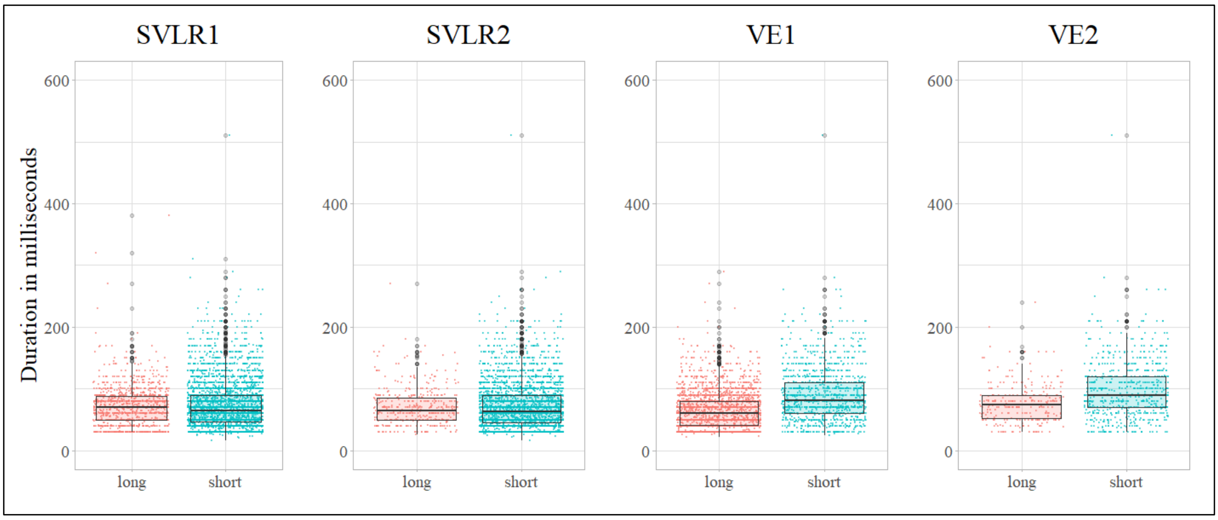


Figure 33. Boxplots and jitterplots of raw vowel duration of the lexical set STRUT for the categories SVLR1, SVLR2, VE1 and VE2.

almost identical (*SVLR2* long contexts: 70.76 ms; *SVLR2* short contexts: 70.99 ms) and the *SVLR1* short contexts are on average slightly longer than the *SVLR1* long environments (*SVLR1* long contexts: 70.98 ms; *SVLR1* short contexts: 73.01 ms). I also checked for the effects of the *SVLR1* and *SVLR2* categories across the different genders, age groups and regional backgrounds, but I could not find any striking distributions. The plots and average values therefore indicate that the lexical set STRUT is generally not affected by Aitken's Law.

The boxplots for the categories *VE1* and *VE2* show a different picture. For both categories, vowel duration is clearly shorter in the long environments than in the short environments. Vowel duration in the *VE1* long contexts (mean: 65.60 ms) is on average more than 24 ms shorter than in the *VE1* short contexts (mean: 90.11). Likewise, the *VE2* long contexts (77.01 ms) are on average more than 20 ms shorter than the *VE2* short contexts (mean: 97.31 ms). These shortening effects are also relatively stable across the regions, genders and age groups. The lexical set STRUT is therefore shorter when followed by voiced consonants than by voiceless consonants. This clearly contradicts the VE and this unusual effect is also clearly visible when investigating the average durations for the different postvocalic consonant environments (see Table 33).

Table 33. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set STRUT. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categorizations.

Postvocalic environments	Average vowel duration
laterals	61.37 ms
nasals	59.85 ms
voiced fricatives	76.36 ms
voiceless fricatives	82.61 ms
voiced plosives	77.28 ms
voiceless plosives	94.82 ms
pauses	132.39 ms

The vocalic durations are clearly shorter for the voiced fricatives than for the voiceless fricatives. Likewise, voiceless plosives are clearly longer than voiced plosives and interestingly, plosives generally trigger longer STRUT pronunciations than fricatives. Vowels before laterals are shorter with 61.37 ms and the shortest STRUT vowels on average are followed by nasals (59.85 ms). These distributions do not only contradict the postvocalic VE but also the lengthening effect concerning the manner of postvocalic consonants (House & Fairbanks, 1953, p. 108) (see subsection 3.1.2). The duration of STRUT is clearly influenced by the postvocalic context, but the lengthening effects do not correspond

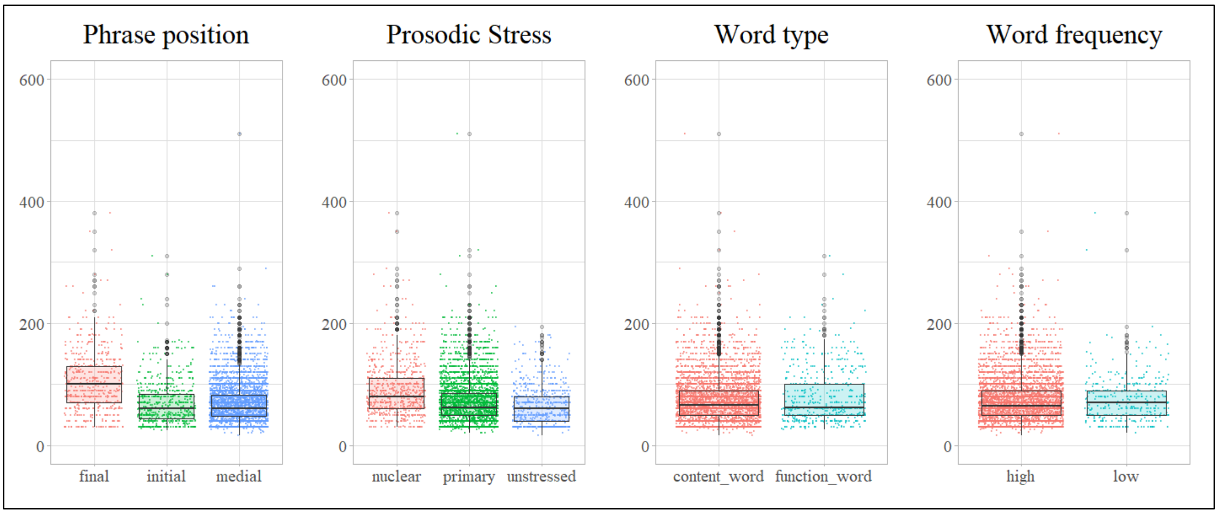


Figure 34. Boxplots and jitterplots of raw vowel duration for the lexical set STRUT separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

to previous findings. Similar to other vowels, however, STRUT is lengthened before pauses (132.39 ms).

This effect can also be seen in Figure 34: tokens in phrase-final positions are clearly longer than tokens in phrase-medial or phrase-initial positions. Whereas phrase-final STRUT tokens have an average duration of 103 ms, phrase-medial and phrase-initial tokens have average durations of 68.39 ms and 69.14 ms, respectively. Hence, STRUT is clearly influenced by the effect of *constituent-final lengthening* (see subsection 3.1.8). Similarly, STRUT is also influenced by *lexical* and *prosodic stress* (see subsections 3.1.6 and 3.1.7). STRUT vowels in primary stressed syllables (mean duration: 69.88 ms) are longer than vowels in unstressed syllables (mean duration: 62.23 ms) but not as long as vowels in nuclear stressed syllables (mean duration: 92.05 ms). This distribution is clearly visible in the boxplots (see Figure 34).

The durational differences between content words and function words as well as between high frequency words and low frequency words are less pronounced. The boxplot representing function words is a bit higher than the boxplot representing content words. The difference between the boxplots for word frequency is even weaker. As for the average durations, content words (mean: 71.02 ms) are roughly 8 ms shorter than function words (mean: 79.55 ms) and low frequency words (mean: 75.61 ms) are roughly 4 ms longer than high frequency words (mean: 71.47 ms). While these distributions do marginally correspond with the effect of lexical frequency, they are not in line with the effect of the lexical category (see subsection 3.1.5).

The plots for the *syllable phone count* and the *word syllable count* of STRUT do not show a clear trend (see Figure 35). While syllables with one phone and monosyllabic words have the longest STRUT

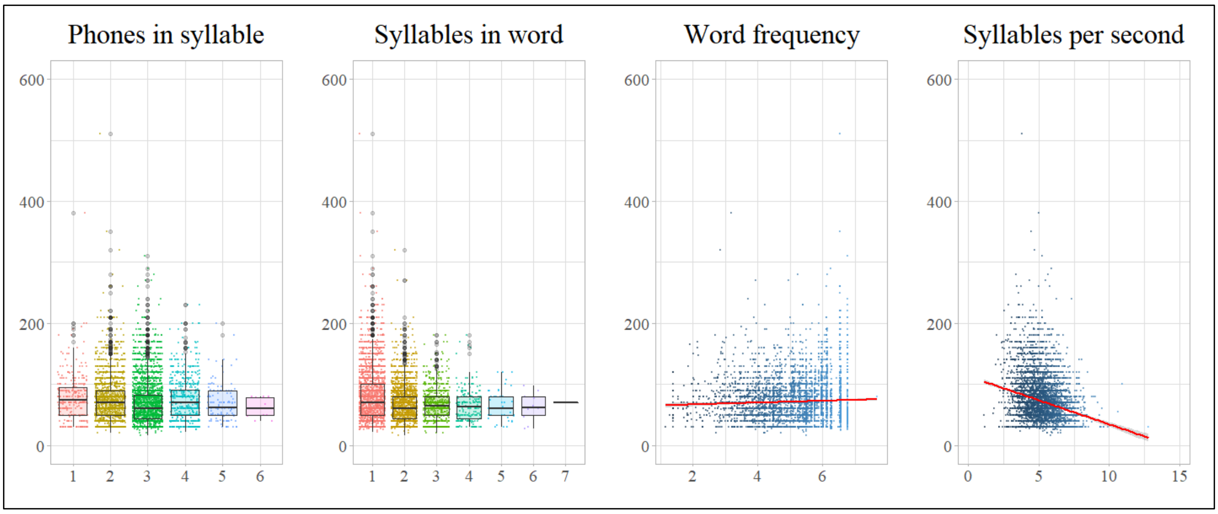


Figure 35. Plots of raw vowel duration of the lexical set STRUT separated for the linear variables number of phones in syllable, number of syllables in word, word frequency and syllables per second.

tokens on average, the other tokens with higher *phone syllable counts* and *word syllable counts* show no clear downward trend in duration. Overly long pronunciations become less frequent in higher *phone syllable counts* and *word syllable counts*, but the average durations do not steadily decrease. Hence, it does not seem that STRUT is affected by *intrasyllabic compression* (see subsection 3.1.3) or by *polysyllabic shorten*ing (see subsection 3.1.4).

Similar to KIT, the plot for *word frequency* does not show a clear trend either. The slope of the red regression line has a very slight positive trend. This would mean that STRUT tokens with a high frequency tend to be slightly longer in duration than STRUT tokens with a low frequency. This trend is, however, not in line with the average durations of high- and low frequency words and it would also contradict the influence of *lexical frequency* (see subsection 3.1.5). It is therefore unclear whether the lexical set is influenced by *word frequency* in a significant way.

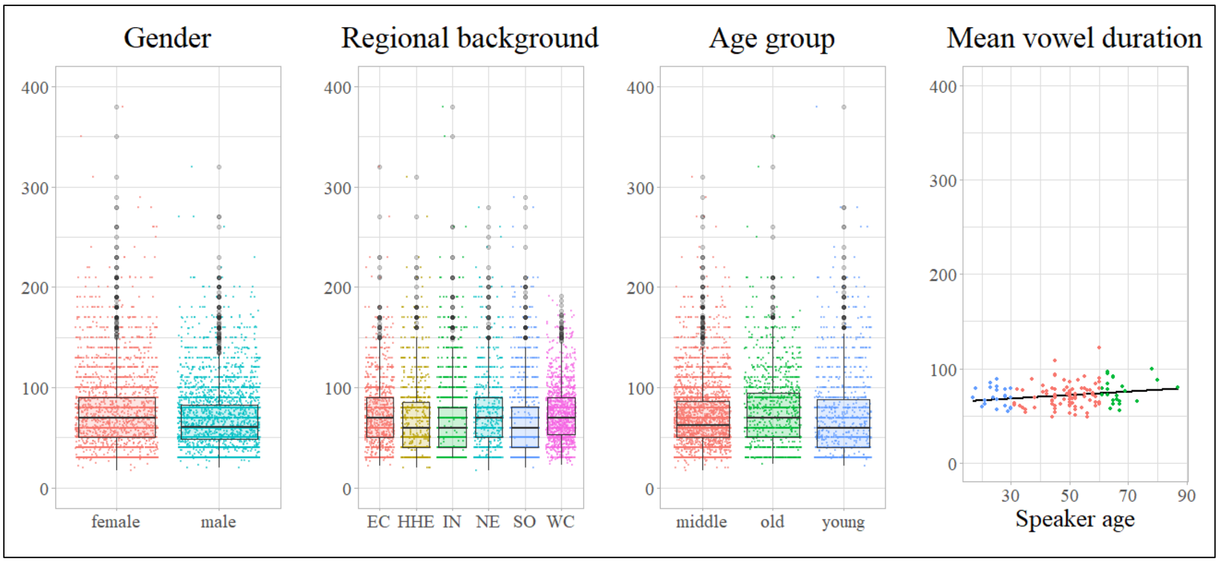


Figure 36. Plots of raw vowel duration of STRUT separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

The effect of the local articulation rate on STRUT is clearer. Similar to the other vowels, STRUT becomes increasingly shorter in higher articulation rates. The negative slope of the red regression line shows that STRUT tokens tend become shorter if the *syllable per second* rate is higher. This is in line with the effect of *tempo* (see subsection 3.1.9).

The plots for the social variables show only little variation. On average, female speakers (mean: 74.78 ms) produce slightly longer STRUT tokens than male speakers (mean 68.92 ms) and the boxplot is therefore a bit higher (see Figure 36). Vowel duration also mildly differs across the different regions with an average duration range from 67.33 ms (*Insular*) to 75.61 ms (*Northeast*). Yet, the durational variation within the regions and genders does also vary greatly, so it is difficult to conclude that the duration of STRUT is, for instance, generally shorter in the Northern Isles than in the Northeast of mainland Scotland. Similar to the other vowels, there is a stronger durational discrepancy for the age groups. The older speakers produce STRUT tokens which are on average 8 ms longer than the STRUT pronunciations of the other age groups. The boxplot for the old age group is also higher. This trend goes hand in hand with the mean vowel duration. The slope of the black regression line is mildly positive which indicates that the duration of STRUT increases with a higher age of the speakers. This tendency goes hand in hand with the mean articulation rate (see section 5.1). Older speakers generally tend to speak more slowly so the STRUT durations are a bit longer for the older speakers as well.

Similar to the other vowels, I also fitted several linear mixed effects models for the lexical set STRUT. I included all possible interactions and I found interactions that significantly improved the model fit. The conditional R^2 values are higher than the ones for KIT as they lie in the range between 0.31 and 0.42. The models on the effects of Aitken's Law all excluded the categories *SVLR1* and *SVLR2* because they did not return significant effects. Hence, the models confirm that the lexical STRUT is not affected by the phonological or morphological conditioning of Aitken's Law. The categories *VE1* and *VE2* were, however, significant and the best model could be fitted with the *VE1* categorization. A summary of the model can be found in Table 34.

The *VE1* context, the *phrasal position*, *stress*, *style*, the *articulation rate* (syllable per second) as well as the *age group* are significant fixed factors in the model. There are also significant two-way interactions between the *VE1* short contexts and other extralinguistic factors as well as three-way interactions between the VE1 short environments and male speakers from different regions. All other variables were dropped in the model building process. Thus, for instance, the *regional background* or *gender* has no significant influence on vowel duration but there are interactions for the VE across specific regions and genders. The influence of the *syllable phone count*, the *word syllable count* and the *word frequency* do not reach statistical significance in the model. This is in line with the distributions in the corresponding plots because it was not possible to detect clear durational variation for these variables (see Figures 34 and 35).

*Table 34. Best linear mixed model for the lexical set STRUT fit by maximum likelihood (marginal $R^2 = 0.20$; conditional $R^2 = 0.42$). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ VE1 + (VE1 * reg * gender) + (VE1 * age_group) + position + stress + style + syl_per_sec + age_group + (1 | speaker) + (1 | syl_label) + (1 | word_label).*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	2947.6	3184.7	-1435.8	2871.6	3757

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-4.6325	-0.5757	0.0260	0.5901	4.9983

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.01714	0.1309
syl_label	(Intercept)	0.01463	0.1209
speaker	(Intercept)	0.01245	0.1116
Residual		0.11127	0.3336

Intercept: VE1long, regEast_central; genderfemale; positionfinal; stressnuclear; stylescripted, agegroupmiddle

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	4.813e+00	123.161	5.745e-02	< 2e-16	***
VE1short	2.497e-01	1.28366	4.949e-02	5.32e-07	***
age_groupold	1.160e-01	1.12298	3.610e-02	0.001571	**
age_groupyoung	-8.768e-02	0.91605	3.526e-02	0.013962	*
positioninitial	-2.155e-01	0.80616	2.890e-02	1.11e-13	***
positionmedial	-1.691e-01	0.84443	2.189e-02	1.42e-14	***
stressprimary	-1.729e-01	0.84119	1.765e-02	< 2e-16	***
stressunstressed	-2.392e-01	0.78724	5.068e-02	2.65e-06	***
styleunscripted	-9.822e-02	0.90644	2.512e-02	0.000132	***
syl_per_sec	-5.996e-02	0.94180	5.994e-03	< 2e-16	***
VE1short:regNorth_east	1.298e-01	1.13863	6.502e-02	0.045941	*
VE1short:regWest_central	-2.774e-01	0.75774	5.555e-02	6.18e-07	***
VE1short:gendermale	-2.320e-01	0.79297	5.514e-02	2.65e-05	***
regSouth:gendermale	-1.918e-01	0.82546	8.564e-02	0.026671	*
VE1short:age_groupyoung	1.215e-01	1.12916	3.461e-02	0.000454	***
VE1short:regHighland:gendermale	2.144e-01	1.23914	8.656e-02	0.013288	*
VE1short:regInsular:gendermale	3.112e-01	1.36506	8.174e-02	0.000143	***
VE1short:regSouth:gendermale	1.747e-01	1.19086	8.435e-02	0.038439	*
VE1short:regWest_central:gendermale	2.787e-01	1.32142	7.587e-02	0.000243	***

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

The *VE1* short contexts are significant when compared to the intercept and the estimates reveal a positive effect size of 28 percent. This means that STRUT vowels in the *VE1* short contexts are significantly longer than those in *VE1* long contexts. This finding is in line with the distribution seen in Figure 33, but it completely contradicts the VE (see subsection 3.1.2): The *VE1* short environments are not shorter, but significantly longer than the *VE1* long environments.

Similar to the other vowels, STRUT is also affected by *constituent-final lengthening* (see subsection 3.1.8) and *stress* (see subsections 3.1.6 and 3.1.7). The vocalic nuclei of syllables in phrase-medial positions are roughly 16 percent and vowels in phrase-initial syllables are almost 20 percent shorter than vowels in phrase-final position (=intercept). Likewise, unstressed syllables are more than 20 percent and primary stressed syllables are almost 16 percent shorter than syllables under nuclear stress (= intercept).

Style is another variable which has a significant influence on the duration of STRUT. Overall, the short monophthong is roughly 10 percent shorter in unscripted speech forms than in scripted speech (= intercept). This finding is in line with the expectations (see subsection 4.1.2): in scripted speech, speakers often read a text and have more time to focus on the pronunciation of words. In unscripted speech, speakers have less time to plan their conversational contributions and they might therefore reduce or shorten many vowel pronunciations. The vowels in unscripted speech might therefore be shorter than the vowels in scripted speech.

Vowel duration in STRUT is also affected by the local articulation rate. For each additional syllable produced in the timeframe of 1 second, vowel duration tends to become almost 6 ms shorter. This demonstrates the effect of *tempo* (see subsection 3.1.9): STRUT becomes shorter if the speed of speech is higher.

The only extralinguistic variable which returns significant effects without interactions is the *age group* of the speakers. When compared to the intercept (= *age group middle*), the *age group old* produces 12 percent longer vowels. The effect of the *age group young* also reaches significance: the young speakers produce roughly 9 percent shorter vowels than the middle-aged speakers. Vowel duration in STRUT is therefore affected by the age group and there is an inverse trend that duration decreases when age increases.

Apart from that, there is also a two-way interaction between the *age group young* and the *VE1 short* contexts. The corresponding estimate reveals that especially younger speakers produce long vowels in *VE1 short* contexts. The difference between *VE1 long* and *short* environments is therefore particularly strong among the young speakers.

The *VE1 short* contexts are also particularly long for the male speakers as there is a significant interaction between the two variables *VE1* and *gender*. The influence of gender can also be seen in several three-way interactions. The male speakers from the *Highlands and Hebrides*, from the regions *Insular*, *Southern* and *West central* produce significantly long vowels in *VE1 short* contexts when compared to the intercept. The difference for the other regions does not reach significance across the male speakers. Apart from that, there is also a general two-way interaction between the male speakers and the region *South*. The estimate reveals that, in general, *Southern* male speakers produce shorter STRUT tokens.

Overall, the models and plots have shown that STRUT is not affected by Aitken's Law. There is no clear trend in the plots (see Figure 33) and the two categories *SVLR1* and *SVLR2* did not return significant effects in the model building. This is generally in line with the results of previous investigations, including Aitken's (1981) classification which stated that STRUT is not an SVLR-affected vowel. However, a remarkable finding of this study is that STRUT is significantly longer before voiceless consonants than before voiced consonants in spoken SSE. This pattern is stable in both fricative and plosive environments and the duration is also especially short before nasals. The durations therefore show the influence of an "anti-Voicing Effect" (Stuart-Smith et al., 2019). While McClure's

(1977, pp. 12–13) carrier sentence readings revealed that STRUT is one centisecond longer before /t/ and /s/ than before /d/ and /z/ and while (McKenna, 1988, p. 107) found /ʌ/ to be slightly shorter before /d/ than before /t/, these measurements did not represent spontaneously spoken SSE. The present findings, however, demonstrate that the effect of postvocalic voicing and manner is also reversed in the short monophthong /ʌ/ in naturally spoken SSE. Apart from that, the short monophthong STRUT is clearly affected by other sociolinguistic and prosodic factors. This contradicts many impressionistic accounts on Scottish vowel duration (see subsection 3.2.1) which claimed that STRUT is invariably short.

5.2.3 DRESS

The lexical set DRESS represents the short monophthong /ɛ/ in words such as <bed> and <get>. A statistical overview of the vocalic durations of DRESS can be found in Table 35 and a visualization can be found in Figure 37.

Table 35. Statistical summary of vowel duration for the lexical set DRESS.

Token number:	13714
Average duration:	91.76 ms
Standard deviation:	47.27 ms
Minimal duration:	10.01 ms
Maximal duration:	730.00 ms
1st quantile:	60.00 ms
Median:	90.00 ms
3rd Quantile:	112.94 ms

The 13714 DRESS tokens have an average duration of 91.76 ms and a standard deviation of 47.27 ms. DRESS is clearly the longest short monophthong, but it also has a very high standard deviation when compared to the other short vowels (see section 5.1). The range of the absolute durations is also very

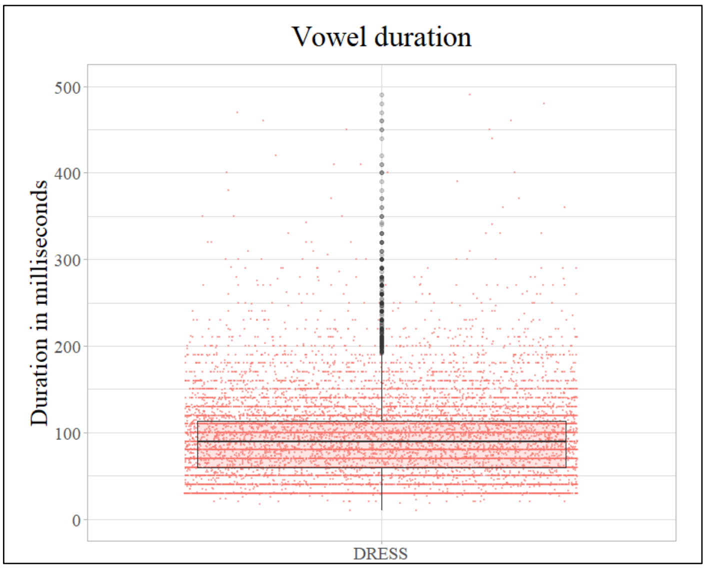


Figure 37. Vowel duration of the lexical set DRESS in milliseconds.

broad and comprises values between 10.01 (minimum) and 730.00 ms (maximum). Most of the pronunciations lie in the range between roughly 60 and 113 ms. The overall duration of DRESS is much shorter in this study than in the reference studies by McClure (1977) and McKenna (1988). The most likely reason for this difference is that this study incorporates naturally spoken language. The older reference studies elicited DRESS tokens in word list and carrier sentence readings, so the differences in speech style might lead to different vowel duration ranges.

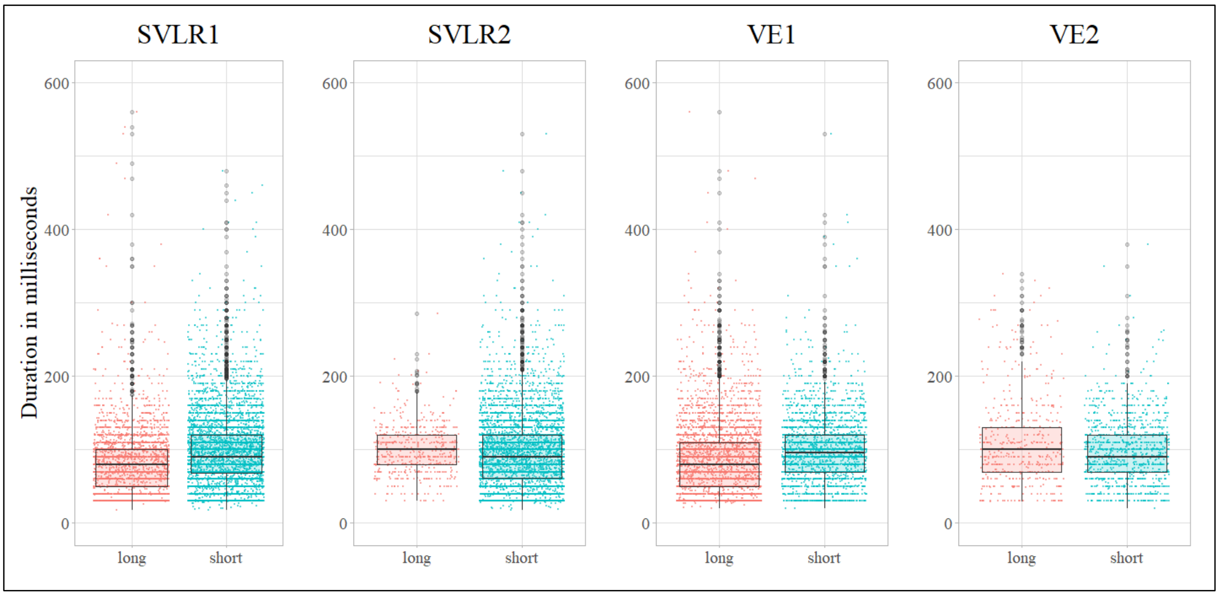


Figure 38. Boxplots and jitterplots of raw vowel duration of the lexical set DRESS for the categories SVLR1, SVLR2, VE1 and VE2.

The boxplots for the SVLR and VE categorizations (see Figure 38) show unusual distributions. As for the *SVLR1* categorization scheme, the short contexts are clearly longer than the long environments. The red boxplot for the *SVLR1* long contexts lies below the turquoise boxplot of the short environments and the average durations differ substantially as well (*SVLR1* long: 82.37 ms; *SVLR1* short: 97.60 ms). The influence of the phonological conditioning of Aitken’s Law does therefore not seem to operate in the usual way in the lexical set DRESS. The respective plots for the *SVLR2* categorization show a completely different picture. Here, the *SVLR2* long environments are generally longer than the *SVLR2* short environments. This is reflected in a higher boxplot and a longer average duration (*SVLR2* long: 100.16 ms; *SVLR2* short: 95.00 ms). The morphological conditioning of Aitken’s Law therefore seems to have a positive effect on the duration of DRESS. It is nevertheless interesting that the *SVLR2* long contexts lead to an increase in vowel duration while the *SVLR1* long environments lead to a decrease. Hence, there must be a sharp differentiation between the phonological and morphological conditioning of Aitken’s Law in the lexical set DRESS.

The categorizations for the VE show a similar distribution. As for *VE1*, the long contexts are also clearly shorter than the short contexts (mean *VE1* long: 87.60 ms; mean *VE1* short: 99.99 ms). The distribution is therefore similar to the one of the *SVLR1* categorization scheme. Nevertheless, the *VE2* long contexts are longer than the *VE2* short environments. This can be seen in the slightly higher boxplot and in the increased average duration (*VE2* long: 111.52 ms; *VE2* short: 98.13 ms). While the durational difference is relatively small, one can conclude that DRESS is slightly longer before voiced than before voiceless plosives. Yet, it is also striking that the *VE2* short contexts, which represent voiceless plosives only, are relatively long. It is therefore worth investigating the DRESS durations in different postvocalic contexts (see Table 36).

Table 36: Average vowel durations sorted for different postvocalic consonant contexts for the lexical set DRESS. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categorizations.

Postvocalic environments	Average vowel duration
laterals	85.85 ms
nasals	89.92 ms
voiced fricatives	74.70 ms
voiceless fricatives	106.72 ms
voiced plosives	106.36 ms
voiceless plosives	95.75 ms
pauses	133.99 ms

Table 36 reveals an unusual distribution: DRESS tokens with postvocalic voiceless fricatives are much longer (mean: 106.72 ms) than their equivalents in postvocalic voiced fricative contexts (mean: 74.70 ms). In fact, on average, the shortest DRESS tokens can be found in the voiced fricative environments and the second longest DRESS tokens in the voiceless fricative environments. This not only contradicts the lengthening effects of Aitken's Law but also those of the VE (see subsection 3.1.2). This might also be the reason why the *SVLR1* and *VE1* long contexts are shorter than the *SVLR1* and *VE1* short contexts. Voiced fricatives are considered to be long environments for both Aitken's Law and the VE, but the actual duration of DRESS is very short in these environments. The *SVLR2* and *VE2* categorizations exclude (monomorphemic) voiced fricatives, so the *SVLR2* and *VE2* long contexts are

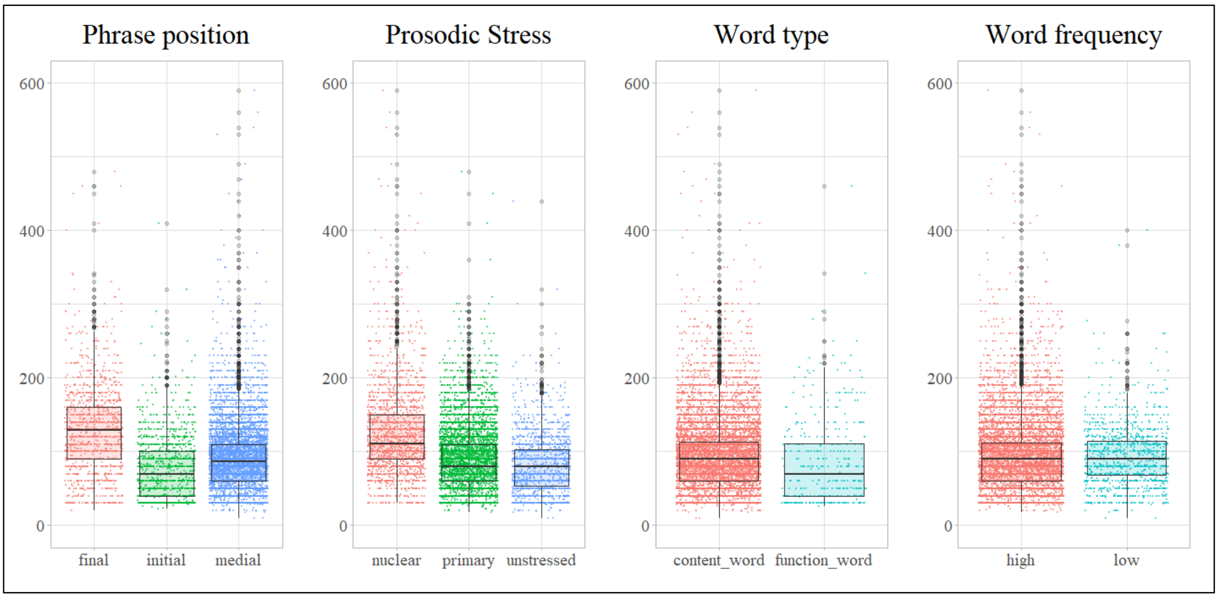


Figure 39. Boxplots and jitterplots of raw vowel duration for the lexical set DRESS separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

consequently longer than the *SVLR2* and *VE2* short contexts. The shortening effect of postvocalic voiced fricatives does not correspond with previous findings (McClure, 1977; McKenna, 1988) and vowel duration in the lexical set DRESS does therefore not operate in the usual way. Furthermore, the phonological conditioning of Aitken's Law does not seem to apply and the VE only seems to operate in plosive environments. Apart from that, Table 36 also indicates the effect of *constituent-final lengthening* (see subsection 3.1.8) due to the increased duration before pauses.

The boxplots for the categorical intralinguistic variables *phrasal position*, *stress*, *word type* and *word frequency* (see Figure 40) show the more regular patterns that can also be found across the other short monophthongs. The duration of DRESS clearly increases in phrase-final positions. Phrase-final tokens have the highest average duration (129.69 ms) followed by phrase-medial (mean: 89.13 ms) and phrase-initial tokens (mean: 77.36 ms). Interestingly, there seems to be a difference between phrase-medial and phrase-initial tokens which is also visible in the boxplots. For most other vowels, phrase-medial and phrase-initial vowels are on the same level, but for DRESS, duration is slightly longer in phrase-medial positions. The lexical set DRESS is clearly affected by *constituent-final lengthening* (see subsection 3.1.8).

The short monophthong is also affected by stress. DRESS tokens in nuclear stressed syllables are substantially longer (mean: 125.66 ms) than DRESS tokens in primary (mean: 85.66 ms) or unstressed syllables (mean: 82.54 ms). This distribution can also be seen in the boxplots. It is clear that DRESS is affected by *prosodic stress* (see subsection 3.1.7) and there are signs that also *lexical stress* (see subsection 3.1.6) has an influence, but the durational difference between lexically stressed and unstressed vowels is not very pronounced.

As for word type, it is clear to see that content words are longer than function words. The average duration is also clearly longer (content words: 91.92 ms; function words: 85.98 ms). This corresponds

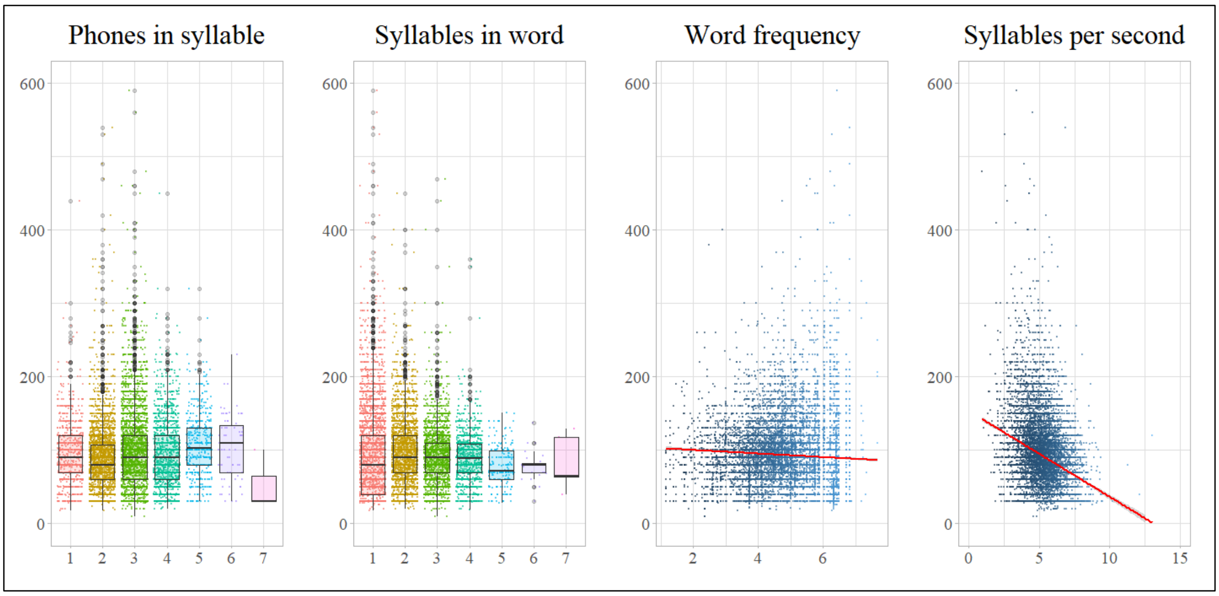


Figure 40. Plots of raw vowel duration of the lexical set DRESS separated for the linear variables number of phones in syllable, number of syllables in word, word frequency and syllables per second.

with the effect of the *lexical category* (see subsection 3.1.5), function words carry mostly grammatical information and they are therefore often reduced and shorter than content words. The difference between high and low frequency words is less pronounced. The boxplot for the low frequency words is slightly higher and also the average duration is a bit higher (low frequency words: 93.00 ms; high frequency words: 91.64 ms). Thus, whereas the overall duration is in line with the effect of *lexical category*, the effect of *lexical frequency* is not reflected in the boxplots (see subsection 3.1.5).

The plots for the linear intralinguistic variables show a mixed picture (see Figure 41). While long vowel pronunciations become less frequent in higher *syllable phone counts* and *word syllable counts*, the average duration does not steadily decrease with a higher number of phones or syllables. It is therefore unclear whether DRESS is affected by *intrasyllabic compression* (see subsection 3.1.3) or *polysyllabic shortening* (see subsection 3.1.4).

As for linear *word frequency*, the slope of the regression line is slightly negative which indicates that frequent words tend to be shorter than non-frequent words. This corresponds with the binary word frequency plot in Figure 39 and also with the general effect of *lexical frequency* on vowel duration (see subsection 3.1.5).

The effect of *tempo* (see subsection 3.1.9) is detectable as well: DRESS tokens become shorter in higher articulation rates. The red regression line has a clear negative slope which demonstrates that vowel duration becomes shorter if a speaker produces more syllables per second.

The plots for the social variables indicate that vowel duration differs across some groups (see Figure 41). As for *gender*, the boxplot for the female speakers is clearly higher than the one for the male speakers. The average duration of the female speakers is also higher (female speakers: 95.06 ms; male speakers: 88.75 ms). Vowel duration also differs for the *regional backgrounds*. One can also notice that

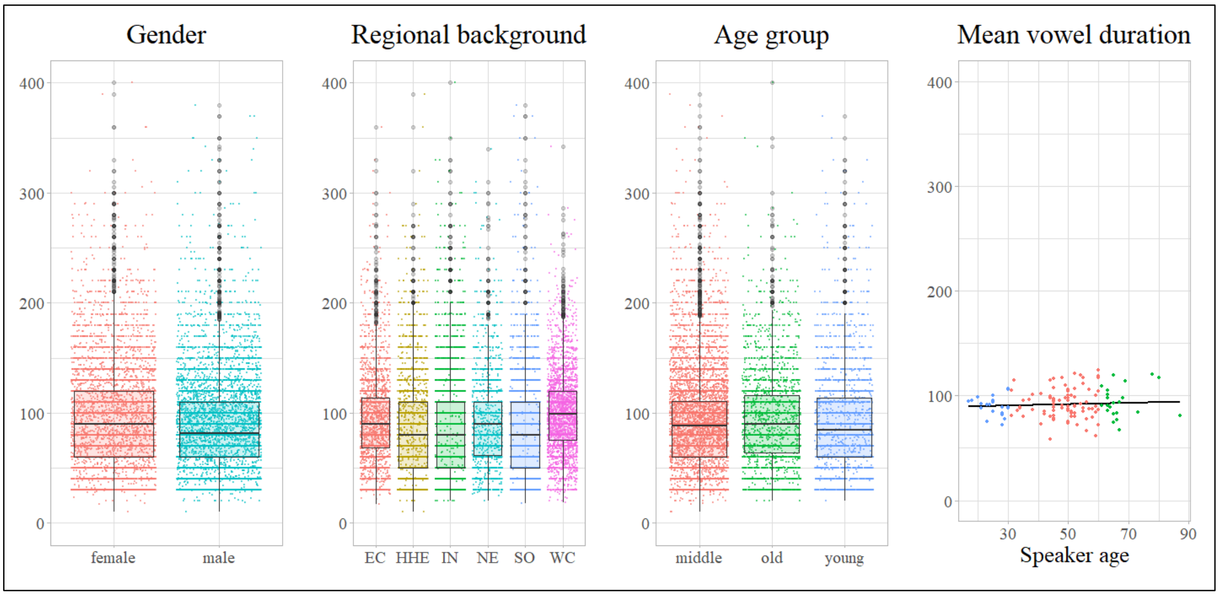


Figure 41. Plots of raw vowel duration of DRESS separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

the region *West central* has a much higher boxplot than the other regions. The DRESS tokens from this region also have the highest average duration with a value exceeding 100 ms. The boxplot for the region *East central* is also relatively high. The boxplots of the other regions are below *West central* and *East central*. In contrast to this, the durational difference between the *age groups* is not very pronounced. While the boxplot for the *age group old* is on the highest level, the other boxplots are not far apart. The average durations are also on a similar level (*age group old*: 93.80 ms; *age group middle*: 91.11 ms; *age group young*: 92.05 ms) and the slope of the regression line for the mean vowel duration comes close to zero. It is therefore unclear whether age has a significant influence on vowel duration in DRESS.

The linear mixed effects modeling for the lexical set DRESS generated models with conditional R^2 values in the range of 0.38 and 0.42. I tested for all possible interactions but there were only a few cases in which the interactions improved the accuracy of the models. In all model building processes, the fixed factor *age group* was dropped because it returned no significant effects. The models therefore corroborate the observation (see Figure 41) that the age groups have no significant influence on the duration of DRESS. Apart from that, the categorical variables *VE2* and *SVLR2* were also dropped in all models. The differences between the *SVLR2* and *VE2* levels were not very pronounced in the plots (see Figure 38) and the models confirm that the long and short contexts are not significantly different. Nevertheless, the best model was built on the *SVLR2* dataset and a summary of the model can be found in Table 37.

Significant fixed factors include the *phrasal position, stress,* the articulation rate (*syllable per second*), the *regional background* and there is one significant interaction between the *phrasal position* and *stress*. The other factors turned out to be insignificant, so there is no influence of the variables *gender, age group, style, SVLR2, syllable phone count, word syllable count*, or *word frequency*. The plots have already indicated vowel duration does not systematically vary for the *syllable phone count, word syllable count*, or *word frequency* (see Figure 40) and the models corroborated this observation. The postvocalic lengthening effect of the voiced plosives (*VE2*) is also insignificant which means that the VE does not operate in DRESS. The *VE1* short contexts were clearly longer than the long contexts and the difference between the *VE2* environments is not statistically significant. Therefore, the lengthening effect of postvocalic voiced consonants does not affect DRESS. The same can also be said for Aitken’s Law. The *SVLR2* categorization was dropped from the respective models and the *SVLR1* short contexts are also longer than the *SVLR1* long contexts. Thus, Aitken’s Law does not operate in DRESS either, especially since vowel duration is particularly short before voiced fricatives. It is also irrelevant whether DRESS is produced by male or female speakers (*gender*) or whether it occurs in scripted or unscripted speech forms (*style*).

