Univerzita Karlova,
Filozofická fakulta
Ústav anglického jazyka a didaktiky

Bakalářská práce

Nela Bradíková

Duration of stressed and unstressed syllables in British and American political debates

Trvání přízvučných a nepřízvučných vokálů v britských a amerických politických debatách

Praha, 2024

doc. Mgr. Radek Skarnitzl, Ph.D.

Acknowledgements

I would like to thank my supervisor doc. Skarnitzl for his engaging teaching in the first year of my BA that initiated my interest in phonetics. Additionally, I would like to thank him for his help with choosing a topic for my final thesis as well as all his immense support throughout the creation of this thesis.

Prohlášení autora

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 5.8.2024 ……..

Nela Bradíková

I have no objections to the BA thesis being borrowed and used for study purposes.

Souhlasím se zapůjčením bakalářské práce ke studijním účelům.

Abstract and Key Words

The aim of this thesis is to investigate segmental duration as an acoustic correlate of word stress in connected spontaneous speech in English. To lay the groundwork for this investigation, the theoretical part provides wide description of different aspects of the following three associated topics – speech rhythm, word stress and segmental duration. The practical study then aspires to reexamine the notion that stressed vowels are longer in duration than unstressed vowels which has been previously concluded from the analysis of controlled material and attest to whether the longer duration of stressed vowels in comparison to unstressed vowels is relevant for connected spontaneous speech. The analysed material is compiled from the recordings of connected speech of 8 British and 8 American speakers engaging in political debates. The analysis consists of the extraction of temporal information of stressed and unstressed vowels that is subsequently analysed in comparison between the two varieties of English and in terms of other aspects shown to affect segmental duration such as phonological vowel length or the nature of the following segment. The results of the study indicate that stressed vowels are longer in duration than unstressed vowels in connected speech of both examined varieties of English. The results further attest to inherent phonological length as a factor affecting segmental duration and suggest that the previously established notion of *pre-fortis shortening* proves to be more pertinent in monosyllabic words.

Key words: speech rhythm, words stress, acoustic correlates, vowel duration, connected speech

Abstrakt a klíčová slova

Cílem této práce je prozkoumat segmentální trvání a jeho korelaci se slovním přízvukem ve spontánní souvislé řeči v angličtině. Teoretická část je věnována osvětlení příslušné problematiky a poskytuje tak široký popis aspektů tří souvisejících témat – rytmu řeči, slovního přízvuku a segmentálního trvání. Praktická část usiluje o prověření předchozího závěru, že přízvučné vokály mají delší trvání než nepřízvučné vokály, který byl vyvozen zkoumáním čtené či jinak kontrolované řeči, a snaží se zjistit, zda je delší trvání přízvučných vokálů oproti trvání nepřízvučných vokálů platné i pro spontánní souvislou řeč. Analyzovaný materiál obsahuje nahrávky politických debat 8 britských a 8 amerických mluvčí. Analýza spočívá v extrakci temporálních informací o přízvučných a nepřízvučných vokálech a následném porovnání mezi oběma varietami angličtiny. Dále jsou v analýze temporálních informací zahrnuty další faktory, které prokazatelně ovlivňují trvání segmentů, jako je fonologická délka samohlásky nebo následující segment. Výsledky studie ukazují, že přízvučné vokály mají delší trvání než nepřízvučné vokály v souvislé řeči obou zkoumaných variet angličtiny. Dále výsledky potvrzují inherentní fonologickou délku jako faktor ovlivňující segmentální trvání a naznačují, že zkracování vokálů před fortisovými samohláskami je platné především v jednoslabičných slovech.

Klíčová slova: rytmus řeči, slovní přízvuk, akustické koreláty, vokalické trvání, souvislá řeč

Table of Contents

List of abbreviations

BrE	British English
AmE	American English
IQR	interquartile range
AVD	absolute vowel duration

List of Tables/Figures

1. Introduction

The main objective of this final thesis is to further examine vowel duration as a pertinent acoustic correlate of stress in British and American English. As duration has been identified as an acoustic correlate in previous research of controlled material and stressed vowels have been established as being longer in duration than unstressed vowels, the aim of this study is to attest to whether these two previously established notions can be applied to connected spontaneous speech in English. In a wider scope, this thesis is concerned with the notion of speech rhythm, the concept of English as a stress-based language as well as segmental duration and the many factors that have a hand in influencing it.

In order to create the foundation for the research part of this thesis, the theoretical background chapter will focus on a variety of aspects of the three abovementioned concepts – speech rhythm, word stress and segmental duration. The first section will present the development of research and discussion of speech rhythm in the past several decades with the focus on isochrony in production as well as in perception and the conceptualization of English as a stress-based language. Moreover, this section will discuss rhythmicality in speech while also presenting the reinterpretation of isochrony in speech in addition to alternative approaches to (ir)rhythmicality. As this thesis focuses on stressed and unstressed vowels, the next section will focus on the notion of stress through multiple viewpoints – production, perception, phonology and most importantly through acoustic parameters. In the discussion of acoustic parameters, the significance of duration as an acoustic correlate of stress will be established along with other acoustic characteristics. Lastly, the theoretical chapter will entail the description of factors other than stress that may influence segmental duration.

As has been suggested, the research part then examines the duration of stressed and unstressed vowels. The analysis was done on the recordings of connected speech of 8 British and 8 American speakers and entailed extracting data of the exact durations of vowels along with careful listening analysis and subsequent marking of vowels as stressed or unstressed. The temporal data will be presented in comparison between the two varieties as well as in terms of other aspects that affect segmental duration such as phonological vowel length or the nature of the following segment.

2. Theoretical background

2.1 Speech rhythm

When it comes to rhythm, it is apparent that time is crucial. Temporal organisation is essential for identifying patterns, defining structure, or finding (ir)regularity in different areas of human behaviour. Apart from speech, the two areas that are often discussed in connection to rhythm are music and dance as these two human activities usually come to mind first when talking about rhythm (Volín, 2010). The discourse on rhythmicality and rhythmic patterns of both music and dance seems to be mainly unanimous, and both activities could be, for the most part, described as isochronous, as opposed to speech, which only "may involve isochrony" (Ravignani & Madison, 2017). Nevertheless, the concept of isochrony seems to be at the start of the discussion of not only rhythm in general but also speech rhythm and therefore needs to be commented on.

2.1.2 Isochrony

In the simplest terms, isochrony can be defined as "a rhythmic pattern where all intervals have *equal duration*" (Ravignani & Madison, 2017, pp. 2). Since equal duration is improbable for many areas of research, this specific definition is often termed *idealised isochrony* and is rarely used. Therefore, when we discuss isochrony, we generally refer to *empirical isochrony* "where all intervals have ***roughly*** *equal* duration" (Ravignani & Madison, 2017, pp. 2).

In terms of speech, isochrony is a determining concept for one of the two general approaches to rhythm that Nolan & Jeon (2014) identify as *coordinative* or *periodic* rhythm. The authors further describe c*oordinative* rhythm, in opposition to *contrastive* rhythm (which will be discussed in Section 2.1.6), as a type of rhythm that "implies both repetition of a pattern and regularity of the interval taken by each repetition" and "therefore entails […] *isochrony*, meaning that a given repeated element or structural grouping of elements (e.g. syllable or foot) should always occupy the same time span" (pp. 2).

Which element or structure is isochronous and therefore being periodically repeated distinguishes between so-called *stress-timing* and *syllable-timing*. Although the absolute version of this dichotomy has been generally disputed and is now widely rejected due to the lack of acoustic evidence of exact isochronous patterns in the rhythm of speech, the influence of the distinction between these two rhythmic patterns on subsequent research has been so vast that it is only appropriate to comment on the concept and the following rejection of it.

2.1.3 Stress-timing and syllable-timing

Cauldwell (2002) notes that the most definitive and influential definition of *stress-timing* and *syllable-timing* has been that of Abercrombie (1967). Abercrombie's definition focuses on the rhythmic differences among languages and proposes that the majority of the world's languages fall into two mutually exclusive categories based on two types of isochrony:

> As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In … *syllable-timed* rhythm … the syllables recur at equal intervals of time – they are isochronous. In … *stress-timed* rhythm ... the stressed syllables are *isochronous* … there is considerable variation in syllable-length in a language spoken with a stress-timed rhythm whereas in a language spoken with a syllable-timed rhythm the syllables tend to be equal in length… (Abercrombie, 1967, pp. 97-98, as cited in Cauldwell, 2002).

The concept of isochronous patterns in English has previously been explored by some of Abercrombie's predecessors, notably Classe (1939) and Pike (1945). Although Abercrombie expands on their research, broadening his theory of isochronous timing across languages, his conclusions about isochrony are vastly different. Notably, Classe (1939) fails to find perfect isochrony. While he does not dismiss the concept and considers that isochrony might be a subjective phenomenon, his formulation of isochrony is "rather careful" as he defines it as "just a tendency to speak in rhythmic units that are perceived as isochronous" (Lehiste, 1977, pp. 253). Cauldwell (2002) also finds another point of comparison between the three researchers and notes that while Abercrombie does not distinguish between speech styles, Pike (1945) sees *stress-timing* as evident only in poetry and possibly in some prose, and Classe (1939) heavily focuses on distinguishing prose and verse as speech styles relevant for isochrony and speech as lacking the necessary conditions for isochronous chains.

Ultimately, the most relevant information for this thesis is the fact that in accordance with the aforementioned theories, English has been defined as a *stress-timed* language, a label that is assigned to the language to some extent to this day. Other languages considered as *stress-timed* are, for example, German or Arabic, whereas languages such as French, Spanish, or Italian have been classified as *syllable-timed* (Fletcher, 2010, pp. 552). A third categorisation, as mentioned by Low (2015), called *mora-timing,* has been proposed by researchers (pp. 126). Since only few languages are recognised as *mora-timed* languages and scarcely any of the relevant research for this thesis refers to this type of rhythmic pattern, *mora-timing* need not be discussed further.

In his discussion of Abercrombie's *stress-* and *syllable-timing* hypothesis, Cauldwell (2002) highlights the fact that it is not a single hypothesis but a collection of interdependent hypotheses, specifically: "(a) all languages fall into one of two mutually exclusive categories: stress-timed or syllable-timed; (b) in stress-timed languages, stresses occur at equal time-intervals (stress-isochrony); (c) in syllable-timed languages, syllables occur at equal time intervals (syllable-isochrony); (d) syllable-length varies in stress-timed languages, but not in syllable-timed languages; (e) inter-stress-intervals vary in length in syllable-timed languages, but not in stress-timed languages" (pp. 3). As has been mentioned, the strict differentiation between *stress-timed* and *syllable-timed* languages has now been generally rejected. Although the subsequent research in rebuttal of the *stress-* and *syllable-timing* hypothesis has inevitably included research on both *syllable-* and *stress-timed* languages, due to the interlinking of the six sub-hypotheses, only one of the two types of isochrony needs to be refuted in order to disprove the existence of either stress-isochrony or syllable isochrony. Therefore, the following section, in which the evidence against the hypothesis will be discussed, will primarily focus on research of *stress-timed* languages and more specifically on studies that examine English, as that is the focus of this thesis.

2.1.4 Evidence against contrastive rhythm and stress-isochrony

In the discussion of the evidence against "pure isochrony", several authors mainly focus on the work of scholars from the 1980s onwards, such as Roach (1982), Dauer (1983) or Couper-Kuhlen (1990) (as cited in Cauldwell, 2002; or Fletcher, 2010). Although these studies are indisputably important for the discussion of isochrony and also will be commented on, the rejection of isochrony (or at least doubt concerning the notion, as demonstrated by Classe, 1939) had become a matter of discussion even earlier.

As has been previously described, for a language to be identified as *stress-timed*, stresses in said language should occur in equal-time intervals. In other words, the duration of interstress intervals in *stress-timed* languages should be roughly equal, and the variation of said duration should be minimal. Thus, the primary measurement of most studies trying to examine *stress-timing* is unsurprisingly the temporal information on *interstress intervals*.

Low (2015), however, notes that when researchers measured the interstress intervals in *stress-timed* languages, namely in English (Shen & Peterson, 1962; Bolinger, 1965, as cited in Low, 2015), they failed to find evidence of roughly equal timing. While the studies of Shen & Peterson (1962) and Bolinger (1965), both cited in Lehiste (1977), examined continuous (although read) sentences, studies have also been done on "rhythmically controlled" material.

Lehiste (1977) notes on O'Connor's 1965 study, in which he recorded a limerick with a rhythm as strict as possible and measured the duration between the 15 occurred stress groups. The results showed a difference of 88 ms between the shortest and longest stress group, and O'Connor, therefore, concluded that "physical isochrony was clearly not present even under these very favorable conditions" (Lehiste, 1977, pp. 254).

The study of isochrony and *stress-timing* continued into successive decades. Both Cauldwell (2002) and Low (2002) note the respective studies of Roach (1982) and Dauer (1983), which furthermore address the issue of languages as strictly *stress-timed* or *syllable-timed* through the examination of interstress interval length.

Roach (1982) conducts his experimental work on recordings of spontaneous speech of six speakers – three speak one of the following *stress-timed* languages (English, Russian, Arabic), and the other three are speakers of one of the following *syllable-timed* languages (French, Telugu and Yoruba). The study aimed to confirm two of Abercrombie's claims – firstly, the claim that there is considerable variation in syllable length in *stress-timed* languages, whereas the syllable of a *syllable-timed* language tends to be equal in length; and secondly, the claim that stressed pulses are unevenly spaced in *syllable-timed* languages (Abercrombie, 1967, as cited in Roach, 1982). However, the results of Roach's study support neither of these two claims. Moreover, Roach's results show that the *stress-timed* group of languages displayed greater variability in the length of interstress variables. As Cauldwell (2002) notes, the fact that the results are opposite to the expectations, it could be suggested that the interstress intervals differentiate between the *stress-* and *syllable-timed* languages in the reverse direction than Abercrombie in his hypothesis. However, Roach (1982) regards the results simply as a basis for the rejection of Abercrombie's claims and concludes that all languages, as well as one speaker on different occasions, display both sorts of timing. Nevertheless, a language will have a type of timing which generally predominates.

Dauer (1983) also examined interstress interval length across multiple languages, namely English, Spanish, Italian and Greek. As characterised by Low (2015), the results of the study do not demonstrate that interstress intervals are more equal in *stressed-timed* languages (pp. 126), and therefore, the study arrives at the same conclusion as that of Roach (1982). The results do not display empirical evidence for the claim that rhythmic patterns can be found by measuring timing units found in speech. Regarding the difference between *stress-timed* and *syllable-timed* languages, Dauer (1983) arrives at the conclusion that the difference "has to do with differences in syllable structure, vowel reduction, and the phonetic realization of stress

and its influence on the linguistic system" (pp. 51) rather than stress- or syllable-isochrony. Moreover, due to the lack of evidence for isochronous temporal patterns, Dauer goes on to support the terms *stress-based* and *syllable-based* languages over the previously used terms *stress*- and *syllable-timed* languages.

2.1.5 Stress-isochrony and perception

So far, the discussed studies on stress- and syllable-isochrony have been chiefly concerned with the production of speech. The tendency to examine the production of isochrony is only fair since Abercrombie (1967) proposed that "every language is **spoken** with one kind of rhythm or the other" (pp. 97-98, as cited in Cauldwell, 2002). However, since the aforementioned studies failed to support the notion of isochrony in the production of speech, many scholars turned to a reinterpretation of isochrony into the area of perception.

One of the scholars focusing on the perception of isochrony is Couper-Kuhlen, who was mentioned at the beginning of the previous section. As described by Cauldwell (2002), Couper-Kuhlen's 1993 study, focused solely on examining English, is concerned with the identification of so-called isochronous chains through hearers' perception. The participants identified 48 isochronous chains, but some stretches of speech (36% of all syllables) did not form isochronous chains. According to the result of the study, Couper-Kuhlen does not completely support the notion of English as stress-timed. On the one hand, she states that "English speech is not uniformly isochronous over extended periods of time" (Couper-Kuhlen, 1993, pp. 48, as cited in Cauldwell, 2002), but on the other hand, she notes that just as significantly, the examined passage cannot be wholly concluded as unisochronous either. Therefore, as summarised by Cauldwell (2002), the study arrives at the conclusion that "English is not isochronous when viewed from the macroperspective of the entire temporal extent of a spoken text but from the microperspective of the internal characteristics of each of the 48 chains it is isochronous" (pp. 5).

While Couper-Kuhlen's (1993) study highlights the possibility of identifying isochrony at different levels of examination, Lehiste (1977) proposes the notion of isochrony as a perceived and imposed phenomenon. Principally, she suggests that "it is […] quite likely that the listener imposes a rhythmic structure on sequences of interstress intervals in spite of the fact that their durational differences are above the perceptual threshold" (Lehiste, 1977, pp. 258). Lehiste (1977) further notes that this imposing of structure seems to be a fairly general phenomenon, which is a position that is echoed even in recent research. In their overview of isochrony in human behaviour, Ravignani & Madison (2017) reveal that "when confronted with

isochronous sequences where intervals have been slightly jittered, humans will tend to regularize the intervals and perceive the whole sequence as isochronous" (pp. 2). The authors further report that "the roots of the human propensity for isochrony are clearly found in our biology" (pp. 3). Considering the remarks presented in these two studies, it seems appropriate to come to the conclusion that although speech cannot be objectively labelled as isochronous, humans seem to be inclined to perceive speech as rhythmical.

2.1.6 Contrastive rhythm and functional irrythmicality of speech

While Nolan & Jeon (2014) note on the fact that "disobliging acoustic data could be circumvented by attributing isochrony to perception" (pp.2), their discussion and subsequent conclusion does not subscribe to such relegation of isochrony. As has been mentioned in Section 2.1.2, the authors also introduce the concept of *contrastive* rhythm as an alternative view to the previously discussed *coordinative* rhythm and the adjacent notion of isochrony. The authors define *coordinate* rhythm as "a view which sees rhythm in the alternation of stronger and weaker elements" (pp. 2). As the authors further specify, it is important to note that while these stronger and weaker elements may be defined by their difference in duration, the view is otherwise non-temporal, as the contrastive elements are not synchronised to an external clock. Due to the defining opposition of stronger and weaker elements, the authors note that, for a language such as English, it is only natural to map this definition of rhythm onto the alternation of stressed (strong) and unstressed (weak) syllables. Since it has been generally agreed upon that speech does not appear to have *coordinative* rhythm, the authors finally conclude that out of the two conceptualizations of rhythm, *contrastive* rhythm has more potential in the description and modelling of speech.

However, Nolan & Jeon (2014) ultimately reach an alternative conclusion in their approach to speech rhythm. In regard to the alternation of stronger and weaker elements in English, the authors argue that upholding the concept of alternation requires the dismissal of the actual phonetic properties of the sequences of syllables that occur. Furthermore, in a number of languages, which do not display such strong cues of stronger elements such as English, it would be especially difficult to identify the alternation of prominence. The authors therefore conclude that speech seems to display neither *coordinative*, nor *contrastive* rhythm. This ultimately leads Nolan & Jeon (2014) to the proposal of an alternative hypothesis in which they suggest that speech is not inherently rhythmical.

Cauldwell (2002) reaches a somewhat similar conclusion regarding speech rhythm. After the rejection of speech being isochronous (both in terms of production and perception),

the author reaches the position that spontaneous speech is irrythmic due to the fact that "it occurs in a series of short bursts, […] most of which (close to 90%) are too short to trigger perceptions of rhythmicality" (pp. 16). However, the author defines two ways in which rhythms may occur and that is either as *coincidental* or *elected* (pp. 16). The main determinant of rhythmicality are the decisions of speakers regarding the adopted wording and its division into units in speech. While *coincidental* rhythmicality can be defined only as a side effect, *elected* rhythmicality requires conscious attention.

Nevertheless, spontaneous speech is otherwise characterised by irrythmicality and the perception of speech (as discussed in the previous section) by rhythmicality. Cauldwell (2002) proposes that the balance of irrythmic production and rhythmic perception provides "a necessary tension for effective communication" (pp. 19). The author hypothesises that if the rhythms of speech were not ever-changing, the speakers may not hold the attention of the hearers because they would be distracted by the pattern of an established rhythm and therefore distracted from the attention of meaning. Conclusively, due to the non-occurrence of continued rhythm, spontaneous speech can be argued as functionally irrythmical.

2.1.8 Speech rhythm in English

The previous sections included a manifold of views regarding the discourse on speech rhythm. Hence it is only appropriate to try to define the rhythm of speech in English in terms what has been discussed. As has been mentioned, the clear-cut division of languages into the two categories of *stress-timed* and *syllable-timed* languages has now been generally rejected. If we adopt Roach's (1982) approach that one of the two types of timing will predominate in a language, it is evident that in English *stress-timing* is the predominant of the two. Consequently, English remains to be discussed as a *stress-timed* or *stress-based* in many different contexts and it is therefore necessary to examine what exactly the notion of stress entails.

2.2 Word stress in English

As has been established in the previous section, stress is an important factor in the rhythm of speech in English. Although there seems to be a lack of consensus on the nature of speech rhythm itself, it is clear that an important element in the rhythm of speech in English is the alternation of stressed and unstressed syllables. Stress (or lack thereof in unstressed syllables) can be studied and defined through four fundamental viewpoints: production, perception, acoustics, and phonology, and the following chapter will focus on framing and discussing stress from each perspective.

To begin our discussion of stress, it seems helpful to position ourselves in terms of on what level of phonetic description will the discussion be situated. The characterisation of speech rhythm was positioned entirely on the level of prosody; therefore, the discussed characteristics are applicable to sentences or even larger units of speech. As we are moving on to stress, although the discussion remains on the suprasegmental level, the focus shifts from larger to smaller units of speech as when we discuss stress, the discussion revolves around syllables and whether they are stressed or lack stress. However, in the discourse on acoustic correlates (Section 2.2.2), both the research and description presented in this section focuses for the most part on vowels and therefore descends to the segmental level of speech.

2.2.1 Production and perception of stress

The first viewpoint from which stress can be characterised is production. Roach (2009) notes that it is generally assumed that the production of stress, i.e., stressed syllables, depends on more muscular energy being used by the speaker than the speaker uses in the production of unstressed syllables. Using more muscular energy in context means that "when we produce stresses syllables, the muscles that we use to expel air from the lungs are more active, producing higher subglottal pressure" (pp. 73). Furthermore, it is probable that other parts of the vocal apparatus are likewise more active (Roach, 2009).

As Ashby and Maidment (2005) note, it is important to remember that stress is a relational feature. In other words, the characteristics (of any kind – perceptual, acoustic etc.) of a stressed syllable are only relevant and telling when they can be compared to the characteristic of an unstressed syllable. For a syllable to be perceptually determined as stressed, a combination of factors need to be involved. This collection of factors is usually discussed as a joint concept of *prominence* defined by length, loudness and pitch (Ashby & Maidment, 2005). A stressed syllable is therefore perceived as "longer and louder than its [unstressed] neighbours and may be marked by some pitch movement or new level of pitch" (pp. 156). Each of these three factors

subsequently have a corresponding measurable acoustic characteristic which will be discussed in the following part of this chapter. Roach (2009) also suggests the change in the quality of the vowel as another factor that can add to a syllable being perceived as more prominent. This can be illustrated by the fact that the most frequently occurring vowels in weak (therefore unstressed) syllables are /ə, ɪ, i, u/, and therefore the prominence of stressed syllables is enhanced as they occur in a "sea" of these weak vowels.

2.2.2 Acoustic correlates of stress

As has been previously indicated, the third viewpoint, from which the discussion regarding stress can be conducted, is the acoustic description of stress. As Gorgon & Roettger (2017) note, stress and its acoustic correlates have been a recurring subject for phonetic research for some time. In their cross-linguistic survey of studies, the authors assemble research concerning stress and its acoustic parameters in 75 languages and deduce that the area of investigated acoustic dimensions "can be coarsely broken down into four categories: duration, fundamental frequency, intensity, and spectral characteristics" (Gordon & Roettger, 2017, pp. 4). These four respective categories can be applied to the discussion of acoustic correlates of stress in English and will now be commented on with detail.

2.2.2.1 Duration

As mentioned in the discussion of perception of stress, a key factor in determining whether a syllable is stressed or unstressed is its perceived length (not to be confused with phonological length). Moreover, stressed syllables are generally perceived to be longer. In other words, we can discuss length as the perceptual concept (as defined in the previous section) and duration as its corresponding acoustic measurement.

Duration has been already measured and identified as the acoustic correlate of stress in one of the earliest studies on acoustic parameters of stress done be Fry (1955). In this study, Fry (1955) selected "a group of English words in which a change of function from noun to verb is commonly associated with a shift of stress from the first to the second syllable" (Fry, 1955, pp. 765), for example words such as *object*, *permit* etc. Although the research is in part focused on perception, the study also measured vowel durations. Due to the material being pairs of two syllable words with shifting stress, duration was mostly analysed in terms of duration ratios between the first vowel and the second vowel of a specific word, and the author ultimately concluded duration to be an effective cue that shows major differences with a shift of stress.

However, vowel duration and its correlation to stress has also been the subject of more recent research. In their study of stress in Southern British English, Eriksson & Heldner (2015) identify duration as one of the functions of stress. In comparison to Fry (1955), their researched material is not only more extensive in the amount of the material but also includes different speaking styles. The material consists of the recordings of 31 speakers and includes recordings of semi-spontaneous speech and readings of a word list and a list of phrases. Their results indicate significant difference between the duration of unstressed and stressed syllables with the duration of stressed syllables having longer mean durations that unstressed syllables. The study also notes on the differences between primary and secondary stress which will be further discussed in Section 2.2.3. In conclusion, the study suggests that vowel duration is a reliable correlate of stress.

Moreover, many studies, such as Huggins (1972, 1975), Klatt (1975, 1976), Umeda (1975, 1976) and Crystal & House (1982, 1988a, 1988b, 1990) (as cited in Fletcher, 2010), have researched the effect of not only stress but also other factors such as vowel quantity or word and phrase position on segmental duration. Generally, it is important to not regard stress as the only determining factor in vowel duration but consider other contexts which will be further discussed in Section 2.3.

Finally, it may be informative to specify that the research on vowel duration and stress in English examines almost exclusively vowel duration. However, in several languages, such as Estonian or Welsh, stress is successfully distinguished only in consonant and not vowel duration (Gordon & Roetger, 2017).

2.2.2.2 Fundamental frequency

The second acoustic parameter that correlates to stress is fundamental frequency. Fundamental frequency (f_0) can be defined as "the rate of the vibration of the vocal folds" and moreover "the rate at which the speech pressure waveform repeats" (Ashby & Maidment, 2005, pp. 154). Ultimately, it is connected with the perceived pitch of the speech signal – the higher the f_0 of the sound, the higher we perceive the sound.