*Table 37. Best linear mixed model for the lexical set DRESS fit by maximum likelihood (marginal R^2 = 0.21; conditional R^2 = 0.42). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ (position * stress) + position + stress + syl_per_sec + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label).*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	1590.4	1678.8	-779.2	1558.4	1830

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.9752	-0.4969	0.0453	0.5787	4.5815

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.01795	0.1340
syl_label	(Intercept)	0.00849	0.0921
speaker	(Intercept)	0.01432	0.1197
Residual		0.11584	0.3403

Intercept: positionfinal; stressnuclear; regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.502e+00	245.1780	6.689e-02	< 2e-16	***
positioninitial	-5.393e-01	0.583170	7.887e-02	1.10e-11	***
positionmedial	-2.590e-01	0.771845	4.570e-02	1.69e-08	***
stressprimary	-3.048e-01	0.737241	5.473e-02	2.95e-08	***
syl_per_sec	-9.351e-02	0.910725	9.251e-03	< 2e-16	***
regHighland	6.345e-03	1.006364	5.033e-02	0.89989	
regInsular	2.866e-02	1.029071	5.336e-02	0.59235	
regNorth_east	1.112e-02	1.011186	5.183e-02	0.83042	
regSouth	-2.088e-01	0.811595	5.104e-02	8.24e-05	***
regWest_central	3.194e-02	1.032455	4.569e-02	0.48600	
positioninitial:stressprimary	2.487e-01	1.282390	9.377e-02	0.00806	**
positionmedial:stressprimary	6.000e-02	1.061840	5.966e-02	0.31465	

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

DRESS is, however, affected by *stress*. Tokens in primary stressed syllables are more than 26 percent shorter than tokens in nuclear stressed syllables (= intercept). DRESS vowels in unstressed positions are not listed in the model output because they were filtered out for the *VE2* categorization scheme (see subsection 4.4.2). The vowel is therefore clearly affected by *prosodic stress* (see subsection 3.1.7), but the influence of *lexical stress* (see subsection 3.1.6) remains unclear.

Similarly, the short monophthong is also affected by the *phrasal position*. Vowels in phrase-initial position are 41.69 percent and vowels in phrase-medial syllables are 22.82 percent shorter than DRESS tokens in phrase-final positions. Hence, *constituent-final lengthening* (see subsection 3.1.8) is also prevalent in the short monophthong DRESS. It is also striking that there seems to be a difference between phrase-initial and phrase-medial positions because the effect sizes of the estimates are very different. Furthermore, the *phrasal position* also interacts with *stress*. The interaction between primary stressed syllables and phrase-initial positions is significant and the effect size is positive. This means that, when compared to the intercept, vowel duration significantly increases in primary stressed syllables in phrase-initial positions.

Like the other vowels, DRESS is also affected by *tempo* (see subsection 3.1.9). For every additional syllable produced in the timeframe of 1 second, the duration of DRESS tends to become roughly 9 percent shorter. The duration of DRESS therefore varies depending on the local articulation rate.

Apart from that, the duration of DRESS is also affected by the *regional background* of the speakers. Specifically, speakers from the dialect region *Southern* produce significantly shorter DRESS vowels than speakers from the other regions. Only the region *Southern* returns significant p-values and the estimates reveal that the duration tends to be roughly 19 percent shorter than the DRESS tokens of the speakers from the region *East central* (= intercept). While the plots indicated that vowel duration is higher for the regions *East central* and *West central* (see Figure 41), the significantly shorter duration for the region *Southern* only became apparent via the linear mixed effects modeling.

Overall, the models and plots have shown that DRESS is neither affected by the phonological nor by the morphological conditioning of Aitken's Law. Unlike other vowels, DRESS is clearly shortened before voiced fricatives and this contradicts many previous observations (House & Fairbanks, 1953; McClure, 1977; McKenna, 1988). The short monophthong is also not affected by the VE. DRESS is generally shorter in SVLR and VE long contexts than in SVLR and VE short contexts. Similar to the vowel STRUT, the findings do therefore suggest the influence of an "anti-Voicing Effect" (Stuart-Smith et al., 2019). DRESS is clearly shorter before the voiced fricatives than before the voiceless fricatives in connected speech. It seems that the duration of DRESS behaves differently in naturally-occurring language than in a fixed experimental setting (House & Fairbanks, 1953; McClure, 1977; McKenna, 1988).

5.3 Long Monophthongs

This section comprises the findings for the long SSE monophthongs GOOSE, FLEECE, THOUGHT, FACE, GOAT and CAT. Whereas studies on vowel formats usually exclude tokens below 50 ms (Fruehwald, 2013, p. 117; Tanner et al., 2020, p. 5), the present study will only exclude tokens below 40 ms to avoid reduced realizations. This study investigates naturally occurring language and very short realizations may therefore occur frequently. In addition, the present study investigates vowel duration so the exclusion of many short tokens could possibly hide specific durational distributions. I want to provide a full account of vowel duration in 21st century SSE, so I want to include as many tokens as possible. Yet, it is very likely that tokens shorter than 40 ms are reduced schwas and they should therefore be excluded from the analysis.

5.3.1 GOOSE

The lexical set GOOSE represents the high back vowel /u/. This monophthong is often fronted in Scottish English and therefore transcribed with the symbol /ʉ/ (Wells, 1982, p. 402). The vowel formant measures in this study also found a fronted vowel quality (see section 5.1). The Basic Scottish Vowel System (Abercrombie, 1979) (see section 2.3) does not distinguish between the lexical sets FOOT and GOOSE. SSE only incorporates one vowel phoneme which represents words such as <school> and <good>. I will therefore adopt Abercrombie's (1979) categorization for the present study. An overview of the duration of GOOSE is listed in Table 38 and visualized in Figure 42.

Table 38. Statistical summary of vowel duration for the lexical set GOOSE. Tokens below 40 ms were excluded.

Token number:	6351
Average duration:	92.33 ms
Standard deviation:	55.55 ms
Minimal duration:	40.00 ms
Maximal duration:	830.00 ms
1st quantile:	57.24 ms
Median:	77.49 ms
3rd Quantile:	110.00 ms

The 6351 GOOSE tokens have an average duration of 92.33 ms and a standard deviation of 55.55 ms. It must be added, however, that many tokens (N=2555) were shorter than the 40 ms threshold and they were therefore excluded from the analysis. The longest GOOSE token is 830 ms long and the shortest tokens lie at the 40 ms threshold. The median is 77.49 ms and most of the data is in the range between 57.24 ms and 110.00 ms. The overall duration of GOOSE is shorter than the measurements by McClure (1977) and McKenna (1988). However, this is not surprising because the present study investigated naturally spoken language and not read speech elicited in an experimental setting.

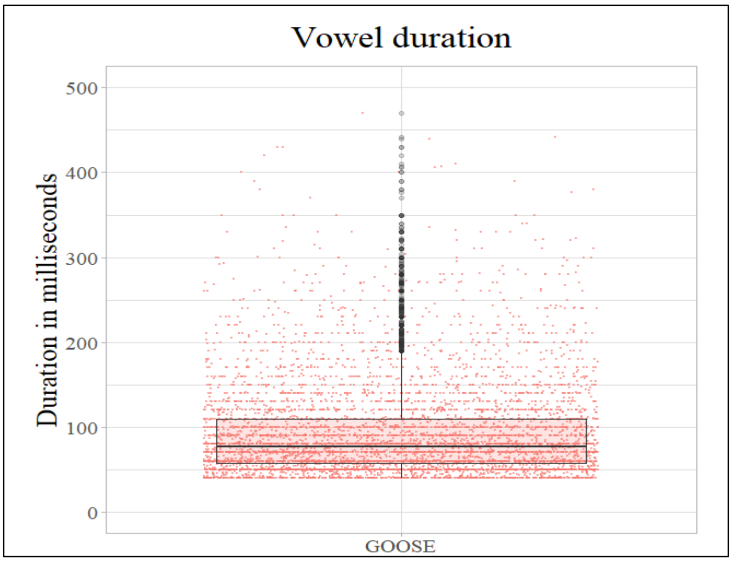


Figure 42. Vowel duration of the lexical set GOOSE in milliseconds. Durations below 40 ms were excluded.

The boxplots for the categorizations *SVLR1*, *SVLR2*, *VE1* and *VE2* indicate that Aitken's Law operates in GOOSE (see Figure 43). The boxplots for the *SVLR1* and *SVLR2* long contexts are clearly higher than the boxplots of the respective short environments. The average durations are also longer in the long contexts (*SVLR1* long: 103.08 ms; *SVLR1* short: 79.15 ms; *SVLR2* long 88.05 ms; *SVLR2* short 80.16 ms). As for the VE, the *VE1* long contexts are slightly longer than the *VE1* short contexts.

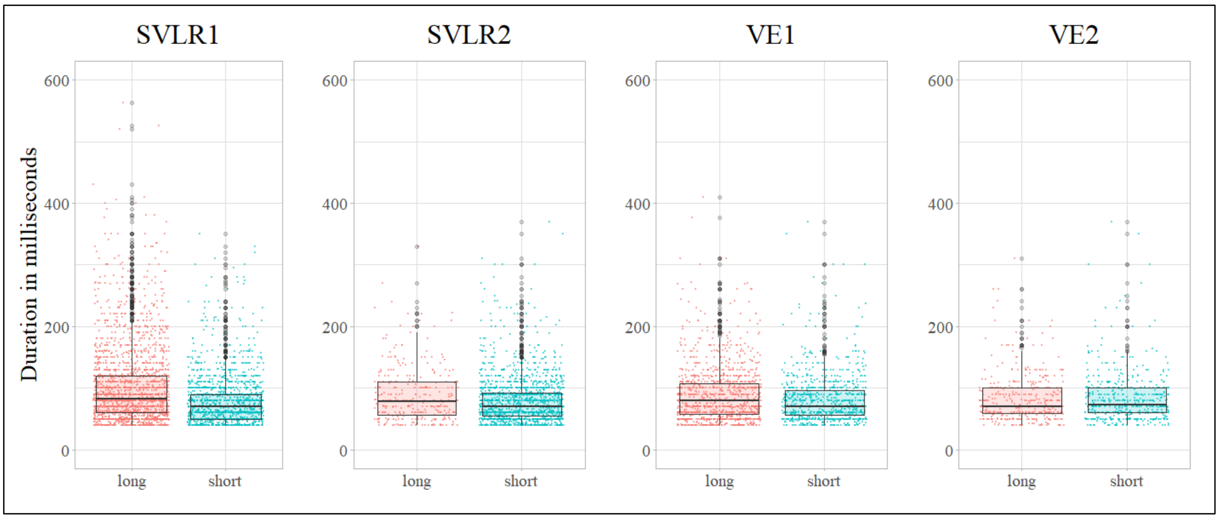


Figure 43. Boxplots and jitterplots of raw vowel duration of the lexical set GOOSE for the categories SVLR1, SVLR2, VE1 and VE2.

However, the difference between the two categories is clearly weaker than the difference for the SVLR long and short environments. The durational difference between the averages of *VE1* long (mean: 88.27 ms) and *VE1* short (mean: 82.01 ms) amounts to only 6.26 ms. The *VE2* boxplots show no durational difference: both boxplots are on the same level. The average durations are roughly identical and interestingly, the *VE2* short contexts (mean: 83.80 ms) are even a bit longer than the *VE2* long contexts (mean: 83.59 ms). This means that GOOSE is not longer in voiced plosive contexts than in voiceless plosive contexts and this clearly contradicts the VE. The average durations in the different postvocalic consonantal contexts show a similar picture (see Table 39).

Table 39. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set GOOSE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categorizations.

Postvocalic environments	**Average vowel duration**
laterals	88.24 ms
nasals	77.76 ms
voiced fricatives	98.50 ms
voiceless fricatives	82.82 ms
voiced plosives	78.33 ms
voiceless plosives	79.87 ms
pauses	173.42 ms

Overall, voiceless plosives lead to longer GOOSE tokens than voiced plosives. The vowels followed by nasals are also remarkably short and have a lower average duration than the plosive contexts. This clearly contradicts the VE and the lengthening effect of the manner of the consonantal context (see subsection 3.1.2). The difference between the long and short fricative contexts is much more pronounced (15.68 ms) and thus indicates the phonological conditioning of Aitken’s Law. The vowels followed by laterals are shorter than those in voiced fricative contexts but longer than the GOOSE tokens in most other environments. The increased average vowel duration before pauses clearly shows the effect of *constituent-final lengthening* (see subsection 3.1.8). The plots and average durations therefore indicate that GOOSE is affected by Aitken's Law but not by the VE.

I further plotted the SVLR contexts for the different intra- and extralinguistic variables to see whether the effect of Aitken's Law varies in specific groups or contexts. The plots indicated that the SVLR is stable across the different *age groups* and *genders*. *SVLR1* and *SVLR2* long contexts were consistently longer than the respective short contexts for the male and female speakers as well as for the age groups *old*, *middle* and *young*. The SVLR plots for the *regional background*, the *phrasal position* and *stress* show more variation (see Figure 44).

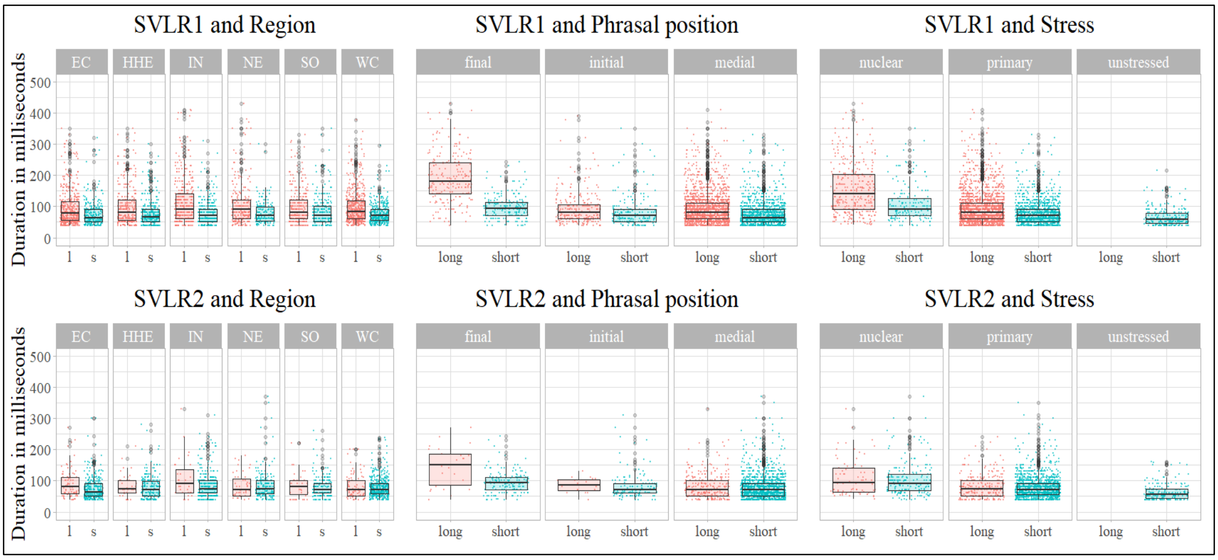


Figure 44. Boxplots and jitterplots of raw vowel duration of the lexical set GOOSE for the categories SVLR1, SVLR2 separated for the variables regional background, phrasal position and stress.

While the lengthening effect in the *SVLR1* categorization is relatively stable across the different regions, the difference for the *SVLR2* contexts is much weaker overall. The plots therefore indicate that GOOSE is affected by the phonological conditioning of Aitken's Law, but it is not entirely clear whether the vowel is significantly affected by the morphological conditioning of the SVLR. At the same time, the durational difference between the *SVLR2* long and short environments is more pronounced in the dialect region *Insular* than in the other regions. The effect is particularly weak in the *Northeast* because the boxplots for the *SVLR2* long and short environments are on a very similar level in that region. In addition, the *Northeast* is the only region in which the average value of the *SVLR2* short environments (85.39 ms) exceeds the mean duration of the *SVLR2* long environments (84.00 ms). It is therefore unlikely that the morphological conditioning of Aitken's Law applies in the Northeast of Scotland.

The plots for Aitken's Law in different *phrasal positions* show striking differences. Not only are the tokens generally longer in phrase-final positions, but the plots also show that the difference between the long and short contexts is much more pronounced phrase-finally than in the other phrasal positions. The overall difference between the *SVLR1* long and short contexts is 99.68 ms in phrase-final tokens but only 16.67 ms in phrase-medial and 9.71 ms in phrase-initial positions. Hence, it seems that the effects of Aitken's Law are amplified in prepausal positions which corresponds to findings in earlier studies (Chevalier, 2019; Rathcke & Stuart-Smith, 2016).

The situation is partly comparable for the variable *stress*. The difference between the *SVLR1* long and short contexts is clearly stronger in nuclear stressed syllables than in primary stressed syllables. This would indicate that SVLR effects are stronger in prominent prosodic positions. However, the durational difference in the *SVLR2* contexts is not much greater in nuclear stressed GOOSE tokens than in primary stressed GOOSE vowels. The plots show that nuclear stressed tokens are generally longer than vowels in primary stressed or unstressed syllables but the difference between long and short contexts is only more pronounced for the SVLR1 categorization. Unstressed syllables are generally SVLR short environments, so the plots do therefore not show any SVLR long tokens in unstressed positions.

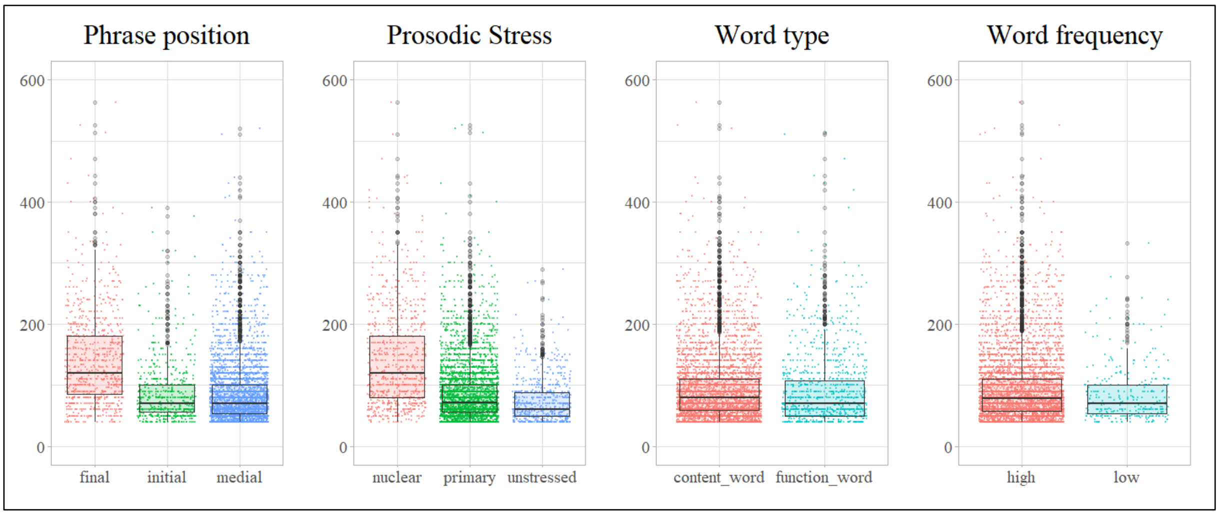


Figure 45. Boxplots and jitterplots of raw vowel duration for the lexical set GOOSE separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

The plots for the categorical intralinguistic variables clearly show the influence of *constituent-final lengthening* (see subsection 3.1.8) and *prosodic stress* (see subsection 3.1.7). As observed in Figure 45, GOOSE vowels are clearly longer in phrase-final positions than in phrase-medial or phrase-initial positions. Likewise, nuclear stressed syllables carry much longer GOOSE tokens than primary stressed or unstressed syllables.

The difference between the levels of the *lexical category* and *frequency* is not as strong. The boxplot for the content words is slightly higher than the boxplot representing the function words. This corresponds to the effect of *lexical category* (see subsection 3.1.5) because function words are often reduced and therefore shorter in connected speech. In contrast to this, the boxplot for the high frequency words is slightly higher than the one for the low frequency words. This is contradictious to the effect of *lexical frequency* because frequent words tend to be shorter than infrequent words (see subsection 3.1.5). The plots do not provide a clear indication of whether GOOSE is affected by *lexical category* and *frequency* (see subsection 3.1.5).

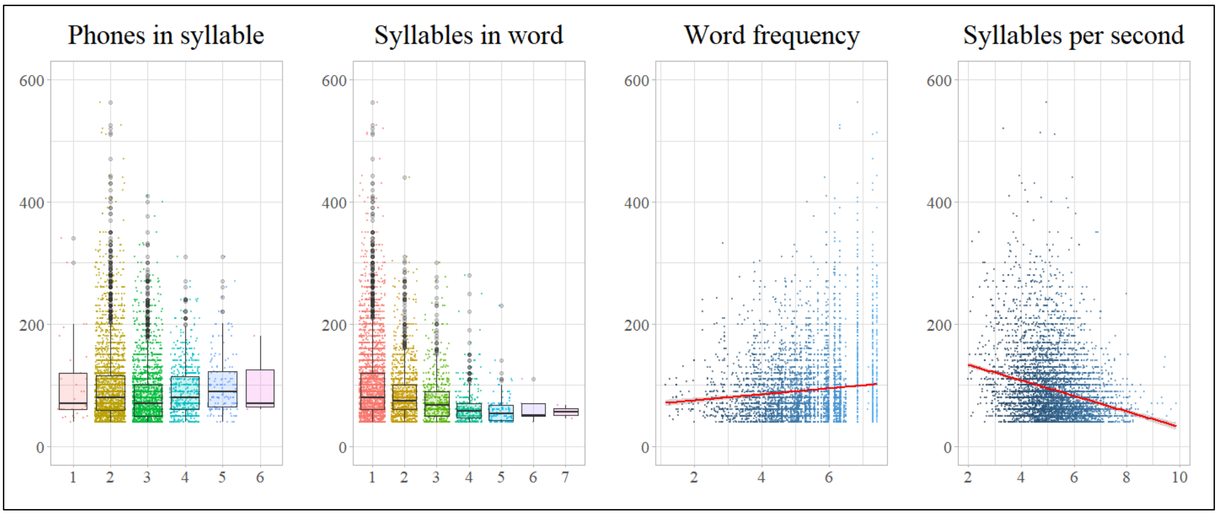


Figure 46. Plots of raw vowel duration of the lexical set GOOSE separated for the linear variables number of phones in syllable, number of syllables in word, word frequency and syllables per second.

The plots for the linear intralinguistic variables (see Figure 46) show that the local articulation rate has a strong influence on the duration of GOOSE. Similar to the other vowels, the red regression line has a clear negative slope which indicates that GOOSE tokens become shorter in high articulation rates and longer in low articulation rates. Hence, the duration of the monophthong is influenced by *tempo* (see subsection 3.1.9).

The plot also shows an influence of *lexical frequency*. In contrast to most other vowels, the regression line in the word frequency plot has a positive slope which indicates that vowel duration increases with a higher word frequency. This trend corresponds with the categorical plot on word frequency (see Figure 45), but it contradicts previous findings on the relationship between vowel duration and *lexical frequency* (see subsection 3.1.5).

As for *intrasyllabic compression*, the corresponding boxplots representing the number of phones in a syllable do not follow a general trend. While long GOOSE tokens become less frequent in syllables with many phones, the average duration of the GOOSE tokens does not steadily decrease with a higher *syllable phone count*. Rather, the boxplots and the average values rise and fall irregularly. It is therefore unclear whether GOOSE is significantly conditioned by the effect of *intrasyllabic compression* (see subsection 3.1.3).

The boxplots for the *word syllable count* show a more regular distribution. Here, the duration of GOOSE generally decreases with a higher *word syllable count*. The only exceptions are the tokens in words with six and seven syllables. The average duration of GOOSE vowels in words with six syllables (64.09 ms) is higher than the equivalent mean duration of GOOSE in words with seven syllables (56.63 ms). However, the GOOSE dataset comprises only five words with six syllables and only two tokens with seven syllables (e.g. <homosexuality>), so the unusual distribution could be a result of the very low token numbers. Despite this irregularity, the plots indicate that GOOSE is affected by *polysyllabic shortening* (see subsection 3.1.4).

I also plotted the duration of GOOSE for the extralinguistic variables *gender*, *regional background* and *age group* and I also plotted the mean vowel duration by the speakers' age. However, the plots did not show any specific distribution and this is why they are not shown. The duration of GOOSE was similar for the different regions and age groups and the mean vowel duration did not increase with a higher age of the speakers. As for *gender*, the boxplot for the male speakers is slightly higher than the boxplot for the female speakers but the respective average durations do not show a great difference (mean female speakers: 91.42 ms; mean male speakers: 93.12 ms).

While the plots and average durations provide a detailed overview of the duration of GOOSE in different contexts and groups, the linear mixed effects models give evidence of whether the durational differences reach statistical significance. The conditional R^2 values for the models lie in the range between 0.37 and 0.40. This means that the models can explain up to 40 percent of the variation in the dataset. In all models, the fixed factors *age group*, *region*, *style* and *word frequency* were excluded. This means that the duration of GOOSE is not significantly affected by the *regional background* of the speakers, nor by their *age group*. The duration of GOOSE is not significantly different in scripted or unscripted speech and the word frequency has no significant effect either. In contrast to this, the *VE1* categorization returned significant effects: GOOSE tokens are roughly 15 percent shorter in *VE1* short contexts than in *VE1* long contexts. In addition, there is a significant interaction between *VE1* short contexts and primary stressed syllables. When compared to the intercept, the *VE1* short contexts are roughly 17 percent longer which means that the durational difference between the *VE1* long and short environments is greater in nuclear stressed syllables than in primary stressed syllables. This indicates the operation of the VE and an interaction between the VE and *prosodic stress*. However, the respective models for the *VE2* sub-dataset excluded the *VE2* categories because the difference between the *VE2* long and short contexts did not reach statistical significance. This means that there is no significant durational difference between GOOSE tokens in voiced and voiceless plosive contexts. The models therefore corroborate the observation that the boxplots for the *VE2* long and short environments in Figure 43 are on the same level. Consequently, it is very unlikely that the VE operates in GOOSE because the *VE2* categorization scheme was explicitly designed to have a clear separation between the phonological conditioning of Aitken's Law and the VE (see subsection 4.4.2). The plosive contexts of the *VE2* represent stable phonetic contexts with the same manner of articulation and the absence of lengthening in the voiced environments can be assigned to the operation of Aitken's Law. In addition, the models for Aitken's Law included the *SVLR1* and *SVLR2* categories as significant predictor variables. The best model was fitted with the *SVLR1* categorization scheme and a summary of the model output can be found in Table 40.

*Table 40. Best linear mixed model for the lexical set GOOSE fit by maximum likelihood (marginal R^2 = 0.26; conditional R^2 = 0.40). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ SVLR1 + (SVLR1 * position) + position + stress + num_syl_word + num_pho_syl + syl_per_sec + gender + (1 | speaker) + (1 | syl_label) + (1 | word_label).*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	3495.4	3595.3	-1731.7	3463.4	3798

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.9535	-0.6572	-0.0618	0.5663	4.1788

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.01412	0.11881
syl_label	(Intercept)	0.01037	0.10185
speaker	(Intercept)	0.00626	0.07912
Residual		0.13327	0.36506

Intercept: SVLR1long; positionfinal; stressnuclear; genderfemale

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.941e+00	380.1615269	8.224e-02	< 2e-16	***
SVLR1short	-4.789e-01	0.6194558	4.845e-02	< 2e-16	***
positioninitial	-5.792e-01	0.5603190	3.647e-02	< 2e-16	***
positionmedial	-5.210e-01	0.5939037	2.849e-02	< 2e-16	***
stressprimary	-3.299e-01	0.7189825	1.902e-02	< 2e-16	***
stressunstressed	-3.491e-01	0.7053566	4.385e-02	5.08e-15	***
num_syl_word	-5.938e-02	0.9423505	1.348e-02	1.56e-05	***
num_pho_syl	-5.681e-02	0.9447701	2.051e-02	0.00614	**
syl_per_sec	-9.785e-02	0.9067799	6.455e-03	< 2e-16	***
gendermale	4.707e-02	1.0481995	1.966e-02	0.01833	*
SVLR1short:positioninitial	4.758e-01	1.6092451	5.603e-02	< 2e-16	***
SVLR1short:positionmedial	3.867e-01	1.4720456	4.337e-02	< 2e-16	***

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

Significant fixed factors are the *SVLR1* context, the *phrasal position*, *stress*, the *word syllable count*, the *syllable phone count*, the local articulation rate in terms of *syllables per second* and the *gender* of the speakers. The model also includes significant interactions between the *SVLR1* short contexts and the phrasal positions.

The model estimates show that GOOSE tokens are roughly 38 percent shorter in *SVLR1* short contexts than in *SVLR1* long environments. This clearly indicates that the long monophthong is affected by the phonological conditioning of Aitken's Law. The interactions with the *phrasal position* further corroborate the observation that the effect size of Aitken's Law is stronger in phrase-final syllables than in non-final syllables. The interactions show that, when compared to the intercept, *SVLR1* short contexts are clearly longer in phrase-initial and phrase-medial syllables. This means that the durational difference between *SVLR1* long and short tokens is less pronounced in these non-final contexts. The same observation could be made in Figure 44: the difference between the *SVLR1* long and short contexts is much greater in final syllables.

The model output further specifies the influence of *constituent-final lengthening* (see subsection 3.1.8). GOOSE tokens are roughly 43 percent shorter in phrase-initial and approximately 40 shorter in phrase-medial syllables than in phrase-final positions. This trend was already visible in the plots (see Figure 45) and the model confirms the significant influence of this variable.

A similar trend can be observed for the variable *stress*: when compared to the intercept, the vowels are almost 30 percent shorter in unstressed syllables and roughly 28 percent shorter in primary stressed syllables than in nuclear stressed syllables. Hence, the model output shows that GOOSE is affected by *prosodic stress* (see subsection 3.1.6). It is, however, unclear whether GOOSE is also significantly affected by *lexical stress* because the effect sizes for the unstressed and primary stressed syllables are very similar overall.

The model output also shows that the duration of GOOSE is affected by *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4). For each additional phone in a syllable and for each additional syllable in a word, the duration of GOOSE decreases by roughly 6 percent. While a clear *intrasyllabic compression* effect could not be detected in the plot (see Figure 46), the model output testifies that the *syllable phone count* does have a significant influence on the duration of the vowel.

The model further specifies that the local articulation rate has a significant influence on the duration of GOOSE. The estimates reveal that, for every additional syllable produced in the timeframe of one second, the monophthong is shortened by roughly 9 percent. Hence, GOOSE is shorter in fast speech and longer in slow speech and this corresponds to the effect of *tempo* (see subsection 3.1.9).

The only extralinguistic factor which reaches statistical significance is the *gender* of the speakers. The male speakers produce roughly 4 percent longer GOOSE tokens than the female speakers. This trend was not observable in the plots and the average durations of GOOSE by male and female speakers do not strikingly differ. The variable *gender* was also excluded in all the other models with the *SVLR2*, *VE1* and *VE2* variables. Consequently, male speakers produce longer GOOSE tokens, but the difference is marginal and does not correlate with any other variable in the dataset.

Overall, the plots and models show that GOOSE is affected by Aitken's Law. The durational difference between SVLR long and short contexts reaches statistical significance and this difference is also visible in the plots. The effects of the SVLR are overall stronger in phrase-final syllables than in non-final positions. While male speakers tend to produce slightly longer GOOSE tokens overall, there is no interaction between Aitken's Law and other extralinguistic predictor variables. This means that the SVLR is stable in GOOSE across the different regions and age groups in the dataset. The findings for GOOSE are therefore in line with most previous research. In the impressionistic accounts on Scottish vowel duration (see subsection 3.2.1), GOOSE has been described as a vowel that is affected by Aitken's

Law (Abercrombie, 1979; Aitken, 1981; Dieth, 1932; Grant, 1931; Grant & Dixon, 1921; G. Watson, 1923; Wettstein, 1942; Wölck, 1965; Zai, 1942). The subsequent empirical studies could find evidence that Aitken's Law operates in the high vowel (Chevalier, 2019; Hewlett et al., 1999; McClure, 1977; McKenna, 1988; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019; Watt & Ingham, 2000). Only the studies by Watt and Yurkova (2007), Llamas et al. (2011) and Warren (2018) reported that GOOSE is rather affected by the VE in the Northeast of Scotland and in Tyneside. Yet, the present study could also find clear evidence of SVLR durational patterns in GOOSE. The *VE2* categorization further reveals that following voiced plosives do not lead to significantly longer GOOSE articulations than following voiceless plosive contexts. The average duration of GOOSE in voiceless plosive contexts is even slightly higher than the average duration of the vowels in the respective voiced plosive environments. The shortest durations of GOOSE are produced when the vowel is followed by nasals. The findings do therefore report a stable SVLR pattern but an unstable VE which mostly corroborates earlier research.

5.3.2 FLEECE

The lexical set FLEECE represents the long high monophthong /i:/ in words such as <these> and <need>. A statistical overview of the vocalic durations of FLEECE can be found in Table 41 and a visualization can be found in Figure 47.

Table 41. Statistical summary of vowel duration for the lexical set FLEECE. Tokens below 40 ms were excluded.

Token number:	9146
Average duration:	100.42 ms
Standard deviation:	53.91 ms
Minimal duration:	40.00 ms
Maximal duration:	540.00 ms
1st quantile:	63.32 ms
Median:	88.30 ms
3rd Quantile:	120.00 ms

FLEECE has an average duration of 100.42 ms and a standard deviation of 53.91 ms. FLEECE is therefore clearly longer than the short high monophthong KIT (see subsection 5.2.1). The token number is 9146 and most of the pronunciations lie in the range between 63.32 ms and 120 ms. Due to the filtering process, the shortest FLEECE realizations are 40 ms long and the longest token lasts 540 ms. Similar to the other vowels, the duration of FLEECE in this study is shorter than the FLEECE durations in McClure (1977) and McKenna (1988). The most likely explanation for this difference is that this investigation studies naturally occurring language only and speakers have less time to pronounce words in spontaneous conversations than in word list or carrier sentence readings.

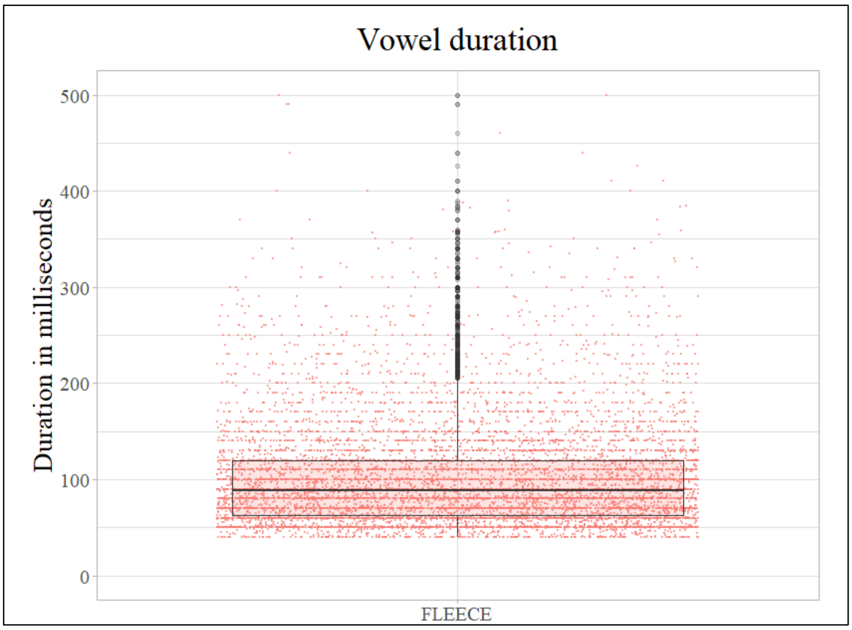


Figure 47. Vowel duration of the lexical set FLEECE in milliseconds. Durations below 40 ms were excluded.

The plots representative for the VE and SVLR categories show that long contexts are always longer than the short contexts (see Figure 48). There is great variation overall, but the red boxplots of the long environments are always higher than the turquoise boxplots of the short environments. The difference between the *SVLR2* long and short contexts is particularly pronounced. There are relatively few *SVLR2* long tokens, but they are much longer on average than the *SVLR2* short tokens (*SVLR2* long: 115.40 ms; *SVLR2* short: 83.96 ms). The average durations of the *SVLR1* environments also differ greatly (*SVLR1* long: 105.31 ms; *SVLR1* short: 87.10 ms). The same is also true for the VE categorizations but the durational difference between the long and short contexts is not as widespread as the ones for the SVLR

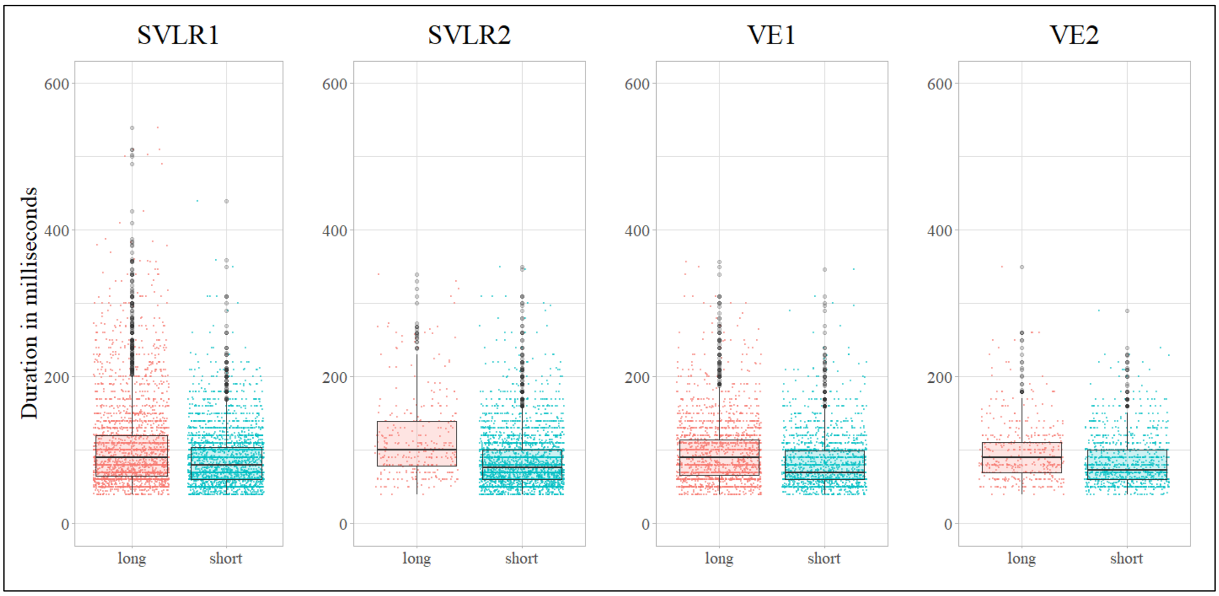


Figure 48. Boxplots and jitterplots of raw vowel duration of the lexical set FLEECE for the categories SVLR1, SVLR2, VE1 and VE2.

categorizations (*VE1* long: 96.05 ms; *VE1* short: 82.33 ms; *VE2* long: 96.76 ms; *VE2* short: 82.93 ms). The trend of postvocalic consonantal lengthening and shortening also becomes apparent when checking the average durations for the different postvocalic contexts (see Table 42).

Table 42. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set FLEECE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categorizations.

Postvocalic environments	Average vowel duration
laterals	94.81 ms
nasals	83.00 ms
voiced fricatives	115.96 ms
voiceless fricatives	85.79 ms
voiced plosives	94.86 ms
voiceless plosives	83.49 ms
pauses	182.49 ms

As seen in Table 42, the voiced plosive contexts are clearly longer than the voiceless plosive environments which corresponds to the effects of the VE. According to Aitken's Law, vowels in these contexts should have comparable durations. Nevertheless, vowels in voiced fricative contexts are clearly longer than in voiceless fricative environments. While this distribution fits to both Aitken's Law and the VE, the durational difference between the voiced and voiceless fricative contexts (30.17 ms) is much greater than the durational difference between the voiced and voiceless plosive environments (11.37 ms). This could be a sign that Aitken's Law operates alongside the VE in FLEECE (McMahon, 1991). The voiceless fricatives are also almost as short as the voiceless plosive contexts and the vowel durations

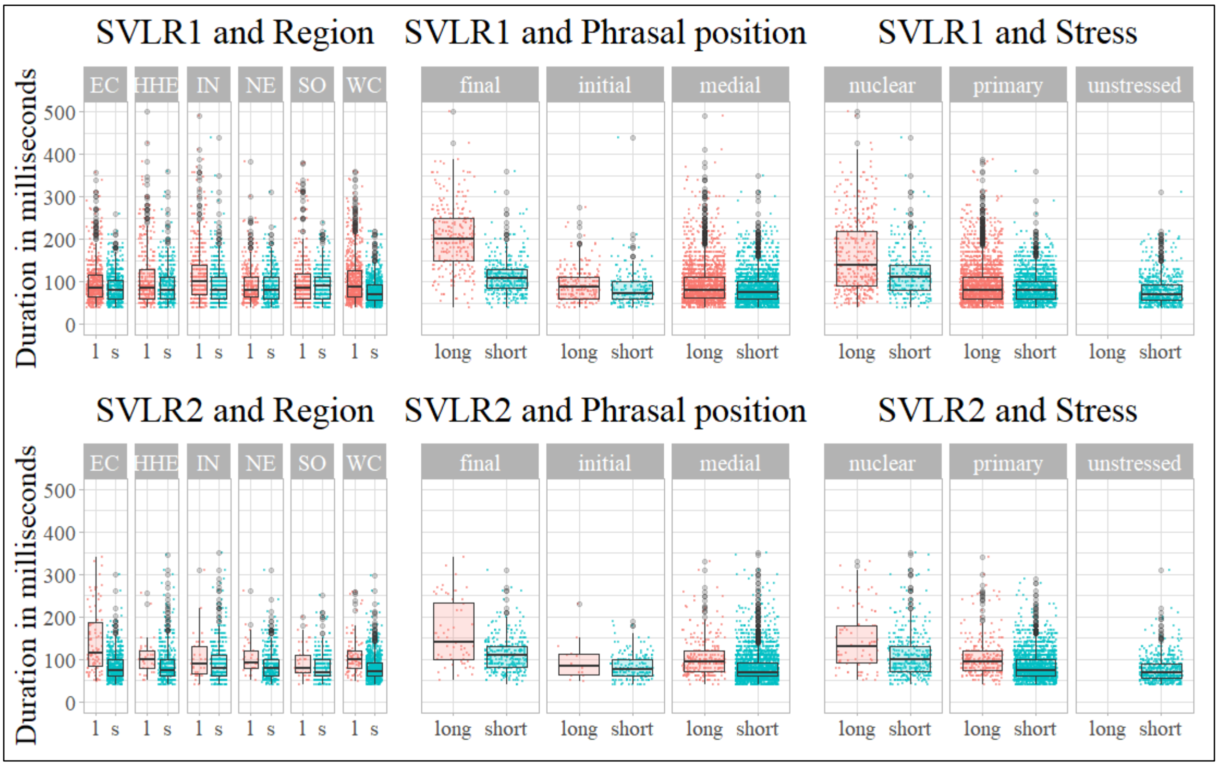


Figure 49: Boxplots and jitterplots of raw vowel duration of the lexical set FLEECE for the categories SVLR1, SVLR2 separated for the variables regional background, phrasal position and stress.

before nasals are very short as well. Hence, postvocalic voiceless fricative, voiceless plosive and nasal environments lead to shorter vowel durations which corresponds to the effects of the SVLR. As this long monophthong has often been categorized as an SVLR-affected vowel, I further plotted the SVLR categories for different extra- and intralinguistic variables to see whether Aitken's Law varies for specific groups and contexts. The plots for the *age groups* and *genders* did not reveal specific distributions; the proportions of the *SVLR1* and *SVLR2* categories are similar for the different *age groups* and *genders*. Similar to the proportions in GOOSE (see subsection 5.3.1), more interesting distributions could be found for the variables *regional background*, *phrasal position* and *stress* (see Figure 49).

As seen in Figure 49, the effect size of Aitken's Law varies in different regions. Whereas there is a stark difference between the *SVLR1* long and short contexts in the regions *East central*, *Insular* and *West central*, the durational difference is much weaker for the regions *HHE* and especially *Northeast* and *Southern*. The respective average values reveal that the *SVLR1* long contexts are generally longer than the *SVLR1* short contexts across all regions, but the durational differences vary greatly. For example, the durational difference between the *SVLR1* long and short context is the highest for the region *West central* (26.58 ms) and the smallest for the *Northeast* of Scotland (6.69 ms). This distribution corresponds to earlier findings: Warren (2018) as well as Watt and Yurkova (2007) found out that SVLR effects are very weak in the Northeast, but most other studies in the Western central belt found robust effects of Aitken's Law in FLEECE (Chevalier, 2019; McClure, 1977; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, Hewlett, & Turk, 1999). The morphological conditioning of Aitken's Law (*SVLR2*) appears to be more regular across the different regions. The *SVLR2* long contexts are longer than the *SVLR2* environments across all dialect backgrounds. Only the region *East central* stands out because the boxplot for the *SVLR2* long contexts is much higher and larger than the respective boxplot for the *SVLR2* short contexts. This tendency is also reflected in the average values: The *SVLR2* long contexts of the Eastern central belt are almost 60 ms longer than the respective *SVLR2* short contexts. It is therefore possible that there are significant interactions between Aitken's Law and the regional background of the speakers.