The aforementioned study conducted by Eriksson & Heldner (2015) examined fundamental frequency in two different ways. Firstly, the authors measured fundamental frequency level and their results of the suggest that stressed vowels have significantly higher f_0-level values than unstressed vowels. Secondly, the study examined fundamental frequency variation which the authors define as the "Standard Deviation of f_0-level" (Eriksson & Heldner,

2015, pp. 42). However, the results show that the effect of stress and stress level on fundamental frequency variation is minimal. In conclusion, out of the two parameters, only fundamental frequency level can be considered a successful correlate of word stress.

2.2.2.3 Intensity

As has been previously mentioned, stressed syllables can be perceived as louder than unstressed syllables and the corresponding acoustic parameter to this detection is intensity. Alongside duration, intensity was another parameter examined in the aforementioned early study of Fry (1955). Similarly to duration, intensity was also analysed in terms of ratios as the overall intensity varied between speakers but also between the two vowels in the specific words. The results of the author's analysis of disyllabic words which undergo stress shift indicate that intensity is also a "cue for the judgement of stress", however the "duration ration is a more effective cue than intensity ratio" (Fry, 1955, pp. 768).

Since Fry's first study on intensity and stress, intensity measurements have however diversified and started to encompass other factors. As Gordon & Roettger (2017) note, some studies have included "non-frequency-dependent measures of intensity (e.g. mean, peak, midpoint)" (pp. 5) while others may target measures of frequency-sensitive intensity. Moreover, some studies examining stress in American English, namely Lieberman (1960) and Beckman (1986) (as cited in Gordon & Roettger, 2017) have studied intensity through the *intensity integral*, which, albeit not frequency sensitive, incorporates another parameter – duration. Therefore, it is important to note what specific measures studies employ, especially if comparing research.

2.2.2.4 Spectral characteristics

The last category of acoustic parameters identified by Gordon & Roetger (2017) is the category of spectral characteristics, most often examined through the first two formant measurements. Some languages may include formant measurement as a corelate to a larger extent and may observe changes in vowel quality as important for the study of stress; however, for English, the discussion of vowel quality and formants is only partially relevant in terms of stress-induced vowel reduction. It is generally known that, in English, unstressed vowels typically undergo reduction which also involves a reduction in quality to schwa. This process will be further commented on in Section 2.2.2.5.

However, Eriksson & Heldner (2015) discuss in their research the measurement *spectral emphasis* which can be generally described as "an acoustic feature reflecting the relative

intensity in the higher frequency bands" (Heldner, 2001, pp. 1). It is important to note that while it is a spectral measurement, *spectral emphasis* largely correlates to intensity discussed in the previous section. For their study on stress, the authors define *spectral emphasis* more specifically as "as the difference in dB between the Sound Pressure Level (SPL) of the full spectrum and the SPL of f_0 in each segment" (Eriksson & Heldner, 2015, pp. 42). The results show a significant difference between stressed and unstressed vowels that is most evident in semi-spontaneous speech (as opposed to reading lists). The authors subsequently suggest that *spectral emphasis* plays a significant role in English. However, the authors also note that other studies have mostly been done on American English and there may be a possibility that varieties differ in this respect.

Moreover, Eriksson & Heldner also note on the fact that the three discussed acoustic parameters have been previously ranked 1) *f0-level,* 2) *duration,* 3) *intensity* in terms of significance in the earlier studies. In contrast, the authors suggest that "if we go by the degree of explained variance [= statistical method accounting dispersion of the dataset], our ranking is quite different – Spectral Emphasis (17.5%), Duration (14.2%) and f0-level (7.3%)" ((Eriksson & Heldner, 2015, pp. 44). Nevertheless, all three parameters can evidently be considered acoustic correlates of stress, although to a variably debatable extent.

2.2.3 Levels of stress

As has been indicated in the previous section, while the two-level differentiation between stressed and unstressed syllables may be sufficient in certain contexts, it is ultimately possible to differentiate at least one intermediate level of stress in English. As Laver (1994) notes, based on the "the graded difference of prominence that characterizes individual syllables" (pp. 516) in certain tri- and longer polysyllabic words, such as *resignation* or *systematic*, we can distinguish three levels of stress: primary stress, secondary stress and unstressed. To give an example, Roach (2009) illustrates the difference between primary and secondary stress on the word *anthropology* /ˌænθrəˈpɒlədʒi/ in which the first syllable carries secondary stress (marked by the low mark) and the third syllable carries primary stress (marked by the upper mark) (pp. 87). Unstressed syllables are then all other syllables in the words, i.e. second, third and fourth. As has been discussed in the Section 2.2.1, one of the factors that contributes to syllables being perceived as more prominent is pitch and therefore intonation. Moreover, intonation (or the peak in intonation, tonic accent) seem to be a determining factor for primary and secondary stress. As Ladefoged & Johnson (2015) note "in longer words containing two stresses, the apparent difference in the levels of the first and the second stress is really due to

the superimposition of an intonation pattern" (pp. 121). Therefore, the placement of primary stress on the third syllable in the previously mentioned examples is largely due to the fact that the syllable can be understood as the tonic syllable in which the movement of intonation reaches a peak.

The notion of more than two levels of stress in English is also supported by acoustic evidence. Eriksson & Heldner (2017) have shown that secondary stress is evidenced by multiple acoustic characteristics. Notably, duration has shown "a significant stepwise increase from unstressed to secondary stressed to primary stressed" (pp. 44) – this stepwise increase can be demonstrated by the authors' results of the mean durations of vowels that were respectively 53, 66, and 79 ms for unstressed, secondary stressed and primary stressed vowels.

Moreover, spectral emphasis has also proven to differentiate between all three levels, although "the difference between primary and secondary stressed is smaller than that between unstressed and stressed" (pp. 44). As has been discussed, the authors find that f_0-level indicates significant difference between stressed (primary and secondary compiled) and unstressed vowels; however, the difference between primary and secondary stressed vowels is significant only for male speakers, whereas it does not prove to be significant for female speakers. The authors however conclude that this fact can be explained "by the fact that the difference between unstressed and stressed (i.e. primary and secondary stressed pooled) is larger for the male speakers than for the female speakers" (pp. 42).

However, the authors also note on other research, Yuan et al. (2008, as cited in Eriksson & Heldner, 2015), whose results indicated f_0 as a significant difference between primary and secondary stresses vowels, whereas duration showed no difference. Nevertheless, it can be concluded that research shows, although quite inconclusively, that the perceptual difference in prominence between primary and secondary stress does correspond to differences in acoustic characteristics.

2.2.4 Phonological stress and placement of stress

Stress has been so far defined by the differences in or exaggeration of phonetic parameters through which a syllable can be assigned a level of prominence. This conceptualisation of stress can be otherwise specified as phonetic stress. In other words, *phonetic stress* can be defined as "a gradient phenomenon, and the phonetic realization of any syllable can be said to show a greater or less degree of stress relative to the manifestation of some other syllable" (Laver, 1994, pp. 511). However, as Laver (1994) further notes, another

approach to stress is regarding it as a phonological property of the syllable and understanding the distinction of stressed and unstressed syllables as different levels of *phonological stress*. The phonological lens then views the placement of stress on a particular syllable as a defining property of that word, which is a notion that can also be called *lexical stress*.

Lexical stress is a defining concept for the placement of stress in English. In the simplest terms, English is "sensitive to the lexical class of a word, that is, whether it is a noun, a verb, an adjective and so on" (Ashby & Maidment, 2005). This fact can be illustrated by the so-called word class pairs that fall into two different word classes based on stress placement. According to Roach (2009), these words generally consist of prefix + stem and follow the following rule:

> If a pair of prefix-plus-stem words exist, both members of which are spelt identically, one of which is a verb and the other of which is either a noun or an adjective, then the stress is placed on the second syllable of the verb but on the first syllable of the noun or adjective (pp. 87).

Therefore, we can observe pairs such as object as a noun with a stressed first syllable /ˈɒbdʒekt/ and object as a verb with a stressed second syllable /əbˈdʒekt/. Another example would be perfect as an adjective /ˈpɜːfekt/ and as a verb /pəˈfekt/ (Roach, 2009, pp. 87).

While other languages can be clearly labelled as having either *fixed stress* on specific syllables or *variable stress* where its placement cannot be predicted, the placement of stress in English cannot be defined in such simple terms. As has been mentioned, one of the ways of defining English is as a *lexical stress language*. Therefore, stress is placed depending on the word class but also general composition of the word and therefore English stress is 'fixed' to an extent (in a different sense than *fixed stress languages* – stress is fixed on a specific syllable according to each word, not a specific syllable within a word, i.e. first, penultimate, last etc.). However, it is hard to make general conclusions. For instance, Roach (2009) presents a set of rules regarding the placement of stress based on the morphological composition of words. Nevertheless, he concludes that seeing stress patterns as unchanging would be wrong and notes that stress position may vary either because of an adjacent word or simply because speakers do not agree on the placement of stress in certain words (such as controversy, which can be either pronounced /ˈkɒntrəvɜːsi/ or /kənˈtrɒvəsi/). Moreover, it is also important to note that the placement of stress may vary depending on the variety of English that is being discussed.

2.2.5 Stress in connected speech

In the previous sections, stress has been considered as the contrast between stressed and unstressed syllables within individual words. Normal speech however does not happen by speaking isolated words. In connected speech, speech segments as well as suprasegmentals are subsequently subject to certain connected speech processes that change them and as a result are often not fully realised.

How a word occurs and is pronounced in isolation is also called *citation form*. In terms of stress, citation form can be further characterised as a form when "at least one syllable is fully stressed and there is no reduction of the vowel quality" (Ladefoged & Johnson, 2015, pp. 115). Moreover, the realisation of a word by a citation form can be considered as a phonetically full form of the word. In connected speech however, many changes may take place and one of those is the reduction of unstressed vowels. Unstressed vowels are, as has been already extensively discussed, are shorter in duration that stressed vowels which is a fact that also contributes to the reduction. However, the reduction may also extend into the actual phonetic realisation the word and for many words we can discuss the difference between the so-called *strong form* and *weak form* of the specific word. As Roach (2009) notes, most of these words that have both of strong and weak forms belong to a category of function words which do not have a full dictionary meaning. In connected speech, the vowels in these words are subsequently often reduced to a schwa – examples include words such as *for* (/fɔː/ - strong form, /fə/ - weak form); *than* (/ðæn/ - strong, /ðən/ - weak). However, it is important to note that weak forms do not differ from strong forms only through vowel reduction, but some words may also undergo elision of consonants, such as *her* (/hɜːr/ - strong, /hər/ and /ər/ - weak) (Roach, 2009).

In summary, it can be noted that the pronunciation of words in connected speech may be reduced. The reduction is generally tied to the weakening of stress and therefore the vowels in unstressed syllables may be often reduced to a schwa or a vowel close to schwa. Additionally, the stress in connected speech depends not only on lexical stress but also on intonation movement in a sentence (Ladefoged & Johnson, 2015). In other words, we can conclude that the difference between stressed and unstressed vowels can be defined by the fact that "if [the vowel] is stressed, it can be at the center of an intonational pitch change so that it receives a tonic accent, which might be said to raise it to a more primary level of stress. If [the vowel] is unstressed, it can have a full vowel or a reduced vowel" (Ladefoged & Johnson, 2015, pp. 260).

2.3 Segmental duration

To conclude the theoretical background section and fully set the ground for the practical part of this thesis, the last topic than needs to be discussed is what other factors apart from stress influence the segmental duration. Although numerous studies have reported on vowel duration of both vowels and consonants, the following discussion will understandably, due to the subject of this thesis, focus primarily on the factors affecting vowels duration.

One of the most notable summaries of the influential factors on segmental duration is the research of Klatt (1976). In this study, the author identifies not only the main phonetic factors that influence segmental duration but also notes on the non-phonetic factors that may influence the durational structure of sentences such as physiological or extralinguistic factors. Klatt's summary will serve as a guiding hand for the structure of this subchapter while also being supplemented by other research.

To begin this part of the discussion, it seems informative to briefly mention some extralinguistic factors or factors related to the higher levels of linguistics. Of these extralinguistic factors, it can be noted that factors such as speaker mood or physical conditions may affect the overall speaking rate and therefore also segmental duration. As shown by Williams & Stevens (1972, as cited in Klatt, 1976), actors have demonstrated a very slow speaking rate when emoting anger and a somewhat slower speaking rate when expressing fear and sorrow. Moreover, changes in speaking rate may complexly influence the durational patterns of sentences and slowing speech down induces extra duration of pauses while increasing speaking rate is accompanied by phonetic and phonological simplification along with the durational shortening of vowels (Klatt, 1976). Another level of factors may be discourse-level factors, although Klatt (1976) notes on the lack of quantitative data. Nevertheless, studies such as Lehiste (1975b, as cited in Klatt, 1976) note on the fact that during reading, the last sentence of a paragraph tends to be longer in duration that other sentences.

Studies have however examined many factors in the syntactic level as well as word and phonetic/phonological levels. These factors are often termed positional factors and, as the term suggests, consider the position of the measured phonemes within a word or within a sentence. The first positional condition that may be considered is whether the observed vowel occurs in a monosyllabic or polysyllabic words. In studying vowel durations of AmE vowels, Umeda (1975) notes on the differences between the mean values of the durations of individual vowels in a monosyllabic condition and a polysyllabic condition. While her results do not suggest an overall tendency among all vowels, it can be noted that each vowel, except for the diphthong

/aʊ/, showed a difference in its mean vowel duration between the two conditions. For instance, the mean duration of /e/ was greater in polysyllabic positions than in monosyllabic conditions while most of the mean durations of other vowels demonstrated the opposite. Moreover, other factors, such as the prepausal or pre-fortis position or phonological length, have been shown to have extensive influence on segmental duration and will now be discussed in greater detail.

2.3.1 Prepausal lengthening

The first phenomenon that will be discussed further and that has been observed in regard to variation of segmental duration is *prepausal lengthening*, especially at clause and phrase boundaries. The general presumption for this kind of lengthening is that "fluent pauses may occur within sentences, especially between words that are not syntactically related" (Klatt, 1971, as cited in Klatt, 1976, pp. 1211). However, research also suggests that prepausal lengthening also occurs at phrase and clause boundaries when there is no physical pause present in the acoustic signal. (Klatt, 1976).

Phrase-level effects on segmental duration have been evidenced for instance by Martin (1970, as cited in Klatt, 1976) whose research showed that segments in spontaneous just before major grammatical constituent boundaries tend to be lengthened. Further evidence has been presented by Klatt (1975, as cited in Klatt, 1976) who observed lengthening at the ends of noun phrases, including the boundary between a noun phrase and a verb phrase, and at the end of conjoined or embedded clauses, and found that the increase in vowel duration was 30% (as averaged over all phrase boundaries in the corpus).

Moreover, it is also possible to observe lengthening at word boundaries and research suggests that word-final syllables seem to be somewhat longer in duration, even in non-phrase-final positions (Oller, 1973, Klatt, 1975, as cited in Klatt 1976). Another study that examined *prepausal lengthening* is that of Crystal & House (1988b) who observed the effect "in which vowels preceding syntactic pauses – phrase markers, sentence markers, etc. – are longer than in other location" (pp. 1577). The authors further note on their results in measuring word-final vowels that the lengthening factors are strongly affected by the presence or absence of a final consonant. As the authors modify the resulting average duration according to the following factors, they observe "a 10%-20% decrease if there is a final consonant, a 40% increase if there is a final consonant and a following pause, and a 90% increase if there is a following pause without a final consonant" (pp. 1577).

In other words, the duration of word-final vowels is evidenced to be lengthened when the vowel is not followed by a consonant and followed by a pause; or when the vowel is followed by a consonant and the word is followed by a pause. However, when a vowel is in a word-final syllable but followed by a consonant and not by a pause, the vowel is generally shorter that in other instances. A similar effect has also been observed by Umeda (1975) in unstressed vowels in word-final syllables. In conclusion, as the results of the discussed studies have shown, *prepausal lengthening* proved to be an influential factor on segmental duration.

2.3.2 Effect of postvocalic consonants on vowel duration

As noted by Klatt (1976), the duration of vowels before voiceless consonants tends to be shorter than before voiced consonants. A marked durational difference was observed for instance by House & Fairbanks (1953, as cited in Klatt, 1976) whose results showed that the difference between durational values between vowels before voiceless consonants and voiced consonants in a phrase-final position is about 50-100 ms. Klatt also subsequently notes that further research suggests that the difference in non-phrase-final syllables in not as large, as it seems to be only 10-20 ms. From this viewpoint that sees the voiceless consonant as the dominant factor in the process, the effect is often called "pre-fortis clipping" (e.g. Roach, 2009).

However, if we view the voiced consonant as the dominant factor in this preconsonantal phenomenon, the term *lengthening-before-voicing* effect is also used (Crystal & House, 1988a). The research of Crystal & House (1988a) has further emphasized the importance of considering multiple factors simultaneously. Their results suggest that short vowels before voiced and voiceless obstruents seem to be generally of equal length. However, when their position is specified as prepausal, short vowels before voiced obstruents are longer that before voiceless obstruents. Nevertheless, the effect of the following consonant of vowel duration can be concluded to be significant.

2.3.3 Inherent phonological duration

Another factor that needs to be considered and when discussing segmental duration and the last one to be discussed is the fact that "each phonetic segment has its own intrinsic or inherent phonological duration" (Klatt, 1976, pp. 1213). For example, it has been observed that the duration of voiceless fricatives is generally about 40 ms longer than that of corresponding voiced fricatives (Klatt, 1976).

However, the difference in inherent phonological duration is chiefly apparent in the duration of vowels, namely between short and long vowels, and diphthongs. In their study on

vocalic durations in American English, Crystal & House (1988c) examine the duration of individual vowels in recorded readings of three fast and three slow talkers in which they group the examined vowels into the following categories:

> (a) short: [ɪ], [e], [ʌ], [ʊ]; (b) long: [i], [e], [æ], [ɑ], [ɔ], [o], [u]; (c) diphthongs: [aɪ], [aʊ], [ɔɪ]; (d) schwa: [ə]; (e) *r*-colored: [ɝ]; (f) other: [ɨ], [ʉ] (pp. 265).

Their general analysis of vowels (with unspecified context, no other factors taken into consideration) suggest that "vowel categories display so-called inherent durational characteristics" (pp. 267) – mean duration of long vowels was 130 ms, of short vowels 72 ms, and of diphthongs 176 ms. The authors also considered the durations of these categories with regard to stress and concluded that while "stressed vowels are longer than unstressed vowels, [the] vowel categories maintain their relative durations independent of stress" (pp. 267).

Since the presented study is a study of American English, it is important to note that other varieties in English could display different durational values. Nevertheless, it is evident that there is a durational difference between vowels in terms of their inherent phonological duration.

3. Material and Method

As has been indicated in the introduction of this thesis, the aim of this study was to investigate vowel duration in connected spontaneous speech. The material used for this analysis was provided by my supervisor but originally had been obtained from publicly accessible political debates and consists of the recordings of 16 speakers in total: 8 British English speakers and 8 American English speakers, each variety is represented by 4 male speakers and 4 female speakers. The material consists of connected spontaneous speech of approximately 200 words per speaker, each recording of one speaker is roughly 60 - 100 seconds.

The recordings of British English speakers were selected from the BBC programme Westminster Hour (www.bbc.co.uk/programmes/b006s624) and the American English material comes from the archives of the C-SPAN network (www.c-span.org). Neither BrE nor AmE speakers display any significant variation in terms of their accents – the BrE speakers are speakers of the South British English variety, while the accent of AmE speakers can be categorised as General American.

Word stress has been shown to be a significant factor affecting vowel duration (as discussed in Section 2.2.4). However, previous studies that have examined the correlation of word stress and vowel duration (e.g. Eriksson & Heldner, 2015) have only analysed controlled or semi-spontaneous speech. This analysis therefore aims to examine whether what we already know about stress and vowel duration also applies to spontaneous connected speech.

Firstly, the base hypothesis of this analysis is that the duration of stressed vowels is longer than the duration of unstressed vowels. However, the presence or absence of stress may not be the only factor important for vowel duration and therefore other segmental and prosodic factors need to be considered. A wide variety of factors both on the segmental and prosodic level could possibly be examined; however in line with the scope of a thesis of this character, this analysis will focus mainly on segmental factors and therefore examine how factors such as the nature of the following segment, phonological vowel length or the context of number of syllables affect vowel duration. Lastly, this thesis will analyse these factors while comparing the data between the British English and American English varieties.

The recordings were already automatically segmented and annotated to individual Praat textgrids from previous analyses. Firstly, this automatic segmentation needed to be manually corrected, therefore any errors were removed and more importantly all word boundaries and the boundaries of vowels were accurately adjusted.

Secondly, all vowels were marked as stressed/unstressed through listening analysis and manual labelling. Due to the nature of the material being recordings of spontaneous speech (not recited wordlists or otherwise prepared/read text), the analysis could not be done solely based on predicted lexical stress. In other words, stress was not determined based on canonical forms of words, but by listening to actual realizations, which included entire deaccented words, as well as stress placed on less likely words. Determining whether a vowel is stressed or unstressed proved to be quite challenging and unclear at times, any ambiguity or indistinctness was therefore always discussed with my thesis supervisor. An example of the resulting textgrid can be seen in Figure 1 where the lowest tier contains the said word, the individual phonemes are indicated in the middle tier and the vowels are marked as stressed or unstressed in the top tier.

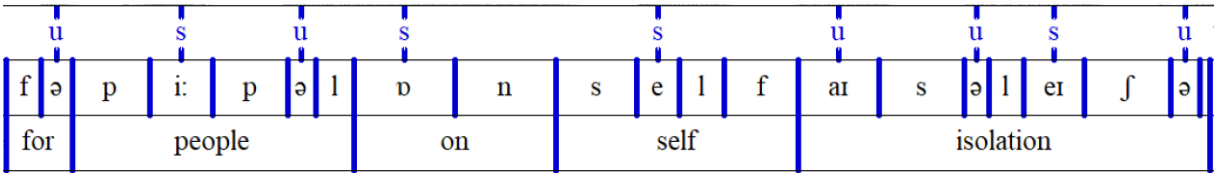

Fig. 1 example of an analysed segment of speech in Praat

Vowel duration was then extracted from the material with a Praat script along with corresponding information of each vowel token such as stress, the quality of the vowel, the word containing the vowel or number of syllables in the word, as well as phonological length of the vowel, language variety and speaker id.

Lastly, in order to be able to analyse the possible effect of segmental context, the following segment for each vowel token needed to be determined. Each vowel token was therefore assigned information on whether the following segment is a sonorant, a fortis consonant or a lenis consonant, or whether the vowel is part of an open syllable and therefore does not have a following segment.

In total, the analysis calculates the vowel duration of 4927 different vowel instances. Around 50% of those vowel tokens are vowels in monosyllabic words. The rest of the material then contains 632 disyllabic word tokens, 260 trisyllabic word tokens, 74 quadrisyllabic word tokens and only 16 pentasyllabic word tokens. The summarised number of vowel instances and word tokens can be found in Table 1.

	number of syllables	*word tokens*	*vowel instances*
monosyllabic words	1	2507	
polysyllabic words	2	632	1264
	3	260	780
	4	74	296
	5	16	80

Table 1 extent of the analysed material

4. Results

4.1 Relative vowel duration

In accordance with common practice and along with the fact that the analysis looks at the vowel duration in speech of multiple speakers, the data first needed to be standardised in order to eliminate any isochronous variation. There are multiple possibilities for the scaling of data when it comes to the rhythm of speech, one of which is for instance the commonly used normalization with respect to speech rate. Ultimately a simpler method of ratio of vowel duration to word duration was chosen as it is sufficient and appropriate for the scope of this thesis. The assumed need for standardization can be demonstrated by the following example in Figure 2 where even visually the fact that the durations of the two unstressed vowels are noticeably shorter than the duration of the stressed vowel can be identified. In absolute numbers, the durations of the vowels are the following: while the duration of the first schwa equals to 37.8 ms and that of the second schwa equals to 38.2 ms (which are both unstressed), the duration of the stressed ʌ is 87 ms. Although the difference between the durations of stressed and unstressed vowels is presumed, it was deemed appropriate to standardise the difference and put it into relative numbers (through the previously mentioned ratio of vowel duration to word duration). The two schwa durations then make up 14.9% (first schwa) and 15.1% (second schwa) of the word duration, while the duration of the stressed ʌ comes to 34.3 % of the complete duration of the word.

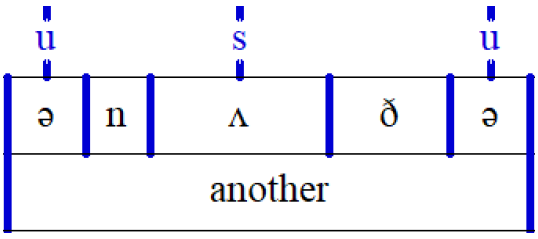

Fig. 2 Praat textgrid of a specific realization of the ‘another’

As can be seen in Figure 3, the resulting relative vowel duration seems to contradict the hypothesis of stressed vowels having longer duration than unstressed vowels. The results indicate that while the median of the relative vowel duration of stressed and unstressed vowels is around the same value, the interquartile range (IQR) of the relative vowel duration of unstressed vowels is considerably greater that of stressed vowels. Similar result can be seen when contrasting the relative vowel duration between the two varieties in Figure 4 with the slight difference between the two median values of relative vowel duration of stressed and unstressed vowels in American English.

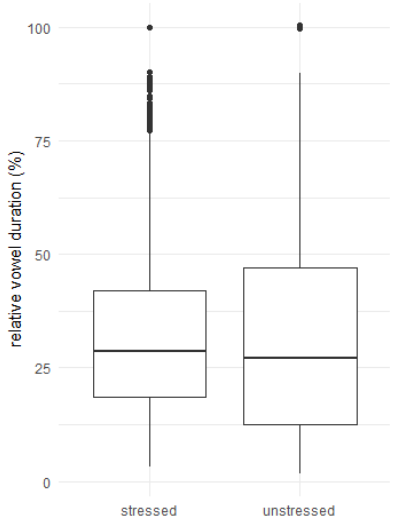

Fig. 3 relative vowel duration

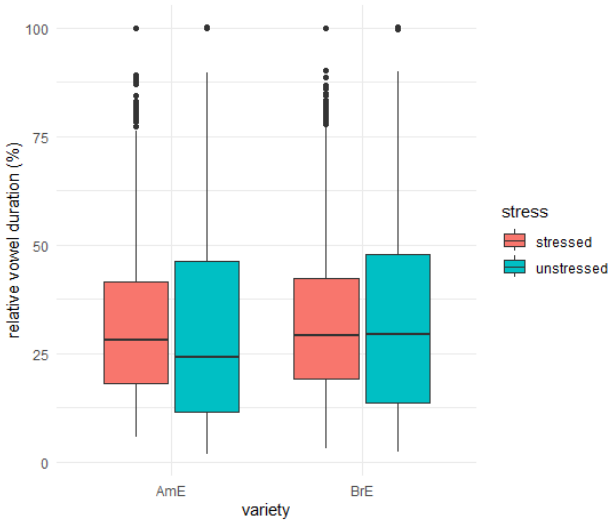

Fig. 4 relative vowel duration in comparison between AmE and BrE

Since the results shown in these two figures indicate notably different results than expected, it is appropriate to investigate what factors could have affected the overall relative vowel duration. As has been discussed in Section 2.3, one of the factors that may influence vowel length is the number of syllables in a word. This condition seemed as an appropriate starting point, therefore the relative durations were separately analysed for monosyllabic and polysyllabic words which proved to be critical for the analysis of overall relative vowel duration. If monosyllabic words are excluded and only the relative vowel duration of polysyllabic words is calculated, the resulting values are completely different. As the results in Figure 5 demonstrate, the relative vowel duration in polysyllabic words is generally greater than that of unstressed vowels in both varieties without any marked difference between the two varieties. As the IQRs of both varieties demonstrate, the relative duration of stressed vowels is generally

approximately 1.5 times longer than the relative duration of unstressed vowels with the median of relative stressed vowel duration estimated around 20% and the relative unstressed vowel duration around 12% of the word duration. The considerable difference between the durations of stressed and unstressed syllables in further demonstrated by the median values of relative vowel duration across both varieties. The relative vowel duration of both stressed and unstressed vowels shows very distant maximum values and numerous outliers that will be discussed later in the context of absolute vowels duration and other factors.