The plots further indicate interactions between Aitken's Law and the *phrasal position*. While *SVLR1* and *SVLR2* long contexts are generally longer than the respective short environments in all positions, the durational difference between the long and short environments is much greater in phrase-final positions. The boxplots for the tokens in phrase-final positions are generally higher, but the difference between the short and long contexts of Aitken's Law is clearly more pronounced in *phrase-final* positions than in *phrase-medial* or *phrase-intial* positions. This is also reflected in the average values: the mean difference between the *SVLR1* long and short contexts is 95.80 ms in *phrase-final* positions, but only 11.04 ms in *phrase-medial* and 8.60 ms in *phrase-initial* positions. Likewise, the average *SVLR2* difference in *phrase-final* position amounts to 49.90 ms but only to 24.55 ms in *phrase-medial* and to 15.67 ms in *phrase-initial* positions.

A similar distribution can also be found for the intralinguistic variable *stress*. The difference between the boxplots of the SVLR contexts is greater in the nuclear stressed syllables than in the primary stressed syllables. The average durations for the respective categories reveal a difference of 45.48 ms between the *SVLR1* long and short contexts in nuclear stressed syllables and 10.48 ms for the corresponding environments in primary stressed syllables. In other words, the durational difference between the *SVLR1* long and short contexts is greater in syllables under nuclear stress. Similarly, the difference between the *SVLR2* long and short contexts amounts to 34.77 ms in nuclear stressed syllables and 24.58 ms in primary stressed syllables. There are no boxplots or values for SVLR long contexts in unstressed syllables because unstressed syllables are generally considered to be short in the context of Aitken's Law. The boxplots and average values indicate interactions between Aitken's Law and *stress* as well as between Aitken's Law and *phrasal position* in the monophthong FLEECE. The durational contrast between SVLR short and long vowels is increased in nuclear and phrase-final contexts which fits to the results of previous studies on Scottish vowel duration in uncontrolled speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016) (see subsection 3.2.3).

Figure 49 already revealed an influence of *constituent-final lengthening* and *prosodic stress* because vowel duration was generally longer in these contexts. The same proportions can also be seen in the plots in Figure 50: FLEECE tokens are much longer in phrase-final positions and under nuclear stress than in phrase-initial, phrase-medial, primary stressed or unstressed syllables. Interestingly, the boxplot for the unstressed syllables is slightly higher than the boxplot for the primary stressed syllables. The average value for FLEECE in unstressed syllables (mean: 95.87 ms) also exceeds the one of primary stressed syllables (mean: 93.33 ms), so it seems that there is no significant influence of *lexical stress* (see subsection 3.1.6). Nevertheless, the influence of *prosodic stress* (see subsection 3.1.7) and *constituent-final lengthening* (see subsection 3.1.8) on FLEECE is certain. The plots representing *word*

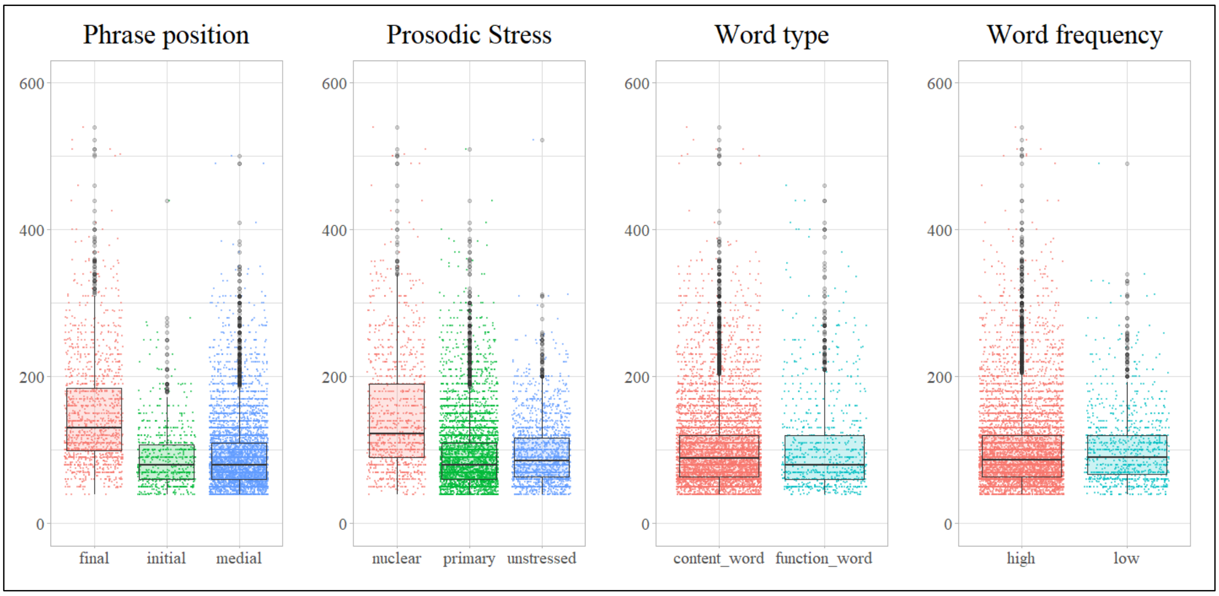


Figure 50. Boxplots and jitterplots of raw vowel duration for the lexical set FLEECE separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

type and *word frequency* are similar to the ones of the other vowels: the boxplots are on a relatively equal level. The average values are also very close, so it does not seem that FLEECE is strongly affected by the influence of *lexical category* or *lexical frequency* (see subsection 3.1.5).

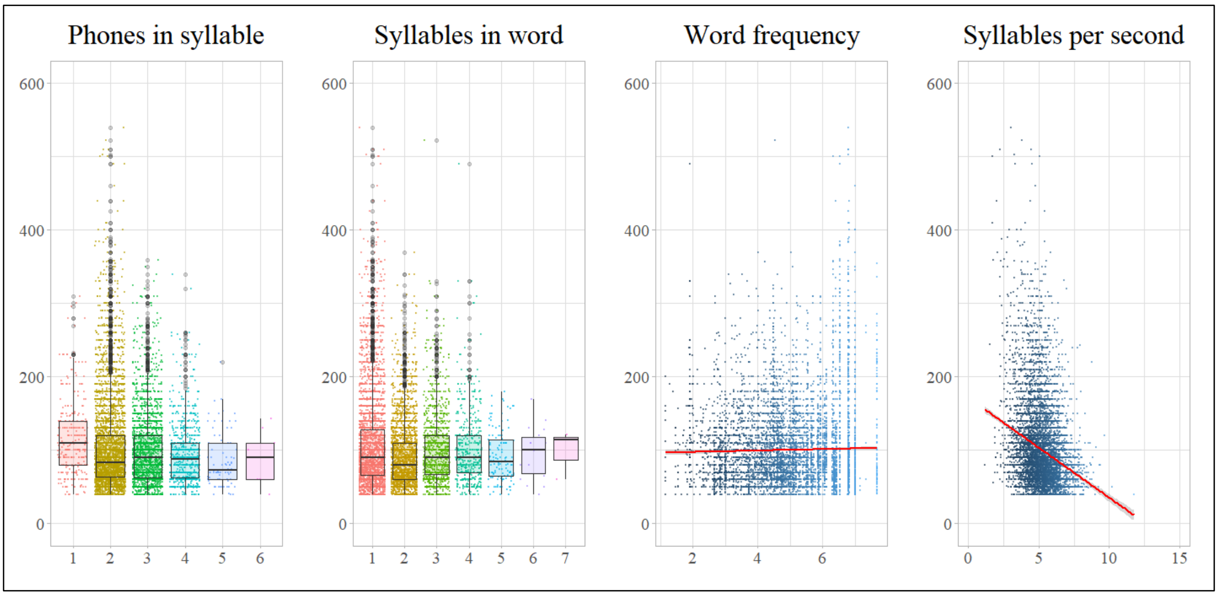


Figure 51. Plots of raw vowel duration of the lexical set FLEECE separated for the linear variables number of phones in syllable, number of syllables in word, word frequency and syllables per second.

The plots for the linear intralinguistic variables are comparable to the ones of other vowels. It is clear that FLEECE is affected by *tempo* (see subsection 3.1.9): long FLEECE durations become increasingly rare in higher articulation rates. The red regression line has a clear negative slope, so speakers generally produce shorter FLEECE vowels if their articulation rate is higher.

Like the categorical plot (see Figure 50), the linear plot on word frequency does not show a clear trend. The slope of the regression line comes close to zero and it is therefore unclear whether high frequency words are different than low frequency words in terms of their duration. The influence of *lexical frequency* on FLEECE is therefore unclear.

The influence of *polysyllabic shortening* (see subsection 3.1.4) on FLEECE is unclear as well. The boxplots do not show a clear negative trend. It is true that very long FLEECE tokens become less frequent in higher *word syllable counts*, but the average durations and the height of the boxplots do not steadily decrease.

The boxplots representative for the *number of phones in a syllable* follow a more regular pattern. The height of the boxplots steadily decreases except for the last boxplot (6 phones in a syllable). The average durations also show a clear negative trend, only the average duration for tokens in syllables with six phones is higher than the average duration for tokens in five-phone syllables. Yet, this category is represented by nine tokens only (e.g. <streets>, <screens>), so the low number of tokens may lead to unusual durational distributions. Overall, the plot indicates that FLEECE is affected by the effect of *intrasyllabic compression* (see subsection 3.1.3): the more phones in a syllable, the shorter the pronunciation of FLEECE.

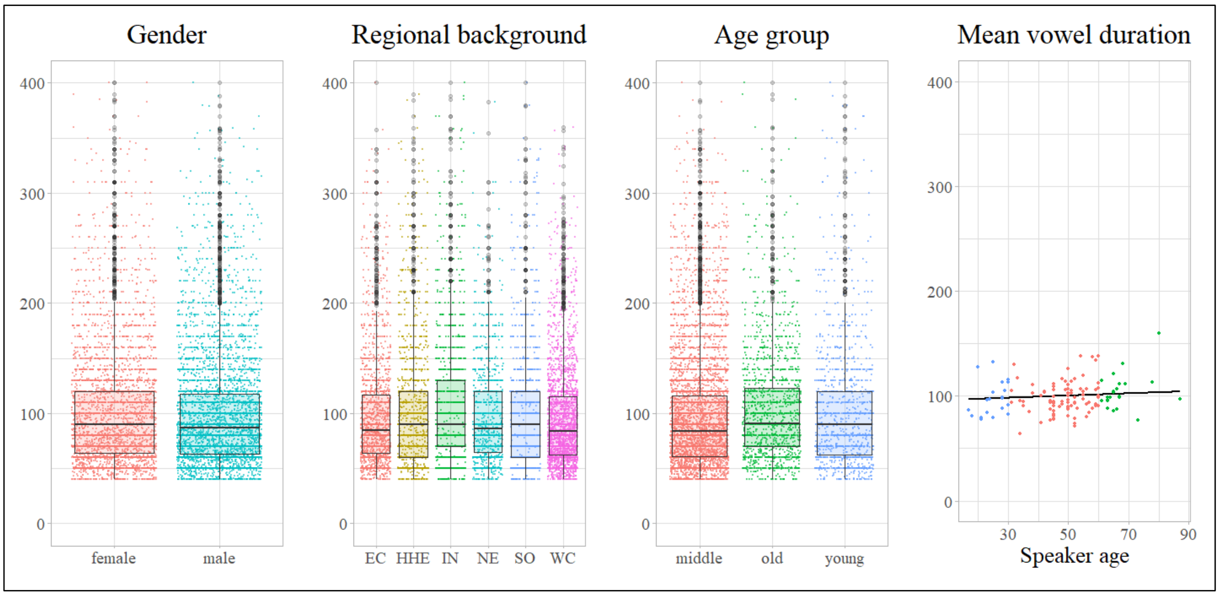


Figure 52. Plots of raw vowel duration of FLEECE separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

The plots for the social variables follow a very regular pattern (see Figure 52). Male and female speakers produce similar FLEECE durations, so the influence of the variable *gender* seems to be limited. The green boxplot representing the older speakers is slightly higher than the boxplots of the other age groups and the black regression line in the mean vowel duration plot has a very mild positive slope. Thus, it is possible that older speakers produce slightly longer FLEECE vowels, but the difference is not very pronounced when compared to the other age groups. As for the *regional backgrounds*, one can see that the boxplot for the dialect group *Insular* is a bit higher than the other boxplots, but vowel duration varies greatly across all regions. It is therefore unclear whether the duration of FLEECE is significantly affected by the *regional background*.

As for inferential statistics, all models excluded the fixed factors *age group*, *word frequency*, *gender*, *style* and *number of syllables in word* because they returned no significant effects. This means that the duration of FLEECE does not significantly vary for the different genders and age groups. It does not matter whether the vowels occur in scripted or unscripted speech and the duration of FLEECE does also not significantly vary in terms of the *lexical frequency* (see subsection 3.1.5). The number of syllables in a word has no significant effect either. This tendency was already visible in the corresponding plot (see Figure 51) in which one could not observe that FLEECE becomes shorter in higher *word syllable counts*. Hence, FLEECE is not significantly affected by *polysyllabic shortening* (see subsection 3.1.4). The conditional R^2 values of the different models lie in the range between 0.47 and 0.49. The fit of the models is therefore higher than the model fits of the short monophthongs. Furthermore, the range of the R^2 values is on a comparable level for all models. The best model was fit for the *SVLR1* categorization and a summary is given in Table 43.

*Table 43. Best linear mixed model for the lexical set FLEECE fit by maximum likelihood (marginal R^2 = 0.23; conditional R^2 = 0.49). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ (SVLR1 * position * stress) + (SVLR1 * reg) + position + stress + num_pho_syl + syl_per_sec + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label)*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	3573.1	3779.4	-1755.6	3511.1	5697

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-4.2564	-0.6425	-0.0128	0.6054	4.1032

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.02742	0.16559
syl_label	(Intercept)	0.00944	0.09716
speaker	(Intercept)	0.01127	0.10617
Residual		0.09362	0.30598

Intercept: SVLR1long; positionfinal; stressnuclear; regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.890e+00	361.36878	6.559e-02	< 2e-16	***
SVLR1short	-4.298e-01	0.6506195	5.025e-02	< 2e-16	***
positioninitial	-5.880e-01	0.5554475	6.595e-02	< 2e-16	***
positionmedial	-4.824e-01	0.6172849	3.702e-02	< 2e-16	***
stressprimary	-2.384e-01	0.7878980	3.986e-02	2.36e-09	***
stressunstressed	-2.940e-01	0.7452805	9.051e-02	0.00117	**
regSouth	-1.043e-01	0.9009169	4.254e-02	0.01508	*
num_pho_syl	-6.862e-02	0.9336835	1.560e-02	1.40e-05	***
syl_per_sec	-9.050e-02	0.9134710	4.804e-03	< 2e-16	***
SVLR1short:positioninitial	2.931e-01	1.3405682	9.193e-02	0.00144	**
SVLR1short:positionmedial	3.439e-01	1.4103976	5.300e-02	9.48e-11	***
SVLR1short:regHighland	7.495e-02	1.0778298	3.180e-02	0.01848	*
SVLR1short:regNorth_east	7.374e-02	1.0765305	3.166e-02	0.01990	*
SVLR1short:regSouth	7.323e-02	1.0759753	3.495e-02	0.03622	*

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

Significant fixed factors are the *SVLR1* context, the *phrasal position*, *stress*, the *regional background*, the *syllable phone count* and the local articulation rate (*syllables per second*). There are further interactions between Aitken's Law and the *phrasal position* and between Aitken's Law and the *regional background*. An interaction between *SVLR1* and *stress* and an interaction between *stress* and *phrasal position* did not reach statistical significance because the p-values fall just above the 0.05 threshold. Like the other models with the other categorization schemes (*SVLR2*, *VE1*, *VE2*), the model excluded the fixed factors *age group*, *word frequency*, *gender*, *style* and *number of syllables in word*. The plots have already shown that the levels of these variables do not show clear differences and the models corroborated the observation that these variables do not affect the duration of FLEECE in a significant way.

Yet, the *SVLR1* environment has a significant influence on the duration of FLEECE. *SVLR1* short contexts are almost 35 percent shorter than *SVLR1* long environments. The best model for the *SVLR2* categorization scheme returned that *SVLR2* short contexts are significantly shorter than *SVLR2* long contexts. This means that FLEECE is affected by the phonological and morphological conditioning of Aitken’s Law.

FLEECE is also significantly affected by the *phrasal position*. Tokens in phrase-initial syllables are almost 45 percent and tokens in phrase-medial syllables are roughly 38 percent shorter than FLEECE tokens in phrase-final syllables (=intercept). This means that FLEECE is clearly affected by *constituent-final lengthening* (see subsection 3.1.8). The model also returns significant interactions between the *SVLR1* short contexts and phrase-initial and phrase-medial syllables. The positive effect sizes reveal that, when compared to the intercept, *SVLR1* short contexts are longer in non-final positions. This indicates that the durational difference between *SVLR1* long and short tokens is less pronounced in phrase-initial and phrase-medial positions. In other words, the effects of Aitken’s Law are stronger in phrase-final positions and this trend could also be observed in the corresponding plots (see Figure 49). The models therefore corroborate the observation that the SVLR interacts with prosody and that its effects are amplified in phrase-final positions (Rathcke & Stuart-Smith, 2016, p. 416).

The long monophthong is further influenced by *stress*. The vowels are roughly 25 percent shorter in unstressed syllables and roughly 21 percent shorter in primary stressed syllables than in nuclear stressed positions. This shows that FLEECE is influenced by *prosodic stress* (see subsection 3.1.7) in connected speech. Interactions between Aitken’s Law and *stress* did not reach statistical significance, so the SVLR is not significantly stronger in nuclear stressed syllables.

FLEECE is also affected by *tempo* (see subsection 3.1.9). The local articulation rate influences the duration of the vowel and the model estimates that the monophthong is almost 9 percent shorter for every additional syllable that is produced in the timeframe of one second. This fits to the overall trend: vowels become shorter in fast speech and longer in slow speech.

The model output further confirms that FLEECE is influenced by *intrasyllabic compression* (see subsection 3.1.3). The linear mixed model estimates that the duration of the vowel decreases by roughly 7 percent for every additional phone in a syllable. The vowel is therefore generally longer in simple syllable structures (e.g. <eat>) than in complex syllables with many phones (e.g. <street>).

As for the extralinguistic variables, the duration of FLEECE is also affected by the *regional background*. The model specifies that the duration of FLEECE is generally shorter in the dialect region *Southern*. When compared to the intercept (dialect region *East central*), speakers from the South of Scotland produce roughly 10 percent shorter FLEECE vowels. This trend was not directly visible in the corresponding plot (see Figure 52). Apart from that, the model also returns significant interactions between Aitken’s Law and three dialect regions, namely *Northeast*, *Southern* and *HHE*. The estimates reveal that, when compared to the intercept, the *SVLR1* short contexts are roughly 7 percent longer in these regions. This means that the durational difference between the *SVLR1* long and short contexts is less pronounced in these dialect areas. Figure 49 already demonstrated that there is not a strong difference between the *SVLR1* long and short contexts in the *Northeast*, *Southern* Scotland and *HHE* and the model output confirms that Aitken’s Law is relatively weak in these regions.

The models for the VE included the *VE1* and *VE2* categorization as significant fixed factors in all model constellations. This means the lexical set FLEECE is influenced by the effect of postvocalic consonant voicing (see subsection 3.1.2). Especially the *VE2* categorization delivers strong evidence for this claim because it only includes FLEECE tokens voiced and voiceless plosive contexts. The effect sizes of the VE categories are, however, less strong than the effect sizes of the SVLR categories. Some interactions also indicate that the VE is stronger in the dialect regions *Southern* and *Northeast* but weaker in the Northern Isles (= dialect region *Insular*). This distribution corresponds with earlier findings. Warren (2018) as well as Watt and Yurkova (2007) found robust VE effects in the Northeast of Scotland and Llamas et al. (2011) found the same in the Northeast of England. The studies on the Northern Isles, however, found out that Aitken's Law is firmly established, especially among children with parents from the same region (Scobbie, 2005; van Leyden, 2002).

The plots and models have shown that FLEECE is affected by the phonological and morphological conditioning of Aitken's Law. Vowel duration is increased in SVLR long contexts and significantly shorter in SVLR short contexts. Similar to GOOSE (see subsection 5.3.1), the effects of Aitken's Law are stronger in phrase-final syllables as there is a significant interaction between the SVLR and the *phrasal position*. The interactions with the regions indicate that the SVLR is weaker outside the Central Belt of Scotland, specifically in the *Northeast*, *HHE* and in the dialect region *Southern*. While FLEECE is affected by Aitken's Law, the model also revealed a significant influence of the VE. The VE is consistent for all age groups and genders, the VE is especially strong in the Northeast and South of Scotland. The dialect region *Insular*, in contrast, does not show a strong effect size of the VE. Furthermore, while FLEECE is generally longer before voiced than before voiceless plosives and fricatives, the duration of this monophthong is clearly shortened before postvocalic nasals. This distribution fits into the overall trend in this investigation that postvocalic nasals lead to short vowel articulations in SSE.

5.3.3 THOUGHT

The lexical set THOUGHT represents the back rounded vowel /ɔ/ in Scottish English. According to the Basic Scottish Vowel System (Abercrombie, 1979), SSE does not distinguish between /ɒ/ and /ɔ/ so the words <cot> and <caught> would be pronounced similarly. In other words, SSE does not differentiate between the lexical sets LOT and THOUGHT (see section 2.3). The description and transcription of this vowel have not always been consistent in the literature. Dieth (1932) and Wölck (1965), for instance, transcribe the vocalic nucleus of the word <loch> with the symbol /o/. Wettstein (1942, p. 3) distinguishes between the vowels /o/ and /ɒ/ in words such as <before> and <hall>. Aitken (1981) uses the symbol /o/ for words such as <cot> and specifies that this is a short monophthong affected by the SVLR (vowel 18). Aitken uses the same symbol with a lengthening mark /oː/ for the lexical set GOAT (see subsection 5.3.5). The durational measurements of THOUGHT from this investigation are summarized in Table 44 and visualized in Figure 53.

Table 44. Statistical summary of vowel duration for the lexical set THOUGHT. Tokens below 40 ms were excluded.

Token number:	7762
Average duration:	112.64 ms
Standard deviation:	55.79 ms
Minimal duration:	40.00 ms
Maximal duration:	701.00 ms
1st quantile:	73.26 ms
Median:	100.00 ms
3rd Quantile:	135.76 ms

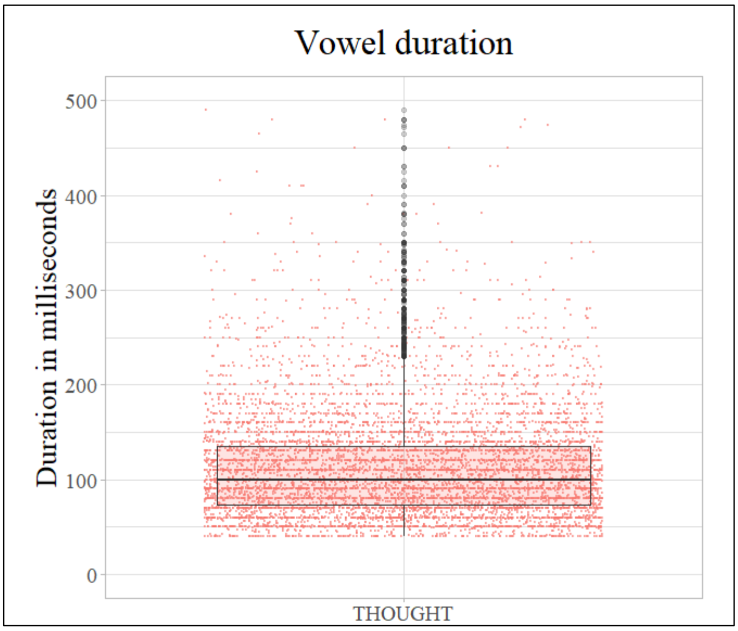


Figure 53. Vowel duration of the lexical set THOUGHT in milliseconds. Durations below 40 ms were excluded.

The 7762 THOUGHT tokens have an average duration of 112.64 ms and a standard deviation of 55.79 ms. The shortest tokens are 40 ms long and the longest THOUGHT pronunciation lasts 701 ms. Most of the articulations are in the durational range between 73.26 ms and 135.76 ms and the median lies at 100 ms.

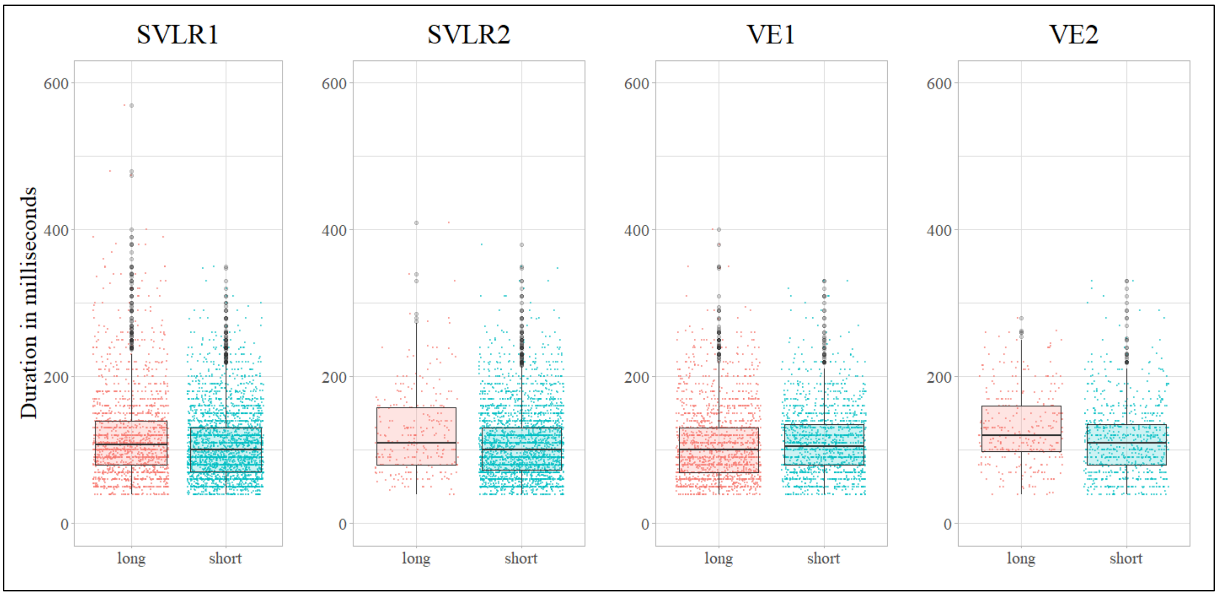


Figure 54. Boxplots and jitterplots of raw vowel duration of the lexical set THOUGHT for the categories SVLR1, SVLR2, VE1 and VE2.

The plots for the categories *SVLR1*, *SVLR2*, *VE1* and *VE2* show some unusual distributions. While vowels are longer in the long environments for the *SVLR1*, *SVLR2* and *VE2* categorization schemes, this trend is reversed for *VE1*: THOUGHT vowels in are shorter in *VE1* long contexts than in *VE1* short environments. This trend is unusual, especially since the *VE2* long contexts are clearly longer than the *VE2* short contexts. The *VE1* environments incorporate the plosives of the *VE2* contexts, but the distributions are nonetheless completely different for both categories. Similarly, the *VE1* and *SVLR1* categorization schemes partially overlap (see subsection 4.4.2), so it is interesting that both have completely different distributions as well. The average durations also reflect this trend. The *SVLR1*, *SVLR2* and *VE2* long contexts are on average clearly longer than the respective short environments (*SVLR1* long: 119.44 ms; *SVLR1* short: 107.04 ms; *SVLR2* long: 123.67 ms; *SVLR2* short: 108.30 ms; *VE2* long: 129.88 ms; *VE2* short: 112.11 ms). Only the *VE1* long environments (mean: 108.32 ms) are shorter than the *VE1* short environments (mean: 111.01 ms).

The average durations for the different postvocalic contexts provide more detailed information on why the distributions are the way they are (see Table 45). First, the mean values show that vowels are clearly longer in voiced plosive than in voiceless plosive contexts. This trend is congruent with the VE and could already be observed in the *VE2* plots (see Figure 54). However, it is interesting to see that, except for postvocalic pauses, following voiced plosives lead to the longest average vowel duration in THOUGHT. Vowels followed by voiced plosives are clearly longer than THOUGHT tokens followed by voiced fricatives. The overall durational difference between the plosive contexts (23.19 ms) clearly exceeds the difference for the voiced and voiceless fricative contexts (5.43 ms). This distribution shows that there is no "additional lengthening" that can be attributed to the SVLR (McMahon, 1991, p. 44). The phonological conditioning of Aitken's Law does therefore not seem to operate in THOUGHT. Another interesting distribution can be found for the nasal contexts: vowels with following nasals are clearly the shortest on average. This does not only go against the VE but also against the lengthening effect of the manner of the postvocalic consonantal context (see subsection 3.1.2). The short THOUGHT tokens before nasals are also the reason why the *VE1* long contexts are on average shorter than the *VE1* short contexts because they pull down the mean value of the *VE1* long environments. At the same time, nasal contexts are classified as *SVLR1* short environments, so this might be another reason why the *SVLR1* long contexts are longer than the *SVLR1* short contexts. Apart from this, Table 45 shows that THOUGHT is affected by *constituent-final lengthening* (see subsection 3.1.8) due to the very long durations before pauses.

Table 45. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set THOUGHT. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	**Average vowel duration**
laterals	113.97 ms
nasals	97.77 ms
voiced fricatives	116.52 ms
voiceless fricatives	111.09 ms
voiced plosives	129.14 ms
voiceless plosives	105.95 ms
pauses	200.54 ms

I further plotted the SVLR and VE environments for the different intra- and extralinguistic variables to see whether the effects of the VE or the SVLR vary in specific groups or contexts. Overall, the *SVLR2* and *VE2* effects are stable across the different regions, age groups and genders. The long contexts are generally longer than the short contexts for these extralinguistic categories. More interesting distributions can be found for the *SVLR1* and *VE1* categorizations, especially if they are separated for the different regions, phrasal positions and stress patterns (see Figure 55).

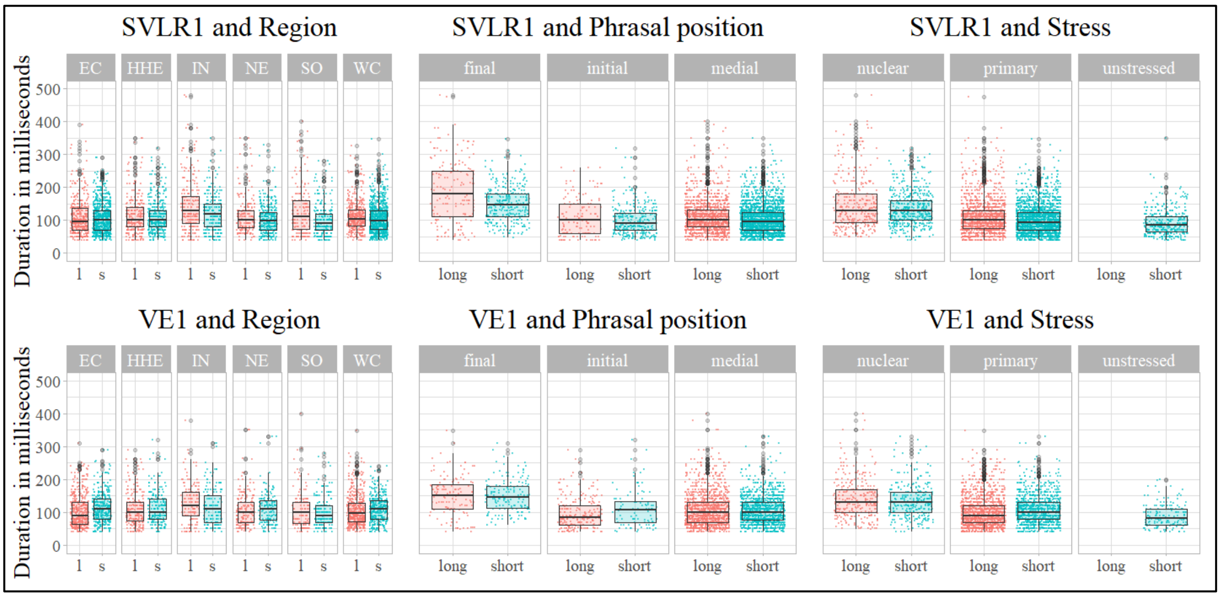


Figure 55. Boxplots and jitterplots of raw vowel duration of the lexical set THOUGHT for the categories SVLR1 and VE1 separated for the variables regional background, phrasal position and stress.

The *SVLR1* patterns are relatively stable across the regions. *SVLR1* long contexts generally lead to longer THOUGHT tokens than *SVLR1* short contexts. However, it is also visible that the speakers from the dialect region *Insular* produce longer THOUGHT vowels overall. The boxplots are clearly higher than the boxplots of the other regions. As for *VE1*, there are different distributions across the different regions. The long contexts in the regions *East central*, *HHE*, *Northeast* and *West central* are clearly shorter than the *VE1* short contexts. Yet, the distribution in the dialect region *Insular* and *Southern* show a more regular trend: the long environments are longer than the short environments. It seems that these dialect regions follow a more stable VE trend than the other regions. One can also see clear discrepancies between the *SVLR1* and *VE1* categories when divided for the *phrasal positions*. The plots show that the durational difference between the *SVLR1* long and short contexts is clearly higher in the phrase-final positions than in phrase-initial or medial syllables. This difference is not given for the *VE1*: the boxplots for the long and short environments are highly similar and the corresponding average values for the phrase-final positions are almost identical (*VE1* long: 150.41 ms; *VE1* short: 150.08 ms). A similar trend can be observed for the phrase-medial tokens. The boxplots are on the same level and the corresponding mean values are highly similar as well (*VE1* long: 106.06 ms; *VE1* short: 105.93 ms). The distribution is different, however, in the phrase-initial positions. Here, the *VE1* long contexts are clearly shorter than the *VE1* short contexts. I filtered the dataset for the *VE1* tokens in phrase-initial positions to check whether any particular variable (e.g. specific words) might be responsible for this trend but failed to find correlations. The distributions therefore indicate an "anti-Voicing Effect" (Stuart-Smith et al., 2019) in THOUGHT in phrase-initial positions. As for stress, *SVLR1* long environments are slightly longer than *SVLR1* short environments in nuclear stressed and primary stressed syllables. It is clear that nuclear stressed syllables are on the whole longer than primary stressed or unstressed syllables, but the discrepancies between the *SVLR1* levels do not vary greatly. The plots for the *VE1* categories and *stress* show that *VE1* long contexts are slightly longer than *VE1* short contexts in nuclear stressed syllables. This trend is, however, reversed in primary stressed syllables where the duration of the *VE1* short environments exceeds the duration of the *VE1* long environments. This could also indicate an "anti-Voicing Effect" (Stuart-Smith et al., 2019) in THOUGHT in primary stressed syllables.

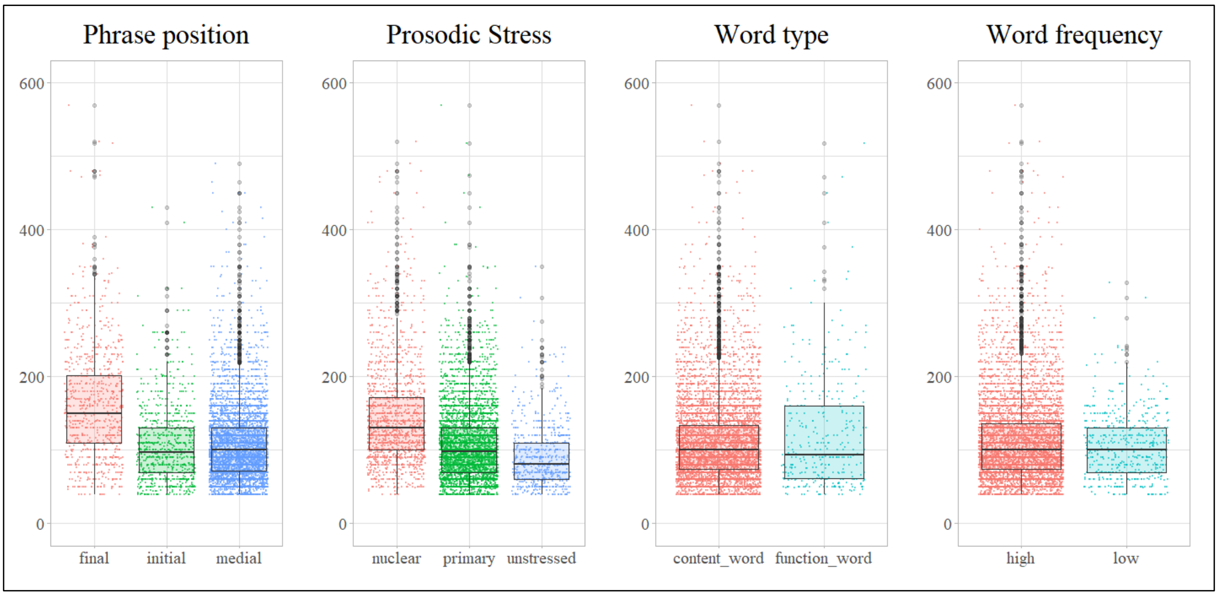


Figure 56. Boxplots and jitterplots of raw vowel duration for the lexical set THOUGHT separated for the categorical variables phrase position, prosodic stress, word type and word frequency.

The plots for the categorical intralinguistic variables (see Figure 56) are similar to the respective plots of the other vowels. THOUGHT is clearly longer in final positions than in non-final positions. Similarly, the vowel is also longer under nuclear stress than in primary stressed or unstressed syllables. The boxplot for the primary stressed syllables is also higher than the one for the unstressed syllables, so there could be an influence of both *prosodic* and *lexical stress* (see subsections 3.1.6 and 3.1.7). The plot for word type shows that function words have more variable durations than content words because the turquoise boxplot is much wider. The respective average values reveal that function words (mean vowel duration: 122.98 ms) are generally longer than content words (mean vowel duration: 112.16 ms). This distribution is not in line with previous studies (see subsection 3.1.5): function words are often reduced and therefore shorter than content words. A similar observation can be made for *word frequency*. The boxplots do not show a clear trend but the average values reveal that high frequency words (mean: 113.19 ms) are longer than low frequency words (mean: 105.61 ms). This is also not in line with previous studies (see subsection 3.1.5).

The linear plot for *word frequency* shows a similar distribution (see Figure 57). The regression line has a positive slope which indicates that high frequency words are longer than low frequency words. The plot for *tempo* shows a more regular trend that can also be observed in the other vowels of this study: the red regression line has a clear negative slope, so THOUGHT tokens are longer in slow speech and shorter in fast speech. The plot for the number of syllables in a word follows a general trend: THOUGHT tokens become shorter with a higher *word syllable count*. The only exceptions are the words

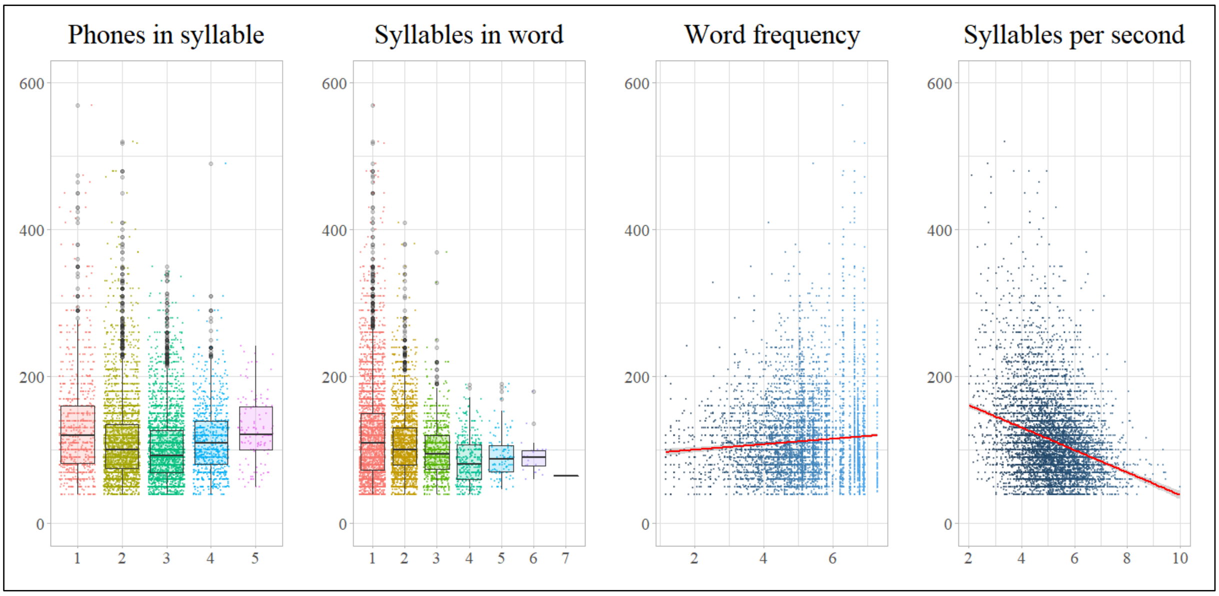


Figure 57. Plots of raw vowel duration of the lexical set THOUGHT separated for the linear variables number of phones in syllable, number of syllables in word, word frequency and syllables per second.

with five and six syllables where the average durations and the height of the boxplots increase. However, these words are represented by fewer tokens so the low token numbers could lead to unusual durational values. Despite these irregularities, it is likely that THOUGHT is affected by *polysyllabic shortening* (see subsection 3.1.4). The plots for the *syllable phone count* do not follow a consistent pattern. There is a negative trend from syllables with one phone to syllables with three phones, but then the duration increases again. The vowel clearly gets longer in syllables with four phones and in syllables with five phones. Hence, it does not seem that THOUGHT is affected by *intrasyllabic compression* (see subsection 3.1.3).

I also plotted the duration of THOUGHT for the different extralinguistic variables *gender*, *regional background* and *age group*. The plots revealed that male speakers produce slightly longer THOUGHT tokens than female speakers. Furthermore, the THOUGHT pronunciations from the *Insular* dialect group are generally longer and the pronunciations from the *Southern* speakers are slightly shorter than the rest. As for *age group*, the old and young speakers produce longer THOUGHT tokens than the middle-aged speakers.

Similar to the other vowels, I also fitted several linear mixed effects models with the different SVLR and VE categorizations. All models excluded the variable *word frequency* which means that the duration of THOUGHT is not significantly influenced by *lexical frequency* effects (see subsection 3.1.5). The plots already revealed that there is not a great difference between the high and low frequency words and the model outputs clarify that this variable has no significant influence. The conditional R^2 values of the different models lie in the range between 0.48 and 0.53. The best model was fit with the *VE2* categorization scheme and a summary of the model output is given in Table 46.

*Table 46. Best linear mixed model for the lexical set THOUGHT fit by maximum likelihood (marginal R^2 = 0.23; conditional R^2 = 0.53). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ VE2 + position + stress + style + num_syl_word + num_pho_syl + syl_per_sec + (VE2 * gender) + (VE2 * age_group) + gender + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label)*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	679.4	799.8	-315.7	631.4	1092

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-4.0048	-0.5684	0.0293	0.5737	3.5612

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.01868	0.1367
syl_label	(Intercept)	0.01448	0.1203
speaker	(Intercept)	0.01817	0.1348
Residual		0.07933	0.2817

Intercept: positionfinal; stressnuclear; stylescripted; genderfemale; regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.878e+00	356.9464689	1.198e-01	< 2e-16	***
positioninitial	-2.897e-01	0.7485036	4.594e-02	4.25e-10	***
positionmedial	-2.471e-01	0.7810511	3.323e-02	2.14e-13	***
stressprimary	-1.819e-01	0.8337212	2.330e-02	1.45e-14	***
styleunscripted	-7.707e-02	0.9258258	3.511e-02	0.02944	*
num_syl_word	-5.280e-02	0.9485685	1.811e-02	0.00389	**
num_pho_syl	-7.960e-02	0.9234862	2.594e-02	0.00266	**
syl_per_sec	-8.767e-02	0.9160617	1.064e-02	5.02e-16	***
gendermale	1.171e-01	1.1241895	4.605e-02	0.01145	*
regSouth	-1.936e-01	0.8239948	6.077e-02	0.00180	**
VE2short:gendermale	-9.576e-02	0.9086784	4.422e-02	0.03055	*
VE2short:age_groupyoung	1.443e-01	1.1552663	6.435e-02	0.02510	*

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

Significant fixed factors in the model output are the *phrasal position*, *stress*, *style*, the *syllable phone count*, the *word syllable count*, the local articulation rate (*syllables per second*), the *gender* as well as the *regional background* of the speakers. The *VE2* categorization and the *age group* did not reach statistical significance on their own, but the model includes interactions between *gender* and *VE2* as well as between the *age group* and *VE2*. All other interactions did not return significant effects and were therefore excluded from the model in the stepwise regression procedure.