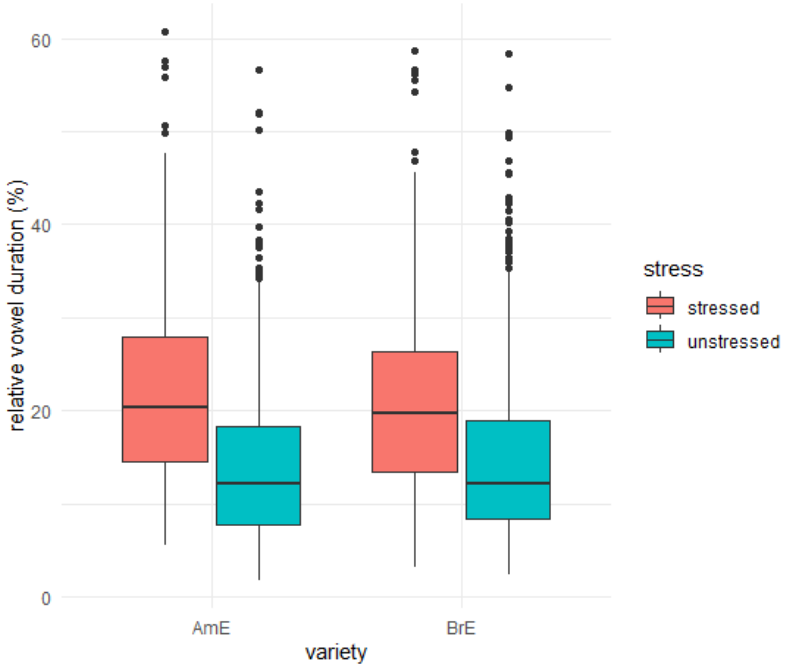

Fig. 5 relative vowel duration of polysyllabic words

On the other hand, if we look at Figure 6 that shows the relative vowel duration of only monosyllabic words, the results are again contradictory as in Figure 3 and 4. The explanation lies in the overall sound structure of the monosyllabic words. All of the monosyllabic words of course contain only one vowel, however most of them will contain more than one consonant. Many of the monosyllabic words do contain only one consonant such as *in*, *to*, *the*, and some have two consonants, for example *have*, *that*, *but*; however, many of the words also contain consonant clusters such as the words *groups*, *trace* or *press*. Most unstressed monosyllabic words will be grammatical words (*as*, *the*, *have* etc.) that are generally phototactically simpler than lexical words. On the other hand, the majority of stressed monosyllabic words will be lexical words that often contain consonant clusters. This then leads to the fact that when the relative vowel duration of stressed vowels is calculated, the ratio of vowel duration to the duration of the entire word is smaller than the ratio estimated by the relative vowel duration of

unstressed vowels simply because the stressed words contain more phonemes than unstressed words. The maximum outliers can also be explained by their phonemic structure, since the relative vowel duration value equal to 100% is the relative vowel duration of words consisting of only one vowel such as *a* or *I* and in British English *are* or *were* (due to the non-rhotic nature of the variety).

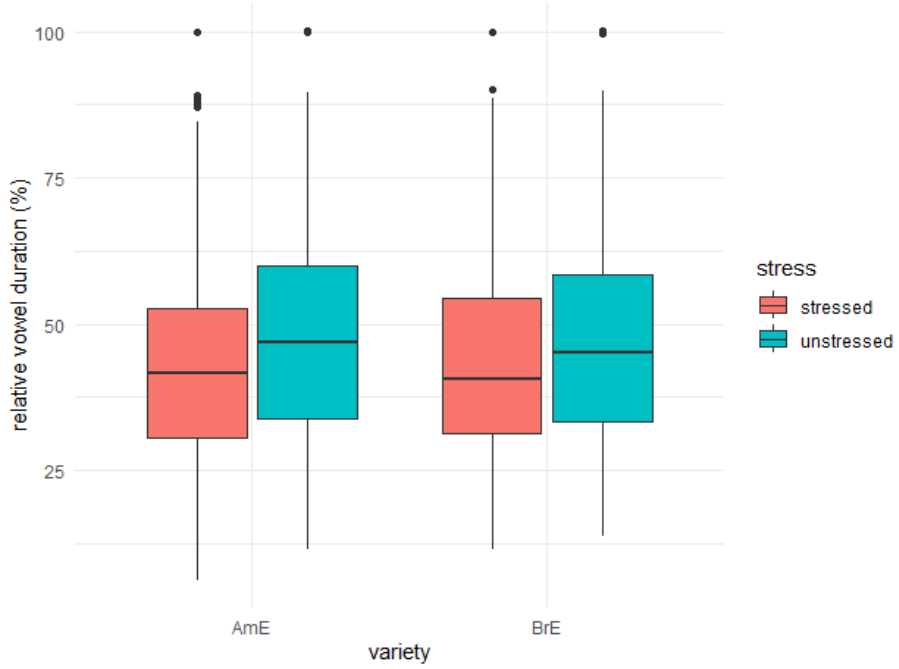

Fig. 6 relative vowel duration of monosyllabic words

As has been mentioned in the description of the material, monosyllabic words make up 2507 of the total 4927 measured vowels, making up about 50% of all vowel tokens. Consequently, when the overall relative vowel duration is calculated (as pictured in Figures 3 and 4), the results are skewed towards those of monosyllabic words. Relative vowel duration is therefore indicative only when differentiating between mono- and polysyllabic words and for that reason the rest of the analysis works only with absolute vowel duration.

4.2 Absolute vowel duration

As has been explained in the previous section, an analysis of absolute vowel duration was subsequently executed. From now on when vowel duration is discussed, it will exclusively refer to absolute vowel duration in ms, unless explicitly stated otherwise. For future brevity, absolute vowel duration will also be abbreviated as AVD.

4.2.1 Effect of language variety

Firstly, the analysis examining absolute vowel duration focuses on the comparison of British and American English. As Figure 7 shows, when looking at the overall absolute duration of stressed and unstressed vowels, there is no considerable difference between British and American English. In both varieties, the vowel duration of stressed vowels is longer than that of unstressed vowels and the interquartile ranges of vowel duration of both stressed and unstressed vowels fall approximately around the same values without any considerable differences. As the figure indicates, the IQR of duration of stressed vowels is approximately 70 – 140 ms, while half of the unstressed vowels have a duration of 40 – 80 ms.

Where the two varieties show difference of some note is in their respective maximum outliers. While both varieties display vowel duration more than two times that of the respective medians, the maximum outliers are more extreme in British English in stressed vowels as well as in unstressed vowels. The maximum vowel duration of a stressed vowel in AmE is less than 350 ms while the maximum vowel duration of a stressed vowel in BrE is reaching almost 450 ms. The difference between maximum outlier of unstressed vowel duration is smaller but the BrE maximum also reaches a higher value of just over 350 ms. Interestingly, the maximum vowel duration of a BrE unstressed vowel is longer than the maximum vowel duration of an AmE stressed vowel. These significantly protruding values are most likely affected by other segmental or prosodic context and will be discussed individually later on in the analysis.

Nevertheless, the median values decisively indicate that the duration of stressed vowels is substantially longer than that of unstressed vowels across both varieties. The median values of stressed vowels in both varieties are around 100 ms, whereas the median values of unstressed vowels are slightly above 50 ms, indicating that generally, the duration of stressed vowels is almost 2 times the duration of unstressed vowels.

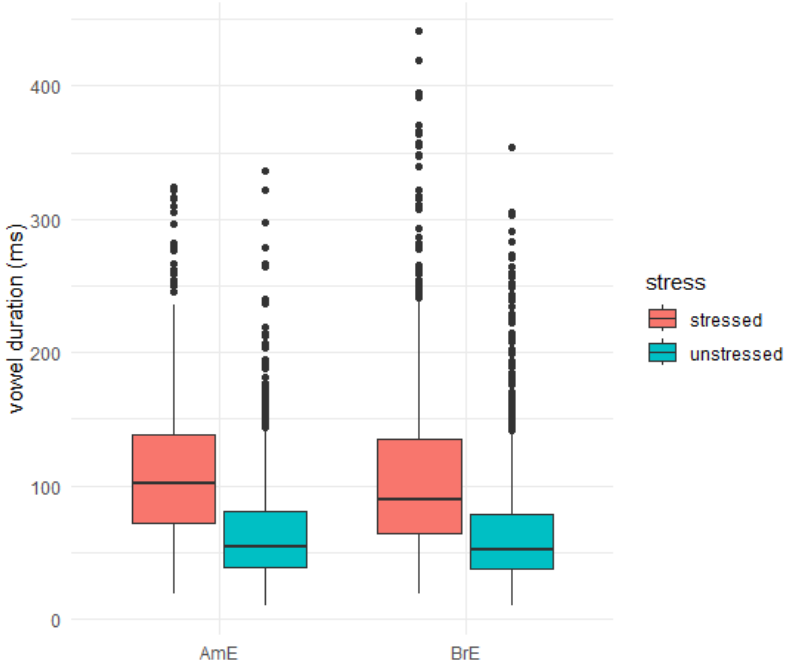

Fig. 7 AVD in AmE in comparison to BrE

4.2.2 Variety between individual speakers

Another comparison that can be made is the comparison between the vowel duration in the speech of individual speakers. As can be seen in Figure 8, there is no prominent difference between the speakers. The duration of stressed vowels is longer than the duration of unstressed vowels consistently across the speech of all speakers. Similarly, the median values of both stressed and unstressed vowels as well as most interquartile ranges are generally in accordance with the overall values discussed in the previous paragraph. However, some smaller differences can be identified in the overall vowel duration ranges of individual speakers. For example, the British F2 speaker seems to have the greatest range of vowel duration (excluding outliers) while the American M4 speaker appears to have the smallest range of vowel duration. Figure 8 also indicates which speakers have demonstrated the maximum vowel duration. The two greatest values of the duration of stressed vowels can be assigned to the same speaker, British speaker F3. The same speaker also displays the maximum outlier for the duration of unstressed vowels.

Therefore, it can be concluded that this particular dataset does not include any notable difference between individual speakers in general and the resulting vowels duration values can be regarded as objective and not results of idiosyncratic variation. However, the previously observed small differences between speakers can partially suggest resolutions on “personal” speaking rates of the individual speakers. For example, if we compare the median value of the duration of stressed syllables between the British F1 and F2 speakers, it can be concluded the

BrE F1 speakers has a generally faster speaking rate than BrE F2 speaker. Similar comparisons could be further made between other speakers with regard to not only the median values but also ICQ ranges of both stressed and unstressed vowels durations across both varieties.

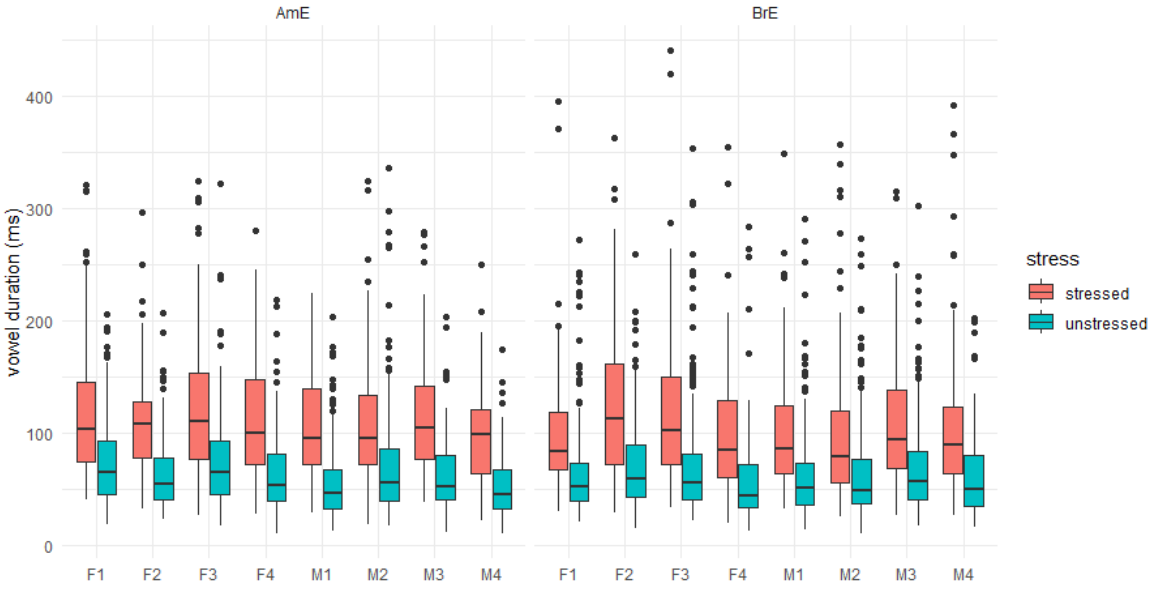

Fig. 8 AVD in the speech of individual speakers

4.2.3 Effect of the number of syllables

Since the analysis of monosyllabic and polysyllabic words has proven to be informative when discussing relative vowel duration, it is only appropriate to analyse whether it is an important factor for absolute vowel duration as well.