When compared to the intercept, THOUGHT tokens are roughly 22 percent shorter in phrase-medial and 25 percent shorter in phrase-initial positions than in phrase-final syllables. This trend was already observable in the corresponding plot (see Figure 56) and the model output confirms the influence of *constituent-final lengthening* (see subsection 3.1.8) in THOUGHT.

The variable *stress* also has a significant influence on the duration of THOUGHT because tokens in primary stressed syllables are roughly 16 percent shorter than tokens in nuclear stressed syllables.

Unstressed syllables are not found in the model output because they are excluded in the *VE2* categorization scheme.

In contrast to most other vowels in this study, THOUGHT is also significantly affected by the *style* of speech. THOUGHT tokens are approximately 8 percent shorter in unscripted speech forms than in scripted speech forms (= intercept). This is in line with the expectations because the language is more spontaneous in unscripted speech and speakers have less time to plan and carefully articulate their conversational contributions. Nevertheless, it must be said that the *style* was not significant in the models with the *SVLR2* and *VE1* categorization schemes. The influence of this variable should therefore not be overestimated, but *style* definitely has a significant influence on the production of THOUGHT in postvocalic plosive contexts.

The monophthong is also affected by the *syllable phone count* and the *word syllable count*. THOUGHT is approximately 5 percent shorter for every additional syllable in a word and 8 percent shorter for every additional phone in a syllable. Hence, the effects of *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* influence the duration of THOUGHT. While the influence of *intrasyllabic compression* was not clearly detectable in the corresponding plot (see Figure 57), the model outputs for the *SVLR2*, *VE1* and *VE2* categorizations all confirmed that this variable is significant.

Like all other vowels, THOUGHT is also affected by the local articulation rate. The monophthong is shortened by roughly 8 percent for every additional syllable produced in the timeframe of one second. THOUGHT is therefore clearly affected by *tempo* (see subsection 3.1.9).

The extralinguistic variable *gender* also has a significant influence on the duration of THOUGHT. Similar to GOOSE (see subsection 5.3.1), the male speakers produce longer THOUGHT tokens overall. The effect size estimates that the vowels of the male speakers are roughly 12 percent longer than the vowels of the female speakers. Nevertheless, the other models with the *SVLR1*, *SVLR2* and *VE1* sub-datasets excluded the variable *gender* in the final models. This means that male speakers produce longer THOUGHT vowels in plosive contexts, but the vocalic durational difference between male and female speakers does not reach statistical significance in other environments.

Another significant extralinguistic variable is the *regional background* of the speakers. Specifically, speakers from the dialect region *Southern* produce roughly 12 percent shorter THOUGHT tokens than the speakers form the region *East central* (= intercept). The dialect region *Southern* was also significant in all other models, so there is a general trend that speakers from *Southern* Scotland produce shorter THOUGHT tokens. This corresponds to the pattern that *Southern* speakers produce relatively short vowels overall (see section 5.1).

The interactions reveal that *VE2* short tokens are shorter for male speakers and longer for younger speakers when compared to the intercept. This means that the durational difference between the *VE2* long and short contexts is less pronounced for the young speakers than for the middle-aged speakers (=intercept) but more pronounced for the male speakers than for the female speakers (=intercept). Nevertheless, the *VE2* category itself does not reach statistical significance in the model, so there is no significant difference between the *VE2* long and short contexts overall. This distribution is surprising because this trend was not observable in the plots.

Overall, the plots and models revealed that the lexical set THOUGHT does not follow the durational patterns of the SVLR or the VE. The *SVLR1* and *SVLR2* long environments are generally longer than the corresponding short environments, but the longest THOUGHT tokens can be found in voiced plosive environments and not in fricative contexts. The great durational difference in the plosive contexts suggests an influence of the VE, but the *VE1* short tokens are longer than the *VE1* long environments and the difference between the *VE2* levels does not reach statistical significance. Moreover, vowels are extremely short when followed by nasals and this is not in line with the VE (see subsection 3.1.2). To conclude, the categorization schemes of the SVLR and the VE do not fit to the durations of THOUGHT in different contexts. The monophthong THOUGHT is, however, affected by prosodic factors, namely the *phrasal position*, *stress*, *tempo, polysyllabic shortening* and *intrasyllabic compression.*

The timing patterns for THOUGHT are not easy to categorize in the present study as the vowel is not clearly influenced by the VE or by Aitken’s Law. This stands in contrast to Aitken (1981) who claimed that THOUGHT is affected by the SVLR. The previous empirical studies by McClure (1977), McKenna (1988) Watt and Yurkova (2007) as well as Warren (2018) could, however, find VE-related lengthening effects in this vowel but no effects of Aitken’s Law. Scobbie, Hewlett, and Turk (1999) as well as Scobbie, Turk, and Hewlett (1999) could find no evidence of the SVLR either and the studies by Watt and Ingham (2000) and Scobbie (2005) conclude that the SVLR effects are less consistent in THOUGHT than in FLEECE, GOOSE or PRICE. The latest study by Stuart-Smith et al. (2019) could only find SVLR-related timing effects in the dialect region *Northern*. The results by Stuart-Smith et al. (2019) therefore contradict the findings by Warren (2018) and Watt and Yurkova (2007). In short, the state of research on the quantity of THOUGHT was contradictory. The present study found out that neither the SVLR nor the VE operate in this vowel in spoken SSE.

5.3.4 FACE

The lexical set FACE is a monophthong in Scottish English and represents the vowel /e/ in words such as <way> and <make>. The vocalic durations of FACE of this study are summarized in Table 47 and visualized in Figure 58.

Table 47. Statistical summary of vowel duration for the lexical set FACE. Tokens below 40 ms were excluded.

Token number:	9394
Average duration:	113.00 ms
Standard deviation:	57.34 ms
Minimal duration:	40.00 ms
Maximal duration:	780.00 ms
1st quantile:	72.94 ms
Median:	100.00 ms
3rd Quantile:	138.00 ms

The average duration of FACE is 113 ms with a standard deviation of 57.34 ms. The minimal duration is 40 ms due to the corresponding durational threshold and the longest pronunciation lasts 780 ms. The dataset comprises 9394 FACE tokens and most of the measurements lie in the range between 72.94 ms and 138 ms. The median is 100 ms.

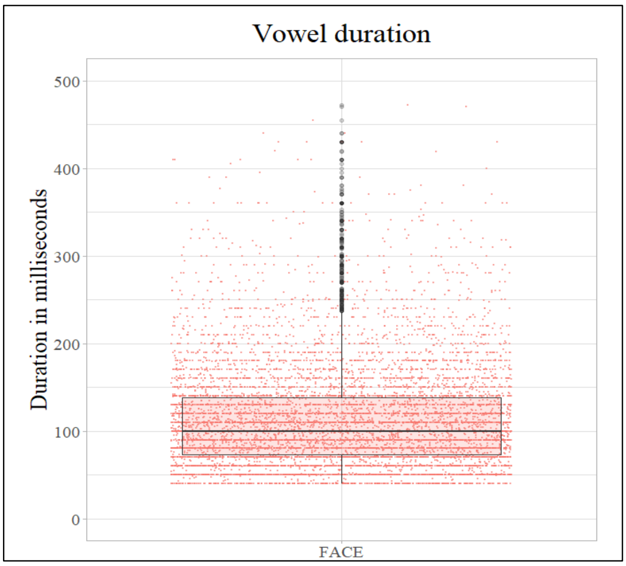


Figure 58. Vowel duration of the lexical set FACE in milliseconds. Durations below 40 ms were excluded.

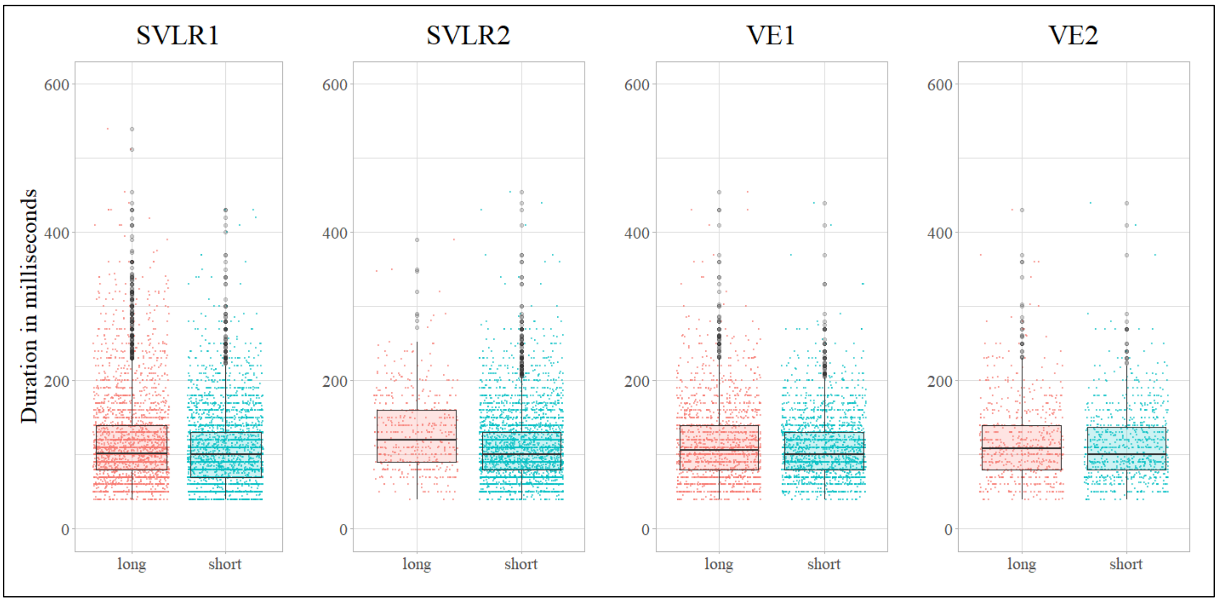


Figure 59. Boxplots and jitterplots of raw vowel duration of the lexical set FACE for the categories SVLR1, SVLR2, VE1 and VE2.

The plots for the SVLR and VE categories show a regular distribution (see Figure 59). The long environments lead to slightly longer vowel pronunciations than the corresponding short environments. The difference is more pronounced for the morphological conditioning of Aitken's Law (*SVLR2*) than for the other categorizations. This is not only reflected in the boxplots but it can also be seen in the respective average durations. The durational difference between the average values of the long and short contexts amounts to roughly 10 ms for the *SVLR1* and approximately 20 ms for the *SVLR2* categorization schemes (*SVLR1* long: 118.58 ms; *SVLR1* short: 108.16 ms; *SVLR2* long: 130.80 ms; *SVLR2* short: 110.46 ms). The difference between the mean values of the *VE1* and *VE2* classification is only 5 ms (*VE1* long: 115.27 ms; *VE1* short: 110.15 ms; *VE2* long: 117.29 ms; *VE2* short: 112.06 ms). This means that the long contexts are generally longer than the short environments, but the differences are often very slight, especially in the VE contexts.

Table 48. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set FACE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	**Average vowel duration**
laterals	98.74 ms
nasals	110.58 ms
voiced fricatives	124.06 ms
voiceless fricatives	103.25 ms
voiced plosives	108.77 ms
voiceless plosives	106.57 ms
pauses	192.32 ms

The average values for the respective postvocalic consonantal contexts indicate that FACE might be affected by Aitken's Law (see Table 48). Apart from postvocalic pauses, voiced fricative contexts lead to the longest FACE durations overall. The average duration for the voiceless fricative contexts is approximately 20 ms shorter. This indicates the phonological conditioning of Aitken's Law as voiced fricatives are considered SVLR long contexts. At the same time, the difference between the plosive contexts is marginal: overall, voiced plosive contexts are only 2 ms longer than voiceless plosive contexts. The FACE tokens followed by laterals are relatively short and the vowels followed by nasals are also clearly shorter than those in voiced fricative contexts. This also corresponds to Aitken's Law because the voiced and voiceless plosive environments as well as the nasals and the lateral /l/ are considered short SVLR contexts.

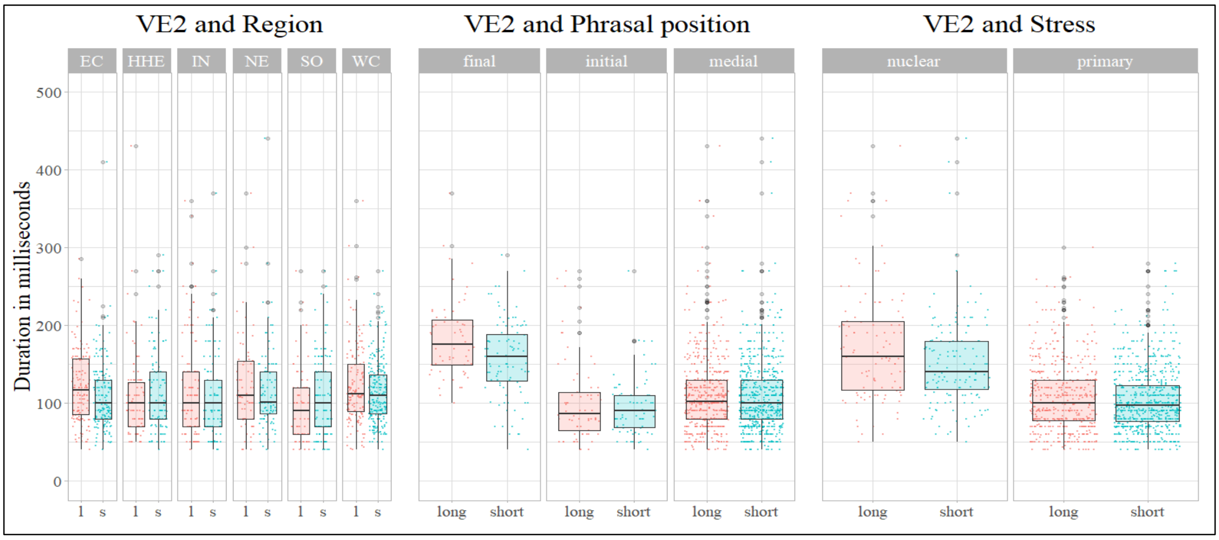


Figure 60. Boxplots and jitterplots of raw vowel duration of the lexical set FACE for the category VE2 separated for the variables regional background, phrasal position and stress.

I further plotted the SVLR and VE categories for different extra- and intralinguistic variables to see whether the effect sizes differ in specific groups or contexts (see Figure 60). The *SVLR2* long and short contexts are relatively stable across the different *regional backgrounds*, *age groups* and *genders*. *SVLR2* long contexts are also consistently longer in than *SVLR2* short environments in different *phrasal positions* and *stress* contexts. This indicates that the morphological conditioning of Aitken's Law is stable in FACE. The distributions are also regular for the *SVLR1* and *VE1* categorizations. More varying distributions can be found for the *VE2* category, especially when it is subdivided for the *regional background*, *phrasal position* and *stress* (see Figure 60). For instance, the boxplots of the *VE2* long contexts are shorter than the boxplots of the respective short contexts in the regions *HHE* and *Southern*. The regions *East Central*, *Insular, Northeast* and *West central* follow a more regular pattern in which postvocalic voiced plosives cause longer FACE tokens than following voiceless plosives.

The *VE2* strength also varies in the different *phrasal positions*. First, the vowel is generally longer in final positions than in non-final positions. Second, the *VE2* long contexts are only noticeably longer than the *VE2* short contexts in phrase-final positions. The long and short environments are on a highly similar level in phrase-initial and phrase-final syllables. This indicates that possible VE-related lengthening effects are more pronounced in prepausal environments.

The *VE2* levels also show variation with respect to *stress*. The tokens are overall longer in nuclear stressed syllables than in primary stressed syllables. At the same time, the durational difference between the levels is greater under nuclear stress than under primary lexical stress. This distribution also shows that the influence of the VE might be amplified in prominent prosodic positions.

The plots for the intralinguistic variables (see Figure 61) show further influences on the duration of FACE. The vowel is clearly longer in phrase-final positions which further corroborates the influence of

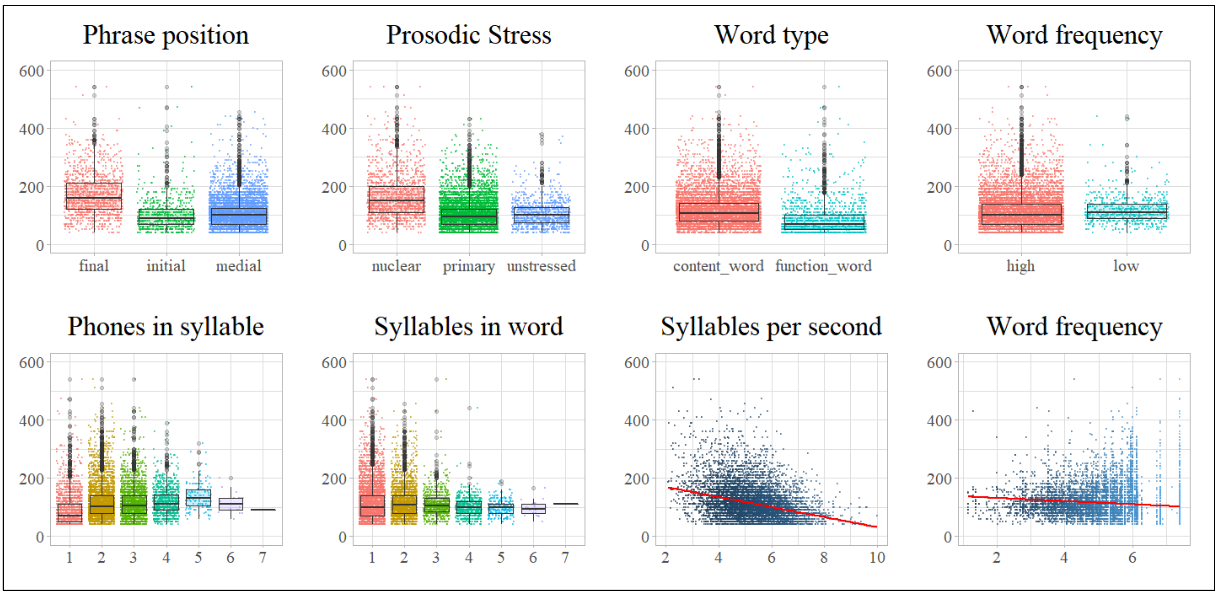


Figure 61. Plots of raw vowel duration for the lexical set FACE separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

constituent-final lengthening (see subsection 3.1.8). The monophthong is also longer in syllables under nuclear stress than in primary or unstressed syllables. *Prosodic stress* (see subsection 3.1.7) does therefore influence the duration of FACE as well. The tokens in primary stressed and unstressed syllables do not show great durational differences, so it is unclear whether FACE is also affected by *lexical stress* (see subsection 3.1.6). The plot for the *word type* shows that content words are generally longer than function words. This coincides with the effect of *lexical category* (see subsection 3.1.5) because function words are frequently shortened in spoken language. While the categorical plot for *word frequency* does not show a specific distribution, the linear plot for the same variable indicates that high frequency words are slightly shorter than low frequency words. The regression line has a small negative slope, so the duration should marginally decrease with a higher *lexical frequency* (see subsection 3.1.5). It is, however, unclear if this effect is significant because the differences are not very pronounced. The influence of the variable *tempo* (see subsection 3.1.9) is stronger. FACE tokens become increasingly shorter if the local articulation rate is higher. This fits into the overall trend (see section 5.1): vowels become shorter if the speech is faster. The plots for the *syllable phone count* and the *word syllable count* provide a mixed picture. Long FACE tokens become less frequent in words with many syllables and in syllables with many phones. However, the average values and boxplots do not steadily decline, especially not with a higher *syllable phone count*. On the contrary, vowel duration rather increases if the *number of phones in a syllable* gets higher. This distribution is not in line with the effect of *intrasyllabic compression* (see subsection 3.1.3). The durations of FACE in the plot for the *word syllable count* follow a more regular pattern but it is also not clear whether the vowel is affected by *polysyllabic shortening* (see subsection 3.1.4).

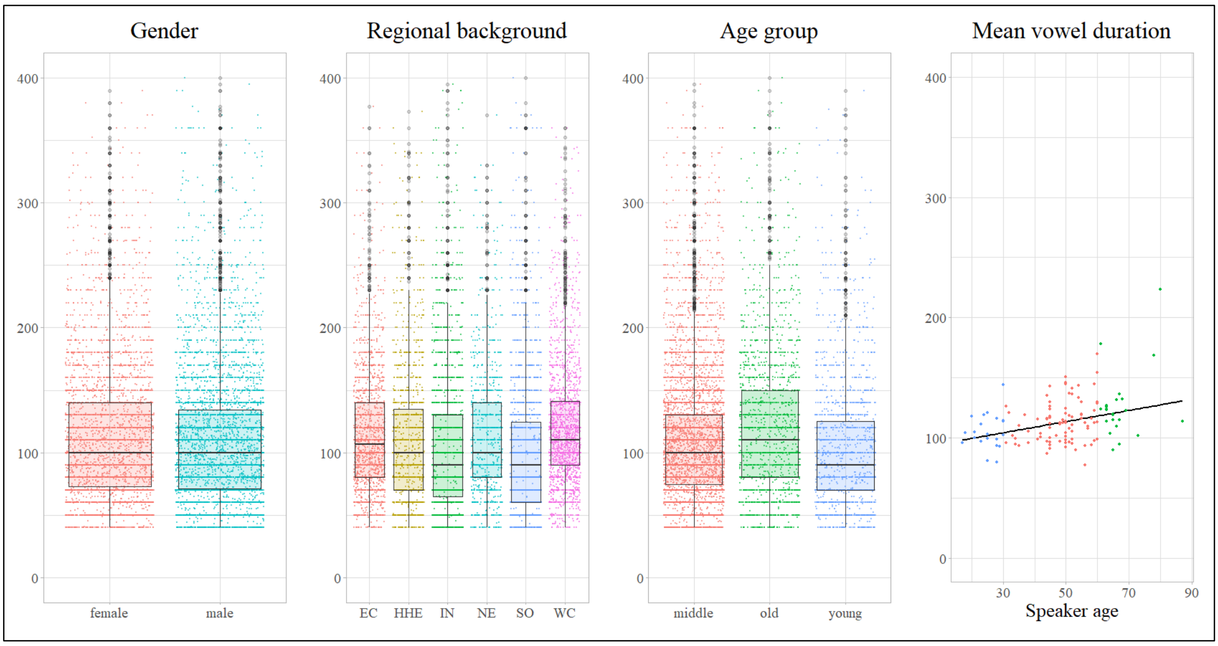


Figure 62. Plots of raw vowel duration of FACE separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

The plots for the social variables (see Figure 62) corroborate many observations that could already be made in Figure 26. The older speakers produce longer FACE tokens overall and also the mean vowel duration increases with a higher age. The female speakers produce slightly longer vowels and the regions do also differ in terms of their FACE durations. Especially the speakers from the *Southern* Scotland produce shorter vowels which is in line with the overall trend that the vowels are relatively short in that region (see section 5.1).

The conditional R^2 values of the mixed effects models for the lexical set FACE lie in the range between 0.41 (best *SVLR1* model) and 0.49 (best *VE2* model). The categorization schemes *SVLR1*, *SVLR2*, *VE1* and *VE2* all returned significant effects in the corresponding models. This indicates that both Aitken's Law and the VE operate in FACE. The *SVLR1* categorization further interacts with the variables *phrasal position*, *stress* and *regional background*. The estimates reveal that the phonological effect of Aitken's Law is stronger in phrase-final positions and in syllables under nuclear stress, but weaker in the regions *Southern* and *West central*. The best model fit was, however, obtained for the model with the *VE2* categorization scheme and a summary of the model output is given in Table 49.

*Table 49. Best linear mixed model for the lexical set FACE fit by maximum likelihood (marginal $R^2 = 0.22$; conditional $R^2 = 0.49$). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ VE2 + position + stress + style + syl_per_sec + (VE2 * position) + (VE2 * stress) + (VE2 * gender) + gender + (1 | speaker) + (1 | syl_label) + (1 | word_label)*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	762.1	847.0	-365.0	730.1	1477

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.7605	-0.5942	0.0488	0.5664	5.2878

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.020	0.141
syl_label	(Intercept)	0.005	0.075
speaker	(Intercept)	0.017	0.132
Residual		0.078	0.279

Intercept: VE2long; positionfinal; stressnuclear; stylescripted; genderfemale

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.820e+00	337.0218	7.046e-02	< 2e-16	***
VE2short	-3.068e-01	0.735818	6.808e-02	7.21e-06	***
positioninitial	-4.031e-01	0.668265	5.811e-02	6.05e-12	***
positionmedial	-3.058e-01	0.736504	4.708e-02	1.14e-10	***
stressprimary	-3.300e-01	0.718919	3.415e-02	< 2e-16	***
styleunscripted	-6.625e-02	0.935894	2.998e-02	0.02843	*
syl_per_sec	-8.783e-02	0.915912	8.245e-03	< 2e-16	***
gendermale	-8.995e-02	0.913974	3.576e-02	0.01253	*
VE2short:positionmedial	1.636e-01	1.177723	5.891e-02	0.00556	**
VE2short:stressprimary	9.714e-02	1.102013	4.430e-02	0.02851	*
VE2short:gendermale	6.622e-02	1.068465	3.312e-02	0.04576	*

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

Significant fixed factors are the variables *VE2*, the *phrasal position*, *stress*, *style*, the local articulation rate in terms of *syllables per second* as well as the *gender* of the speakers. There are also interactions between *VE2* and the *phrasal position*, *VE2* and *stress* as well as between *VE2* and the *gender* of the speakers.

The estimates reveal that *VE2* short tokens are approximately 26 percent shorter than the *VE2* long tokens (= intercept). This shows that the VE is in operation in the vowel FACE in plosive contexts. Postvocalic voiced plosives lead to significantly longer FACE tokens than postvocalic voiceless plosives.

The monophthong is also affected by *constituent-final lengthening* (see subsection 3.1.8). Phrase-initial tokens are roughly 33 percent and phrase-medial tokens are 26 percent shorter than phrase-final tokens. The duration of FACE clearly increases if the syllable is positioned at the end of an utterance.

Likewise, FACE is also significantly longer when the vowel is positioned in a nuclear stressed syllable. The model output estimates that primary stressed syllables are approximately 28 percent shorter than FACE tokens under nuclear stress. This distribution was already visible in the plot and the model confirms that FACE is affected by *prosodic stress* (see subsection 3.1.7). Unstressed syllables are excluded in the VE2 categorization scheme so it is unclear whether FACE is also affected by lexical stress (see subsection 3.1.6).

Similar to the other vowels, FACE is also affected by *tempo* (see subsection 3.1.9). The estimates of the local articulation rate (*syllables per second*) reveal that the duration of the vowel decreases by 8.41 percent for every additional syllable that is produced in the timeframe of one second. This is in line with the overall trend (see section 5.1) and with the corresponding plot (see Figure 61): the faster the speech, the shorter the vowels.

Unlike most other vowels in this study, FACE is also significantly affected by the *style* of speech. According to the model output, the vowel is 6.42 percent shorter in unscripted speech forms than in scripted speech forms. This is in line with the expectation (see subsection 4.1.2) because unscripted speech is often faster than scripted speech. However, the variable *style* only reaches statistical significance in the postvocalic plosive contexts. The other models with the *SVLR1*, *SVLR2* and *VE1* categorization schemes excluded the variable *style* because it did not return significant effects. This means that FACE is only shorter in plosive contexts of unscripted speech.

Another unusual distribution can be found for *gender*. According to the estimates, the male speakers produce tokens which are overall 8.61 percent shorter than the tokens of the female speakers. Yet again, this finding might only be relevant for the plosive contexts because the best *SVLR1* and *SVLR2* models did not incorporate *gender* as a significant variable.

Apart from that, there are several significant interactions which include the *VE2* short contexts. When compared to the intercept, the *VE2* short tokens are roughly 17 percent longer in phrase-medial positions and therefore more similar to the corresponding *VE2* long contexts in phrase-medial positions. This means that the durational difference between the *VE2* long and short tokens is less pronounced than the durational difference between the *VE2* levels in phrase-final positions. This trend was also visible in the corresponding plot (see Figure 60) and the model confirms that the VE-related lengthening effect is significantly amplified in phrase-final positions.

The interaction between the *VE2* short contexts and primary stressed syllables describes a similar relationship. The short contexts are 10 percent longer, so the durational difference between the *VE2* levels is less pronounced in the primary stressed syllables than in the nuclear stressed syllables. This discrepancy was also visible in the corresponding plots (see Figure 60): the difference between the long and short contexts is greater in nuclear stressed syllables than in primary stressed syllables.

The last significant interaction involves the *VE2* and *gender*. Male speakers do not only produce shorter FACE tokens overall, but the discrepancy between the long and short contexts is also weaker than the durational difference of the levels for the female speakers. The *VE2* effect is therefore weakened for the male speakers in postvocalic plosive environments.

The plots and models have shown that both the VE and Aitken's Law operate in the long monophthong FACE. The vowels are clearly longer in voiced fricative contexts (*SVLR1* - phonological conditioning of Aitken's Law) and before morpheme boundaries (*SVLR2* – morphological conditioning of Aitken's Law). At the same time, the models also conveyed that the long and short contexts of the categorizations *VE1* and *VE2* lead to significantly different vocalic durations. This shows that FACE is also affected by the VE. Apart from that, FACE is also influenced by the intralinguistic variables *tempo, stress* and *phrasal position.* The variables *gender* and *style* have a further significant influence on FACE in the plosive environments. In accordance with previous studies, the models and interactions have shown that SVLR- or VE-related lengthening effects are amplified in prominent prosodic positions (Chevalier, 2019; Rathcke & Stuart-Smith, 2016).

The SVLR effects in FACE are surprising because most previous investigations could not report patterns of Aitken's Law in this monophthong. The impressionistic accounts on Scottish vowel duration (see subsection 3.2.1) note that this vowel varies in terms of its quantity in different postvocalic contexts (Dieth, 1932; Grant, 1931; Grant & Dixon, 1921; Murray, 1873; G. Watson, 1923; Wettstein, 1942; Wölck, 1965; Zai, 1942). McClure (1977) found that FACE is not influenced by the VE and the effects of Aitken's Law are less strong in this vowel than in /i/ or /u/. Scobbie (2005) as well as Watt and Ingham (2000) conclude that /e/ is less strongly affected by the SVLR and McKenna (1988) and Warren (2018) emphasize that this vowel is rather influenced by the VE than by Aitken's Law. Watt and Yurkova (2007) found that, in contrast to the other vowels, FACE conforms most consistently to both the VE and the SVLR and the latest study by Stuart-Smith et al. (2019) also concludes that the monophthong is affected by Aitken's Law only in the Northeast of Scotland. Stuart-Smith et al. (2019) found further evidence of an "anti-Voicing Effect" effect in this monophthong. The findings of the present investigation are fully in line with the results by Watt and Yurkova (2007): FACE is significantly longer in phonological and morphological SVLR long contexts than in the respective short contexts, but the monophthong is also longer in VE long environments than in the respective short environments. The study could find no incidence of an “anti-Voicing Effect” because voiced consonants lead to longer FACE tokens than voiceless consonant contexts. The difference between the fricative contexts is, however, much more pronounced than the difference between the plosive contexts.

5.3.5 GOAT

GOAT represents the long monophthong /o/ in Scottish English. In contrast to RP or SSBE, this vowel is not diphthongized, so the words <no> and <go> are pronounced with a relatively stable vowel quality in SSE. A summary of the vowel duration of GOAT can be found in Table 50 and a visualization of the data is displayed in Figure 63.

Table 50. Statistical summary of vowel duration for the lexical set GOAT. Tokens below 40 ms were excluded.

Token number:	6708
Average duration:	119.10 ms
Standard deviation:	70.32 ms
Minimal duration:	40.00 ms
Maximal duration:	1120.00 ms
1st quantile:	70.00 ms
Median:	100.00 ms
3rd Quantile:	141.40 ms

The 6708 GOAT tokens have an average duration of 119.10 ms and a relatively high standard deviation of 70.32 ms. The minimum duration is 40 ms and the longest GOAT token is 1120 ms long. Most of the durations lie in the range between 70 ms and 141.40 ms and the median is located at 100 ms.

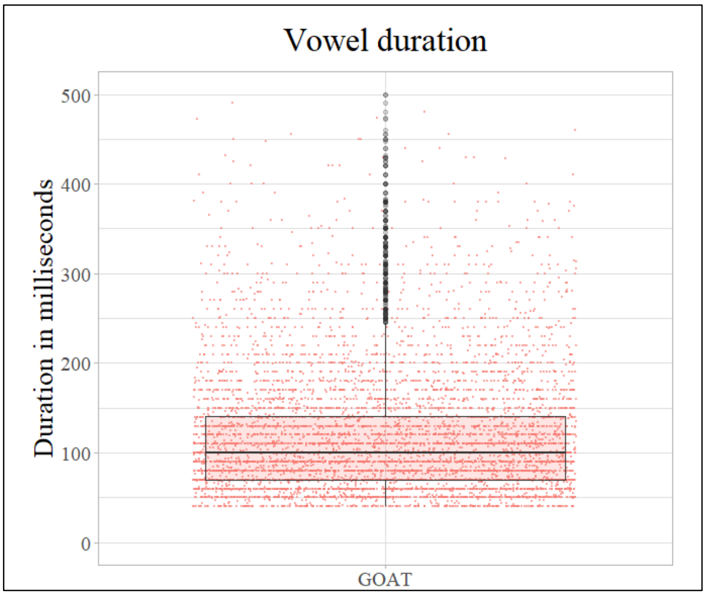


Figure 63. Vowel duration of the lexical set GOAT in milliseconds. Durations below 40 ms were excluded.

The plots for the *SVLR1*, *SVLR2*, *VE1* and *VE2* categories show regular distributions (see Figure 64). The vowel durations in the long contexts are longer than those in the short contexts across all categorization schemes. One can see that the difference between the boxplots is slightly more pronounced for the SVLR categories than for the VE categories, but all long contexts lead to longer GOAT tokens. The plots therefore indicate that GOAT is affected by the SVLR and the VE. It is, however, unclear whether the durational differences reach statistical significance.

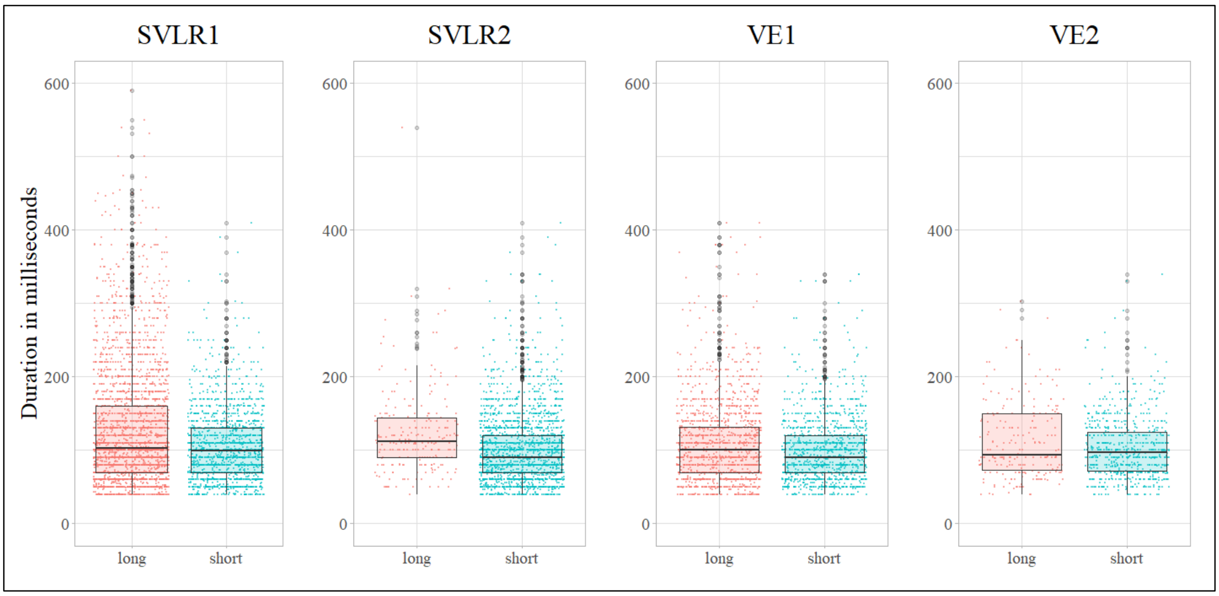


Figure 64. Boxplots and jitterplots of raw vowel duration of the lexical set GOAT for the categories SVLR1, SVLR2, VE1 and VE2.

The average durations for the postvocalic consonant environments also indicate that GOAT might be affected by Aitken’s Law (see Table 51). Apart from postvocalic pauses, GOAT tokens are on average the longest in voiced fricative contexts. The ratio of the fricative contexts is also much higher than the ratio of the plosive contexts. Whereas voiced plosive environments are only 3.83 milliseconds longer than voiceless plosives, the durational difference between the voiced and voiceless fricatives amounts to 15.74 ms. GOAT tokens are relatively long when followed by laterals but remarkably short in nasal environments. These distributions indicate that GOAT might be influenced by the phonological conditioning of Aitken’s Law.

Table 51. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set GOAT. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	Average vowel duration
laterals	115.52 ms
nasals	103.01 ms
voiced fricatives	117.18 ms
voiceless fricatives	101.44 ms
voiced plosives	105.95 ms
voiceless plosives	102.12 ms
pauses	205.24 ms

I also plotted the SVLR variables for different intralinguistic and extralinguistic factors and found some interesting distributions for the variables *regional background*, *phrasal position* and *stress* (see

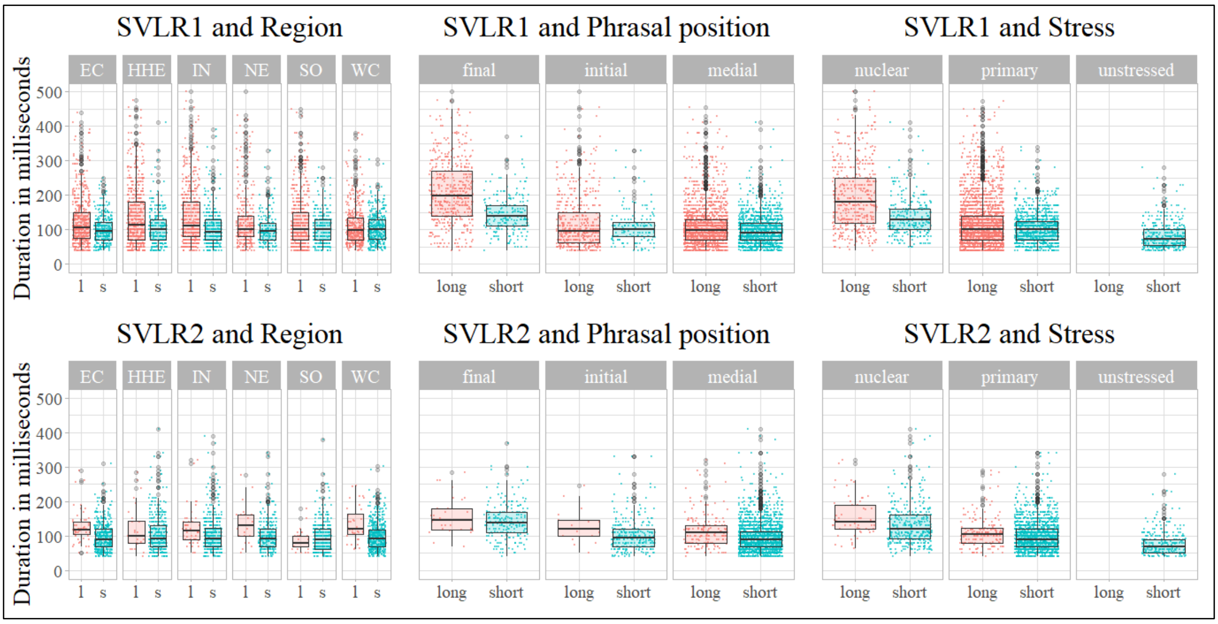


Figure 65. Boxplots and jitterplots of raw vowel duration of the lexical set GOAT for the categories SVLR1 and SVLR2 separated for the variables regional background, phrasal position and stress.

Figure 65). The boxplots for the *SVLR1* long environments are clearly higher than the boxplots of the short contexts in the dialect regions *HHE* and *Insular*. The boxplots for *West central*, however, are on a very similar level. It thus seems that the phonological conditioning of Aitken's Law affects GOAT differently in different regions. The plots also show that there is a clear interaction between Aitken's Law and the variables *phrasal position* and *stress*. The difference between the *SVLR1* long and short contexts is clearly more pronounced in phrase-final positions and in syllables under nuclear stress. The boxplots for the initial and medial positions as well as the boxplots for primary stress are on a very similar level. This indicates that the phonological effects of Aitken's Law are more pronounced in prominent prosodic positions. Apart from that, the duration of GOAT itself is also generally higher in nuclear and phrase-final syllables. The plots for the morphological conditioning of Aitken's Law (*SVLR2*) show a different picture. As for the regions, the effects of Aitken's Law are particularly strong in the *Northeast* and in the dialect area *West central*. In contrast, the regional background *Southern* has shorter GOAT tokens in *SVLR2* long contexts than in *SVLR2* short contexts. This indicates that the morphological conditioning of Aitken's Law does not operate in GOAT in *Southern* Scotland. Contrary to the *SVLR1*, the difference between the *SVLR2* long and short contexts does not strikingly vary according to the *phrasal position* or *stress*. The *SVLR2* long contexts are longer than the short contexts, but the durational difference does not vary greatly in phrase-final, phrase-medial, or phrase-initial positions nor in syllables under nuclear or primary stress. The morphological conditioning of Aitken's Law does therefore not seem to vary in different prosodic positions but vowel duration is generally longer in phrase-final and nuclear stressed syllables.

The same trend can be seen in the plots for the intralinguistic variables (see Figure 66): phrase-final and nuclear stressed syllables are clearly longer than non-final, primary stressed, or unstressed GOAT tokens. The plots therefore indicate that GOAT is affected by *constituent-final lengthening* (see subsection 3.1.8) and *prosodic stress* (see subsection 3.1.7). While primary stressed syllables are slightly longer than unstressed syllables, it is unclear whether GOAT is also significantly affected by *lexical stress* (see subsection 3.1.6) in spoken SSE.

Similar to the other vowels, GOAT is also clearly influenced by *tempo* (see subsection 3.1.9). The *syllables per second* plot shows that GOAT tokens become shorter in higher articulation rates. The regression line has a clear negative slope which fits into the overall trend: vowels become shorter if the speech is faster.

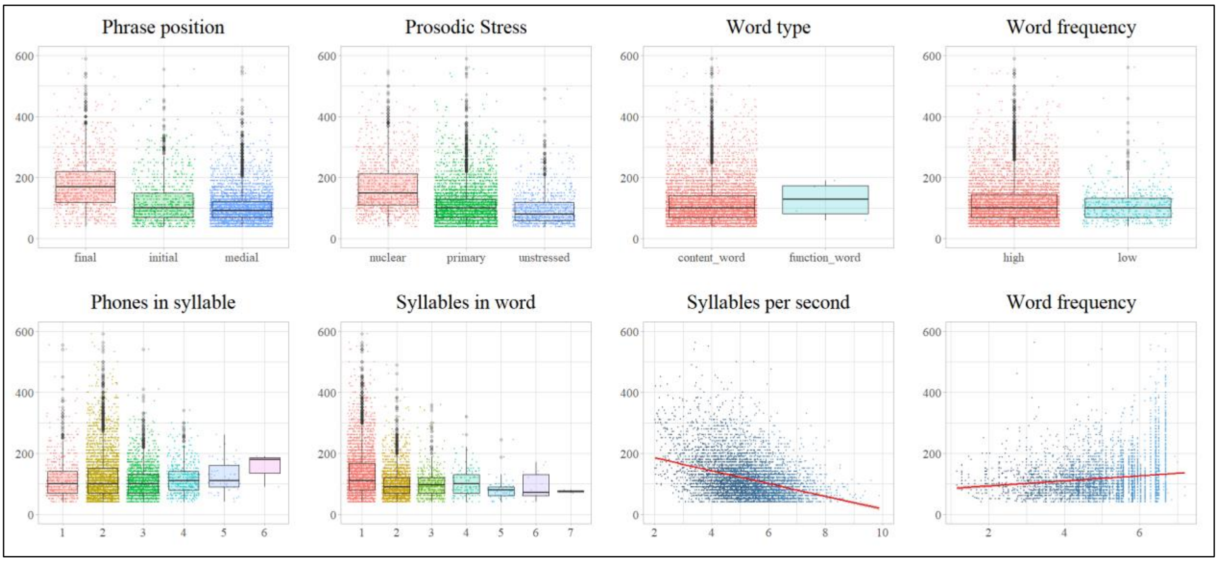


Figure 66. Plots of raw vowel duration for the lexical set GOAT separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

The plots for word type and word frequency show unusual distributions. The boxplots indicate that GOAT tokens are longer in function words than in content words. However, the dataset only comprises four function words so the durations might not be fully representative. The categorical plot for word frequency shows that high and low frequency words have similar durations because both boxplots are on the same level. The linear plot for the same variable indicates that high frequency words are slightly longer than low frequency words because the red regression line has a mildly positive slope. These distributions are therefore clearly not in line with the results of previous studies on the effect of *lexical category* and *lexical frequency* (see subsection 3.1.5). Function words and high frequency words are frequently reduced and they should therefore be shorter than content words and low frequency words in connected speech. The results of this study, however, show different distributions for GOAT.