In Figure 9, it is evident that the overall difference between monosyllabic and polysyllabic words is also significant for absolute vowel duration as it has been to relative vowel duration. The vowel duration of stressed vowels in monosyllabic words is indicated to be greater than in polysyllabic words in both varieties. The median value of the duration of stressed vowels slightly decreases with each added syllable up until quadrisyllabic words. In pentasyllabic words, the median value of the duration of stressed vowels remains around the same value as quadrisyllabic words in American English, whereas in British English, the median value of vowels in pentasyllabic words is actually greater than that of vowels in quadrisyllabic words. However, the stagnation/increase probably does not demonstrate any general tendencies as the material includes only a small number of both quadrisyllabic and pentasyllabic words.

The maximum duration values and a large number of outliers can be assigned to stressed vowels in monosyllabic words which can similarly as with relative vowels duration be

explained by the fact that monosyllabic words make up about 50% of the material and the inconsistencies are presumably a result of other segmental or prosodical factors that will be discussed in the later part of this analysis.

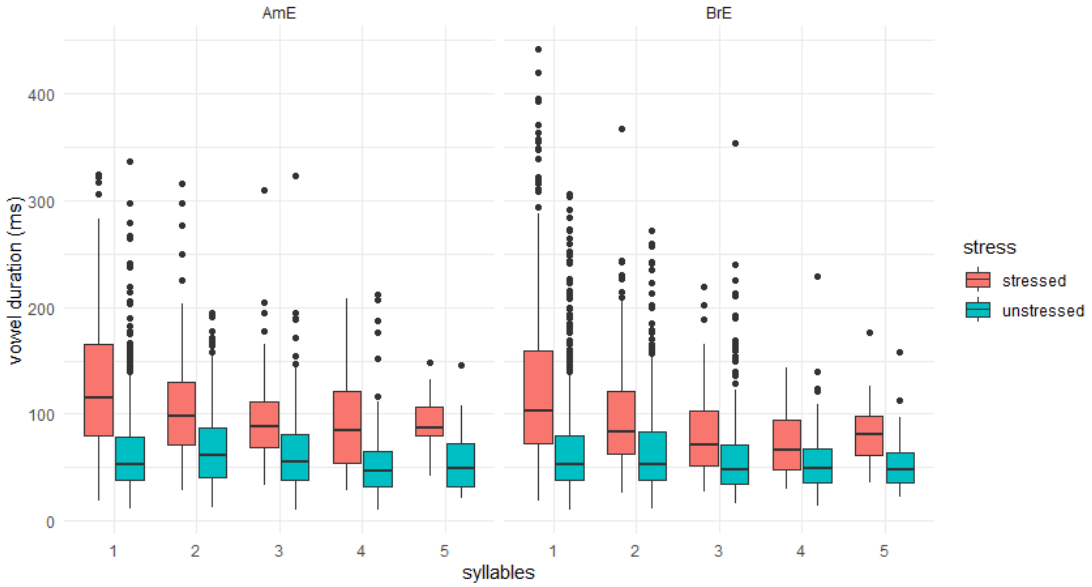

Fig. 9 AVD in terms of number of syllables and language variety

In terms of overall duration of unstressed vowels, there seems to not be much variation between monosyllabic and polysyllabic words. It can be noted that similarly to vowels in stressed monosyllabic words, vowels in unstressed monosyllabic words also display a large number of outliers. The disproportion in number of outliers can be explained again explained by the constitution of the material.

4.2.4 Effect of phonological vowel length

As has been discussed in Section 2.3.3, inherent phonological length is a significant factor that affects segmental duration and therefore was also considered as a factor in this analysis. Ultimately, the vowels were categorised simply into two categories and the categorization has been applied as follows: vowels / ɪ, e, æ, ɒ, ʌ, ʊ, ə/ were labelled as short and vowels /iː, ɜː, ɑː, ɔː, uː/ as well as all diphthongs /eɪ, ɔɪ, aɪ, eə, ɪə, ʊə, əʊ, and aʊ/ were labelled as long.

As can be seen in Figure 10, there is a marked difference in vowel duration between both long and short stressed vowels as well as between long and short unstressed vowels. There is also a slight difference between the two varieties of English as the influence of vowel length on vowel duration is greater in British English.

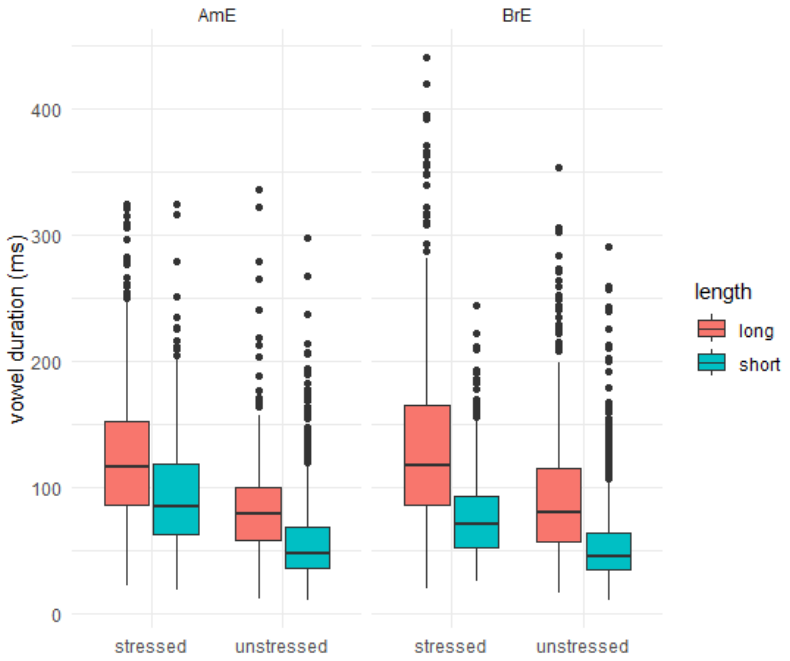

Fig. 10 AVD in terms of vowel length and language variety

Figure 10 shows a considerable difference between the durations of long and short stressed vowels in British English where the full range of duration (minimum to maximum) of stressed long vowels is essentially double the range of unstressed vowels. The duration of stressed long vowels is indicated to be longer than that of stressed short vowels mainly by the median values which are around 120 ms for the duration of stressed long vowels and around 70 ms for the duration of stressed short vowels. The difference is further evident in the IQR values since the duration values of long stressed vowels are situated between approx. 90 – 160 ms and in comparison, the duration values of short stressed vowels fall between approx. 50 – 95 ms. The difference in duration between long and short stressed vowels is also evident in American English, only it is slightly smaller in contrast which is less evident in the IQR values. However, the median values of the durations of stressed long vowels are similar between the two varieties (approx. 120 ms) while the median of the duration of stressed short vowels in AmE (approx. 90 ms) is slightly bigger than in BrE.

Unstressed long and short vowels also show difference in their duration in both varieties. Similarly to stressed vowels, the median values of the duration of unstressed vowels are similar between the two varieties – about 80 ms for the duration of long vowels and approximately 50 ms for short vowels. As the graph suggests, the difference is again between the IQR values of the two varieties in which British English, specifically in long vowels, shows greater range. The IQR of the duration of short unstressed vowels is comparable between BrE and AmE and the

values fall approx. between 40 – 60 ms. However, the IQR of the duration of long unstressed vowels in BrE lies approximately between 60 – 110 ms but for AmE, the top value of the IQR is approximately 20 ms less that in BrE.

In analysing the relation of vowel duration and vowel length, the results already partially illustrate the nature of some of the maximum outliers. As Figure 10 suggests, the maximum vowel duration and the first few following highest values correspond to the values of stressed long vowels, indicating that the vowel length proves to be an influential factor for the more protruding values as well.

4.2.5 The influence of the following segment

The last factor that was considered in this analysis is the influence of postvocalic consonants on absolute vowel duration. Studies have previously shown (as discussed in Section 2.3.1) that the duration of vowels followed by a voiceless consonant often tends to be shorter than the duration of vowels followed by a voiced consonant – a phenomenon sometimes called *lengthening-before-voicing* (Crystal & House, 1982) but more commonly *pre-fortis clipping* (e.g. Roach, 2009).

Figure 11 shows the differences between the duration of vowels followed by a fortis consonant, a lenis consonant or a sonorant as well as the duration of open syllables (i.e. no segment followed the vowel, the vowel is either at the end of the word or the word consists of only the vowel). The results are again contrasted between the two language varieties.

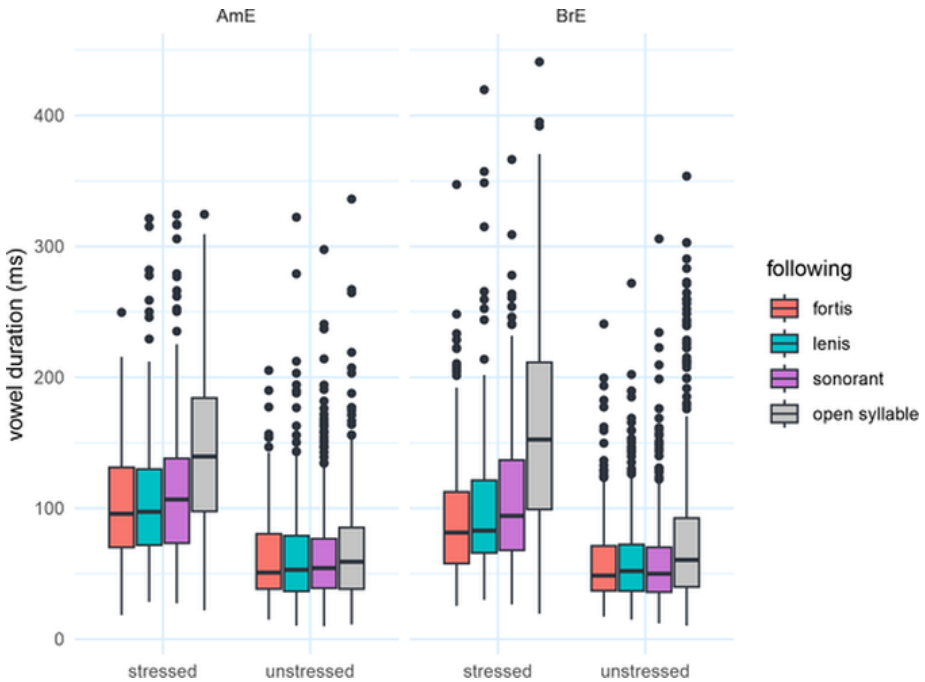

Fig. 11 AVD in terms of the following segment and language variety

When we compare unstressed vowels with a lenis consonant in the postvocalic position and unstressed vowels followed by a fortis consonant, the results do not display any great difference in absolute duration since both the median and IQR values are similar between the two groups. Similarly, the comparison of the duration of pre-fortis and pre-lenis unstressed vowels to the duration of unstressed vowels followed by a sonorant does not produce any prominent differences. Unstressed vowels show a small difference in the duration of open syllables which are slightly longer in duration in comparison to the previously described duration of closed syllables. Again, the difference is slightly more evident in the British variety in which unstressed open vowels also display a more extensive variety in values and result in more outliers than AmE unstressed open syllables.

Stressed syllables on the other hand do show some difference in the duration of individual vowels groups as well as between the two varieties. As Figure 11 illustrates, the British variety shows greater differences between the four vowel groups than the American variety. The data suggests that the BrE pre-fortis stressed vowels may be slightly shorter in their duration that the BrE pre-lenis vowels, as indicated by the IQR, although the median values of the duration of pre-fortis and pre-lenis stressed vowels are around the same value. However, a slightly more prominent difference in the duration of British stressed vowels is between pre-sonorant vowels and vowels followed by a fortis or lenis consonant, again indicated by the median values and the IQR. Similarly to unstressed vowels the most marked difference is between the duration of open and closed BrE syllables. The duration of stressed BrE vowels seem to be substantially longer than that of closed syllables which can be demonstrated most notably by the difference in the median values of the individual groups. Specifically, the median of the duration of pre-sonorant stressed vowels (which seem to have the longest duration) is close to 100 ms, whereas the median of the duration of open syllable stressed vowels is just over 150 ms. Stressed BrE open syllable vowels also show the greatest range of duration, and the overall maximum outlier is a BrE open syllable stressed vowel.

In contrast, the differences between the stressed vowels of the American variety are, apart from open syllables, more similar to the previous comparison of unstressed vowels than to the differences between British stressed vowels. As Figure 11 illustrates, there is no substantial difference in the duration of pre-fortis and pre-lenis vowels and while the median of the duration of stressed vowels followed by a sonorant is slightly above those of stressed pre-fortis and pre-lenis consonants, stressed pre-sonorant vowels are also not substantially longer in duration. Although slightly less major than in BrE stressed vowels, the duration of stressed

open syllable vowels is also longer than the duration of stressed closed syllable vowels. To draw a comparison to BrE, the median of the duration of pre-sonorant stressed vowels is just above 100 ms, whereas the median of the duration of open syllable stressed vowels is about 140 ms. It can be also said that in American English, stressed vowels in closed syllables are generally slightly longer in duration than in British English as indicated by the median values which are all comparatively higher in AmE than in BrE.

Since almost half of the data comprises of monosyllabic words and additionally because monosyllabic words have been shown to differ, this subset of measured vowels tokens was subsequently analysed separately. Monosyllabic words also included substantially more open syllable tokens than polysyllabic words as they made up about a third of monosyllabic words.

As we can see in Figure 12, analysing vowel duration of vowels in monosyllabic words independently does yield different results, especially in stressed vowels. The duration of unstressed vowels is generally similar between the two varieties, as indicated by the graph. The absence of difference between these groups however can be further illustrated by the number of upper quartile outliers, especially in British English. These outliers indicate that the duration of unstressed open syllable vowels can potentially be considerably longer, although as were outliers commented previously, the notably large difference may be due to other prosodic factors.

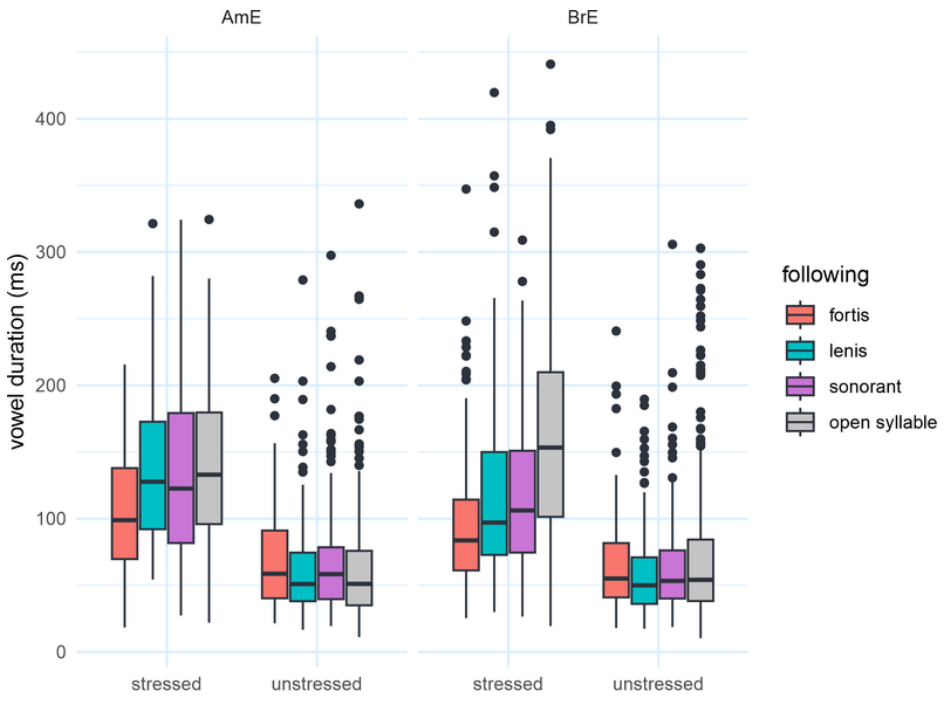

Fig. 12 AVD of vowels in monosyllabic words in terms of the following segment and language variety

However, contrary to the notion of pre-fortis shortening, unstressed vowels in monosyllabic words in both varieties can be observed as slightly longer in duration to the other groups as suggested by the difference in median values across both varieties. This difference seems to be slightly more evident in the American variety since the IQR of AmE pre-fortis vowels indicates higher values.

On the other hand, the resulting vowel durations of vowels in monosyllabic stressed words support the notion of pre-fortis shortening. As Figure 12 shows, both varieties display shorter durations of pre-fortis vowels. Although British English shows greater difference in the IQR of pre-fortis vowels and pre-lenis vowels, American English on the other hand displays greater difference in the median values of the pre-fortis and pre-lenis vowel durations – the median value of the duration of pre-fortis AmE stressed vowels is about 100 ms, the median value of the duration of pre-lenis AmE vowels of around 130 ms. As Figure 12 reveals, the American variety also does not display a major difference between the duration of stressed pre-lenis or pre-sonorant vowels and vowels in open syllables.

If we look at the durations of British stressed vowels in monosyllabic words represented in Figure 12, there is also sufficient evidence for pre-fortis shortening. As has been mentioned, the IQR of the duration of pre-fortis vowels (more than in AmE) as well as median values (although not as notably), suggest shorter duration in comparison to the results of pre-lenis and pre-sonorant vowels. Additionally, the British variety, in comparison to American English, shows notable difference in the duration of stressed open syllable vowels to the duration of pre-consonantal vowels. The longer duration of stressed vowels in open syllables here is indicated by both the overall range as well as the IQC. Moreover, the median value of stressed open syllable vowels (approximately 150 ms) is almost 2 times the median value of pre-fortis vowels (approx. 80 ms) and about 1.5 times the median value of pre-sonorant vowels (approx. 100 ms) suggesting that in British English, the duration of stressed open syllable vowels is substantially longer that the duration of stressed pre-consonantal vowels.

4.3 Analysis of individual examples and outliers

In the next section of the analysis, the maximum outliers will be looked at individually and in detail while considering other possible factors.

4.3.1 Longest durations of stressed vowels

The longest duration of a stressed vowel across both varieties can be ascribed to a female British speaker F3. The duration of the vowel is 441 ms and as can be seen in Figure 13, it is contained in the word "say". There are multiple things to be noted about the characteristics of this particular vowel token. Firstly, it is an instance of a diphthong and in line with the previous classification can be labelled as a long vowel – those have been earlier in the analysis shown to have longer durations than short vowels. Secondly and more importantly, the vowels occurs at the end of a prosodic phrase. As has been discussed in the theoretical part of the thesis (in the section 2.3.1), pre-pausal and phrase-final vowels have been shown to be longer in duration by multiple studies (e.g. Klatt, 1976 or Umeda, 1975) and the phrase-final element is a determining fact for the duration of this vowel token. Due to the scope of this analysis and thesis, the prosodic factor has not been marked and thoroughly analysed and will be only commented on through examples. The spectrogram also shows the creaky quality of the vowel, typical for phrase-final segments.

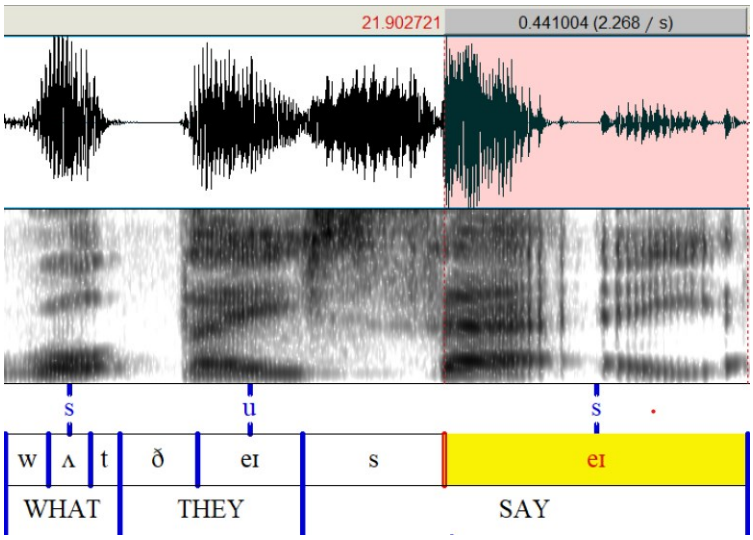

Fig. 13 spectrogram of the prosodic phrase containing the vowel with the longest duration

The second longest vowel occurs in the speech of the same speaker – the vowel is also a long vowel, the vowel /ɑː/ in the word "charge" and its duration is 419.7 ms. The long duration of this vowel can be explained by the same phenomenon as the previously mentioned /eɪ/ as it occurs at the end of a prosodic phrase and additionally before a pause.

The longest duration of a vowel in the American variety occurs in the speech of a male speaker M2 in the word “I” and the duration equates to 324.5 ms, over 100 ms shorter than the longest duration. Equally as the vowel with the longest duration, it is also a diphthong; however, it does not occur in a phrase final, pre-pausal position that could otherwise explain its longer than average duration. In this case, the long duration could be explained by the fact that it occurs at a passage in the recording that consists of shorter consecutive phrases, divided by silent pauses, that overall indicates signs of hesitation – “I am” is after all in most cases an incomplete sentence, in this case it most definitely is. Therefore, the duration of this particular vowel is also influenced by prosodical factor, although different than the ones mentioned previously.

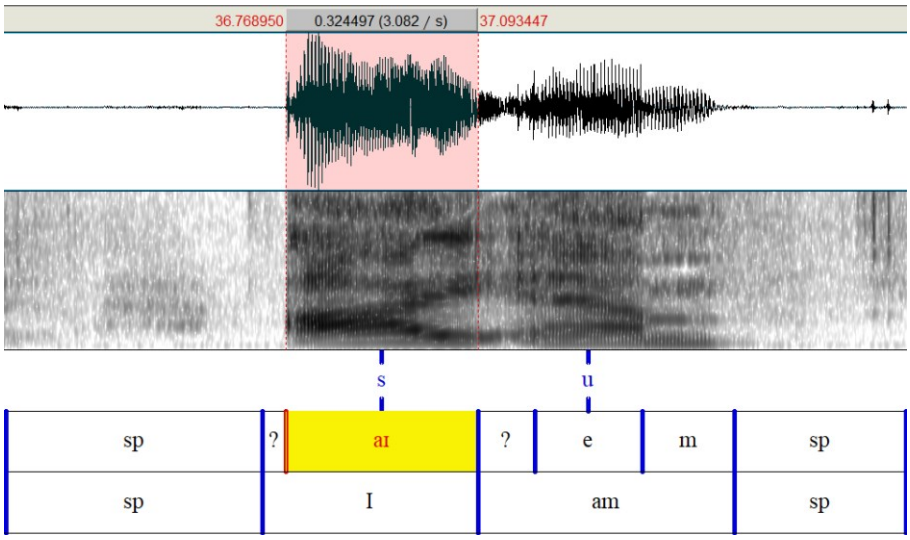

Fig. 14 spectrogram of the prosodic phrase containing the vowel with the longest duration in AmE

4.3.2 Longest durations of unstressed vowels

As with the longest durations of a stressed vowel, the longest duration of an unstressed vowel also occurred in the speech of the British F3 speaker and generally has the same characteristics as the previously discussed vowel in the word “say”. In this instance, it is the final vowel of the word “yesterday” and it is 353.7 ms which is actually longer that the previously discussed longest duration of a stressed vowel in American English. The vowel, similarly to the stressed vowel in “say”, is phrase-final and pre-pausal which contributes to its extended duration.

The longest duration of an unstressed vowel in American English is also spoken by the same speaker as the longest stressed vowel in AmE – American speaker M2. It is also a diphthong in the word “my” and its duration comes to 336.2 ms which is also longer than the

longest stressed vowel spoken by the same speaker. In this case however, the vowel does occur in the phrase-final and pre-pausal position which is what affects its duration.

Since the longest durations of both stressed and unstressed vowels in both varieties occurred in the recording of the same two speakers, it is possible to hypothesise that these long durations are idiosyncratic features of these two speakers. This conclusion seems to be true at least to some extent as both speakers do feature a few times in the first 20 longest durations of individual stressed and unstressed vowels tokens. Specifically, vowels produced by the British F3 speaker make up 3 out of 20 longest durations of stressed vowels and 6 out of 20 longest durations of unstressed vowels, while vowels contained in the recording of the American M2 speaker make up 2 out of 20 longest durations of stressed vowels and also 6 out of 20 longest durations of unstressed vowels. Thus, these long durations could be at least partially seen as a somewhat individual characteristic.

4.3.3 Shortest durations of stressed and unstressed vowels

Finally, it seems appropriate to at least briefly mention the shortest durations of both stressed and unstressed vowels that occurred in the material. Unsurprisingly, the 35 shortest durations are durations of unstressed vowels. The overall shortest duration of a vowel is the unstressed second schwa in the word “governments”, spoken by the female AmE F4 speaker and it equals to only 9.9 ms. The second shortest duration of a vowel is also an unstressed vowel and the shortest duration of a vowel in the British variety that occurs in the speech of the BrE M2 speaker in the word “to” and its duration is 10.3 ms. Both vowels expectedly occur intervocalically, in the middle of a phrase and demonstrate the reduction of unstressed syllables (discussed in section 2.2.5).

The duration of the shortest stressed vowel is therefore thirty-sixth on the list of the shortest vowel durations. It occurs in the speech of the AmE M2 speaker, previously discussed in connection with the longest durations, in the word “health” with the duration of 18.4 ms. In this case, the duration can be discussed in terms of pre-fortis shortening, as the fortis /θ/ influences the vowel despite being preceded by l. Additionally, since the following segment is a liquid consonant, specifically a dark l allophone, the border between the two segments is quite difficult to locate and therefore the duration of the vowel could possibly be measured as slightly longer than the stated 18.4 ms. Nevertheless, the duration is almost twice as long as the shortest vowel duration, which signifies the durational difference between stressed and unstressed syllables, even in the bottom outliers.

5. General discussion and Conclusion

The general objective of the practical part of this thesis was to reexamine the objectively accepted notion that stressed vowels are longer in duration than unstressed vowels. As previous research (e.g. Eriksson & Heldner, 2015) had been done on controlled material, the aim was to attest to whether the duration of stressed vowels can be said to be longer than that of unstressed vowels in connected spontaneous speech. Subsequently, the analysis also aimed to consider the data with respect to other factors that have been evidenced to affect segmental duration.