The plots for the number of phones in a syllable and the number of syllables in a word do not show clear trends. Instead of a shortening, the duration of GOAT slightly increases in syllables with many phones. It is possible that the long durations in multi-phone syllables are a result of low token numbers, but the overall vowel duration does not decrease with a higher *syllable phone count*. It is therefore unclear whether GOAT is significantly affected by *intrasyllabic compression* (see subsection 3.1.3). The plot for the *word syllable count* is more regular because vowel duration shows an overall decrease from monosyllabic words to seven syllable words, but the decrease does not follow a steady pattern. For example, GOAT is longer in four-syllable words than in three-syllable words, so it is not clear whether the vowel is significantly affected by *polysyllabic shortening* (see subsection 3.1.4).

I also plotted GOAT for the different extralinguistic variables *gender*, *regional background*, *age group* and I plotted the *mean vowel duration* against the age of the speakers. The plots did not convey any irregular distributions and this is why they are not shown here. The duration of GOAT therefore appears to be stable across different *genders*, *regional backgrounds* and *age groups* in Scotland.

The linear mixed effects models for GOAT excluded the variables *gender* and *lexical frequency* in all possible constellations. This means that the duration of the vowel does not significantly vary with respect to *lexical frequency* or the *gender* of the speakers. The best model was fit with the *SVLR1* categorization and a summary of the model output can be found in Table 52.

*Table 52. Best linear mixed model for the lexical set GOAT fit by maximum likelihood (marginal $R^2 = 0.32$; conditional $R^2 = 0.41$). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ SVLR1 + (SVLR1 * position) + (SVLR1 * stress) + position + stress + num_syl_word + num_pho_syl + syl_per_sec + (1 | speaker) + (1 | syl_label) + (1 | word_label)*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	4485.3	4588.1	-2226.6	4453.3	4564

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.6928	-0.6308	0.0331	0.6178	3.9746

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.0122	0.110
syl_label	(Intercept)	0.0050	0.0707
speaker	(Intercept)	0.0050	0.0708
Residual		0.1457	0.3817

Intercept: SVLR1long; positionfinal; stressnuclear; genderfemale

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	6.197e+00	491.102	6.843e-02	< 2e-16	***
SVLR1short	-3.443e-01	0.70871	4.868e-02	2.18e-12	***
positioninitial	-4.995e-01	0.60681	2.599e-02	< 2e-16	***
positionmedial	-4.220e-01	0.65571	2.186e-02	< 2e-16	***
stressprimary	-4.133e-01	0.66143	2.025e-02	< 2e-16	***
stressunstressed	-3.882e-01	0.67825	4.148e-02	< 2e-16	***
num_syl_word	-2.693e-02	0.97342	1.293e-02	0.0377	*
num_pho_syl	-4.694e-02	0.95414	1.837e-02	0.0115	*
syl_per_sec	-1.320e-01	0.87635	5.582e-03	< 2e-16	***
SVLR1short:positioninitial	2.319e-01	1.26096	5.251e-02	1.03e-05	***
SVLR1short:positionmedial	2.356e-01	1.26569	3.861e-02	1.14e-09	***
SVLR1short:stressprimary	1.854e-01	1.20371	3.387e-02	4.63e-08	***

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

Significant fixed factors include the *SVLR1* categorization, the *phrasal position*, *stress*, the *number of syllables in a word*, the *number of phones in a syllable* as well as the local articulation rate in terms of *syllables per second*. The model output further includes significant interactions between Aitken’s

Law and the *phrasal position* as well as between SVLR and *stress*. The influence of *style*, the *age group* and *regional background* turned out to be insignificant, so the phonological conditioning of the SVLR does not vary in terms of the speakers' dialect background or age group nor in scripted or unscripted speech forms.

The model corroborates the observation that the duration of GOAT varies with respect to the phonological conditioning of Aitken's Law. Tokens in *SVLR1* short contexts are roughly 30 percent shorter than tokens in *SVLR1* long environments (= intercept). At the same time, the model output also confirms that the effect of Aitken's Law is stronger in prominent prosodic positions. The *SVLR1* short tokens are roughly 26 percent longer in phrase-initial and phrase-medial positions when compared to the intercept. This means that the durational difference between the long and short environments is 26 percent smaller than the respective difference in the phrase-final syllables (= intercept). Likewise, the *SVLR1* short tokens are approximately 20 percent longer in primary stressed syllables when compared to the intercept. Thus, the difference between the tokens in long and short contexts is significantly stronger in nuclear stressed syllables (= intercept).

The model output also shows that the *phrasal position* generally has a significant effect on the duration of GOAT. The model estimates that phrase initial syllables are approximately 40 percent and phrase-medial syllables are 35 percent shorter when compared to the intercept.

Similarly, GOAT is also significantly affected by *stress*. Primary stressed syllables are approximately 34 percent shorter and unstressed syllables are roughly 32 percent shorter when compared to the intercept. This corroborates the influence of *prosodic stress* (see subsection 3.1.7). Yet, the effect size is weaker for unstressed syllables than for the primary stressed syllables, so it seems that GOAT is not affected by *lexical stress* (see subsection 3.1.6).

While the plots did not provide a clear distribution, the model output indicates that the *syllable phone count* and the *word syllable count* have a significant influence on the duration of GOAT. According to the estimates, vowel duration decreases by 2.66 percent for every additional syllable in a word. Likewise, vowels tend to become 4.59 percent shorter for every additional phone in a syllable. This means that the vowel GOAT is affected by *polysyllabic shortening* (see subsection 3.1.4) and *intrasyllabic compression* (see subsection 3.1.3) in the *SVLR1* environments. Nevertheless, the effect sizes are relatively small and the other models with the *SVLR2*, *VE1* and *VE2* categorization schemes all excluded the variable *number of phones in a syllable*. This means that the *syllable phone count* is only significant in phonological environments of Aitken's Law.

Similar to the other vowels, the duration of GOAT is also significantly influenced by the local articulation rate. The estimates reveal that, for every additional syllable produced in the timeframe of one second, the monophthong is shortened by 12.37 percent. Hence, GOAT is shorter in fast speech and longer in slow speech which is in line with the effect of *tempo* (see subsection 3.1.9).

As the other models included the categories *SVLR2*, *VE1* and *VE2* in the respective output, it is possible to say that the monophthong GOAT is affected by the phonological and morphological conditioning of Aitken’s Law as well as by the VE. The durational difference between SVLR long and short contexts reaches statistical significance and this difference is also visible in the plots. The effects of the phonological SVLR are overall stronger in phrase-final and nuclear stressed syllables than in non-final or non-nuclear positions. This shows that the effects of Aitken’s Law on this vowel are amplified in prominent prosodic positions. The extralinguistic variables, however, do not have a great influence on the VE or SVLR, nor to vowel duration itself. Hence, GOAT is affected by a relatively stable VE and SVLR.

This trend was, however, not observed in previous studies. Most previous studies only report a consistent VE, but no previous investigation found clear SVLR effects. The impressionistic accounts on Scottish vowel duration (see subsection 3.2.1) describe that the duration of this vowel varies in different postvocalic contexts (Dieth, 1932; Grant, 1931; Grant & Dixon, 1921; Murray, 1873; G. Watson, 1923; Wettstein, 1942; Wölck, 1965; Zai, 1942). Aitken (1981) does not explicitly state that the monophthong is affected by the SVLR. The subsequent empirical studies provide similar outcomes. McClure (1977) found out that the effects of Aitken’s Law are less strong in GOAT and that the vowel is influenced by a variable VE in his experiment. McKenna (1988) as well as Scobbie, Hewlett, and Turk (1999) conclude that GOAT is affected by the VE and not by the SVLR. The following studies by Watt and Ingham (2000) and Scobbie (2005) also describe that Aitken’s Law is relatively weak in the long monophthong. Watt and Yurkova (2007) as well as Warren (2018) emphasize that GOAT is affected by a strong and consistent VE in the Northeast of Scotland. The most recent study by Stuart-Smith et al. (2019), however, found evidence of an “anti-Voicing Effect” in this vowel. In contrast to previous studies, GOAT is significantly longer before voiced than before voiceless consonants in contemporary SSE. The present investigation has shown that the monophthong is clearly affected by Aitken’s Law but also by the VE.

5.3.6 CAT

The lexical set CAT represents the long open vowel /a/ in SSE. In contrast to RP and SSBE, there is no quality difference between the vocalic nuclei in words such as <bad> and <balm>, so SSE is not affected by the TRAP-BATH split (see section 2.3). A brief statistical summary of the duration of CAT is listed in Table 53 and visualized in Figure 67.

The lexical set CAT is the longest monophthong in this study with an average duration of 119.71 ms. The standard deviation is 51.23 ms and most of the data points lie in the range between 85.51 ms and 142.42 ms. The shortest CAT vowels last 40 ms and the longest token is 600 ms long. The median is 110 ms.

Table 53. Statistical summary of vowel duration for the lexical set CAT. Tokens below 40 ms were excluded.

Token number:	10868
Average duration:	119.71 ms
Standard deviation:	51.23 ms
Minimal duration:	40.00 ms
Maximal duration:	600.00 ms
1st quantile:	85.51 ms
Median:	110.00 ms
3rd Quantile:	142.42 ms

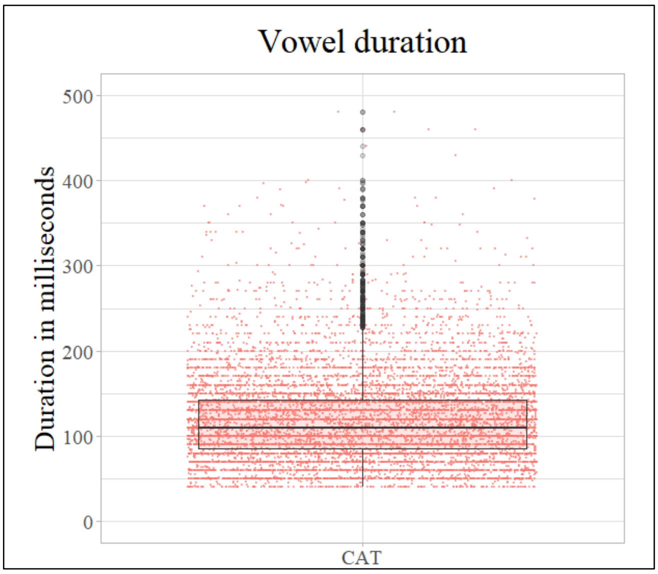


Figure 67. Vowel duration of the lexical set CAT in milliseconds.
Durations below 40 ms were excluded.

The boxplots for the variables representing Aitken’s Law and the VE do not show clear lengthening effects for all categories (see Figure 68). The boxplots for the *SVLR1* and *VE1* are on the same level and do not show striking durational differences. The average durations provide further evidence that CAT is not influenced by the typical SVLR or VE-related lengthening effects: vowel duration is on average

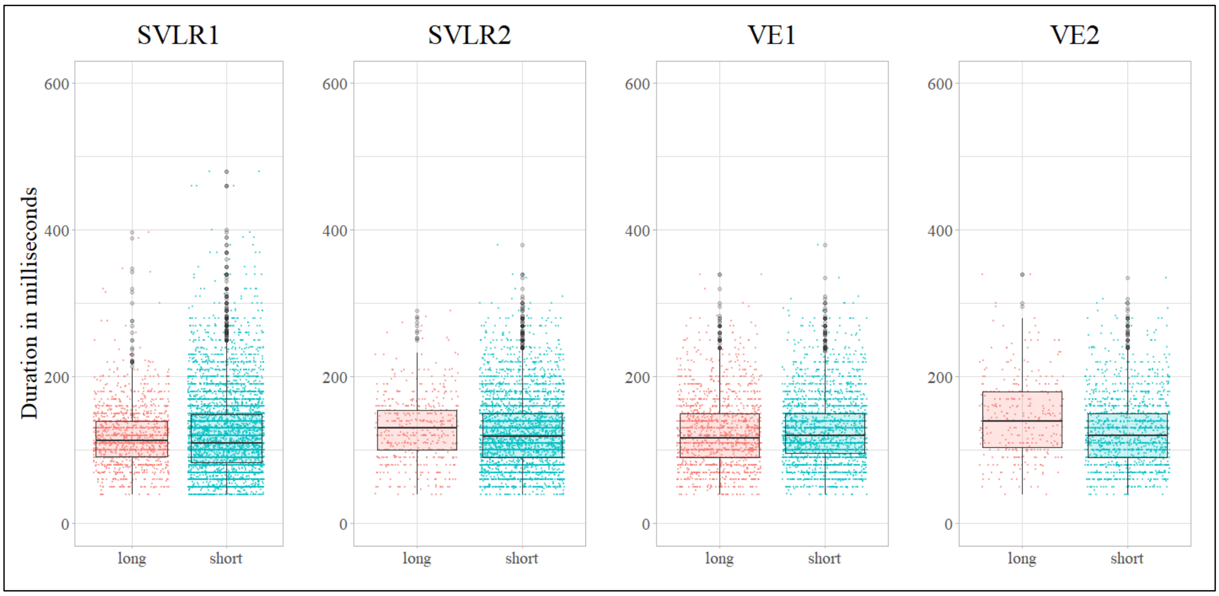


Figure 68. Boxplots and jitterplots of raw vowel duration of the lexical set CAT for the categories SVLR1, SVLR2, VE1 and VE2.

shorter in the long environments than in the respective short contexts for *SVLR1* and *VE1* (*SVLR1* long: 119.18 ms; *SVLR1* short: 121.05 ms; *VE1* long: 123.24 ms; *VE1* short: 125.79 ms). More regular effects can only be found for the *SVLR2* and *VE2* categorizations. The *SVLR2* long contexts are slightly longer than the respective short environments. This can be seen in the boxplots and in the average durations (*SVLR2* long: 128.96 ms; *SVLR2* short: 123.22 ms). Yet, the durational difference is not very great, so it is unclear whether it is statistically significant. A more pronounced durational difference can be seen for the *VE2* categorization. The boxplot of the *VE2* long contexts is clearly higher than the respective boxplot for the short environments and the average durations are almost 20 ms apart (*VE2* long: 144.45 ms; *VE2* short: 124.58 ms). This indicates that CAT is affected by the VE in plosive contexts. More information on the mean duration of CAT in other postvocalic contexts is summarized in Table 54.

Table 54. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set CAT. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	**Average vowel duration**
laterals	112.33 ms
nasals	114.40 ms
voiced fricatives	105.35 ms
voiceless fricatives	128.69 ms
voiced plosives	130.40 ms
voiceless plosives	123.65 ms
pauses	187.32 ms

Overall, vowels are clearly shorter in voiced fricative than in voiceless fricative contexts. This distribution not only contradicts the phonological conditioning of the SVLR, but it is also not in line with the VE. CAT might therefore not be affected by the SVLR or the VE in fricative environments. The nasal and lateral contexts are longer than the voiced fricative contexts but clearly shorter than the plosive environments. In fact, CAT has the longest average duration when it is followed by voiced plosives. The average durations show that, while the VE is established in the plosive contexts, its influence might not be as strong in the other environments. The phonological conditioning of Aitken's Law does not seem to apply at all. Apart from that, CAT is most likely influenced by *constituent-final lengthening* (see subsection 3.1.8) due to the increased duration before pauses.

I also plotted the VE and SVLR categories for different intra- and extralinguistic variables. The long and short contexts have comparable durations across the different *age groups* and *genders*, so it is unlikely that these variables interact with the SVLR or the VE. More interesting distributions could be found for the *SVLR1* and *SVLR2* categorization schemes when plotted for *regional background*, *phrasal position* and *stress* (see Figure 69). The plots show that the phonological conditioning of Aitken's Law (*SVLR1*) does not strikingly differ for the different regions. The *SVLR1* long environments are not strikingly longer than the short environments. However, the region *Northeast* has shorter vowels in *SVLR1* long contexts than in *SVLR1* short environments. A more striking discrepancy can be observed for the *SVLR1* in different *phrasal positions*. The *SVLR1* long contexts are clearly longer than the

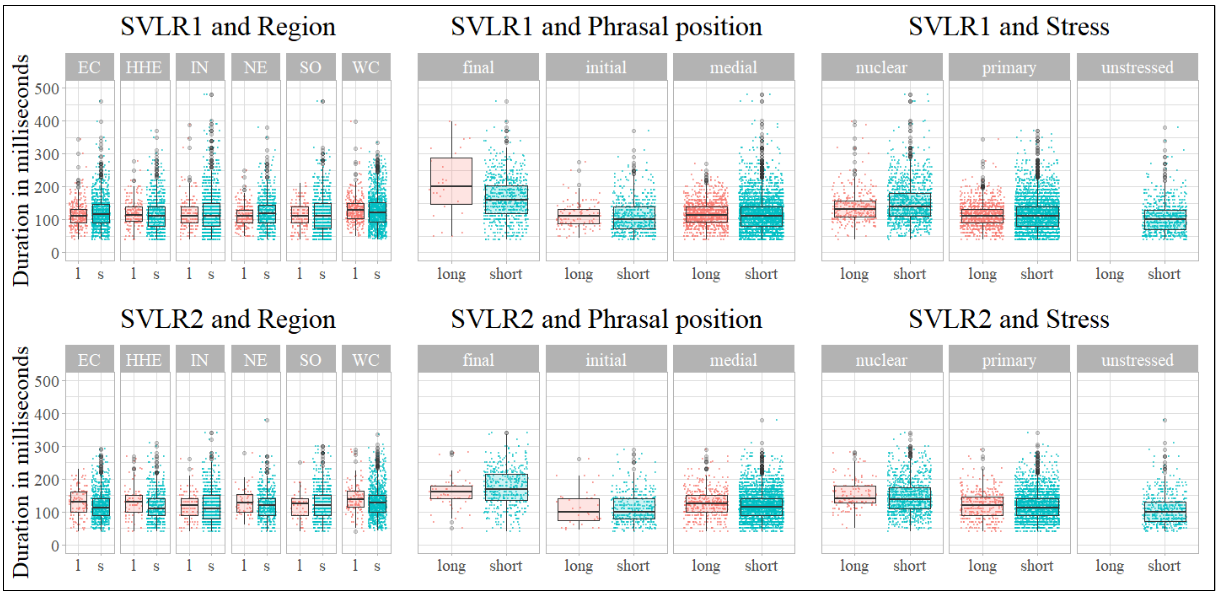


Figure 69. Boxplots and jitterplots of raw vowel duration of the lexical set CAT for the categories SVLR1 and SVLR2 separated for the variables regional background, phrasal position and stress.

respective short contexts in final utterance positions. The duration of CAT does, however, not differ in the long and short environments in initial or medial positions for *SVLR1*. Apart from that, the tokens in final positions are also generally longer than those in non-final syllables. The boxplots for the *SVLR1* and *stress* are very balanced. All boxplots are on a comparable level, the vowels in the long environments are not strikingly longer than the tokens in the short contexts. In fact, the long contexts are slightly shorter than the short contexts in nuclear and primary stressed syllables. The tokens under nuclear stress are generally longer but there is no durational difference that can be attributed to the phonological conditioning of Aitken's Law. The *SVLR2* scheme shows different distributions. The *SVLR2* long contexts are slightly longer than the respective short environments in most dialect regions. Only the regions *Insular* and *Southern* have longer vowels in *SVLR2* short contexts than in *SVLR2* long environments. This indicates that the morphological conditioning of Aitken's Law does not seem to apply in GOAT in Shetland, Orkney and the South of Scotland. The plots for the *phrasal positions* convey irregular distributions. While phrase-final tokens are generally longer than non-final tokens, the *SVLR2* long contexts are clearly shorter than the respective short environments in phrase-final positions. This clearly contradicts Aitken's Law. The long and short contexts are on a very similar level in phrase-initial and phrase-medial positions. Hence, there seems to be no interaction between the *SVLR2* categorization and the *phrasal position*. As for *stress* and as mentioned above, CAT is generally longer in nuclear stressed syllables than in primary or unstressed syllables. The difference between the *SVLR2* levels is, however, not very pronounced. The long contexts are slightly longer than the short environments, but it is unclear whether this difference reaches statistical significance.

The plots for the intralinguistic variables (see Figure 70) corroborate the observation that CAT is affected by *constituent-final lengthening* (see subsection 3.1.8) and *prosodic stress* (see subsection 3.1.7). The CAT tokens in final positions are clearly longer than those in initial or medial positions.

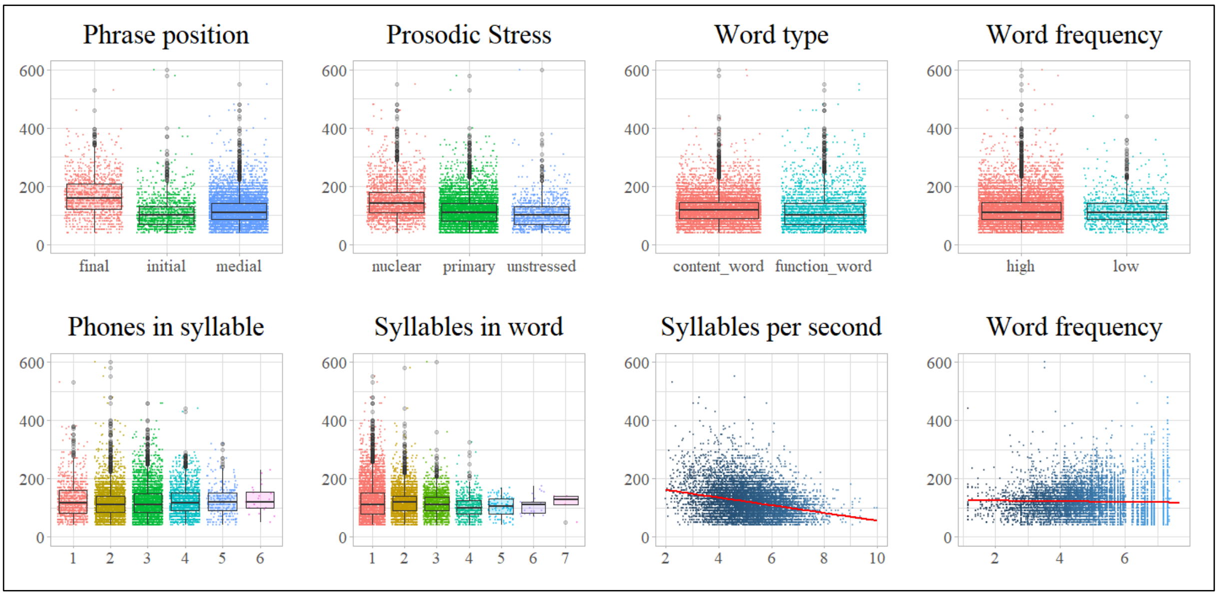


Figure 70. Plots of raw vowel duration for the lexical set CAT separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

Likewise, vowels in nuclear stressed syllables are clearly longer than those in primary or unstressed syllables. The primary stressed syllables are also longer than the unstressed syllables, but it is unclear whether the difference is statistically significant.

Furthermore, the duration of CAT differs with respect to *word type*, the boxplots indicate that content words are slightly longer than function words which corresponds to the effect of *lexical category* (see subsection 3.1.5). The categorical and linear plots for word frequency, however, do not show any specific distributions. The boxplots are on the same level and the slope of the red regression line in the linear plot is constant. The plots therefore suggest that *lexical frequency* (see subsection 3.1.5) does not influence the duration of CAT.

The plots for the *syllable phone count* and *word syllable count* show that long CAT tokens become less frequent in syllables with many phones and in words with many syllables. However, the boxplots and average values do not steadily decrease. It is therefore unclear whether CAT is significantly affected by *intrasyllabic compression* (see subsection 3.1.3) or *polysyllabic shortening* (see subsection 3.1.4).

The durational plots for CAT against the extralinguistic variables *gender*, *regional background* and *age group* did not convey any specific distributions and this is why they are not shown here. The duration of CAT seems to be a bit shorter in the dialect region *Southern* but there is great variation overall. The duration of CAT does not seem to vary in the different *age groups*, *genders* or across the *regional backgrounds*.

The conditional R^2 values of the different linear mixed effects models lie in the range between 0.40 and 0.48. All models excluded the variables *word frequency*, *style* and *age group* which means that CAT is not significantly affected by these variables. The plots already showed no clear effects of *lexical frequency* (see Figure 70) and the models confirm that this variable is not significant. CAT does also not significantly vary in scripted and unscripted speech forms and the durations are similar across different generations in SSE. Furthermore, the variables *SVLR1* and *VE1* turned out to be insignificant in the respective models. This means that CAT is not affected by the phonological conditioning of Aitken’s Law and that the VE does not operate in all consonantal contexts. The best fit was obtained with the model applying the *SVLR2* categorization scheme and a summary of the output can be found in Table 55.

*Table 55. Best linear mixed model for the lexical set CAT fit by maximum likelihood (marginal $R^2 = 0.20$; conditional $R^2 = 0.48$). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ SVLR2 + position + stress + num_syl_word + num_pho_syl + syl_per_sec + (SVLR2 * reg) + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label)*

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	1730.8	1879.4	-842.4	1684.8	4692

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-5.4935	-0.5009	0.0480	0.5720	4.3117

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.0159	0.1262
syl_label	(Intercept)	0.0132	0.1149
speaker	(Intercept)	0.0086	0.0932
Residual		0.0683	0.2613

Intercept: SVLR2long; positionfinal; stressnuclear; regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.851e+00	347.555	6.036e-02	< 2e-16	***
SVLR2short	-8.525e-02	0.91828	3.427e-02	0.01293	*
positioninitial	-2.931e-01	0.74593	2.070e-02	< 2e-16	***
positionmedial	-2.136e-01	0.80768	1.602e-02	< 2e-16	***
stressprimary	-1.359e-01	0.87290	1.044e-02	< 2e-16	***
stressunstressed	-2.530e-01	0.77647	2.070e-02	< 2e-16	***
num_syl_word	-6.062e-02	0.94117	8.033e-03	8.38e-14	***
num_pho_syl	-7.189e-02	0.93063	1.003e-02	2.52e-12	***
syl_per_sec	-6.769e-02	0.93455	4.601e-03	< 2e-16	***
regInsular	-1.710e-01	0.84280	5.534e-02	0.00207	**
SVLR2short:regInsular	1.452e-01	1.15628	4.850e-02	0.00277	**

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

The model output includes the significant fixed factors *SVLR2*, *phrasal position*, *stress*, the *number of phones in a syllable*, the *number of syllables in a word*, the local articulation rate in terms of *syllables per second* as well as the *regional background* of the speakers. The *SVLR2* further interacts with the dialect region *Insular*.

The *SVLR2* categorization reaches statistical significance. *SVLR2* short contexts are 8.18 percent shorter when compared to the intercept, so there is a mild effect that can be attributed to the morphological conditioning of Aitken’s Law. However, this effect is not stable across all regions. The interaction between *SVLR2* short contexts and the dialect region *Insular* shows that the durational difference between the *SVLR2* levels is not significant in Orkney or Shetland.

Apart from that, the model confirms that CAT is significantly influenced by the *phrasal position* and *stress*. Phrase-medial tokens are approximately 20 percent and phrase-initial vowels are roughly 25 percent shorter than phrase-final tokens (= intercept). Hence, the model confirms that CAT is affected by *constituent-final lengthening* (see subsection 3.1.8). Similarly, primary stressed syllables are roughly 13 percent and unstressed syllables are approximately 23 percent shorter than nuclear stressed syllables (= intercept). This means that *prosodic stress* (see subsection 3.1.7) has an influence on the duration of CAT as well. The different effect sizes of primary stressed syllables and unstressed syllables further show that CAT might also be affected by *lexical stress* (see subsection 3.1.6). This tendency was already visible in the corresponding plot (see Figure 70).

The model output further confirms that CAT is significantly influenced by *intrasyllabic compression* (see subsection 3.1.3) and *polysyllabic shortening* (see subsection 3.1.4). According to the estimates, CAT tokens are roughly 6 percent shorter for each additional syllable in a word and approximately 7 percent shorter for each additional phone in a syllable.

The local articulation rate has a significant influence on CAT as well. Vowel duration decreases by 6.55 percent for each additional syllable produced in the timeframe of one second. This fits into the overall trend that vowels become shorter in fast speech (see section 5.1).

The only significant extralinguistic variable is the *regional background* of the speakers. The model output specifies that speakers from the dialect region *Southern* produce roughly 10 percent shorter CAT vowels than the speakers from the region *East central* (= intercept). This distribution is in line with the general trend that *Southern* speakers produce shorter vowels overall (see section 5.1). The other regions do not reach statistical significance in the model output.

The plots and models have shown that CAT is not affected by the phonological conditioning of Aitken's Law. There is only a very mild SVLR effect when vowels are followed by morpheme boundaries. However, the morphological conditioning of the SVLR does not apply in Orkney or Shetland. The VE does only operate consistently in the plosive environments, the effect is, however, not stable in other postvocalic consonant environments. In particular, the vowel is shortest before voiced fricatives which contradicts Aitken's Law and the VE. The timing patterns of CAT are therefore not easy to categorize in contemporary spoken SSE and the results of this study are therefore broadly in line with those of previous investigations (see subsections 3.2.2 and 3.2.3).

5.4 Diphthongs

This section comprises the findings for the SSE diphthongs MOUTH, PRICE and CHOICE. Similar to the long monophthongs (see section 5.3), I will exclude tokens below 40 ms to avoid reduced realizations. Diphthongs are generally considered to be long in English (Roach, 2010, p. 17) and tokens below 40 ms are very unlikely to be realized with a diphthongal quality.

5.4.1 MOUTH

The lexical set MOUTH represents the diphthong /ʌʊ/ in SSE. This vowel has a raised onset when compared to the equivalent diphthong /aʊ/ in RP. A statistical overview of the vocalic durations of MOUTH can be found in Table 56 and a visualization of the data can be found in Figure 71.

Table 56. Statistical summary of vowel duration for the lexical set MOUTH. Tokens below 40 ms were excluded.

Token number:	2992
Average duration:	114.9 ms
Standard deviation:	56.95 ms
Minimal duration:	40.00 ms
Maximal duration:	660.00 ms
1st quantile:	70.00 ms
Median:	108.30 ms
3rd Quantile:	145.90 ms

The 2992 MOUTH tokens have an average duration of 114.9 ms and a standard deviation of 56.95 ms. Most of the measurements lie in the range between 70 and 145.90 ms and the median is 108.30 ms. The shortest pronunciations last for 40 ms and the longest token is 660 ms long. The lexical set MOUTH is therefore the shortest diphthong and the average duration even falls below the mean durations of the monophthongs CAT (see subsection 5.3.6) and GOAT (see subsection 5.3.5).

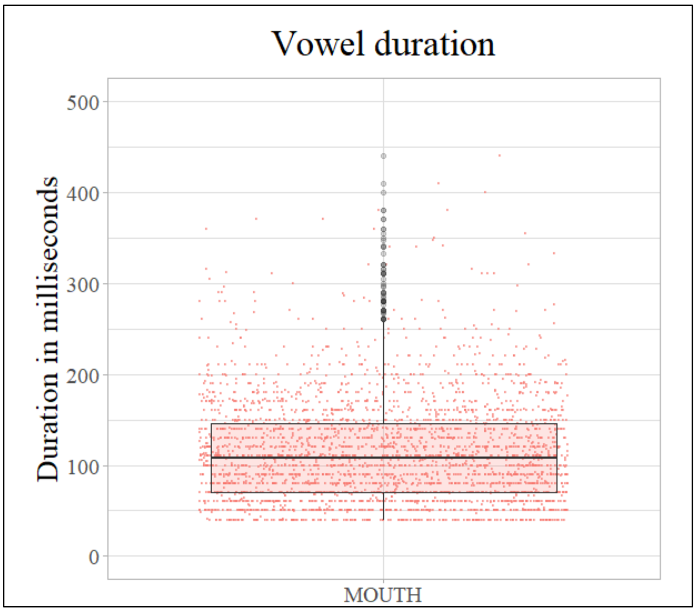


Figure 71. Vowel duration of the lexical set MOUTH in milliseconds. Durations below 40 ms were excluded.

The boxplots for the SVLR and VE categorizations show durational differences for Aitken’s Law but similar durations for the VE long and short contexts (see Figure 72). The boxplot for the *SVLR1* and *SVLR2* long contexts are clearly higher than the boxplots of the respective short environments. The average durations do also vary for the long and short environments (*SVLR1* long: 131.18 ms; *SVLR1* short: 106.52 ms; *SVLR2* long: 122.20 ms; *SVLR2* short: 108.06 ms). The plots and average values thus indicate that MOUTH is affected by the phonological and morphological conditioning of Aitken’s Law.

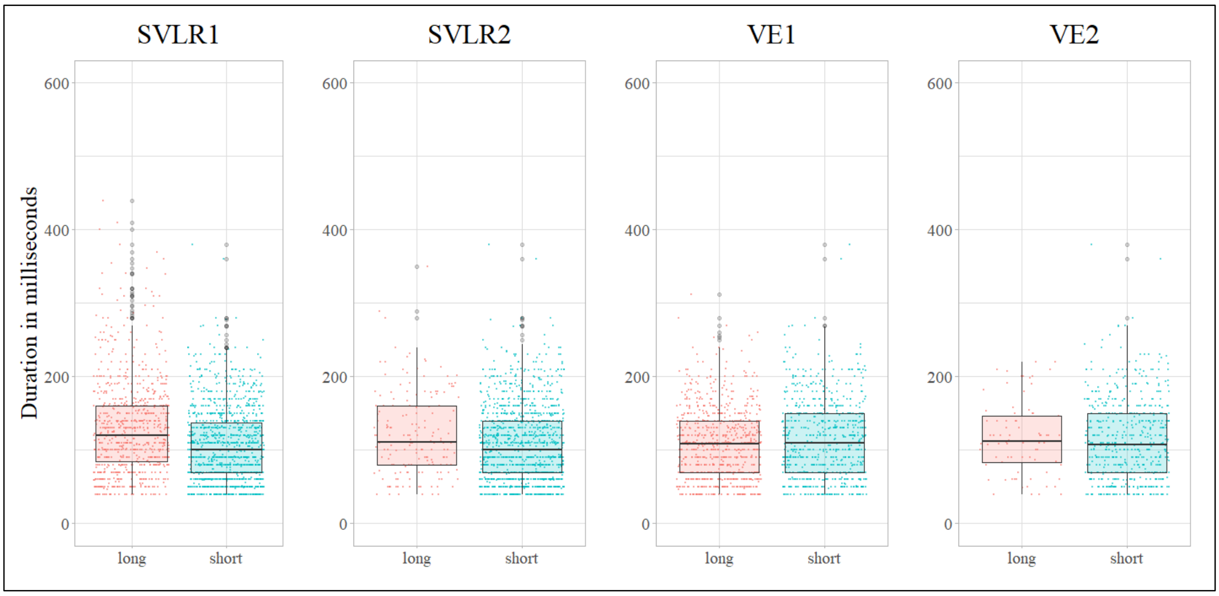


Figure 72. Boxplots and jitterplots of raw vowel duration of the lexical set MOUTH for the categories SVLR1, SVLR2, VE1 and VE2.

The situation is different for the VE contexts. The boxplots for the *VE1* and *VE2* long and short environments are on a similar level and the mean durations reveal that MOUTH is longer in *VE1* short contexts (mean: 114.57 ms) than in *VE1* long contexts (mean: 108.86 ms). Yet, the *VE2* long contexts (mean: 117.07 ms) are nevertheless slightly longer than the *VE2* short environments (mean 113.76 ms), but it is unclear whether this difference is statistically significant. The average durations for the different postvocalic contexts provide more detailed information on why the distributions are the way they are (see Table 57).

Table 57. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set MOUTH. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	**Average vowel duration**
laterals	117.32 ms
nasals	99.11 ms
voiced fricatives	137.22 ms
voiceless fricatives	115.40 ms
voiced plosives	122.65 ms
voiceless plosives	113.94 ms
pauses	186.45 ms

Similar to the other vowels, the longest MOUTH tokens occur before pauses which indicates that the vowel is affected by *constituent-final lengthening* (see subsection 3.1.8). It is also likely that MOUTH is affected by the phonological conditioning of Aitken’s Law due to the following durational differences: first, the voiced fricatives lead to the second-highest average duration of MOUTH. This clearly indicates the SVLR lengthening effect before voiced fricatives. Second, the fricative ratio (21.82 ms) clearly exceeds the plosive ratio (8.71 ms), so the fricatives have a stronger lengthening effect than the plosives. Furthermore, MOUTH tokens are extremely short before nasals which is in line with the

SVLR but contradicts the VE. Particularly short durations before nasals could also be observed in the monophthongs STRUT, GOOSE, FLEECE and THOUGHT (see sections 5.2.2, 5.3.1, 5.3.2 and 5.3.3.). The vowels before laterals are also relatively short when compared to the durations in voiced plosive and voiced fricative contexts. All in all, the average durations indicate that Aitken's Law operates in MOUTH. The VE does, however, not seem to have a big influence on the duration of the vowel. On the contrary, the plots and average duration suggest that MOUTH is shorter in VE long contexts.

The discrepancy between the SVLR and VE environments can also be seen when plotting the categories for different intra- and extralinguistic variables. I found some particularly interesting distributions for the *SVLR1* and *VE1* categories in different *regions*, *phrasal positions* and *stress* patterns (see Figure 73).

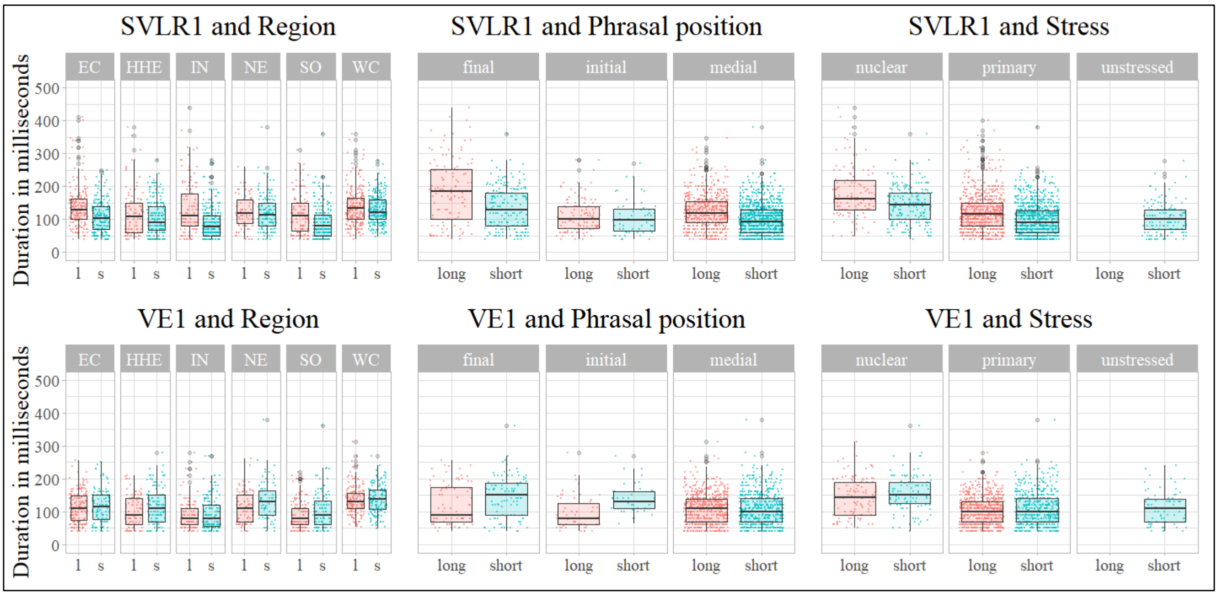


Figure 73. Boxplots and jitterplots of raw vowel duration of the lexical set MOUTH for the categories SVLR1 and VE1 separated for the variables regional background, phrasal position and stress.

The *SVLR1* plots show that the MOUTH tokens are longer in long environments than in short environments across the different regions of Scotland. Yet, the difference between the boxplots is more pronounced in the dialect regions *Insular* and *Southern* than in, for instance, *West central*. The durations of *West central* are also generally longer than those of the other regions. It could be possible that Aitken's Law varies across different regions.

The plots for the *phrasal positions* further indicate that the phonological conditioning of Aitken's Law is stronger in phrase-final positions. The difference between the boxplots is greater in final syllables than in phrase-initial or phrase-medial syllables. Especially the boxplots for the initial positions are on a very similar level. It is therefore possible that Aitken's Law interacts with the *phrasal position*. The *SVLR1* plots for stress show that MOUTH is longer in nuclear stressed syllables than in primary stressed or unstressed syllables. However, the difference between the long and short contexts in nuclear and primary stressed syllables is comparable. The *VE1* plots show the exact opposite distributions when compared to the *SVLR1* plots. The *VE1* short contexts are longer than the *VE1* long environments across all regions. While the difference is less pronounced in the region *Insular*, it is nevertheless clear that *VE1* long contexts lead to shorter durations overall. MOUTH is also shorter in *VE1* long environments in different phrasal positions. The *VE1* long contexts are shorter than the respective short environments in phrase-final and phrase-initial positions, the boxplots are, however, on a similar level in phrase-medial syllables. Whereas the VE lengthening effects are amplified in final positions for many other vowels in this study, the VE contexts for MOUTH have the opposite effect: *VE1* long contexts lead to shorter MOUTH pronunciations than *VE1* short environments, especially in phrase-final syllables. Similar distributions can be observed for the variable *stress*. *VE1* long tokens are clearly shorter than *VE1* short vowels in nuclear stressed syllables. The distributions are more balanced in primary stressed syllables. The plots for the *VE1* therefore suggest that MOUTH is influenced by an "anti-Voicing Effect" (Stuart-Smith et al., 2019), especially in prominent prosodic positions.

The plots for the intralinguistic variables (see Figure 74) corroborate many observations that could already be made in Figure 73. MOUTH is affected by *constituent-final lengthening* (see subsection 3.1.8) and *prosodic stress* (see subsection 3.1.7) because phrase-final and nuclear stressed tokens are clearly longer than non-final or non-nuclear tokens. The boxplots for the primary stressed and unstressed syllables are on a similar level, so it is unlikely that MOUTH is significantly affected by lexical stress (see subsection 3.1.6) in spoken SSE.

The plot for *word type* does not reveal anything because the dataset does not include function words with the vocalic nucleus /ʌʊ/. The boxplots for the categorical variable *word frequency* are on a similar level. The linear plot for *word frequency* does also not reveal any specific distributions because the red regression line is stable. It seems that the duration of MOUTH is not significantly influenced by *lexical frequency* effects (see subsection 3.1.5).