Overall, the results suggest that stressed vowels are indeed longer in duration than unstressed vowels even in connected speech across both examined varieties of English. Due to the scope of the thesis, the analysis did not include any statistical tests which would undoubtedly result in even more informative data. Nevertheless, by analysing the median values of the durations of stressed and unstressed vowels in both varieties, the results suggest that the duration of stressed vowels is close to 2 times the duration of unstressed vowels.

Moreover, the analysis included examination of further factors such as number of syllables, inherent phonological length and the following segment. As has been suggested, the difference between the duration of stressed and unstressed vowels further differed if the inherent phonological length of vowels was considered, and the results indicated that the duration of long vowels was markedly longer than that of short vowels to such an extent that unstressed long vowels showed similar but also slightly longer duration than stressed short vowels. Furthermore, the difference between the median values of the durations of long and short vowels (both stressed and unstressed) suggests similar results as previous research (Crystal & House, 1988c).

Although some research has included the monosyllabic/polysyllabic condition as a factor (e.g. Umeda, 1975), the number of syllables did not prove to have an extensive effect on the overall difference in duration between stressed and unstressed vowels. Whether the vowel occurred in a monosyllabic or a polysyllabic word however proved to be of note when analysing pre-fortis shortening. As Crystal & House (1988a) suggest in their research, pre-fortis shortening is often influenced by other factors. While the authors note prepausal lengthening as significant for pre-fortis shortening in their research (which was not considered in this study), the result of this study indicate that pre-fortis shortening seems to be more extensive in monosyllabic words. The previous factors were considered between the two varieties which sometimes showed different values but similar tendencies.

Apart from not utilising statistical methods, the scope of the thesis also restricted the factors considered and therefore further research could expand the findings in a number of ways. For instance, the study only considered two levels of stress and did not distinguish between primary and secondary stressed vowels. Since the labelling of vowels was done manually through perception, if a word did contain primary and secondary stress and both were perceptually noticeable, both we marked as stressed and therefore included in the data. However, further examination of the three stress levels and the correlation (or lack thereof) in vowel duration would be interesting as research has shown inversely different results, suggesting duration both as distinct and as unimportant for discerning between primary and secondary stress (Eriksson & Heldner, 2017; Yuan et al., 2008, as cited in Eriksson & Heldner, 2015).

More importantly, hugely informative would be the analysis of prosodic information, especially prepausal and phrase-final position. As research has shown (see Section 2.3.1), segmental duration is largely affected when the segment appears at the end of a prosodic phrase or is followed by a pause. Although the factor was not marked in the data and generally not analysed, an attempt to at least illustrate the phenomenon of *prepausal lengthening* was included in the qualitative part of the analysis though the discussion of outliers. Nevertheless, quantitative data analysing additional prosodic information would be highly suitable for subsequent research.

In conclusion, the results of the study indicate that, in connected spontaneous speech, the duration of stressed vowels is longer than that of unstressed vowels. Moreover, the difference in duration is generally consistent across both varieties of English and although several factors influence specific durational values, the overall tendency of stressed vowels being longer than unstressed vowels persists.

6. References

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge University Press.

Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples – Journal of Applied Language Studies*, 2, 1-24.

Crystal, T. H., & House, A. S. (1988a). Segmental durations in connected-speech signals: Current results, *Journal of the Acoustical Society of America*, 83, 1553–1573.

Crystal, T. H., & House, A. S. (1988b). Segmental durations in connected-speech signals: Syllabic stress, *Journal of the Acoustical Society of America*, 83, 1574-1585.

Crystal, T. H., & House, A. S. (1988c). The duration of American-English vowels: an overview, *Journal of Phonetics*, 16, 263-284, https://doi.org/10.1016/S0095-4470(19)30500-5.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics* 11, 51-62.

Eriksson, A., & Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style, *Proceedings of Interspeech* 2015, 41–45.

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver & F. E. Gibon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 523-602). John Wiley & Sons.

Fry, Dennis B. (1955). Duration and intensity as physical correlates of linguistic stress, *Journal of the Acoustical Society of America*, 27, 765–768.

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey, *Linguistics Vanguard*, 3, 20170007.

Heldner, M. (2001). Spectral emphasis as an additional source of information in accent detection.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence, *Journal of the Acoustical Society of America*, 59, 1208–1221.

Ladefoged, P., & Johnson K. (2015). *A course in phonetics*. Wadsworth.

Laver, J. (1994). *Principles of phonetics.* Cambridge University Press.

Lehiste, I. (1977). Isochrony reconsidered, *Journal of Phonetics*, 5, 253–63.

Low, E. L. (2015). The rhythmic patterning of English(es): Implications for pronunciation teaching. In M. Reed & J. Levis (Eds.), *Handbook of English Pronunciation* (pp. 125–139). John Wiley & Sons.

Nolan, F., & Jeon, H.S. (2014). Speech rhythm: a metaphor?, *Philosophical Transactions of the Royal Society*, B369, 20130396.

Umeda, N. (1975) 'Vowel duration in American English', *Journal of the Acoustical Society of America*, 60, 434–445.

Ravignani, A., & Madison, G. (2017). The Paradox of Isochrony in the Evolution of Human Rhythm. *Frontiers in Psychology*, 8, 1820. https://doi.org/10.3389/fpsyg.2017.01820

Roach, P (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.) *Linguistic controversies, Essays in linguistic theory and practice*. (pp. 73 – 79). Edward Arnold.

Roach, P. (2009). *English phonetics and phonology a practical course* (4th ed.). Cambridge University Press.

Volín, J. (2010). On the significance of the temporal structuring of speech. In M. Malá, & P. Šaldová (Eds.), *...for thy speech bewrayeth thee* (pp. 289-305). Filozofická fakulta, Univerzita Karlova v Praze.

10. Resumé

Cílem této bakalářské práce je prozkoumat segmentální trvání a jeho korelaci se slovním přízvukem ve spontánní souvislé řeči v angličtině. Vokalické trvání bylo v předchozím výzkumu předčítaného materiálu nebo semi-spontánní řeči prokázáno jako akustický korelát slovního přízvuku, a tak je záměrem této práce ověřit, zda jsou předchozí zjištění platná a aplikovatelná na souvislou řeč. Konkrétně je tedy cílem prozkoumat, zda mají přízvučné vokály delší trvání než nepřízvučné vokály a vzápětí zohlednit další faktory, které mohou segmentální, a tedy i vokalické, trvání ovlivňovat, jako například inherentní fonetická délka nebo vliv následujícího segmentu.

Teoretická část se věnuje třem tématům, která se k záměru této práce vztahují, a to rytmu řeči, slovnímu přízvuku a faktorům ovlivňující segmentální trvání. První podkapitola tedy uvádí čtenáře do problematiky rytmu řeči a s tím spjatým konceptem izochronie, které je věnována část 2.1.2. V této části je definován koncept izochronie jako schéma, kdy mají všechny intervaly přibližně stejné trvání. Izochronie je poté uvedena jako základní pojem pro pojetí řečového rytmu jako koordinativního rytmu (*coordinative rhythm*). Dále je izochronie představena jako klíčový pojem pro rozdělení jazyků na taktově-izochronní (*stress-timed*) a slabičně-izochronní (*syllable-timed*) jazyky. Výzkumu spjatému s tímto rozdělením jazyků a charakteristikám obou skupin jazyků, s důrazem na taktově-izochronní, jelikož je mezi ně řazena angličtina, je věnována následující část, na kterou navazuje část popisující doklady a studie, které řeč jako rytmickou a izochronní neshledávají. Poté je také koncept izochronie vysvětlen z pozice percepce, kdy výzkum ukazuje, že řeč zřejmě není izochronně produkována, ale pravidelnost v ní percepčně vnímáme. Vzápětí jsou také popsány alternativní koncepce rytmu řeči, které například řeč hodnotí jako funkčně nerytmickou. Konec této podkapitoly definuje angličtinu jako jazyk, ve kterém je rytmus řeči blízký taktové izochronii.

Druhá podkapitola je věnována přízvuku v angličtině. Ten popisuje hned z několika pohledů – z pohledu produkce řeči, percepce řeči a dále z akustického a fonologického hlediska. Důležitým termínem pro diskusi slovního přízvuku je zde uvedena tzv. prominence, která je definována percepcí délky, hlasitosti a výšky řečového signálu. Přízvučné slabiky jsou tak popsány jako slabiky, které jsou percepčně vnímané jako delší, hlasitější a případně mají odlišnou výšku než přilehlé nepřízvučné slabiky. S vnímáním prominence a přízvuku jsou poté spjaté akustické koreláty slovního přízvuk, kterým je věnována sekce 2.2.2. Nejprve je uveden akustický korelát klíčový pro tuto práci, tedy trvání, a dále je prezentován výzkum, který trvání shledává jako korelující s přízvukem. Vokalické trvání v přízvučných a nepřízvučných

slabikách je značně rozdílné a přízvučné slabiky mívají delší trvání než nepřízvučné slabiky. Dále jsou v této části popsány základní frekvence, intenzita nebo spektrální charakteristiky, např. tzv. *spectral emphasis*, jako akustické koreláty přízvuku. V návaznosti na korelaci akustických parametrů s přízvukem jev další části uvedeno rozlišení tří úrovní přízvuku v angličtině na primární přízvuk, sekundární přízvuk a nepřízvučnou úroveň. Tyto úrovně rozlišuje například trvání, ale také umístění intonačního vrcholu. Dále je věnována další část přízvuku z hlediska fonologie a také s umístěním přízvuku. V této části jsou osvětleny pojmy jako fonetický přízvuk, fonologický přízvuk nebo lexikální přízvuk a jednotlivé pojmy jsou zde ilustrovány na příkladech. V závěru této podkapitoly je diskutováno chování přízvuku v souvislé řeči, především s důrazem na redukci nepřízvučných slabik, které jsou typicky redukovány jak v trvání, tak v kvalitě a jsou tak často redukovány na šva nebo vokál šva blízký. Angličtina se dá tedy definovat jako jazyk s lexikálním přízvukem, který rozlišuje mezi několika úrovněmi přízvuku a typicky redukuje nepřízvučné vokály.

Poslední podkapitola teoretické části se následně věnuje dalším faktorům, které mohou segmentální trvání ovlivnit. Nejprve jsou stručně představeny extralingvistické faktory jako například emoční vypjetí nebo fyzický stav mluvčího, které mohou ovlivnit artikulační tempo a tím pádem i trvání segmentů – rychlejší artikulační tempo znamená kratší trvání a naopak. Poté text přechází k lingvistickým faktorům jako jsou prosodické nebo segmentální faktory. Detailně jsou poté popsány tři faktory a jevy prokázány jako ovlivňující segmentální trvání, a to tzv. *prepausal lengthening*, *prefortis shortening* a vliv inherentní fonologické délky, které jsou také analyzovány v praktické části.

V praktické části je nejprve objasněn rozsah a původ zkoumaného materiálu, kdy se jedná o nahrávky politických debat 8 britských a 8 amerických mluvčích. Dále je popsán proces zpracování nahrávek a extrahování temporálních informací a také je představena pracovní hypotéza, které předpokládá, že přízvučné vokály budou mít delší trvání než nepřízvučné vokály. Zároveň je vzato v potaz, že segmentální trvání může být kromě přízvuku ovlivněno dalšími faktory, a tak jsou extrahované temporální informace zohledněny vzhledem k počtu slabik slov, k fonologické délce nebo k následujícímu kontextu.

V následující části jsou prezentovány výsledky, počínaje relativním vokalickým trváním, které se však neprojevilo jako užitečný údaj vzhledem k převaze jednoslabičných slov, které díky své segmentální struktuře vedly k omylnému a nevypovídajícímu pokřivení výsledků. Zbytek praktické části tak popisuje analýzu absolutního vokalického trvání. Nejprve je analyzováno trvání přízvučných a nepřízvučných slabik v porovnání mezi britskou a americkou

angličtinou, kde v obou varietách mají přízvučné vokály delší trvání než nepřízvučné vokály. Následuje tedy analýza ostatních faktorů, počínaje rozdíly mezi jednotlivými mluvčími. Trvání přízvučných a nepřízvučných vokálů se mezi jednotlivými mluvčími příliš nelišilo, avšak bylo možné z výsledků vyhodnotit obecné závěry o artikulačních tempech jednotlivých mluvčí. Dále bylo zkoumáno trvání s ohledem na počet slabik, kdy bylo ukázáno, že trvání vokálů v jednoslabičných přízvučných slovech je lehce delší než v nepřízvučných. Obecně ale nebyly zpozorovány velké rozdíly mezi trváním vokálů v jednoslabičných a víceslabičných slovech. Dalším hodnoceným faktorem byla inherentní fonologická délka a výsledky zde znovu ukázaly, že tento faktor trvání značně ovlivňuje. Jako poslední byl hodnocen vliv následujícího segmentu, kdy při analýze veškerých slov výsledky poukazují na delší trvání vokálů v otevřených slabikách, avšak nevypovídají o vlivu fortisových konsonantů na vokalické trvání, jelikož rozdíly mezi trváním pre-fortisových vokálů a vokálů předcházející lenisovým konsonantům nebo sonorům jsou minimální. Vzápětí byl analyzován vliv následujícího segmentu pouze v rámci jednoslabičných slov, kde je vliv fortisových konsonantů v přízvučných slabikách značný.

V poslední části této práce dochází k diskusi a shrnutí zkoumaných výsledků. Lze finálně konstatovat, že i ve spontánní souvislé řeči mají přízvučné vokály delší trvání než nepřízvučné vokály, a i když je segmentální trvání dále ovlivněno dalšími faktory, tendence delšího trvání přízvučných vokálů oproti nepřízvučným přetrvává.