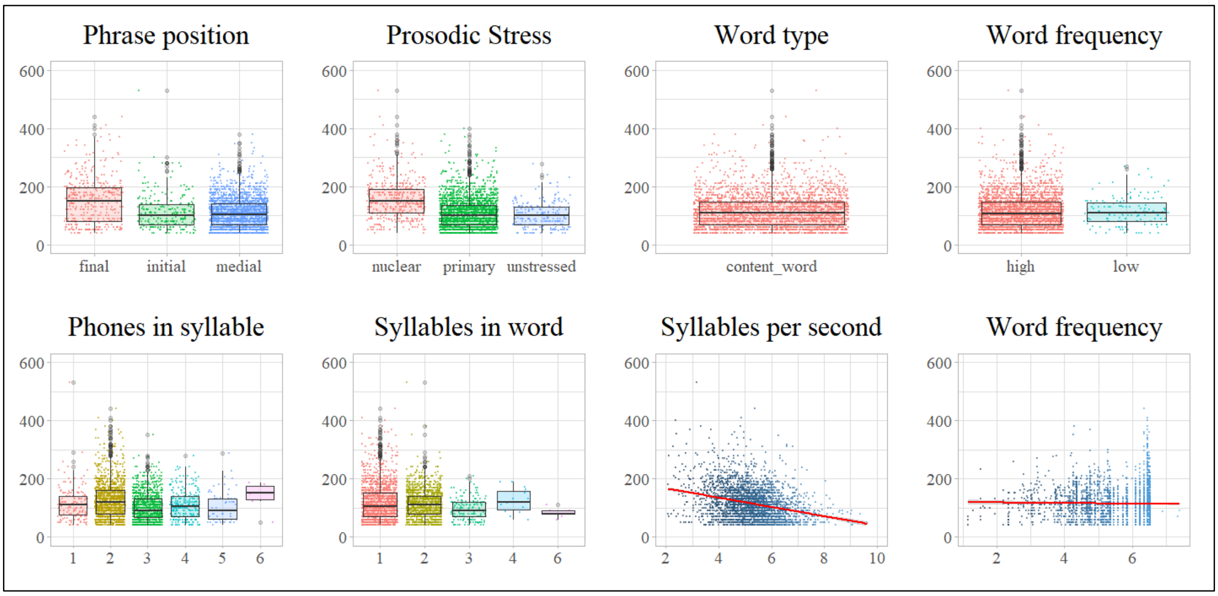


Figure 74. Plots of raw vowel duration for the lexical set MOUTH separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

The vowel is, however, influenced by the local articulation rate. The corresponding plot reveals that long MOUTH tokens become increasingly rare in high articulation rates. This fits into the overall trend: vowels become shorter in fast speech.

The plots for the number of phones in a syllable and the number of syllables in a word do not show a clear tendency. While long MOUTH tokens become less frequent in high *syllable phone counts* and *word syllable counts*, the average durations do not steadily decrease. It is therefore unclear whether MOUTH is affected by *intrasyllabic compression* (see subsection 3.1.3) or *polysyllabic shortening* (see subsection 3.1.4).

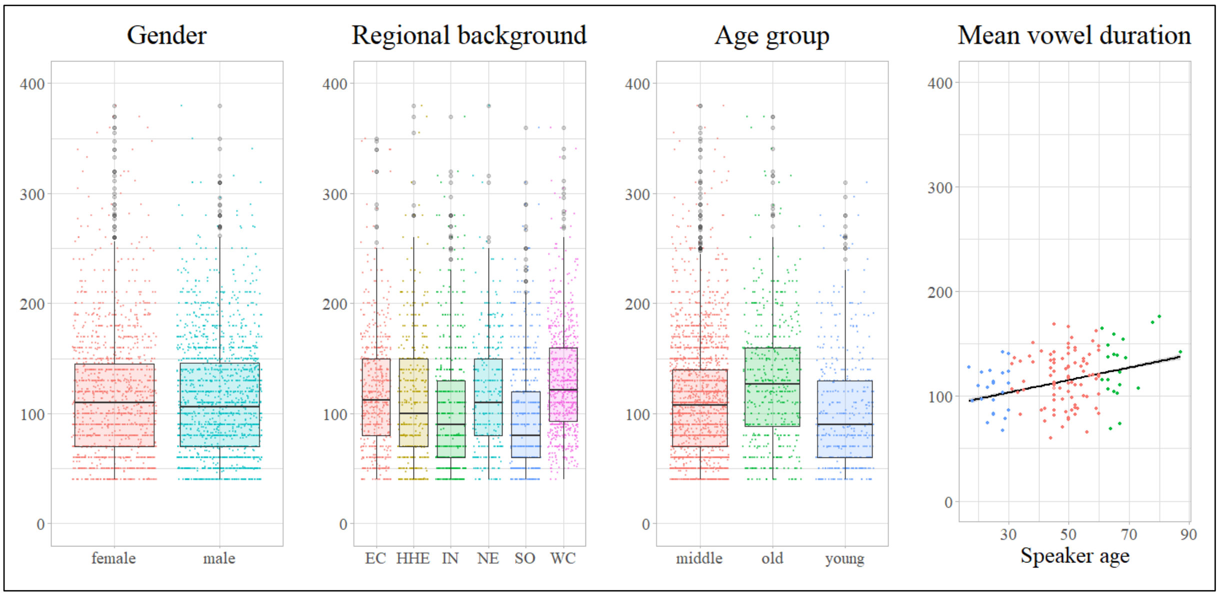


Figure 75. Plots of raw vowel duration of MOUTH separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

I also plotted MOUTH against the extralinguistic categories *gender*, *regional background* and *age group* and I plotted mean vowel duration against the speakers' exact age (see Figure 75). The proportions show that MOUTH is not affected by the *gender* of the speakers, both boxplots are on a similar level. The older age group produces slightly longer vowels than the middle-aged or young speakers. The mean vowel duration also increases with a higher age and this fits well into the overall trend that older speakers produce longer vowels overall (see section 5.1). The only remarkable distribution can be found for the *regional background* of the speakers. The plots indicate that MOUTH is generally shorter in the regions *Insular* and *Southern*.

As for the linear mixed effects modeling, the relatively low token numbers did not allow for many interactions. The token number of 2992 was further subdivided for the *SVLR1*, *SVLR2*, *VE1* and *VE2* sub-datasets, so the frequencies of specific variables and variable combinations could become relatively small. Due to the relatively low token numbers, the random factor *syllable label* was excluded from all models to avoid convergence issues. All models excluded the variables *word frequency*, *age group*, *gender* and the *word syllable count*. The plots already showed that the durations do not strikingly differ with respect to *gender* and *word frequency* (see Figure 74 and Figure 75) and the models confirm that these variables have no significant effect. While older speakers tend to produce slightly longer MOUTH tokens, this difference does not reach statistical significance. The models also confirm that *polysyllabic shortening* (see subsection 3.1.4) does not influence MOUTH because the corresponding variable did not return significant effects in any of the model outputs. As for the postvocalic consonantal contexts, the variables *SVLR2* and *VE2* did not return significant effects. This means MOUTH is not significantly affected by the morphological conditioning of Aitken's Law nor by the VE in plosive contexts. The variable *SVLR1* was significant, however, so MOUTH tokens are significantly shorter in *SVLR1* short contexts than in *SVLR1* long environments. This confirms that the phonological SVLR operates in MOUTH. The variable *VE1* was also significant in the respective model, but the effect sizes revealed that the *VE1* short environments lead to increased durations when compared to the intercept. In other words, MOUTH is significantly longer in *VE1* short contexts than in *VE1* long contexts. The *VE1* models outputs therefore confirm a general "anti-Voicing Effect" in MOUTH. The conditional R^2 values of the different models lie in the range between 0.37 and 0.47 and the output of the best model is summarized in Table 58.

Table 58. Best linear mixed model for the lexical set MOUTH fit by maximum likelihood (marginal R^2 = 0.28; conditional R^2 = 0.47). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ position + stress + style + num_pho_syl + syl_per_sec + reg + (1 | speaker) + (1 | word_label)

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	1332.9	1418.0	-650.5	1300.9	1490

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.1757	-0.6379	0.0336	0.6086	4.6855

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.0119	0.1095
speaker	(Intercept)	0.0327	0.1809
Residual		0.1201	0.3466

Intercept: positionfinal; stressnuclear; stylescripted;regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.70400	300.066	0.09338	< 2e-16	***
positioninitial	-0.16297	0.84961	0.04814	0.000730	***
positionmedial	-0.18427	0.83170	0.02928	4.13e-10	***
stressprimary	-0.28669	0.75074	0.02729	< 2e-16	***
stressunstressed	-0.23231	0.79269	0.04994	4.39e-06	***
styleunscripted	-0.16633	0.84676	0.04022	5.45e-05	***
num_pho_syl	-0.05581	0.94572	0.01652	0.000919	***
syl_per_sec	-0.08524	0.91828	0.01001	< 2e-16	***
regInsular	-0.19276	0.82468	0.06967	0.006562	**
regSouth	-0.15365	0.85757	0.06733	0.024180	*
regWest_central	0.17861	1.19555	0.06093	0.003953	**

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

Significant fixed factors include the *phrasal position*, *stress*, *style*, the *number of phones in a syllable*, the local articulation rate in terms of *syllables per second* as well as the *regional background*. The *SVLR2* categorization was excluded which means that MOUTH tokens in *SVLR2* long contexts are not significantly different from those in *SVLR2* short environments. All possible interactions were dropped in the model building process because they did not return significant effects or improve the model fit.

MOUTH does, however, vary with respect to the prosodic factors *phrasal position*, *stress* and *tempo*. Phrase-initial tokens are roughly 15 percent shorter and phrase-medial tokens are approximately 17 percent shorter than phrase-final tokens (= intercept). This confirms the effect of *constituent-final lengthening* (see subsection 3.1.8). Similarly, primary stressed syllables are 24.93 percent and unstressed syllables are 20.74 percent shorter than nuclear stressed syllables (= intercept). MOUTH is therefore significantly influenced by *prosodic stress* (see subsection 3.1.7). The effect of *lexical stress* (see subsection 3.1.6) is negligible because unstressed syllables have a stronger effect size than primary stressed syllables. Like the other vowels, MOUTH is also influenced by the local articulation rate. The vowel becomes 8.18 percent shorter for each additional syllable produced in the timeframe of one second. This corresponds to the effect of *tempo* (see subsection 3.1.9).

The model output further specifies that MOUTH is affected by *intrasyllabic compression* (see subsection 3.1.3). According to the model output, the diphthong is shortened by 5.43 percent for every additional phone in a syllable. Nevertheless, it must be said that the variable is not significant in the models with the *VE1* and *VE2* categorization schemes. The influence of the *syllable phone count* is therefore only important in the SVLR datasets.

The only significant extralinguistic variable in the model is the *regional background* of the speakers. Specifically, speakers from the dialect regions *Insular* produce 17.54 percent shorter vowels than the people from the region *East central* (= intercept). The MOUTH pronunciations of the speakers from *Southern* Scotland are also 14.25 percent shorter when compared to the intercept. The speakers from the region *West central,* however, produce 19.55 percent longer vowels than the *East central* speakers. The regional differences were already visible in Figure 75, but the model confirms that the regions *Insular* and *Southern* produce significantly shorter vowels and that MOUTH is significantly longer in the region *West central*.

The plots and models have demonstrated that MOUTH is affected by the phonological conditioning of Aitken’s Law in spoken SSE. The effect of the morphological SVLR, however, turned out to be insignificant. In contrast to previous studies, I could not find a significant VE-related lengthening effect in the dataset. On the contrary, the *VE1* classification revealed that the diphthong MOUTH is shorter in *VE1* long contexts than in *VE1* short contexts. Similar to many other vowels in this investigation, MOUTH is particularly short before nasals.

5.4.2 PRICE

The lexical set PRICE represents the diphthongs /ʌɪ/ and /ae/ in SSE. According to Abercrombie (1979), the shorter diphthong /ʌɪ/ is realized in monomorphemic words such as <side> and <tide>. The longer diphthong /ae/ is frequently realized in heteromorphemic words such as <sighed> and <tied>. Most scholars note that there is a perceptual difference between the diphthongs /ʌɪ/ and /ae/ so that Scottish English incorporates a quasi-phonemic contrast (Scobbie & Stuart-Smith, 2008) between words such as <side> and <sighed> (Wells, 1982, p. 305). This corresponds to the morphological conditioning of Aitken's Law: vowels are longer when followed by morpheme boundaries. A summary of the present study's vowel durations of PRICE is given in Table 59 and visualized in Figure 76.

Table 59. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set PRICE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Token number:	9540
Average duration:	125.02 ms
Standard deviation:	60.52 ms
Minimal duration:	40.00 ms
Maximal duration:	690.00 ms
1st quantile:	80.00 ms
Median:	112.40 ms
3rd Quantile:	150.00 ms

The average duration of PRICE is 125.02 ms and the standard deviation is 60.52 ms. The shortest PRICE tokens are 40 ms long and the longest pronunciation lasts 690 ms. Most of the measurements lie in the range between 80 ms and 150 ms and the median is 112.40 ms.

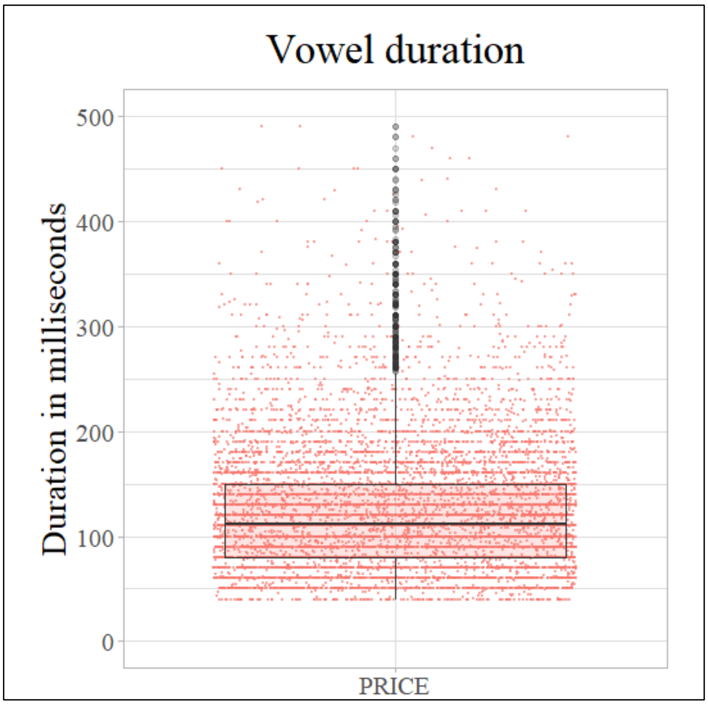


Figure 76. Vowel duration of the lexical set PRICE in milliseconds. Durations below 40 ms were excluded.

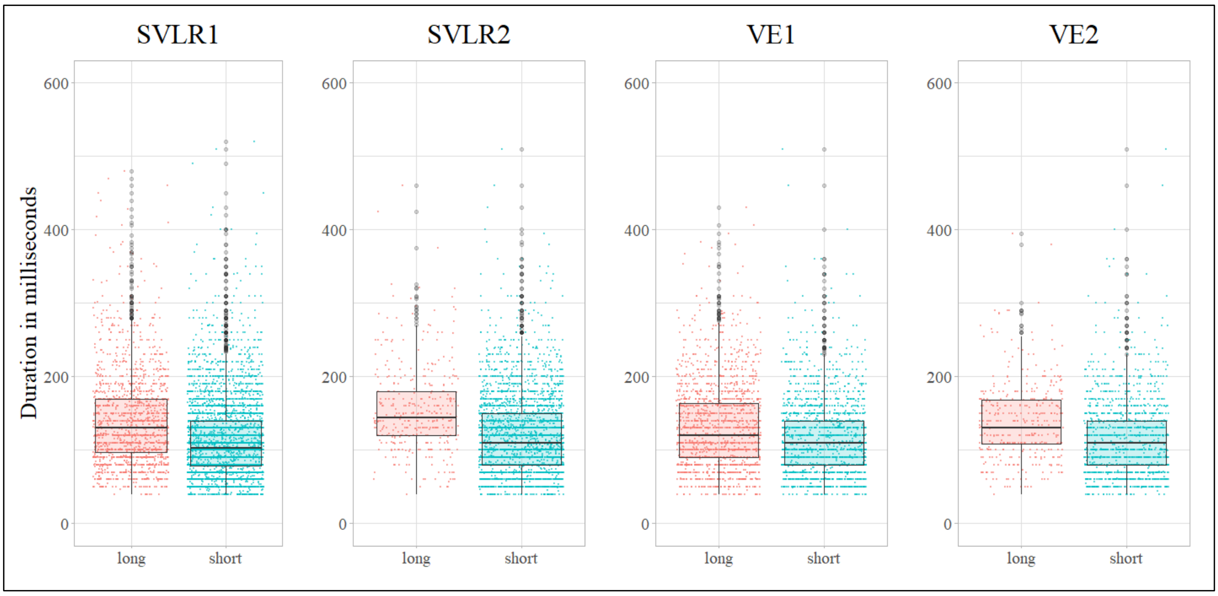


Figure 77. Boxplots and jitterplots of raw vowel duration of the lexical set PRICE for the categories SVLR1, SVLR2, VE1 and VE2.

The plots for the SVLR and VE categories show a very regular pattern (see Figure 77). PRICE is always longer in the long contexts than in the short environments across all categories. The difference between the *SVLR2* levels is especially pronounced: the respective average values are 36.48 ms apart (*SVLR2* long: 155.84 ms; *SVLR2* short: 119.36 ms). This distribution corresponds to previous findings that PRICE is strongly affected by the morphological conditioning of Aitken's Law. All plots in Figure 77 suggest that both the SVLR and the VE operate in PRICE.

A more detailed overview of the average durations in different postvocalic contexts can be found in Table 60. Like the other vowels in this study, PRICE is affected by *constituent-final lengthening* (see subsection 3.1.8) because the vowel is clearly lengthened before pauses. The second longest average duration can be found in voiced fricative contexts. The fricative ratio (41.99 ms) is also more than two times longer than the plosive ratio (16.02 ms) which indicates that the phonological conditioning of Aitken's Law operates in PRICE. Interestingly, following voiceless fricatives have the shortest average duration of all consonantal contexts. The PRICE vowels before nasals are also relatively short which corresponds to Aitken's Law, but contradicts the VE.

Table 60. Average vowel durations sorted for different postvocalic consonant contexts for the lexical set PRICE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	**Average vowel duration**
laterals	122.46 ms
nasals	112.68 ms
voiced fricatives	150.84 ms
voiceless fricatives	108.85 ms
voiced plosives	130.87 ms
voiceless plosives	114.85 ms
pauses	250.65 ms

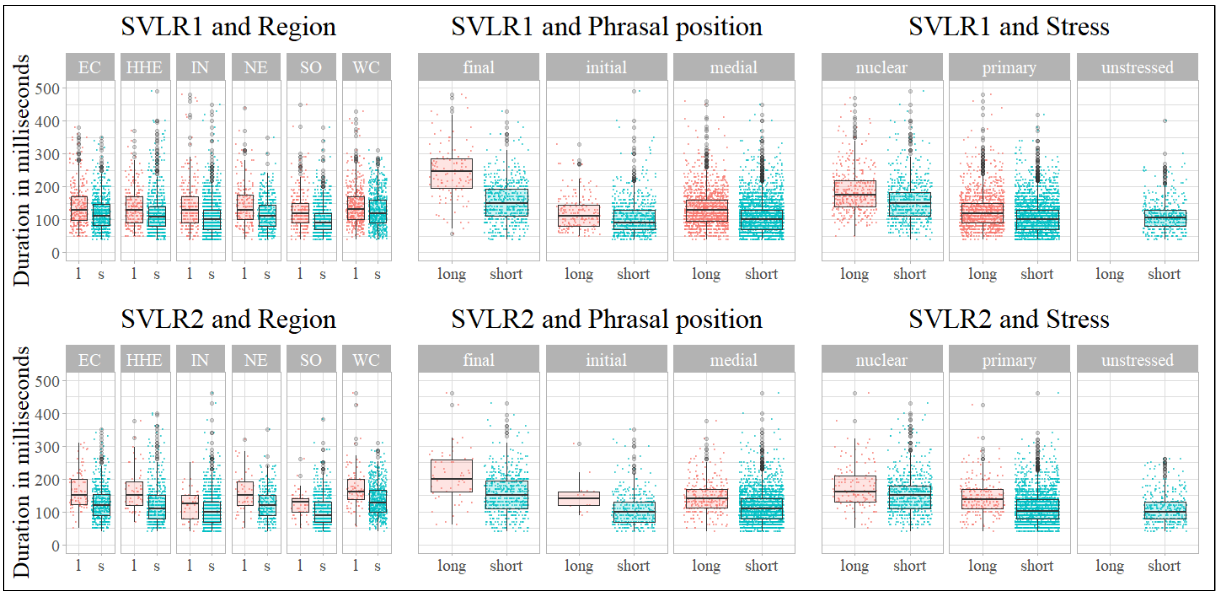


Figure 78. Boxplots and jitterplots of raw vowel duration of the lexical set PRICE for the categories SVLR1 and SVLR2 separated for the variables regional background, phrasal position and stress.

I also plotted the SVLR and VE environments for the different intra- and extralinguistic variables but the distributions are very regular across all contexts (see Figure 78). The PRICE tokens are generally longer in the long environments than in the short contexts. PRICE is generally a bit shorter in the dialect region *Southern* but the durational differences between the *SVLR1* and *SVLR2* long and short contexts are similar across all regions. Thus, Aitken's Law appears to be stable across Scotland. The plots for phrasal *position* and *stress* show that PRICE is generally longer in phrase-final and nuclear stressed syllables. At the same time, the durational difference between the *SVLR1* long and short contexts is much more pronounced in final positions than in phrase-initial or phrase-medial syllables. This indicates that the phonological conditioning of Aitken's Law is amplified in phrase-final positions. The

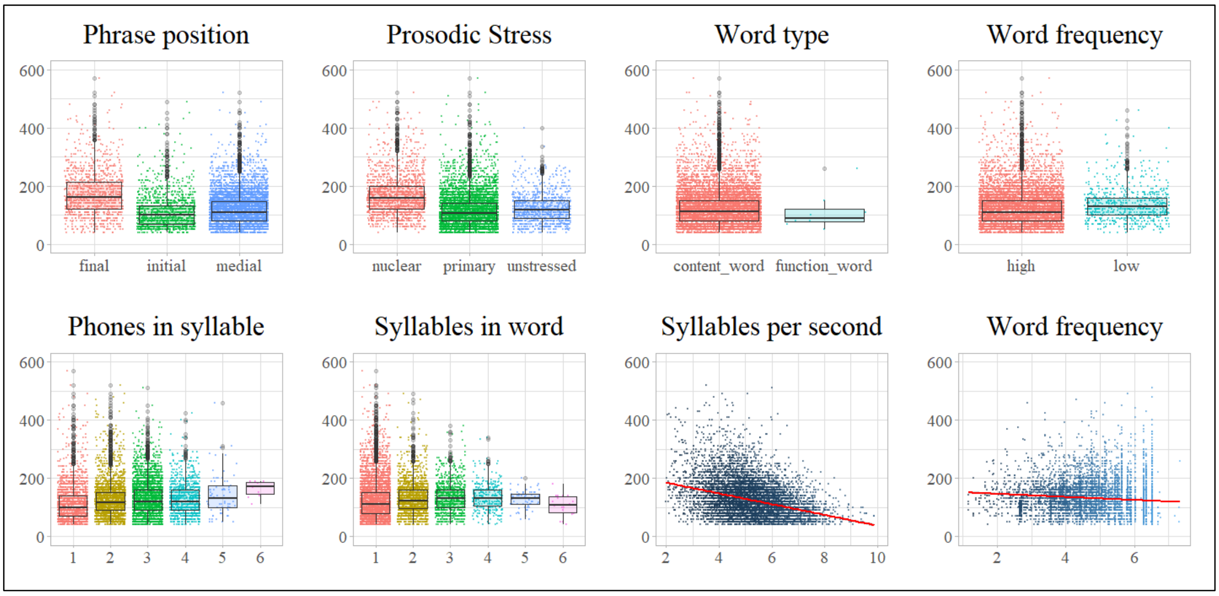


Figure 79. Plots of raw vowel duration for the lexical set PRICE separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

differences for the *SVLR2* categorization scheme are more similar in different phrasal positions. The durational differences for *stress* are more regular for both *SVLR1* and *SVLR2*.

The plots for the intralinguistic variables (see Figure 79) confirm that PRICE is affected by *constituent-final lengthening* (subsection 3.1.8) and *prosodic stress* (subsection 3.1.7) because the diphthong is generally longer in final and nuclear stressed syllables. The influence of *lexical stress* appears to be negligible as unstressed syllables (mean: 122.64 ms) are on average slightly longer than primary stressed syllables (mean: 114.21 ms).

The plots for *word type* and *word frequency* do not show a clear tendency. The boxplot representing content words is slightly higher than the one for function words, but the dataset comprises only very few function words in general. The boxplot for high frequency words is a bit lower than the one for low frequency words and the regression line in the linear plot has a very slight negative slope. This suggests that high frequency words are slightly shorter than low frequency words which corresponds to the effects of *lexical frequency* (see subsection 3.1.5).

The plot for the local articulation rate shows that the duration of PRICE is affected by *tempo* (see subsection 3.1.9). The duration of PRICE decreases in faster articulation rates which is in line with the overall trend in this study (see section 5.1).

It is relatively unclear whether PRICE is affected by the number of phones in a syllable or by the number of syllables in a word. The corresponding plots do not show a clear negative trend. As for the *syllable phone count*, vowel duration generally increases for every additional phone in a syllable. This completely contradicts the effect of *intrasyllabic compression* (see subsection 3.1.3). The plot for the word syllable count does not follow a negative pattern either. The duration of PRICE first increases from monosyllabic words to four-syllable words but then decreases again in words with five or six

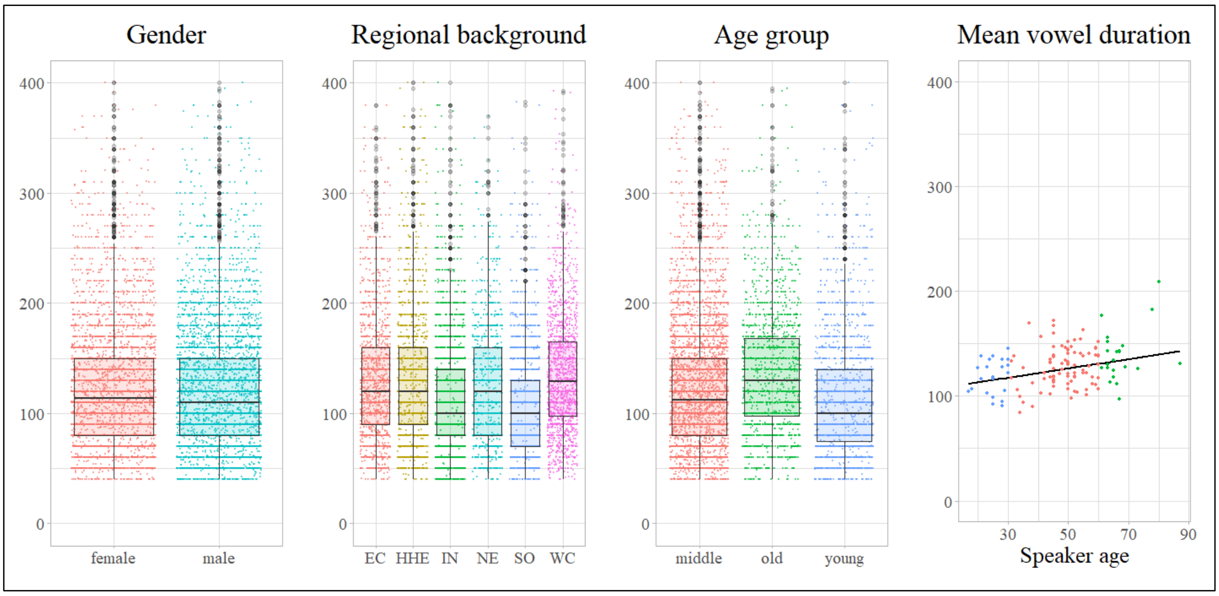


Figure 80. Plots of raw vowel duration of PRICE separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

syllables. It is therefore unclear whether PRICE is significantly influenced by *polysyllabic shortening* (see subsection 3.1.4).

The plots for PRICE across different social variables (see Figure 80) are highly similar to the ones for MOUTH (see Figure 75). One cannot see a difference in the duration of PRICE for the *genders*; both boxplots are on the same level and the respective average durations are almost identical (female: 126.13 ms; male: 124.00 ms). The older speakers produce longer vowels overall. This can be seen in the boxplots and in the increase of mean vowel duration. It is unclear, however, whether this difference is statistically significant. More striking differences can be observed for the *regional backgrounds*. The speakers from the regional background *Southern* and from the dialect region *Insular* generally produce shorter PRICE vowels than the other regions.

The linear mixed effects models for PRICE all excluded the extralinguistic variables *age group* and *gender*. This means that the durational patterns in PRICE are stable irrespective of the *gender* or the *age group* of a speaker. Apart from that, all models included the respective categorization scheme (*SVLR1*, *SVLR2*, *VE1*, *VE2*) which confirms that both Aitken's Law and the VE operate in the diphthong PRICE. Another general observation is that many interactions were excluded during the model building process. The *SVLR1*, *SVLR2*, *VE1* and *VE2* categorizations had only few significant interactions with other variables. The variables *SVLR1* and *phrasal position,* for instance, had a significant interaction which indicated that the difference between *SVLR1* long and short environments is more pronounced in phrase-final positions. Apart from that, the effects of Aitken's Law and the VE are relatively stable and do not vary in specific groups or contexts. The corresponding plots have already shown that the distributions are very regular overall. The long environments were always longer than the respective short environments and the durational differences are similar in different contexts. The model outputs corroborate these observations. The conditional R^2 values of the models range between 0.40 and 0.46 and a summary of the best model is given in Table 61.

Significant fixed factors include the *VE1* categorization, the *phrasal position*, *stress*, *style*, the *syllable phone count*, the local articulation rate (*syllables per second*) and the *regional background* of the speakers. All interactions were excluded in the model building process, so the effect of *VE1* does not significantly vary in different contexts. According to the estimates, PRICE is roughly 9 percent shorter in the *VE1* short contexts than in the *VE1* long environments.

The model outputs further confirm that the diphthong is influenced by *constituent-final lengthening* (see subsection 3.1.8) *stress* (see subsection 3.1.7) and *tempo* (see subsection 3.1.9). PRICE tokens are 15.35 percent shorter in phrase-medial and 19.41 percent shorter in phrase-initial positions when compared to the intercept (= phrase-final tokens). Similarly, when compared to nuclear stressed tokens (= intercept), PRICE is 20.96 percent shorter in primary stressed syllables and 25.28 percent shorter in unstressed syllables. As for the local articulation rate, the model estimates that the diphthong decreases by roughly 9 percent for every additional syllable produced in the timeframe of one second.

Table 61. Best linear mixed model for the lexical set PRICE fit by maximum likelihood (marginal R^2 = 0.26; conditional R^2 = 0.46). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ VE1 + position + stress + style + num_pho_syl + syl_per_sec + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label)

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	2061.3	2172.6	-1012.6	2025.3	3569

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-3.6288	-0.5897	0.0021	0.5665	4.6474

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	0.0204	0.1431
syl_label	(Intercept)	0.0052	0.0721
speaker	(Intercept)	0.0102	0.1011
Residual		0.0915	0.3025

Intercept: VE1 long; positionfinal; stressnuclear; stylescripted;regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.923e+00	373.608	6.609e-02	< 2e-16	***
VE1short	-9.407e-02	0.91022	2.043e-02	4.56e-06	***
positioninitial	-2.157e-01	0.80598	2.195e-02	< 2e-16	***
positionmedial	-1.666e-01	0.84655	1.602e-02	< 2e-16	***
stressprimary	-2.351e-01	0.79049	1.412e-02	< 2e-16	***
stressunstressed	-2.914e-01	0.74722	4.748e-02	1.45e-09	***
styleunscripted	-1.104e-01	0.89543	2.468e-02	1.39e-05	***
num_pho_syl	-5.376e-02	0.94766	1.606e-02	0.000902	***
syl_per_sec	-9.508e-02	0.90929	5.401e-03	< 2e-16	***
regSouth	-1.212e-01	0.88584	3.934e-02	0.002542	**

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

The duration of PRICE also varies significantly in scripted and unscripted speech. The estimates for the variable *style* specify that the diphthong is approximately 10.5 percent shorter in unscripted than in scripted speech forms. The variable *style* was also significant in the *SVLR1*, *SVLR2* and *VE2* models which means that the effect of the variable is consistent in different postvocalic contexts. The shorter durations of PRICE in unscripted speech are in line with the expectations (see subsection 4.1.2) because vowels are often reduced and consequently shorter in these speech forms.

Yet, the model outputs for the variable *syllable phone count* contradict the distributions seen in the corresponding plot (see Figure 79). Whereas the boxplots indicated that the duration of the diphthong increases with a higher *syllable phone count*, the estimates of the model specify that PRICE becomes 5.24 percent shorter for every additional phone in a syllable. Hence, even though there is considerable variability (see Figure 79), PRICE is significantly influenced by *intrasyllabic compression* effects (see subsection 3.1.3) in spoken SSE.

The only extralinguistic variable which turned out significant is the *regional background*. Specifically, speakers from the region *Southern* produce significantly shorter PRICE tokens than speakers from the region *East central* (= intercept). Yet, this difference in duration is not linked to any other variable. Speakers from the South generally produce slightly shorter PRICE diphthongs which corresponds to the overall trend (see section 5.1).

The plots and models for PRICE have shown that the diphthong is affected by both Aitken's Law and the VE. The vowel is longer when followed by voiced fricatives and there is a particular lengthening effect before morpheme boundaries (*SVLR2*). Nevertheless, the duration of the lexical set does also increase before voiced consonants more generally, so the VE does also operate in the diphthong. The SVLR and VE patterns are very stable across different *age groups*, *genders* and *regional backgrounds* in the dataset.

The present investigation could corroborate many previous findings. Sylvester Douglas already noticed in 1775 that the words <pride> and <deny'd> do not rhyme in Scottish English due to vowel quantity differences. The PRICE vowel in <deny'd> is clearly longer than the vowel in <pride> (Douglas, 1991 [1775], p. 151). Aitken (1981) further specified that this lexical set (vowel 1) and the corresponding two realizations are generally affected by the SVLR. While the subsequent empirical study by McClure (1977) found a strong VE in /ʌɪ/ and a variable VE in /ae/, the investigations by Agutter (1988a), McMahon (1991) and Scobbie, Hewlett, and Turk (1999) all conclude that Aitken's Law operates in PRICE. A clear and consistent durational contrast in SVLR short and long environments could also be found by Milroy (1995), Scobbie, Turk, and Hewlett (1999) as well as Watt and Ingham (2000). Pukli (2006) found evidence for the morphological conditioning of the SVLR in PRICE but also a modest VE in monomorphemic words. Watt and Yurkova (2007) also conclude that PRICE is influenced by the morphological SVLR, but they could find not find any evidence for the phonological conditioning of Aitken's Law in Aberdeen English. Warren (2018) could confirm these findings: PRICE is affected by the morphological conditioning of Aitken's Law, but there is a strong VE in monomorphemic contexts in the Northeast. As for Tyneside, Llamas et al. (2011) could confirm the findings by Milroy (1995) that the SVLR operates most consistently in PRICE. In short, most previous investigations confirm that PRICE is affected by the SVLR, especially in heteromorphemic contexts. The results of the present study could find SVLR and VE effects in both monomorphemic and heteromorphemic environments. The diphthong is longer before morpheme boundaries but also before voiced consonants more generally. The difference between voiced and voiceless fricative contexts is especially strong.

5.4.3 CHOICE

The lexical set CHOICE is most commonly represented by the diphthong symbol /ɔe/ in SSE. This phoneme has many allophonic variants in Scottish English (Wells, 1982, p. 406) and Abercrombie (1979, p. 72) notes that this diphthong is characterized by a more centralized onset in SSE. A summary of this study's durational measurements of CHOICE can be found in Table 62 and a visualization of the measurements is displayed in Figure 81.

Table 62. Statistical summary of vowel duration for the lexical set CHOICE. Tokens below 40 ms were excluded.

Token number:	480
Average duration:	140.74 ms
Standard deviation:	59.09 ms
Minimal duration:	40.00 ms
Maximal duration:	430.00 ms
1st quantile:	100.00 ms
Median:	130.00 ms
3rd Quantile:	170.00 ms

CHOICE has the longest average duration of all vowels with 140.74 ms. At the same time, this lexical set is represented by only 480 tokens which is by far the lowest token number for any vowel in this investigation. The shortest CHOICE tokens are 40 ms long and the longest pronunciation lasts 430 ms. Most of the measurements are in the range between 100 and 170 ms and the median is 130 ms. The lexical set CHOICE has a standard deviation of 59.09 ms.

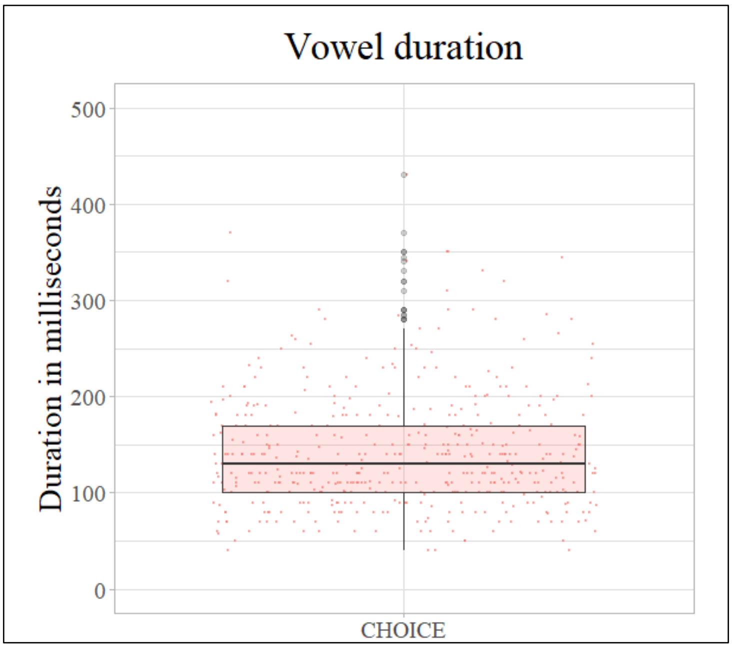


Figure 81. Vowel duration of the lexical set CHOICE in milliseconds. Durations below 40 ms were excluded.

The plots for the variables *SVLR1*, *SVLR2*, *VE1* and *VE2* show some usual distributions (see Figure 82). While the *SVLR1* and *SVLR2* long contexts lead to longer CHOICE tokens than the respective short contexts, the diphthong is clearly shorter in *VE1* long environments than in *VE1* short environments. The average duration of the *VE1* short contexts is 170.55 ms and the mean duration for the respective short contexts is 116.42 ms. This suggests an "anti-Voicing Effect" (Stuart-Smith et al., 2019) in the diphthong. Contrary to this, CHOICE is longer in the *VE2* long environments than in the *VE2* short environments, but it must be taken into consideration that *VE2* is represented by only 16 tokens, five of which represent *VE2* short contexts. The durations for the *VE2* might therefore not be fully representative. A summary of the average durations in different postvocalic contexts is given in Table 63.

Table 63. Average vowel durations and token numbers sorted for different postvocalic consonant contexts for the lexical set CHOICE. The measurements include the duration of function words and proper nouns which are excluded in the VE and SVLR categories.

Postvocalic environments	Average vowel duration
laterals	150.58 ms
nasals	111.79 ms
voiced fricatives	190.05 ms
voiceless fricatives	169.83 ms
voiced plosives	163.67 ms
voiceless plosives	148.06 ms
pauses	271.06 ms

The average durations for the different postvocalic consonant contexts show that CHOICE, like the other vowels in this study, is influenced by *constituent-final lengthening* (see subsection 3.1.8). The longest diphthongs can be found before pauses. The voiced fricatives lead to the second longest durations followed by voiceless fricative contexts. The diphthong is shorter in the plosive contexts and the plosive ratio (15.61 ms) is also slightly smaller than the fricative ratio (20.22 ms). More importantly, the lexical set CHOICE is clearly shortened when followed by nasals. The average duration of the diphthong in postvocalic nasal environments is very short (111.79 ms) when compared to the other values. At the same time, the nasal contexts are represented by many tokens (N = 206). The "anti-Voicing Effect" for the *VE1* category in Figure 82 might therefore be a result of very short CHOICE tokens in nasal contexts. The durations before laterals are also comparatively short.

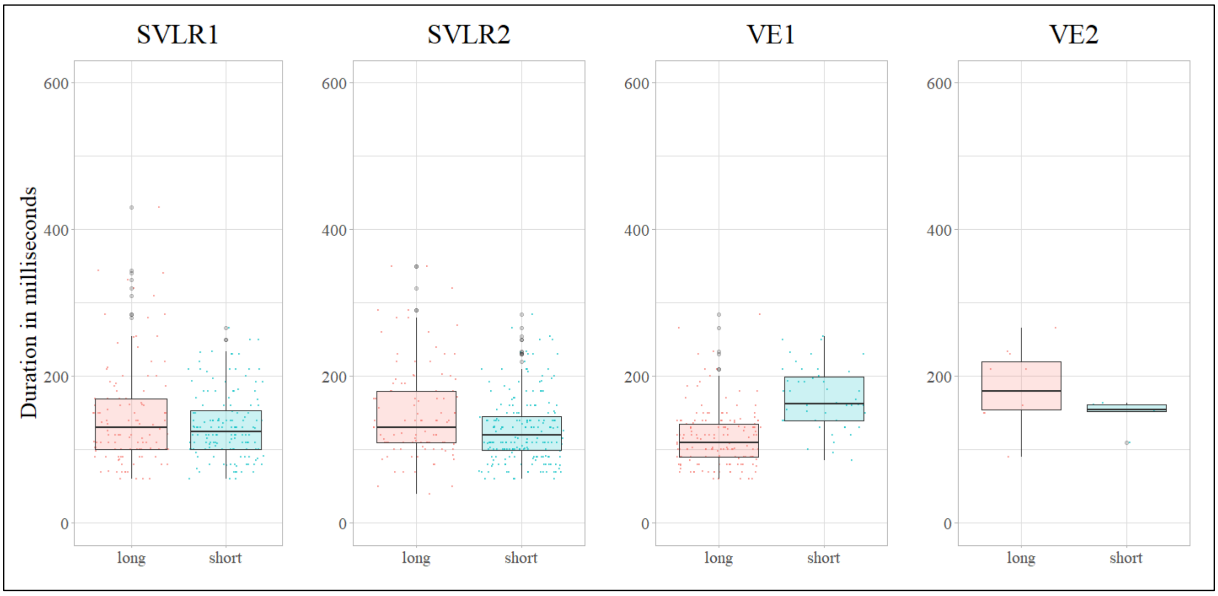


Figure 82. Boxplots and jitterplots of raw vowel duration of the lexical set CHOICE for the categories SVLR1, SVLR2, VE1 and VE2.

I further plotted the SVLR and VE environments for the different intra- and extralinguistic variables to see whether the effects of the VE or the SVLR vary in specific groups or contexts. Some variable combinations were, however, not possible due to low token numbers. Figure 83 provides an overview of the *SVLR1* and *SVLR2* levels separated for the variables *gender*, *style* and *stress*. The plots show that the *SVLR1* and *SVLR2* long contexts are generally longer than the respective short contexts in scripted and unscripted speech as well across the *genders*. Yet, the differences are not very pronounced and the boxplots for the *SVLR1* long and short contexts in unscripted speech are on a very similar level. The proportions are different for the variable *stress*. Here, the difference between the long and short contexts are stronger in nuclear stressed syllables than in primary stress syllables. This indicates that the effects of the SVLR are amplified in prominent prosodic contexts.

The plots for the intralinguistic variables (see Figure 84) show that CHOICE is affected by *constituent-final lengthening* (see subsection 3.1.8) and *prosodic stress* (see subsection 3.1.7). Like the other vowels in the present investigation, CHOICE is lengthened in phrase-final positions and in nuclear stressed syllables. The difference between primary stressed syllables and unstressed syllables is not very pronounced, so it does not seem that the diphthong is strongly influenced by *lexical stress* (see subsection 3.1.6).

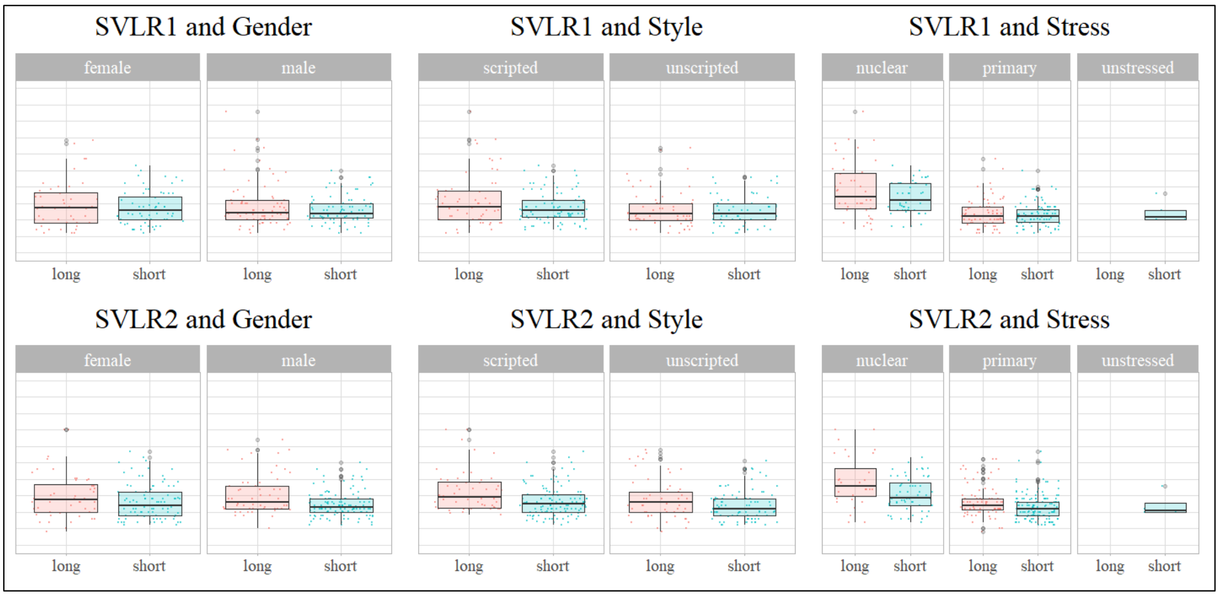


Figure 83. Boxplots and jitterplots of raw vowel duration of the lexical set CHOICE for the categories SVLR1 and SVLR2 separated for the variables gender, style and stress.

The dataset for the diphthong contains no function words, so the influence of *lexical category* (see subsection 3.1.5) cannot be assessed. The categorical and linear plots for *word frequency* do not show a clear trend either. The boxplots are on a similar level but there is more durational variation in the high frequency words because the corresponding boxplot is wider. The slope of the red regression line in the linear plot for word frequency comes close to zero, so it does not seem that CHOICE is significantly affected by *lexical frequency* (see subsection 3.1.5).

The diphthong is, however, clearly affected by the local articulation rate. The plot for the *syllables per second* demonstrates that long CHOICE tokens become less frequent in faster articulation rates. The faster the speech, the shorter the diphthong. This distribution coincides with the overall trend that vowels are generally shorter if the *tempo* (see subsection 3.1.9) of speech is faster.

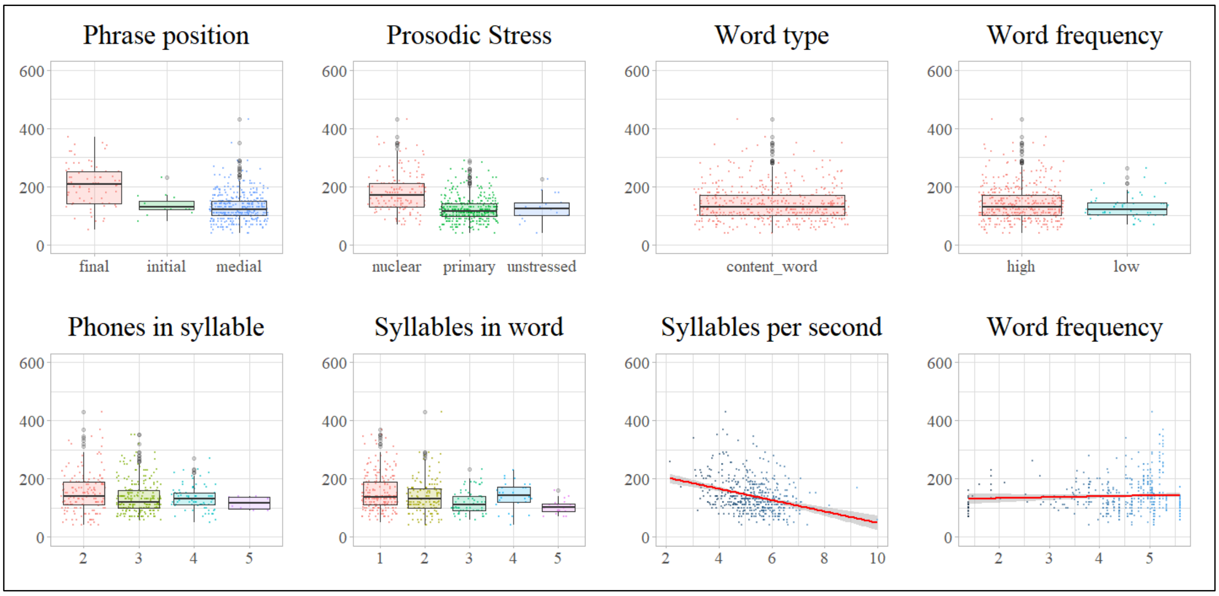


Figure 84. Plots of raw vowel duration for the lexical set CHOICE separated for the intralinguistic variables phrase position, prosodic stress, word type, word frequency (categorical), number of phones in syllable, number of syllables in word, syllables per second and word frequency (linear).

The plot for the *number of phones in a syllable* shows that the duration of CHOICE becomes shorter in syllables with many phones. The height of the boxplots declines and the average durations become shorter with a higher *syllable phone count*. This indicates an influence of *intrasyllabic compression* (see subsection 3.1.3). The plot for the *word syllable count* does not show a steady negative decline in duration. Specifically, words with four syllables are generally longer than words with two, three, or five syllables. This contradicts the effect of *polysyllabic shortening* (see subsection 3.1.4).

The plots for the social variables convey that older speakers produce slightly longer CHOICE tokens (see Figure 85). The boxplot is higher and the mean vowel duration increases with a higher age. The average durations also show that older speakers produce the longest CHOICE tokens overall (age group old: 159.71 ms; age group middle: 136.82 ms; age group young: 137.38 ms). This corresponds to the general trend that older speakers produce longer vowels (see section 5.1).

As for the *regional background*, the plots suggest that speakers from the area *West central* produce longer diphthongs than speakers from the other regions. Nevertheless, there is great durational variability for all regions, so it is unclear whether this difference reaches statistical significance.

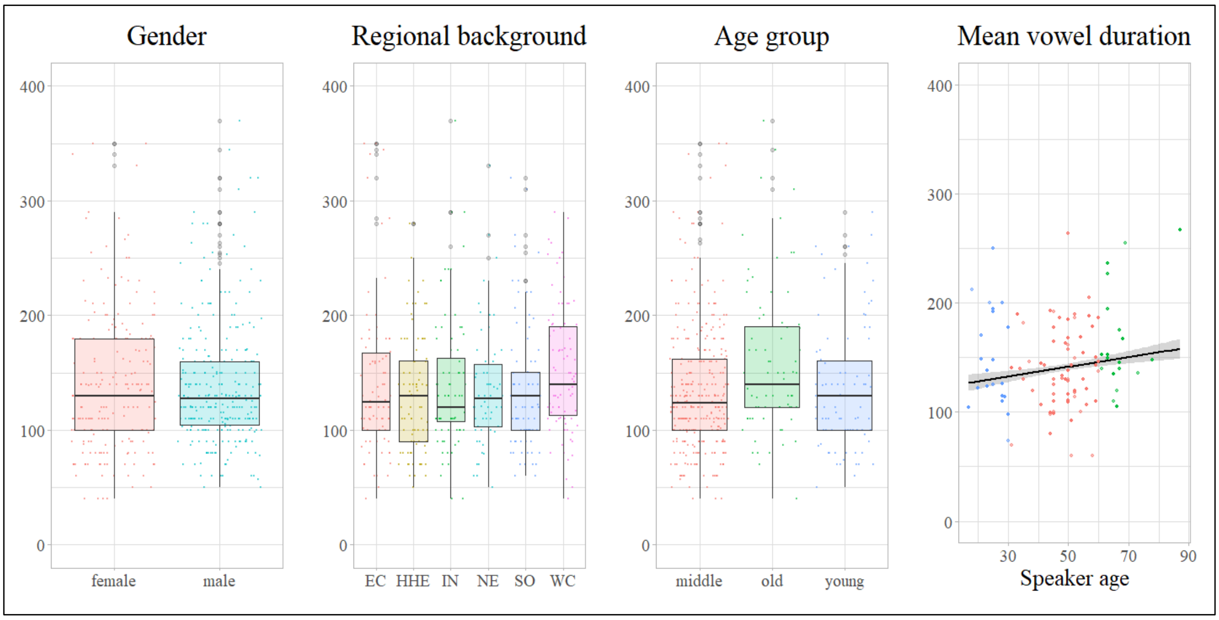


Figure 85. Plots of raw vowel duration of CHOICE separated for the categorical extralinguistic variables gender, regional background and age group as well as a plot with the mean vowel duration (y-axis) and the corresponding speaker's age (x-axis). The coloring of the speakers' age in the mean articulation rate plot is identical with the coloring of the age group plot (age group young: blue, age group middle: red; age group old: green).

The boxplots for *gender* are on a similar level, but the larger boxplot for the female speakers indicates that there is more durational variation. Nevertheless, the average values are almost identical (male: 140.79 ms; female: 140.71 ms), so it is very unlikely that the duration of CHOICE varies significantly across the *genders*.

The linear mixed effects models did not allow for any interactions due to low token numbers. It was not possible to fit a model for the *VE2* categorization scheme because the dataset only includes five voiceless plosive contexts and 11 voiced plosive environments. The models for the *SVLR1*, *SVLR2* and *VE1* categorization schemes all excluded the variables *gender*, *age group*, *lexical frequency*, *style*, the *syllable phone count* as well as the *word syllable count*. This means that the duration of CHOICE does not significantly vary for these variables. The variables *SVLR1* and *SVLR2* were also excluded in the corresponding models because they returned no significant effect. This means that the duration of CHOICE is not significantly affected by the phonological or morphological conditioning of Aitken's Law. The variable *VE1* was significant in the respective model but the effect sizes revealed that *VE1* short contexts are longer than *VE1* short contexts. This confirms that CHOICE is influenced by an "anti-Voicing Effect" (Stuart-Smith et al., 2019). The conditional R^2 values range between 0.54 and 0.61 and they are therefore the highest of all vowels. The best model was fit with the SVLR1 sub-dataset and a summary of the output is given in Table 64.

Table 64. Best linear mixed model for the lexical set CHOICE fit by maximum likelihood (marginal R^2 = 0.31; conditional R^2 = 0.61). T-tests use Satterthwaite's method. The estimates of the fixed effects are exponentiated to facilitate their interpretation. Model formula: log_dur ~ position + stress + syl_per_sec + reg + (1 | speaker) + (1 | syl_label) + (1 | word_label)

	AIC	**BIC**	**logLik**	**deviance**	**df.resid**
	32.6	86.3	-1.3	2.6	250

Scaled residuals:

	Min	**1Q**	**Median**	**3Q**	**Max**
	-2.3258	-0.5320	0.0032	0.6590	3.3269

Random effects:

Groups	Name	Variance	Std.Dev.
word_label	(Intercept)	2.624e-02	0.162002
syl_label	(Intercept)	1.799e-06	0.001341
speaker	(Intercept)	9.205e-03	0.095944
Residual		4.371e-02	0.209058

Intercept: VE1 long; positionfinal; stressnuclear; stylescripted;regEast_central

Fixed effects:

	Estimate	**Estimate** (exp)	**Std. Error**	**p-value**	**Sign. code**
(Intercept)	5.71536	303.4938	0.10689	< 2e-16	***
positioninitial	-0.19928	0.819320	0.09101	0.0296	*
positionmedial	-0.22791	0.796197	0.04611	2.27e-06	***
stressprimary	-0.18423	0.831741	0.03501	3.04e-07	***
stressunstressed	-0.29329	0.745804	0.13326	0.0297	*
syl_per_sec	-0.09385	0.910414	0.01622	2.05e-08	***
regWest_central	0.10845	1.114546	0.06108	0.0798	.

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

The model output incorporates the significant fixed factors *phrasal position*, *stress* and the local articulation rate in terms of *syllables per second*. The variable *regional background* improved the model fit but the durational differences for the regions did not reach statistical significance. The region with the lowest p-value is *West central*.

The model estimates that CHOICE tokens are approximately 18 percent shorter in initial and roughly 20 percent shorter in medial positions than when compared to the intercept (= final positions). This confirms the effect of *constituent-final lengthening* (see subsection 3.1.8). Likewise, the diphthong is 16.83 percent shorter in primary stressed syllables and 25.42 percent shorter in unstressed syllables than in nuclear stressed syllables (= intercept). CHOICE is therefore affected by *prosodic stress* (see subsection 3.1.7) and the strong difference in the effect sizes between the primary stressed and unstressed syllables further indicates an influence of *lexical stress* (see subsection 3.1.6).

Like the other vowels, CHOICE is also significantly affected by *tempo* (see subsection 3.1.9). The model estimates that the diphthong becomes roughly 9 percent shorter for every additional syllable produced in the timeframe of one second. This fits into the overall trend that vowels become shorter in faster articulation rates (see section 5.1).

The models and plots have shown that CHOICE is not significantly affected by the lengthening patterns of Aitken’s Law. The average duration of CHOICE is increased in voiced fricative contexts and before morpheme boundaries, but the durational differences do not reach statistical significance. Yet, the durational distributions do also not fit to the VE either. The vowels are longer in voiced fricative and voiced plosive contexts but there is another trend that contradicts the VE completely: CHOICE is significantly shortened before nasals. According to the effects of postvocalic consonant lengthening (see subsection 3.1.2), nasals should trigger longer vowels than plosives, but this is not the case for CHOICE in SSE. The shortening before nasal consonants can therefore be interpreted as an "anti-Voicing Effect" (Stuart-Smith et al., 2019).

The diphthong CHOICE was not extensively researched in previous studies. Zai (1942) was the first scholar who discusses the quantity of this vowel in Scottish English in detail. He specifies that, in contrast to other vowels of the Morebattle dialect, CHOICE is always long. Aitken (1981) lists two historical predecessors which represent the Modern Scots lexical set CHOICE, namely vowel 9 /oi/ and 10 /əi/ (see Table 5). He states that /əi/ in words such as <avoid>, <join> or <choice> is affected by the SVLR but he is unsure about the status of /oi/. In accordance with Zai's (1942) description, Aitken (1981) suggests that /oi/ is invariably long in some dialects of Scottish English. The other impressionistic accounts (see subsection 3.2.1) do not discuss this vowel. Moreover, CHOICE has also not been investigated by the various empirical studies on Scottish vowel duration. Hence, the present study provided the first detailed findings on the timing patterns in the diphthong. Overall, CHOICE is not affected by the phonological or morphological conditioning of Aitken's Law nor is the diphthong clearly influenced by the VE. Instead, the timing patterns of the diphthong show an "anti-Voicing Effect". The diphthong is particularly short before nasals. A lack of tokens prohibited a closer investigation of CHOICE tokens in voiced and voiceless plosive contexts. Nevertheless, the distributions reveal that the diphthong does not seem to follow the SVLR or VE timing patterns.

5.5 Summary

This chapter has summarized the present study's findings on vowel duration in 21st century SSE. Some vowels adhere to the timing patterns of the SVLR, some are influenced by the VE and there are others which are either influenced by both Aitken's Law and the VE or by none of them. A summary of the results can be found in Table 65.

Consistent SVLR patterns can be found in the lexical sets GOOSE, FLEECE, FACE, GOAT and PRICE. Here, the vowels are generally longer in the phonological and morphological SVLR long contexts and they are shorter in the respective short environments.

In contrast, the study could find no significant SVLR patterns in KIT, STRUT, DRESS, THOUGHT, or CHOICE. The findings do therefore agree with Aitken's (1981) classification that KIT and STRUT are SVLR-unaffected short vowels. Interestingly, the tokens of DRESS are longer in *SVLR1* short contexts than in *SVLR1* long environments. A major reason for this distribution is that DRESS tokens are the shortest on average before voiced fricatives. This does not only contradict the VE but also Aitken's Law.

The present study could also find a significant influence of the VE in the vowels FLEECE, FACE, GOAT and PRICE. GOOSE is not significantly longer before voiced plosives than voiceless plosives but the duration of the vowel is generally significantly longer before voiced consonants. Similar to the latest investigation by Stuart-Smith et al. (2019), the present study could also find evidence for an “anti-Voicing Effect” in STRUT, DRESS, THOUGHT, MOUTH and possibly CHOICE.

Table 65. Summary of the present study’s findings for the lexical sets with Aitken’s (1981) SVLR status, the highest conditional R^2 values, the categorization schemes SVLR1, SVLR2, VE1 and VE2 as well as other important observations.

Lexical Set	**Vowel(s)**	**Aitken’s (1981) SVLR status**	**Highest cond. R^2**	**SVLR1**	**SVLR2**	**VE1**	**VE2**	**Other observation**
KIT	/ɪ/	no	0.36	no	no	no	no	shortest before nasals
STRUT	/ʌ/	no	0.42	no	no	opposite	opposite	shortest before nasals, anti-VE
DRESS	/ɛ/	yes	0.42	opposite	no	opposite	no	shortest before voiced fricatives
GOOSE	/u/	yes	0.40	yes	yes	yes	no	*VE1* significant but not *VE2;* shortest before nasals
FLEECE	/i/	yes	0.49	yes	yes	yes	yes	shortest before nasals
THOUGHT	/ɔ/	regional variation	0.53	weak	weak	opposite	no	shortest before nasals; anti-VE
FACE	/e/	yes	0.49	yes	yes	yes	yes	shortest before laterals
GOAT	/o/	regional variation	0.41	yes	yes	yes	yes	shortest before voiceless fricatives
CAT	/a/	yes	0.48	no	weak	no	yes	shortest before voiced fricatives
MOUTH	/ʌʊ/	yes	0.47	yes	no	opposite	no	shortest before nasals
PRICE	/ʌɪ/ /ae/	yes	0.46	yes	yes	yes	yes	shortest before voiceless fricatives
CHOICE	/ɔe/	yes / regional variation	0.61	no	no	opposite	unclear	shortest before nasals

These findings are closely linked to another trend that becomes apparent when checking the average durations of the vowels in different postvocalic consonant contexts: KIT, STRUT, GOOSE, FLEECE, THOUGHT, MOUTH and CHOICE are the shortest before nasal contexts. This is unusual because vowels are typically longer before nasals than before voiceless plosives and fricatives (House & Fairbanks, 1953, p. 108). Nevertheless, following nasal consonants are considered SVLR short contexts, so these distributions would therefore correspond to the effect of Aitken’s Law.

Another general finding is that all vowels in SSE are significantly affected by the variables *tempo*, *phrasal position* and *stress*. Vowels are generally longer in phrase-final positions and in syllables under nuclear stress. Furthermore, all vowels are affected by the local articulation rate: vowels become shorter if the speed of speech is faster.

6. Discussion

This chapter summarises and discusses the findings of the present study against the background of the research questions (see chapter 1). In section 6.1, I will discuss which vowels are affected by Aitken’s Law and/or by the VE in 21st century SSE. Section 6.2 will focus on sociolinguistic variation in contemporary SSE vowel durational patterns and section 6.3 will elaborate on how prosodic factors influence vowel duration and the timing patterns of the VE and the SVLR. In each section, I will first briefly summarize the main findings for each research question. I will then take a broader approach and discuss how the findings can be interpreted and what general patterns can be observed. A summary can be found in section 6.4.

6.1 Aitken’s Law and the VE in 21st century spoken SSE

The present investigation has shown that some of the vowels in SSE are affected by the durational patterns of Aitken’s Law and/or by the VE. Relatively stable SVLR patterns could be found in GOOSE, FLEECE, PRICE, FACE and GOAT. These vowels are significantly longer in *SVLR1* and *SVLR2* long contexts than in the respective short environments. In contrast, the phonological and morphological SVLR does not affect the short monophthongs KIT, STRUT, or DRESS nor the diphthong CHOICE and the effects are very weak in THOUGHT. The morphological conditioning of Aitken’s Law only mildly affects CAT, but it is not significant in the diphthong MOUTH. Some of the vowels are also affected by the VE: the lexical sets FLEECE, FACE, GOAT and PRICE are generally longer when followed by voiced consonants than by voiceless consonants. No significant VE influences could be detected in KIT and the difference between postvocalic voiced and voiceless plosive contexts (*VE2*) is also not significant in DRESS, GOOSE, THOUGHT, or MOUTH. The vowel CAT shows a mixed picture: while CAT vowels are not significantly longer in the *VE1* long contexts than in the respective short contexts, the *VE2* long contexts trigger significantly longer CAT durations than the *VE2* short environments. Apart from that, the present investigation could also find an “anti-Voicing Effect” (Stuart-Smith et al., 2019) in STRUT, DRESS, THOUGHT, MOUTH and CHOICE for the *VE1* classification. This means that these vowels are generally shorter before voiced consonants than before voiceless consonants. The short monophthong STRUT is also significantly shorter before voiced plosives than before voiceless plosives (*VE2*). The vowels which are affected by a significant “anti-Voicing Effect” are, however, not consistently affected by Aitken’s Law either.

The findings of the present study therefore not only contradict previous experimental studies which found a consistent VE (House & Fairbanks, 1953) but also demonstrate that a strong SVLR effect does not go hand in hand with a weakening of the VE. Many previous investigations expected that a strong SVLR leads to a weakening of the VE because some of the long VE contexts are short SVLR environments (i.e. following nasals, voiced plosives and the liquid /l/) (Chevalier, 2019; Hewlett et al., 1999; Llamas et al., 2011; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; van Leyden, 2002; Watt & Ingham, 2000; Watt & Yurkova, 2007). If those SVLR short contexts condition short vowel durations, this would result in a stronger SVLR effect but a weaker VE at the same time. However, the present study has clearly shown that those vowels which do not show VE-related lengthening effects are also not significantly affected by Aitken's Law. The SVLR does not operate in KIT, STRUT, DRESS, THOUGHT, or CHOICE and these vowels are also not influenced by the VE. Similarly, those vowels which are significantly affected by the SVLR also show VE-related lengthening at the same time. For example, the high vowel FLEECE is clearly affected by the effects of the phonological and morphological SVLR, but the durational difference of FLEECE is also significant between voiced and voiceless plosives (*VE2*). The present investigation has therefore shown that the effects of Aitken's Law and the VE do not cancel each other out. While there is, of course, an overlap between the long SVLR and long VE contexts, the *VE2* classification scheme, which included only voiced and voiceless plosive environments, could demonstrate that there is significant VE-related lengthening in many vowels which are also affected by Aitken's Law at the same time. A juxtaposition of Aitken's Law and the VE might therefore not be accurate to fully grasp the durational patterns in naturally spoken SSE. Whereas previous studies have often interpreted that strong SVLR patterns are equal to weak VE patterns or that a weakening of Aitken's Law means a strengthening of the VE and vice versa, this mutually exclusive relationship might not be fully precise.

Instead, if one puts the SVLR and VE categorizations aside and focuses on the precise postvocalic consonant environments, one can observe an overall trend in SSE: the present study has shown that most of the vowels are shortened before nasal environments. The vowels KIT, STRUT, GOOSE, FLEECE, THOUGHT, MOUTH and CHOICE are shortest when they are followed by nasals. The average durations of these vowels before nasal environments are even shorter than the respective mean durations before voiceless plosives. This contradicts the VE in the way that nasals would be considered VE long environments, but this does not necessarily mean that there are no significant durational differences between the vowels in voiced and voiceless plosive or fricative contexts. In other words, just because the vowel /i/ is shortened in the word <bean> due to the postvocalic nasal environment, this does not mean that the duration of the same vowel does not significantly differ in the words <bead> and <beat> with the postvocalic voiced and voiceless plosive environments. A general feature of SSE vowel duration is therefore that many vowels are shortened before nasals. Postvocalic nasals might therefore be one of the most important environments of the SVLR. Most previous studies have not included following nasals in their analyses (Llamas et al., 2011; McClure, 1977; Pukli, 2006; Scobbie, 2005; van Leyden, 2002; Warren, 2018; Watt & Ingham, 2000; Watt & Yurkova, 2007), but this exclusion could have a strong influence on their findings. For example, the studies in the Northeast of Scotland (Warren, 2018; Watt & Yurkova, 2007) come to the conclusion that the VE is firmly established in that region because there is significant vowel lengthening before tautomorphemic voiced plosive contexts. One could also come to that conclusion in this study if one only addresses the plosive environments: FLEECE, FACE, GOAT, CAT and PRICE are significantly longer before voiced than before voiceless plosives. However, this does not mean that the vowels are unaffected by Aitken's Law, because most of them are also significantly longer in *SVLR1* and *SVLR2* long environments than in the respective short contexts. In contrast to the investigations by Warren (2018), Watt and Yurkova (2007) and many others, the present study's durational difference between the fricative contexts always exceeds the durational difference between the plosive environments; the only exceptions are the SVLR- and VE-unaffected vowels THOUGHT, CAT and DRESS. This means that, for example, GOOSE vowels are clearly longer before voiced fricatives (i.e. <choose>) than before voiceless fricatives (i.e. <loose>), but the vowel is only slightly longer before voiced plosives (i.e. <mood>) than before voiceless plosive contexts (i.e. <loot>).

A general difficulty is, of course, that there is no precise durational threshold "at which the difference between VE- and SVLR-context conditioned lengthening becomes sufficient to establish whether it is valid to talk of an SVLR effect distinct from the VE (…)" (Llamas et al., 2011, p. 1284). In other words, it is not possible to say which durational difference validates the operation of Aitken's Law and/or the VE. It is true that the average durational difference of the vowel FLEECE is greater between voiced and voiceless fricative contexts than between the voiced and voiceless plosive environments (see Table 42), but does this difference constitute the existence or absence of the SVLR and/or the VE in 21st century spoken SSE? While this point is generally open for discussion, the present study and its various statistical models found that there are, indeed, significant durational differences of the vowels in the different postvocalic contexts which are related to Aitken's Law and the VE. There is usually VE-related lengthening, especially when comparing the voiced plosive and voiceless plosive contexts, but the SVLR-related lengthening effect before voiced fricatives and morpheme boundaries is also significant in many vowels. The present study has thereby also demonstrated that it is very important to analyze all the vowels in all possible environments because only this holistic approach provides a full overview of all vowel duration patterns. Likewise, the consonant categorization into SVLR and VE long and short environments is also very important because this can have a profound influence on the findings. Whereas many previous studies used relatively simple SVLR and VE classifications, the present study implemented four different categorization schemes and the distributions can widely differ: the SVLR-affected vowel GOOSE, for instance, is significantly lengthened in the *VE1* categorization but it is not significantly longer in the *VE2* long environments than in the *VE2* short contexts. It is therefore important to investigate all vowels and environments from different perspectives, especially since many of the SVLR and VE short and long contexts overlap. Apart from that, the distributions of the present study show that the vowel durational patterns vary strongly in naturally spoken language and that there is a strong influence of prosodic factors (see section 6.1).

Another difficulty is that not only quantity but also quality defines the status of English long and short vowel phonemes. Many previous studies have shown that the perception of vowel quantity is sometimes not only based on duration but it can also be based on perceived vowel quality differences (Lehiste, 1970, p. 30). I used the Basic Scottish Vowel System (Abercrombie, 1979) as a reference model, but it could be the case that the vowel system in 21st century SSE has changed.

Nevertheless, many of the findings are in line with previous investigations: the short monophthongs /ɪ/, /ʌ/ and /ɛ/ are not significantly affected by Aitken's Law or by the VE (McClure, 1977; McKenna, 1988; Stuart-Smith et al., 2019), but the SVLR is operating in the high front vowel /i/ (Chevalier, 2019; Hewlett et al., 1999; McClure, 1977; McKenna, 1988; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019; van Leyden, 2002; Watt & Ingham, 2000), the high back vowel /u/ (Chevalier, 2019; Hewlett et al., 1999; McClure, 1977; McKenna, 1988; Pukli, 2006; Rathcke & Stuart-Smith, 2016; Scobbie, 2005; Scobbie, Turk, & Hewlett, 1999; Stuart-Smith et al., 2019; van Leyden, 2002; Watt & Ingham, 2000) as well as in the diphthong /aɪ/ (Llamas et al., 2011; McMahon, 1991; Pukli, 2006; Scobbie, Hewlett, & Turk, 1999; Scobbie, Turk, & Hewlett, 1999). Apart from that, the present study has also shown that the SVLR also operates consistently in the monophthongs FACE and GOAT. This is surprising because most other studies found that the patterns of Aitken's Law are relatively weak or even absent in these vowels (McClure, 1977; McKenna, 1988; Scobbie, 2005; Scobbie, Hewlett, & Turk, 1999; Stuart-Smith et al., 2019; Warren, 2018; Watt & Ingham, 2000; Watt & Yurkova, 2007). This means that, overall, Aitken's Law is operating in 21st century spoken SSE, especially in the high and mid vowels /i/, /u/, /e/, /o/ and in the diphthong /aɪ/. The VE is also in operation in these vowels which means that the SVLR and the VE do not cancel each other out and a general feature of 21st century SSE vowel duration patterns is a strong shortening before nasals.

6.2 Sociolinguistic variation in SVLR and VE patterns in 21st century spoken SSE

The sociolinguistic factors affect SVLR and VE timing patterns in some vowels. As for regional variation, for example, the patterns of Aitken's Law are generally less stable in the *South* of Scotland. While Aitken's Law operates in FLEECE, the interaction between *Southern* and the *SVLR1* and *SVLR2* short contexts shows that the effect is less strong in that region (see subsection 5.3.2). The SVLR effect in the South is also less strong in FACE (see subsection 5.3.4). In contrast to this, the influence of the VE is stable in the dialect region *Southern* for both FLEECE and FACE. Apart from that, the phonological conditioning effect of the SVLR is less pronounced in the vowels FLEECE and MOUTH for the region *HHE*. A possible explanation for this might be that Scottish Gaelic, but not Scots, has been the predominant language of the Highlands and Hebrides (see section 2.1). The Highlands and Hebrides have never been traditional Scots speaking areas, so the influence of Aitken's Law might be limited here. In addition, the Highlands and Hebrides have relatively large proportions of residents who were born in England (National Records of Scotland, 2011), which might be another reason for the less stable SVLR patterns. The vowel durational patterns might adhere more to the VE than to Aitken's Law if many speakers in the Highlands are from England. A similar explanation is possible for the weaker SVLR effects in the South of Scotland. Large parts of the population in the South of Scotland were born in England (National Records of Scotland, 2011), so this might be a reason why the SVLR is less stable there. Another possible reason is the proximity of the region to the English border: due to higher contact with speakers from England, the speakers in *Southern* Scotland might converge more to the durational patterns of SSBE (=VE) than to Aitken's Law. In contrast to the less stable SVLR patterns in the Highlands and Hebrides as well as in the South of Scotland, the present study could find relatively consistent patterns in all other regions. This also includes the dialect region *Northeast*, which contradicts the findings of previous studies in that region (Warren, 2018; Watt & Yurkova, 2007). A possible explanation could be that the previous studies have primarily focused on Scots data in mostly controlled speech settings, so it could be that the different durational patterns arise from the discrepancy of the samples; the previous studies investigated Scots in controlled speech but the present investigated SSE in uncontrolled speech. However, the different findings can also be attributed to different methodological approaches: as mentioned in the sections 3.2 and 6.1, Warren (2018) as well as Watt and Yurkova (2007) only investigated vowels in postvocalic plosive and fricative contexts, so they did not analyze vowel duration in all possible environments. The present study, however, included all postvocalic contexts in its analysis and could find a strong shortening before nasals. The postvocalic nasal environments were, however, not investigated by Watt and Yurkova (2007) or by Warren (2018), so this could be another reason for the different findings in these studies.

Apart from the *regional background*, the SVLR and VE patterns also partly differ for the variables *gender* and *age group* in a few vowels. For example, the phonological conditioning of Aitken's Law is less strong for the male speakers in the diphthong MOUTH and the *VE2* effect is also weaker for the male speakers in FACE. In contrast, the *VE2* effect is stronger among the young speakers in the monophthong GOAT. While some social factors have a significant influence on the SVLR and VE patterns in some of the vowels, the influence of sociolinguistic variation is relatively limited. It might be true that the male speakers, for instance, have a weaker *SVLR1* effect in FACE and a weaker *VE2* influence in MOUTH, but the differences between the female and male speakers are not significant across most vowels. The present study therefore contradicts previous suggestions that Aitken's Law could be weakening among younger female speakers (Chevalier, 2019; Milroy, 1995; Rathcke & Stuart-Smith, 2016; Watt & Ingham, 2000) or that is varies significantly in different regions (Warren, 2018; Watt & Yurkova, 2007). The findings of the present study have shown that the SVLR and VE effects are relatively consistent across the *regional backgrounds*, but, in particular, across *genders* and *age groups* in 21st century SSE. This lack of consistent variation in the vowel duration patterns can be interpreted as indicative that SSE is a stable standard variety across the country. The present study has investigated SSE and not Scots, so most of the speech data represents standard language spoken in formal registers (i.e. radio broadcasts, political speeches, broadcast discussions) and this might be the reason why there is little sociolinguistic variation. Whereas the previous studies which focused on uncontrolled speech (see subsection 3.2.3) were investigating predominantly Scots data and found significant sociolinguistic variation (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Warren, 2018), the present study focused on SSE and found stable vowel duration patterns across different *regional backgrounds*, *genders* and *age groups*. The findings therefore support the notion that, in contrast to Scots, which varies significantly in different regions (Johnston, 1997), the phonology of SSE is relatively similar across Scotland (McClure, 1994, p. 79; Stuart-Smith, 2008, p. 48). As discussed in section 2.4, the linguistic situation in Scotland can best be explained by a bipolar linguistic continuum with Scots on the one end and SSE on the other. It might therefore be the case that sociolinguistic variation in Aitken's Law is stronger when people are adhering more to the Scots end of the continuum.

Another possible explanation for the lack of sociolinguistic variation in the SVLR in the present study is that the fine-grained effects of Aitken's Law do not reach the consciousness level of the speakers in connected speech. Whereas some pronunciation features (i.e. rhotic sounds, vowel quality, glottal stops) are more salient and can be used by speakers to index social meaning (Eckert, 2008; Silverstein, 2003), the patterns of Aitken's Law are very subtle and the speakers might therefore not be aware of them. Speakers can consciously or unconsciously front the vowel quality in the word <moon> to the 'Scottish' pronunciation /mʉn/, but the vowel quantity in this word might not be associated with a sociolinguistic variable. By using the fronted /ʉ/ in the word <moon> or the vowel /a/ in the word <trap>, speakers can signal their regional background, but the patterns of Aitken's Law may not be as salient to index social meaning or belonging. At the same time, vowel duration is a linear variable and varies with regard to many intralinguistic factors (see section 3.1). For instance, a vowel in an utterance with a slow articulation rate tends to be longer than a vowel in an utterance with a high articulation rate (see subsection 3.1.9). At the same, time, the present study has also found that the effects of Aitken's Law and the VE often interact with other prosodic factors (see section 6.3). For instance, the vowel GOOSE in the word <move> can be lengthened due to the phonological effect of the SVLR in the voiced fricative context, but this vowel can also be lengthened because its syllable is under nuclear stress in a particular utterance in connected speech (i.e. <he's on the MOVE>). The general variability of vowel duration and the interactions of vowel duration patterns with other prosodic factors may further obscure the timing effects of the SVLR in natural spoken language. As seen in chapter 5, there is great durational variability overall and the average differences between SVLR long and short vowels are only roughly 15 ms (mean *SVLR1* difference) to 21 ms (mean *SVLR2* difference). Given that the just-noticeable difference for vowel sounds is at approximately 50 ms in controlled experimental settings (Labov & Baranowksi, 2006), it is rather unlikely that regular speakers perceive the patterns of Aitken's Law in everyday speech. The present study has exclusively looked at speech production, so it would be the task of a speech perception study to investigate whether the durational patterns are perceivable in naturally spoken language.

The effect of variable speech *style* turned out to be insignificant in many SSE vowels as well. Only the lexical sets STRUT, THOUGHT, FACE, MOUTH and PRICE are significantly shorter in unscripted than in scripted speech. The variable *style* also not interacts with Aitken's Law or the VE, so the influence of this variable is minimal. In other words, the effects of the SVLR and the VE are relatively stable in both scripted and unscripted speech forms. It does not matter if one investigates a speech in the Scottish Parliament, a radio broadcast, or a personal discussion; the effects of Aitken's Law and the VE do not significantly vary in these registers. This finding contradicts the expectation that the vowel durational patterns might be less stable in unscripted speech (see subsection 4.1.2). Yet, the overall effect sizes are comparable with those of recent studies in uncontrolled spoken language (see Table 25 and subsection 3.2.3). Tanner et al. (2020, pp. 8–9) found that the VE effect size lies between 1:1 to 1.09:1 in different Scottish varieties and the present study found similar average distributions (average *VE1* ratio: 0.98:1; average *VE2* ratio 1.10:1). Overall, the effects of Aitken's Law are stronger than those of the VE (average *SVLR1* ratio 1.18:1; average *SVLR2* ratio 1.25:1), but the durational differences are generally smaller than those found in controlled speech settings (Chen, 1970; House & Fairbanks, 1953; Peterson & Lehiste, 1960; Watt & Ingham, 2000). This corresponds to earlier observations that the patterns of Aitken's Law and the VE are weaker in naturally occurring language (Tanner et al., 2020, p. 8). It is noteworthy, however, that Aitken's Law is equally stable in both more scripted and more spontaneous speech because this indicates that the patterns of the SVLR are well-established and stable in 21st century SSE.

While there are no consistent interactions between social factors and Aitken's Law or the VE, vowel duration itself partially varies with respect to the social characteristics of the speakers. This means that the social variables have a limited general influence on vowel duration, but they do not have a consistent influence on the patterns of Aitken's Law or the VE. A general pattern in the dataset of the present study is that older speakers produce slightly longer vowels than the other age groups. The plots have shown that vowel duration tends to be slightly longer in the *age group old* and the durational difference reaches statistical significance in the short monophthongs KIT and STRUT. This increased vowel duration is closely linked to the slower speed of speech among older SSE speakers. Since older speakers have a significantly slower local articulation rate (see section 5.1), they produce slightly longer vowels which corresponds with earlier research on *tempo* (see subsection 3.1.9). The influence of the variable *age group* is, however, not significant if one considers the durations of all vowels combined (see section 5.1). This means that, although older speakers produce longer vowels on average, the durational difference is often not significant. This finding of the present study provides further evidence that the SVLR is stable across the age groups in SSE. Even though older speakers produce slightly longer vowels than the middle-aged or younger speakers, the patterns of Aitken's Law are consistent across all generations investigated.

The present study could therefore show that the influence of sociolinguistic variation on Aitken's Law and the VE is very limited. The SVLR and the VE are relatively stable across different *genders*, *age groups* and *regional backgrounds* in 21st century spoken SSE.

6.3 Prosodic factors and the SVLR and VE in 21st century spoken SSE

The results have shown that prosodic factors have a strong influence on vowel duration in contemporary SSE. I fitted many linear mixed effects models for various vowels and with different categorization schemes, but all models returned significant effects for the variables *tempo*, *phrasal position* and *stress*. In other words, the present study found a clear influence of *prosodic stress* (see subsection 3.1.7), *constituent-final lengthening* (see subsection 3.1.8) and *tempo* (see subsection 3.1.9) in all model constellations. These factors have the strongest and most consistent influence on 21st century SSE vowel duration. Generally, all vowels are longer in phrase-final and nuclear stressed syllables than in non-final or non-nuclear stressed syllables and vowel duration generally depends on the local articulation rate: the faster the speech, the shorter the vowels.

At the same time, the timing patterns of the SVLR and the VE are also affected by these prosodic factors. This means that not only vowel duration in general, but also the effects of the SVLR and VE on vowel duration are influenced by the prosodic factors. In all SVLR-affected vowels, the phonological conditioning of Aitken's Law is amplified in phrase-final positions. This means that the difference between the *SVLR1* long and short contexts is stronger in utterance-final syllables for GOOSE, FLEECE, FACE, GOAT and PRICE. For example, the durational difference between the vowels in the last syllables of the utterances <he's on the *move*> (<move> = *SVLR1* long context due to postvocalic voiced fricative /v/) and <he's got no *proof*> (<proof> = *SVLR1* short context due to postvocalic voiceless fricative /f/) is greater because both syllables are positioned at the end of an utterance. The durational difference between the vowel nuclei in these words would be less pronounced in phrase-medial positions.

In contrast to the phonological conditioning of Aitken's Law (*SVLR1*), the morphological conditioning of Aitken's Law (*SVLR2*) does not show significant interactions with the variable *phrasal position*. The present study found no significant interactions in any of the vowels. This means that the durational difference between the last syllables in utterances such as <We are all *agreed*> (*SVLR2* long) and <We don't like *greed*> (*SVLR2* short) is not greater than when the words are positioned in phrase-medial positions. In short, whereas the phonological conditioning of Aitken's Law interacts with the phrasal position in all vowels, the morphological SVLR conditioning does not. This shows that the phonological and morphological conditioning of Aitken's Law are, indeed, two separate processes because their lengthening patterns behave differently. It may be true that both the phonological and morphological conditioning of Aitken's Law affect the vowels GOOSE, FLEECE, FACE, GOAT and PRICE, but the difference between vowels in short and long contexts is only amplified for the phonological conditioning in phrase-final positions.

Significant interactions between the *phrasal position* could also be found for the *VE1* contexts in FLEECE, CAT and MOUTH as well as between *VE2* categorization and the *phrasal position* in FACE, GOAT, CAT and PRICE. This means that the difference between the *VE1* long and short contexts is significantly greater for the vowels FLEECE, CAT and MOUTH in phrase-final positions than in non-final positions. The durational difference for FACE, GOAT, CAT and PRICE in voiced and voiceless plosive environments is also greater when the syllable is positioned before a pause. The overlap in the interactions for the *VE1* and *VE2* classifications is not surprising because the *VE2* contexts (postvocalic voiced vs voiceless plosives) are also part of the *VE1* categorization (postvocalic voiced vs. voiceless consonants). Likewise, many of the vowels which show interactions between the *VE2* category on the one hand and the *phrasal position* on the other hand also show the same interaction for the phonological conditioning of Aitken's Law, namely FACE, GOAT, PRICE. This provides further evidence for the fact that the durational patterns of Aitken's Law and the VE are linked.

Apart from the strong influence of *constituent-final lengthening* (see subsection 3.1.8) on the SVLR and VE patterns, the durational difference between *SVLR1* long and short environments is also amplified in nuclear stressed syllables for the monophthongs FLEECE, FACE, GOAT and CAT. Hence, the durational difference between the *SVLR1* long and short environments increases when the corresponding syllables are under nuclear stress. The durational difference between vowels in *SVLR1* long and short contexts is less strong in non-nuclear stressed syllables. This means that there is a clear interaction between the phonological conditioning of Aitken's Law and *prosodic stress* (see subsection 3.1.7). The models for the *SVLR2* categorization do, however, not show any significant interactions between the morphological conditioning of Aitken's Law and *prosodic stress*. This means that the durational difference between the syllables <I *brewed* the beer> (*SVLR2* long) and <I *brood* on the problem> (*SVLR2* short) is similar, irrespective of whether the syllables are under nuclear stress or not. The interactions are therefore comparable to those with the variable *phrasal position*: whereas the phonological conditioning of Aitken's Law shows interactions with the variables *phrasal position* and *stress*, the morphological SVLR conditioning does not show any interactions with either of them. This provides further evidence that the phonological and morphological conditioning of the SVLR are two separate processes.

Moreover, the difference between the *VE1* long and short environments is also greater in nuclear stressed syllables for GOOSE, FLEECE, THOUGHT, FACE and GOAT. The contrast between the *VE2* environments is also significantly stronger when FACE is positioned in nuclear stressed syllables. These interactions demonstrate that also the VE-related lengthening effects are often amplified under nuclear stress.

The findings that SVLR and VE patterns are amplified in prominent prosodic positions are in line with the findings of previous studies in connected speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Warren, 2018). This means that Aitken's Law and the VE cannot be seen as isolated processes in connected speech, the durational patterns interact with the factors *phrasal position* and *prosodic stress*. However, a new finding of the present study is that there is a clear difference between the phonological and morphological conditioning of Aitken's Law. Whereas the durational differences between *SVLR1* long and short contexts are amplified in nuclear stressed syllables and/or phrase-final syllables, the differences between the *SVLR2* long and short environments are not. Most previous studies in connected speech have not distinguished between the phonological and morphological conditioning of Aitken's Law in their statistical analyses (see subsection 3.2.3), so this new finding can be attributed to the precise SVLR categorizations used in the present study (see subsection 4.4.2)

Apart from that, the modeling further revealed significant interactions between the *phrasal position* and *stress* in THOUGHT and PRICE for the *SVLR1* categorization as well as in DRESS for the *VE2* categorization. This means that duration increases significantly if these vowels are emphasized with nuclear stress at the end of an utterance.

The other prosodic factors influenced vowel duration less consistently and the SVLR and VE categories did not show any further interactions in the present investigation. Nevertheless, the duration of many vowels varies with respect to the number of syllables in a word and the number of phones in a syllable: significant *intrasyllabic compression* effects (see subsection 3.1.3) could be found in the lexical sets KIT, GOOSE, FLEECE, THOUGHT, GOAT, CAT, MOUTH and PRICE but not in STRUT, DRESS, FACE and CHOICE. *Polysyllabic shortening* (see subsection 3.1.4) only had a significant influence in KIT, GOOSE, THOUGHT, GOAT and CAT, but not in the other vowels. This demonstrates that most vowels tend to become shorter in syllables with many phones and in words with many syllables, but the effects cannot be found across all vowels in contemporary spoken SSE. The data of the present study therefore reveals a varied influence of compensatory shortening in the vowels of 21st century SSE. A general influence of intrasyllabic compression and polysyllabic shortening could also be found in previous studies (Chevalier, 2019; Rathcke & Stuart-Smith, 2016).

The influence of the variable *lexical frequency* (see subsection 3.1.5) is even less important. Significant durational differences between high and low frequency words could only be found in the lexical set KIT. The manually fitted model for all vowels (see section 5.1) returned significant effects for this binary variable, but *lexical frequency* is not significant in the corresponding PrInDT model. While the models only incorporated *lexical frequency* as a binary variable, the linear plots for this variable also did not show a clear trend. This demonstrates that the variable *lexical frequency* does not have a striking and consistent effect on vowel duration in contemporary SSE. While this finding is in line with the study by Rathcke and Stuart-Smith (2016), it contradicts the result of Stuart-Smith et al. (2019) who found a significant influence of lexical frequency in their investigation.

The present study could therefore show that prosodic factors have a strong influence on the timing patterns of Aitken's Law and the VE. The difference between the SVLR- and VE-related long and short contexts are stronger in phrase-final positions and in nuclear stressed syllables. Vowel duration also generally differs with regard to the local articulation rate.

6.4 Summary

This chapter has discussed the findings of the present investigation against the background of the research questions. Stable SVLR patterns could be found in GOOSE, FLEECE, PRICE, FACE and GOAT, but Aitken’s Law does not operate in the short monophthongs KIT, STRUT and DRESS. Consistent VE-related lengthening effects can be found in FLEECE, FACE, GOAT and PRICE. Apart from that, the present study also found evidence of an “anti-Voicing Effect” (Stuart-Smith et al., 2019) in DRESS, THOUGHT, MOUTH, CHOICE and especially STRUT. A general pattern which is also mainly responsible for this “anti-Voicing Effect” is that most vowels are strongly shortened before nasals in SSE. Another general finding of the present study is that the effects of Aitken’s Law and the VE are not mutually exclusive. Most of the vowels that show consistent SVLR patterns are also affected by the VE. The influence of sociolinguistic factors is limited; the VE and Aitken’s Law are relatively stable across different age groups, genders and regions. One explanation for the limited sociolinguistic variation is that the present study has exclusively scrutinized SSE speech in mostly formal contexts. Another explanation is that the patterns of the SVLR and the VE are below the consciousness level so that speakers do not generally use Aitken’s Law or the VE for indexing social meaning in their speech. In contrast to the weak influence of sociolinguistic variables, prosodic factors have a strong influence on the vowel duration patterns in contemporary SSE. The effects of the SVLR and the VE are amplified in prominent prosodic environments, especially in phrase-final and nuclear stressed syllables.

7. Conclusion

The present study has investigated the effects of Aitken’s Law and the VE in 21st century spoken SSE. It is the first study which has analyzed all the vowels of SSE in all possible phonetic contexts on a countrywide scale. Due to contradictive findings in previous studies (see section 3.2), the aim was to find out which vowels are affected by the SVLR / VE in 21st century spoken SSE (RQ1) and in how far the vowel duration patterns are influenced by regional, age- and gender-related variation (RQ2). Furthermore, I also wanted to investigate in how far Aitken’s Law and/or the VE is affected by prosodic factors (RQ3). Following precise data selection and transcription criteria (see section 4.1), I assembled an up-to-date dataset that is balanced in terms of the speakers’ *regional background*, *age* and *gender*. The dataset comprises files from the ICE Scotland corpus (Schützler et al., 2017) but also self-collected data and I made sure that it has comparable proportions of scripted and unscripted spoken language to account for the variable *style* (see section 4.2). In accordance with the transcription criteria (see subsection 4.1.3), I semi-automatically transcribed and force-aligned the data with the help of WebMAUS (Kisler et al., 2017), the MFA 2.0 (McAuliffe et al., 2017) and self-designed scripts in Python (Python Software Foundation, 2021) (see section 4.3). The transcription format includes the most important levels of the prosodic hierarchy (e.g. phoneme, syllable, word, utterance) and it also accounts for all relevant prosodic factors (see section 3.1). The analysis was carried out in R (R Core Team, 2021) and implements descriptive and inferential statistics (see section 4.4). As for inferential statistics, I fitted linear mixed effects models on log-transformed vowel duration with *word*, *syllable* and *speaker* as random factors. I analyzed all the vowels of the Basic Scottish Vowel System (Abercrombie, 1979) both collectively and independently and I incorporated all possible intralinguistic and extralinguistic variables as fixed factors as well as all possible interactions. Furthermore, the analysis included two SVLR and two VE categorizations (see subsection 4.4.2) to investigate vowel duration patterns in detail and I fitted different models for the individual categorization schemes.

Regarding the first research question (RQ1), the analysis has found consistent SVLR patterns in the vowels /u/, /i/, /e/, /o/ as well as in the diphthong /aɪ/. Aitken’s Law does, however, not operate in the short vowels /ɪ/, /ʌ/, /ɛ/ or in diphthong /ɔe/ and the patterns are very weak in the vowels /ɔ/ and /a/. While there are clear SVLR patterns, the present study could also find consistent VE effects in /i/, /e/, /o/ and /aɪ/, but an “anti-Voicing Effect” (Stuart-Smith et al., 2019) in /ɛ/, /ɔ/, /ʌʊ/, /ɔe/ and, in particular, in the short monophthong /ʌ/. As for the second research question (RQ2), the SVLR and VE patterns are only sporadically affected by sociolinguistic variables (see section 6.2), which means that Aitken’s Law is relatively stable across different dialect regions, age groups and genders in 21st century SSE. With regard to the third research question (RQ3), the present study has found that the patterns of Aitken’s Law and the VE are strongly and consistently influenced by prosodic factors (see section 6.2). In particular, the variables *stress* (see subsections 3.1.6 and 3.1.7), *phrasal position* (see subsection 3.1.8) and *tempo* (see subsection 3.1.9) have a significant influence on all vowels. Vowel duration generally increases in nuclear stressed and phrase-final syllables and the duration further depends on the local articulation rate: the faster the speech, the shorter the vowels. In addition, the effects of the SVLR and VE are amplified in prominent prosodic positions which means that the difference between SVLR and VE long and short contexts increases in nuclear stressed syllables and in phrase-final positions.

Another general observation of this study is that many vowels in SSE are shortened before nasal consonants. The vowels /ɪ/, /ʌ/, /u/, /i/, /ɔ/ and the diphthongs /ʌʊ/ and /ɔe/ are clearly shortened before nasals and the average duration of these vowels in nasal environments even falls below the respective mean durations in voiceless plosive contexts. Most previous studies have not included nasal environments in their analyses. However, the findings of the present investigation suggest that nasal contexts are one of the most important environments of Aitken's Law. Apart from this, the present study has also found that most vowels which are affected by clear SVLR lengthening are also significantly influenced by the VE: the vowels /i/, /e/, /o/ as well as the diphthong /aɪ/ are significantly affected by the patterns of both Aitken's Law and the VE. Likewise, those vowels which are not significantly affected by Aitken's Law are also not significantly influenced by the VE. The vowels /ɪ/, /ʌ/, /ɛ/, /ɔ/ and /ɔe/ do not show SVLR- or VE-related durational patterns. This means that the effects of Aitken's Law and the VE do not cancel each other out, but show considerable overlap.

To conclude, the present study has provided an in-depth account of the vowel duration patterns in 21st century SSE. In contrast to all previous studies, this investigation is the first which analyzed all vowels of the Basic Scottish Vowel System (Abercrombie, 1979) in all possible phonetic environments of scripted and unscripted naturally spoken SSE. Unlike most other investigations, the present study used a sample that is balanced in terms of the speakers age, gender and regional background and thus represents the whole of Scotland. The analysis accounted for all relevant prosodic factors that can influence the duration of vocalic intervals and implemented means of inferential statistics. Furthermore, the analysis incorporated different SVLR and VE classifications to investigate the influence of the postvocalic phonetic environment in detail. In short, the present study has revised Aitken's Law for 21st century SSE.

While the principal aim of the present study was to provide answers for the three research questions, there are still many opportunities for future research. The present study has, in accordance with almost all previous studies on the SVLR (see section 3.2), investigated Aitken's Law in speech production. To the best of my knowledge, only the study by Smith and Rathcke (2016) has so far investigated the SVLR in speech perception. As the present study has discussed that the patterns of Aitken's Law might fall below the consciousness level of the speakers in connected speech (see subsection 6.2), future investigations on the perception of SVLR and VE patterns seem rewarding.

Furthermore, the present study has not assessed the speakers' contact with SSBE, hence the influence of dialect contact on SSE vowel patterns is not clear. Following the hypotheses of earlier studies on Glaswegian Scots (Chevalier, 2019; Rathcke & Stuart-Smith, 2016), it could be the case that VE patterns are stronger among those SSE speakers who have frequent contact with SSBE speakers.

Hence, a future investigation including the variable of *SSBE contact* could provide deeper insights into the relationship between dialect contact and vowel duration patterns.

Moreover, in line with the other SVLR and VE investigations in connected speech (Chevalier, 2019; Rathcke & Stuart-Smith, 2016; Stuart-Smith et al., 2019; Tanner et al., 2020; Warren, 2018), the present study has placed vowel duration and the different factors influencing it (see section 3.1) at the center of its analyses. Consequently, the present study has not laid a special focus on vowel quality. I provided an overview of vowel quality in section 5.1 and it showed that, apart from considerable variation, F1 has a significant influence on vowel duration. However, as the first formant generally corresponds to the openness of the mouth, this provides further evidence that vowel duration is affected by tongue height: low vowels tend to be longer than high vowels. As a result, I did not include vowel quality measurements in the models for the individual vowels because I argued that the separate investigation of each single vowel already accounts for the influence of *intrinsic vowel duration* (see subsection 3.1.1). Future investigations, however, could benefit from incorporating more vowel quality measures in their analyses, especially since vowel quality and quantity usually interact in the phoneme structure of English.

Another option for further research would be to carry out a similar study in Southern England with a comparable dataset. As the present investigation has found a strong vowel shortening before nasals in SSE and suggested that this shortening process might be one of the most important features of Aitken's Law (see section 6.1), a subsequent study could carry out a comparable investigation of the vowel duration patterns in SSBE and this would provide clarity as to whether the shortening process before nasals is a specific SVLR-related feature. Of course, such an undertaking would require a comparable dataset with a transcription format that is similar to the one of this investigation, but I am convinced that such large-scale studies will become easier to carry out in the future. When considering the fast developments in speech technology (i.e. automatic speech recognition, forced alignment), I am confident that data preparation and analysis techniques will become easier to carry out in the years to come.

8. References

Abercrombie, D. (1979). The Accents of Standard English in Scotland. In A. J. Aitken & T. MacArthur (Eds.), *Languages of Scotland* (pp. 68–84). W&R Chambers.

Agutter, A. (1988a). The dangers of dialect parochialism: The Scottish vowel length rule. In J. Fisiak (Ed.), *Historical Dialectology: Regional and Social* (pp. 1–21). Mouton de Gruyter.

Agutter, A. (1988b). The not-so-Scottish Vowel Length Rule. In J. M. Anderson & N. Macleod (Eds.), *Edinburgh Studies in the English Language* (pp. 120–132). John Donald Publishers Ltd.

Aitken, A. J. (1962). *Vowel length in modern Scots*. Unpublished handout.

Aitken, A. J. (1975). *The Scottish Vowel-Length Rule*. Unpublished handout.

Aitken, A. J. (1977). How to Pronounce Older Scots. In A. J. Aitken, M. P. MacDiarmid, & D. S. Thomson (Eds.), *Bards and makars: Scottish language and literature: Medieval and Renaissance: A selection of the papers read at the First International Conference on Scottish Language and Literature, Medieval and Renaissance.* University of Glasgow Press.

Aitken, A. J. (1979). Scottish Speech: a historical view, with special reference to the Standard English of Scotland. In A. J. Aitken & T. MacArthur (Eds.), *Languages of Scotland* (pp. 85–118). W&R Chambers.

Aitken, A. J. (1981). The Scottish Vowel-Length Rule. In M. L. Samuels & M. Benskin (Eds.), *So meny people longages and tonges: Philological essays in Scots and mediaeval English presented to Angus McIntosh* (131-157). M. Benskin & M.L. Samuels.

Aitken, A. J. (1984a). Scots and English in Scotland. In P. Trudgill (Ed.), *Language in the British Isles* (pp. 517–532). Cambridge University Press.

Aitken, A. J. (1984b). Scottish accents and dialects. In P. Trudgill (Ed.), *Language in the British Isles* (pp. 94–114). Cambridge University Press.

Anderson, J. M. (1993). Morphology, Phonology and the Scottish Vowel-length Rule. *Journal of Linguistics*, *29*(2), 419–430.

Audacity Development Team. (2019). *Audacity: A Free Digital Audio Editor* (Version 2.3.3) [Computer software]. https://www.audacityteam.org/download/ [Date of access: 12.12.2022]

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5.1), 3048–3058.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [Computer software]. https://catalog.ldc.upenn.edu/LDC96L14 [Date of access: 12.12.2022]

Barnwell, T. P. (1971). *An algorithm for segment durations in a reading machine context*. Technical Report.

Bartón, K. (2022). *MuMIn package* (Version 1.46.0) [Computer software]. https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf [Date of access: 12.12.2022]

Bates, D., Mälcher, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology, 3*, 7–33.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A. M. (1867). *Visible Speech: the Science of Universal Alphabetics: Or, Self-interpreting Physiological Letters, for the Writing of All Languages in One Alphabet. Illustrated by Tables, Diagrams, and Examples*. Simpkin, Marshall. https://books.google.at/books?id=6IQ_AQAAMAAJ [Date of access: 12.12.2022]

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International, 5*(9/10), 341–347.

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* (Version 6.1.08) [Computer software]. http://www.praat.org/ [Date of access 12.12.2019]

Brown, K., & Miller, J. E. (2013). *The Cambridge Dictionary of Linguistics*. Cambridge University Press.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Campbell, E. (2001). Were the Scots Irish? *Antiquity, 75*(288), 285–292.

Carr, P. (1992). Strict Cyclicity, Structure Preservation and the Scottish Vowel-Length Rule. *Journal of Linguistics*, *28*(1), 91–114.

Chen, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica*, *22*(3), 129–159.

Chevalier, F. (2019). On sound change and gender: the case of vowel length variation in Scottish English. *Anglophonia*(27).

Cohn, A. C., Brugman, J., Crawford, C., & Joseph, A. (2005). Lexical frequency effects and phonetic duration of English homophones: An acoustic study. *The Journal of the Acoustical Society of America*, *118*(3), 2036.

Cox, R. A. V. (2011). Gaelic Place-names. In M. Watson & M. Macleod (Eds.), *The Edinburgh companion to the Gaelic language* (pp. 46–62). Edinburgh University Press.

Crystal, D. (2015). *A dictionary of linguistics and phonetics* (6th edition). Blackwell Publlishing.

Crystal, T. H., & House, A. S. (1988a). The duration of American-English vowels: an overview. *Journal of Phonetics*, *16*(3), 263–284.

Crystal, T. H., & House, A. S. (1988b). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America, 83*(4), 1553–1573.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, *88*(1), 101–112.

Di Paolo, M., & Yaeger-Dror, M. (2011). Field methods: gathering data, creating a corpus, and reporting your work. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 7–23). Routledge.

Di Paolo, M., Yaeger-Dror, M., & Beckford Wassink, A. (2011). Analyzing vowels. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 87–106). Routledge.

Dieth, E. (1932). *A Grammar of the Buchan Dialect (Aberdeenshire): Descriptive and Historical*. W. Heffer & Sons Ltd.

Douglas, S. (1991 [1775]. *A treatise on the provincial dialect of Scotland: edited by Charles Jones*. Edinburgh University Press.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476.

Elert, C.-C. (1964). *Phonologic studies of quantity in Swedish : based on material from Stockholm speakers*. Almqvist & Wiksell.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(3), 253–260.

Esposito, A. (2002). On Vowel Height and Consonantal Voicing Effects: Data from Italian. *Phonetica*, *59*, 197–231.

Ewen, C. J. (1977). Aitken's law and the phonatory gesture in dependency phonology. *Lingua*, *41*(3-4), 307–329.

Fruehwald, J. (2013). *The Phonological Influence on Phonetic Change*. University of Pennsylvania. https://repository.upenn.edu/edissertations/862 [Date of access: 12.12.2022]

Gahl, S. (2008). Time and Thyme Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language*, *84*, 474–496.

Giegerich, H. (1992). *English Phonology: An Introduction*. Cambridge University Press.

Gimson, A. C. (1973). *An Introduction to the Pronunciation of English.* Edward Arnold.

Goldman, J.-P., & Simon, A. C. (2020). Prosobox, a Praat Plugin for Analysing Prosody. *Proceedings of the 10th International Conference on Speech Prosody*, 1009–1013.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

Gonzalez, S., Grama, J., & Travis, C. E. (2020). Comparing the performance of forced aligners used in sociophonetic research. *Linguistics Vanguard*, *6*(1), 1–13.

Grant, W. (Ed.). (1931). *The Scottish National Dictionary: Volume 1*. The Scottish National Dictionary Association Ltd.

Grant, W., & Dixon, J. M. (1921). *A manual of modern Scots*. Cambridge University Press.

Gut, U. (2009a). *Introduction to English Phonetics and Phonology*. Peter Lang.

Gut, U. (2009b). *Non-native speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Peter Lang.

Harris, J. (1985). *Phonological variation and change: Studies in Hiberno-English*. Cambridge University Press.

Hay, J. (2011). Statistical analysis. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 198–214). Routledge.

Heffner, R.-M. S. (1937). Notes on the Length of Vowels. *American Speech*, *12*(2), 128.

Hewlett, N., Matthews, B., & Scobbie, J. M. (1999). Vowel duration in Scottish English speaking children. *Proceedings of 14th the International Congress OfPhonetic Sciences (San Francisco, CA)*(3), 2157–2160.

House, A. S., & 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.

Jadoul, Y., Thompson, B., & Boer, B. de (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15.

Johnston, P. (1997). Regional Variation. In C. Jones (Ed.), *The Edinburgh history of the Scots language* (pp. 378–513). Edinburgh University Press.

Johnston, P. (2007). Scottish English and Scots. In D. Britain (Ed.), *Language in the British Isles* (pp. 105–121). Cambridge University Press.

Jones, C. (2002). *The English Language in Scotland: An introduction to Scots*. Tuckwell Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. L. Bybee & P. J. Hopper (Eds.), *Typological studies in language: Vol. 45. Frequency and the emergence of linguistic structure* (pp. 229–254). John Benjamins Publishing Company.

Kamińska, T. E. (1995). *Problems in Scottish English phonology*. *Linguistische Arbeiten: Vol. 328*. Niemeyer.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, *40*(3), 390–402.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 115–132). Academic Press.

Kiełtyka, R. (2003). Aitken's Law: Some Aspects of Scottish Vowel Lengthening. *Studia Anglica Resoviensia*(2), 42–54.

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Klatt, D. H. (1973). Letter: Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, *54*(4), 1102–1104. https://doi.org/10.1121/1.1914322

Kohler, K. (1966). A late eighteenth century comparison of the "Provincial dialect of Scotland" and the "pure dialect". *Linguistics*, *4*(23).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1984). Field Methods of the Project on Linguistic Change and Variation. *Sociolinguistic Working Paper*, *81*, 1–43.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change ; a multimedia reference tool*. Mouton de Gruyter.

Labov, W., & Baranowksi, M. (2006). 50 msec. *Language, Variation and Change*, *18*, 223–240.

Lass, R. (1974). Linguistic Orthogenesis? Scots Vowel Quantity and the English Length Conspiracy. In J. M. Anderson (Ed.), *Historical linguistics: Proceedings of the First International Conference on Historical Linguistics, Edinburgh 2nd - 7th September 1973* (pp. 311–352). North-Holland Pubishing Company.

Laver, J. (1994). *Principles of phonetics. Cambridge textbooks in linguistics*. Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. MIT-Press.

Lehiste, I. (1972). The Timing of Utterances and Linguistic Boundaries. *The Journal of the Acoustical Society of America*, *51*(6B), 2018–2024.

Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.

Lindblom, B. (1967). Vowel duration and a model of lip-mandible coordination. *Dept. For Speech, Music and Hearing Quarterly Progress and Status Report*. https://www.speech.kth.se/prod/publications/files/qpsr/1967/1967_8_4_001-029.pdf [Date of Access: 12.12.2022]

Lindblom, B. (1982). Speech Production. Vowel Duration and a Model of Lip Mandible Coordination. *Journal of the Acoustical Society of America*, *72*, 2039.

Lindblom, B., & Rapp-Holmgren, K. (1971). Reexamining the compensatory adjustment of vowel duration in Swedish words. *Dept. For Speech, Music and Hearing Quarterly Progress and Status Report*, *12*(4), 19–25.

Lisker, L. (1974). On 'explaining'vowel duration variation. *Status Report on Speech Research: A Report on the Stacus and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications*. https://files.eric.ed.gov/fulltext/ED094445.pdf#page=223 [Date of access: 12.12.2022]

Llamas, C., Watt, D., French, P., & Roberts, L. (2011). Effects of the Scottish Vowel Length Rule on Vowel Quantity in Tyneside English. *17th International Congress of Phonetic Sciences*, 1282–1285.

Lodge, K. R. (1984). *Studies in the phonology of colloquial English*. Croom Helm.

Maack, A. (1949). Die spezifische Lautdauer deutscher Sonanten. *STUF - Language Typology and Universals*, *3*(1-6).

MacArthur, T. (1979). The Status of English in and furth of Scotland. In A. J. Aitken & T. MacArthur (Eds.), *Languages of Scotland* (pp. 50–67). W&R Chambers.

MacArthur, T. (1992). The Scots: bilingual or just confused? *World Englishes*, *11*(2/3), 101–110.

Maddieson, I. (1985). Phonetic cues to syllabification. In *Working Papers in Phonetics* (pp. 203–221).

Maguire, W. (2012). English and Scots in Scotland. In R. Hickey (Ed.), *Areal Features of the Anglophone World* (pp. 53–78). Mouton de Gruyter.

Mather, J. Y., & Speitel, H. H. (1986). *The linguistic atlas of Scotland*. Croom Helm.

Max Planck Institute for Psycholinguistics. (2022). *ELAN* (Version 6.0) [Computer software]. https://archive.mpi.nl/tla/elan [Date of access: 12.12.2022]

McAuliffe, M., Coles, A., Goodale, M., Mihuc, S., Wagner, M., Stuart-Smith, J., & Sonderegger, M. (2019). ISCAN: A System for Integrated Phonetic Analyses across Speech Corpora. *Proceedings of the 19th International Congress of Phonetic Sciences*, 1322–1326.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. *Proc. Interspeech 2017*, 498–502. https://www.isca-speech.org/archive/interspeech_2017/mcauliffe17_interspeech.html [Date of access: 12.12.2022]

McClure, J. D. (1977). Vowel duration in a Scottish accent. *Journal of the International Phonetic Association*(7(1)), 10–16.

McClure, J. D. (1994). English in Scotland. In R. Burchfield (Ed.), *The Cambridge History of the English Language: Volume V English in Britain and Oversees: Origins and Developments* (pp. 23–93). Cambridge University Press.

McKenna, G. E. (1988). *Vowel duration in the Standard English of Scotland*. Unpublished M.Litt. thesis.

McMahon, A. M. S. (1991). Lexical phonology and sound change: the case of the Scottish vowel length rule. *Journal of Linguistics*, *27*(1), 29–53.

Meer, P., Fuchs, R., Gerfer, A., Gut, U., & Li, Z. (2021). Rhotics in Standard Scottish English. *English World-Wide*, *42*(2), 121–144.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica*, *41*(4), 215–225.

Milroy, J. (1995). Investigating the Scottish vowel length rule in a Northumbrian dialect. *Newcastle and Durham Working Papers in Linguistics*(3).

Munhall, K., Fowler, C., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*(2), 225–239.

Murray, J. A. H. (1873). *The dialect of the southern counties of Scotland: its pronunciation, grammar, and historical relations ; with an appendix on the present limits of the Gaelic and lowland Scotch, and the dialectical divisions of the lowland tongue ; and a linguistical map of Scotland*.

Asher & Co. https://www.digitale-sammlungen.de/de/view/bsb11160026 [Date of access: 12.12.2022]

Nance, C. (2011). High Back Vowels in Scottish Gaelic, 17–21. https://www.researchgate.net/publication/228766394_High_Back_Vowels_in_Scottish_Gaelic [Date of access: 12.12.2022]

National Records of Scotland. (2011). *Table QS203SC - Country of birth*. https://www.scotlandscensus.gov.uk/search-the-census#/topics/list?topic=Ethnicity,%20Identity,%20Language%20and%20Religion&categoryId=4 [Date of access: 12.12.2022]

National Records of Scotland. (2015). *Scotland's Census 2011: Gaelic report: (part 1)*. https://www.scotlandscensus.gov.uk/media/cqoji4qx/report_part_1.pdf [Date of access: 12.12.2022]

Nieminen, T. (2019). *Praat-textgrids* (Version 1.3.1) [Computer software]. Legisign Software. https://pypi.org/project/praat-textgrids/ [Date of access: 12.12.2021]

Ó Baoill, C. (1997). The Scots-Gaelic interface. In C. Jones (Ed.), *The Edinburgh history of the Scots language* (pp. 551–568). Edinburgh University Press.

Ó Baoill, C. (2011). A History of Gaelic to 1800. In M. Watson & M. Macleod (Eds.), *The Edinburgh companion to the Gaelic language* (pp. 1–21). Edinburgh University Press.

Ohala, M., & Ohala, J. (1992). *Phonetic universals and Hindi segment duration* (Vol. 92). https://www.researchgate.net/publication/221486346_Phonetic_universals_and_Hindi_segment_duration [Date of access: 12.12.2022]

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America*, *54*(5), 1235–1247.

Parmenter, C. E., & Treviño, S. N. (1935). The Length of the Sounds of a Middle Westerner. *American Speech*, *10*(2), 129.

Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*.

Price, G. (1984). *The languages of Britain*. Arnold.

Priva, U. C. (2017). Informativity and the actuation of lenition. *Language*, *93*(3), 569–597. https://www.danielpovey.com/files/2011_asru_kaldi.pdf [Date of access: 12.12.2022]

Pukli, M. (2006). Scottish English and the Scottish Vowel Length Rule: An empirical study of Ayrshire speakers. *PhD Thesis*.

Python Software Foundation. (2021). *Python* (Version 3.9) [Computer software]. http://www.python.org [Date of access: 12.12.2021]

R Core Team. (2021). *R: A language and environment for statistical computing [*Computer software]. R foundation for statistical computing. Vienna, Austria. https://www.R-project.org/ [Date of access: 12.12.2021]

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders, 16*(3), 217–226.

Rathcke, T., & Stuart-Smith, J. H. (2016). On the Tail of the Scottish Vowel Length Rule in Glasgow. *Language and Speech, 59*(Pt 3), 404–430.

Roach, P. (2010). *English phonetics and phonology: A practical course* (4. ed., reprinted.). Cambridge University Press.

Schötz, S. (2007). Acoustic Analysis of Adult Speaker Age. In Christian Müller (Ed.), *Speaker Classification I: Fundamentals, Features and Methods* (pp. 88–107). Springer.

Schützler, O. (2015). *A Sociophonetic Approach to Scottish Standard English*. John Benjamins Publishing Company.

Schützler, O., Gut, U., & Fuchs, R. (2017). New perspectives on Scottish Standard English: Introducing the Scottish component of the International Corpus of English. In S. Hancil & J. C. Beal (Eds.), *Perspectives on Northern Englishes* (pp. 273–302). Mouton de Gruyter.

Schweinberger, M. (2022). *Fixed- and Mixed-Effects Regression Models in R.* The University of Queensland. https://ladal.edu.au/regression.html [Date of access: 12.12.2022]

Scobbie, J. M. (2005). Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: speech production evidence for fine-grained linguistic plasticity. *QMU Speech Science Research Centre Working Paper WP-2*. https://eresearch.qmu.ac.uk/140/ [Date of access: 12.12.2022]

Scobbie, J. M., Hewlett, N., & Turk, A. (1999). Standard English in Edinburgh and Glasgow: the Scottish Vowel Length Rule Revealed. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 230–245). Arnold; Oxford University Press.

Scobbie, J. M., & Stuart-Smith, J. H. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 87–114). Mouton de Gruyter.

Scobbie, J. M., Turk, A., & Hewlett, N. (1999). Morphemes, Phonetics And Lexical Items: The Case of the Scottish Vowel Length Rule. *Proceedings of 14th the International Congress of Phonetic Sciences (San Francisco, CA)*, 1617–1620.

Scotland's Census. (2011). *Languages: Scots*. https://www.scotlandscensus.gov.uk/census-results/at-a-glance/languages/ [Date of access: 12.12.2022]

The Scottish Government. (2015). *Scots Language Policy*. https://www.gov.scot/binaries/content/documents/govscot/publications/factsheet/2015/09/scot s-language-policy-english/documents/ffa3d92c-8ef8-4349-8dbb-70a1455cc33e/ffa3d92c-8ef8 [Date of access: 12.12.2022]-4349-8dbb-70a1455cc33e/govscot%3Adocument

Scottish Parliament. (2022). *Information about the Scottish Parliament copyright licence*. https://www.parliament.scot/about/copyright [Date of access: 12.12.2022]

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication, 23*, 193–229.

Smith, R., & Rathcke, T. (2016). Glasgow gloom or Leeds glue? Dialect-specific vowel duration contrains lexical segmentation and access. *Phonetica, 74*(1), 1–24. http://eprints.gla.ac.uk/116704/ [Date of access: 12.12.2022]

Solé, M.-J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron (Ed.), *Phonology and Phonetics /PP]: 4-4. Laboratory phonology 10* (pp. 607–655). Mouton de Gruyter.

Stuart-Smith, J. H. (2008). Scottish English: Phonology. In B. Kortmann & E. W. Schneider (Eds.), *A handbook of varieties of English: A multimedia reference tool; two volumes* (pp. 47–67). Mouton de Gruyter.

Stuart-Smith, J. H., Macdonald, R., & SPADE Consortium (2019, September 3). *The fate of the Scottish Vowel Length Rule in contemporary Scottish English*, Queen Mary University London/University College London.

Sweet, H. (1877). *A Handbook of Phonetics: including a popular exposition of the principles of spelling reform*. Clarendon Press.

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Tanner, J., Sonderegger, M., Stuart-Smith, J., & Data Consortium, S. (2019). Vowel duration and the voicing effect across English dialects. *Toronto Working Papers in Linguistics*, *41*(1).

Tanner, J., Sonderegger, M., Stuart-Smith, J., & Fruehwald, J. (2020). Toward "English" Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers. *Frontiers in Artificial Intelligence*, *3*, 1–15. https://doi.org/10.3389/frai.2020.00038

Tauberer, J., & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect: some evidence from dialects of north american English. In *INTERSPEECH*. http://www.evanini.com/papers/evanini_INTERSPEECH09a.pdf [Date of access: 12.12.2022]

Taylor, M. V. (1974). The Great Southern Scots Conspiracy: Pattern in the Development of Northern English. In J. M. Anderson (Ed.), *Historical linguistics: Proceedings of the First International Conference on Historical Linguistics, Edinburgh 2nd - 7th September 1973* (pp. 403–426). North-Holland Pubishing Company.

Thomas, E. R. (2011). *Sociophonetics: An introduction*. Palgrave Macmillan.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America, 88*(1), 97–100.

Tsao, Y.-C., & Weismer, G. (1997). Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language & Hearing Research*, *40*(4), 858-866.

Turk, A. E., Nakai, S., & Sugahara, M. (2012). Acoustic Segment Durations in Prosodic Research: A Practical Guide. In Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, & Johannes Schließer (Eds.), *Methods in Empirical Prosody Research* (pp. 1–27). de Gruyter.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*(4), 397–440.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*(2), 171–206.

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, *58*(2), 434–445.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quaterly Journal of Experimental Psychology, 67*(6), 1176–1190.

van Leyden, K. (2002). The relationship between vowel and consonant duration in Orkney and Shetland dialects. *Phonetica*, *59*(1), 1–19.

Warren, D. (2018). *The Scottish Vowel Length Rule in North East Scotland*. Unpublished PhD thesis. University of Aberdeen.

Watson, G. (1923). *The Roxburghshire Word-Book*. Cambridge University Press.

Watt, D., Fabricius, A., & Kendall, T. (2011). More on vowels: plotting and normalization. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 107–118). Routledge.

Watt, D., & Ingham, C. (2000). Durational evidence of the Scottish Vowel Length Rule in Berwick English. *Leeds Working Papers in Linguistics and Phonetics*(8), 205–228.

Watt, D., & Yurkova, J. (2007). Voice onset time and the Scottish Vowel Length Rule in Aberdeen English. *International Congress of Phonetic Sciences 2007*.

Weihs, C., & Weilinghoff, A. (2023). *Using PrInDT for linear mixed effects models in R [forthcoming]*.

Weilinghoff, A. (2019). *Aitken's Law Revised: A large-scale investigation into the Scottish Vowel Length Rule using the ICE Scotland Corpus*. Unpublished Master thesis.

Wells, J. C. (1982). *Accents of English 2: The British Isles.* Cambridge University Press.

Wettstein, P. (1942). *The Phonology of a Berwickshire Dialect*. Schüler s.a.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* [Computer software]. https://ggplot2.tidyverse.org/ [Date of access: 12.12.2022]

Windmann, A., Šimko, J., & Wagner, P. (2015, September 6). Polysyllabic shortening and word-final lengthening in English. In *Interspeech 2015* (pp. 36–40). ISCA.

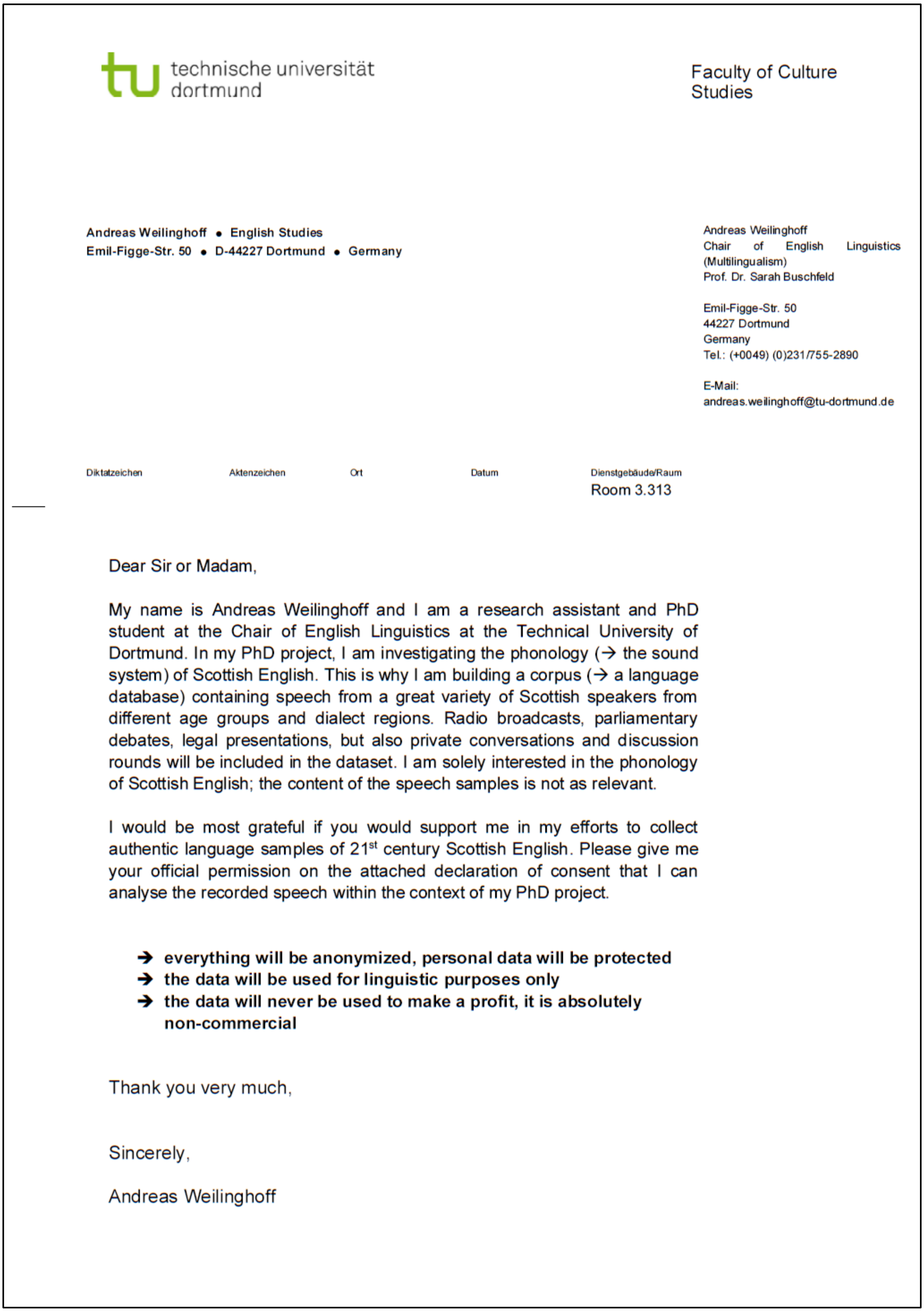
Winter, B. (2020). *Statistics for Linguists: An introduction using R*. Routledge.

Wölck, W. (1965). *Phonematische Analyse der Sprache von Buchan*. Carl Winter Universitätsverlag.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (1999). *The Hidden Markov Model Toolkit (HTK)* (Version 3.4.1.) [Computer software]. https://htk.eng.cam.ac.uk/ [Date of access: 12.12.2022]

Zai, R. (1942). *The Phonology of the Morebattle Dialect: (East Roxburghshire)*. Räber & Company.

9. Appendix



technische universität
dortmund

Faculty of Culture
Studies

Andreas Weilinghoff • English Studies
Emil-Figge-Str. 50 • D-44227 Dortmund • Germany

Andreas Weilinghoff
Chair of English Linguistics (Multilingualism)
Prof. Dr. Sarah Buschfeld

Emil-Figge-Str. 50
44227 Dortmund
Germany
Tel.: (+0049) (0)231/755-2890

E-Mail:
andreas.weilinghoff@tu-dortmund.de

Diktatzeichen	Aktenzeichen	Ort	Datum	Dienstgebäude/Raum
				Room 3.313

Dear Sir or Madam,

My name is Andreas Weilinghoff and I am a research assistant and PhD student at the Chair of English Linguistics at the Technical University of Dortmund. In my PhD project, I am investigating the phonology (→ the sound system) of Scottish English. This is why I am building a corpus (→ a language database) containing speech from a great variety of Scottish speakers from different age groups and dialect regions. Radio broadcasts, parliamentary debates, legal presentations, but also private conversations and discussion rounds will be included in the dataset. I am solely interested in the phonology of Scottish English; the content of the speech samples is not as relevant.

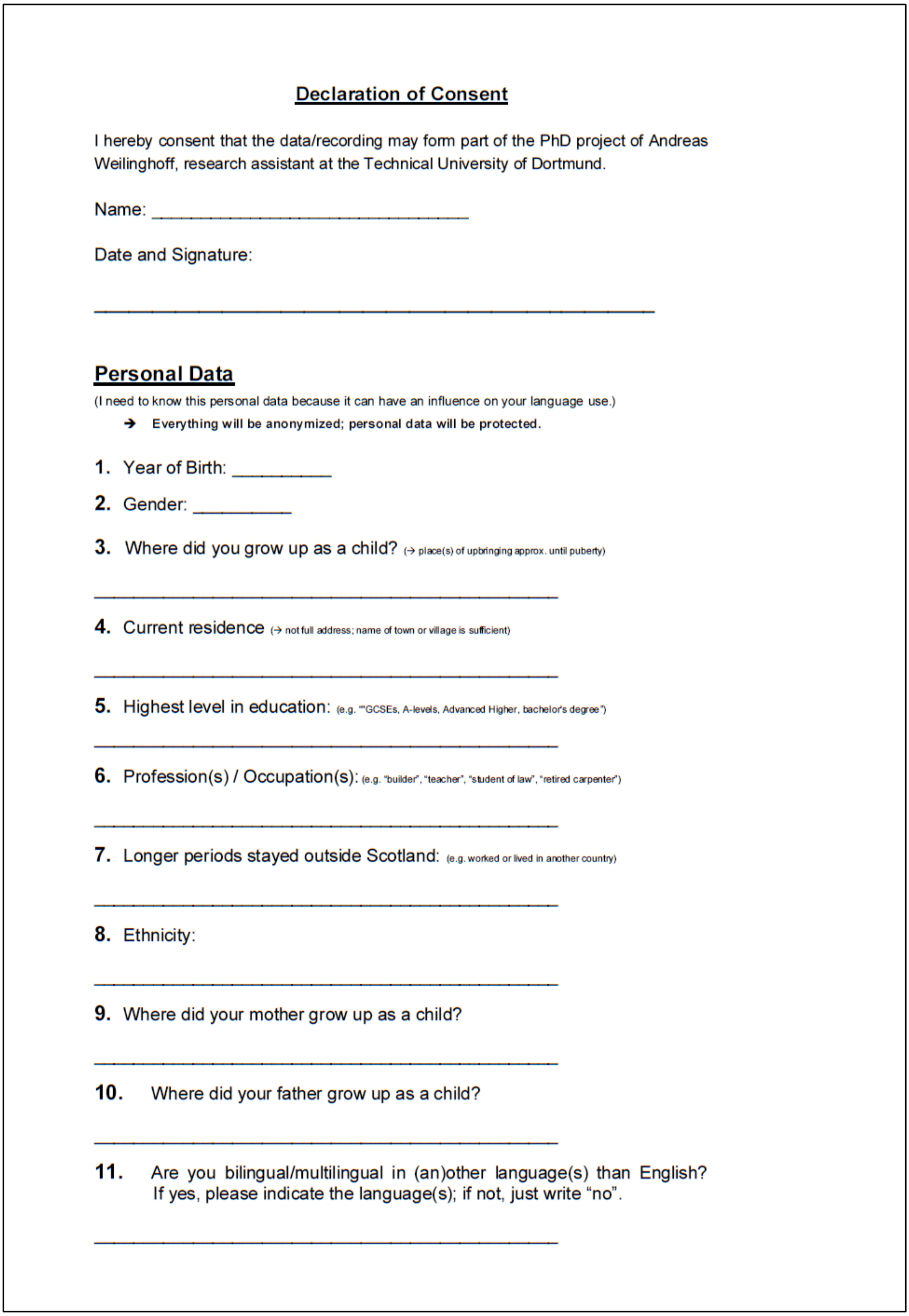
I would be most grateful if you would support me in my efforts to collect authentic language samples of 21st century Scottish English. Please give me your official permission on the attached declaration of consent that I can analyse the recorded speech within the context of my PhD project.

- ➔ **everything will be anonymized, personal data will be protected**
- ➔ **the data will be used for linguistic purposes only**
- ➔ **the data will never be used to make a profit, it is absolutely non-commercial**

Thank you very much,

Sincerely,

Andreas Weilinghoff



Declaration of Consent

I hereby consent that the data/recording may form part of the PhD project of Andreas Weilinghoff, research assistant at the Technical University of Dortmund.

Name: ______________________________

Date and Signature:

____

Personal Data

(I need to know this personal data because it can have an influence on your language use.)

➔ **Everything will be anonymized; personal data will be protected.**

1. Year of Birth: __________

2. Gender: __________

3. Where did you grow up as a child? (→ place(s) of upbringing approx. until puberty)

____

4. Current residence (→ not full address; name of town or village is sufficient)

____

5. Highest level in education: (e.g. "“GCSEs, A-levels, Advanced Higher, bachelor's degree”)

____

6. Profession(s) / Occupation(s): (e.g. “builder”, “teacher”, “student of law”, “retired carpenter”)

____

7. Longer periods stayed outside Scotland: (e.g. worked or lived in another country)

____

8. Ethnicity:

____

9. Where did your mother grow up as a child?

____

10. Where did your father grow up as a child?

____

11. Are you bilingual/multilingual in (an)other language(s) than English? If yes, please indicate the language(s); if not, just write “no”.

____