FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

FONETICKÝ ÚSTAV

DIPLOMOVÁ PRÁCE

DARINA URBANOVÁ

Durational variation of segments in read speech of Czech and British speakers of English

Variabilita trvání segmentů ve čtených textech českých a britských mluvčích angličtiny

PRAHA 2011

Vedoucí práce: doc. PhDr. Jan Volín, Ph.D.

I would like to thank doc. PhDr. Jan Volín, Ph.D. for his patience and energy when reviewing the present paper and for his stimulating comments with regard to the content of the thesis.

I declare that the present MA thesis is my own work for which I used only the sources and literature mentioned. I also declare that the thesis has not been used as part of any different university program or in order to obtain a different or the same university degree.

Prohlašuji, že jsem diplomovou 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.

Praha, 11 August 2011 .....

ABSTRACT: The present study focuses on durational variation of segments in read speech of Czech and British speakers of English. The variation in segmental duration was examined in the speech of three Czech and three British speakers with respect to individual vowel and consonant categories. Further, stress and intonation phrase boundary were explored as possible factors influencing segmental duration.

The following tendencies were observed in the speech of Czech speakers. First, they were found to miss the appropriate proportions between individual vowel categories and realize their long vowels as slightly too short. Second, they lacked an adequate degree of durational contrast in stressed and unstressed vowels when compared to the British speakers. And third, with regard to the boundary, the Czech English vowels and consonants in the last syllable before an intonation phrase boundary were found to lag slightly behind in the degree of pre-boundary lengthening.

The theoretical part of the study focuses on two main topics. The first is foreign accent with regard to accentedness, intelligibility and comprehensibility as well as in connection to its implications and factors which might influence its degree. The second part contains theoretical background about the time domain of speech concerning segmental duration in Czech and English, the area of rhythm, the importance of durational cues in perception and a short survey of factors reported to affect segmental duration. Czech accent in English and the importance of duration in foreign language are also shortly covered.

The present study aims to contribute to a more comprehensive description of Czech English as well as to the search for the proper focus of pronunciation instruction in EFL in the Czech context.

Key words: segmental duration, foreign accent, Czech English, British English, read speech, stress, boundary

ABSTRAKT: Tato práce se zabývá variabilitou trvání segmentů ve čtených textech českých a britských mluvčích angličtiny. Variabilita v trvání segmentů byla v řeči tří českých a tří britských mluvčích zkoumána z hlediska příslušnosti k jednotlivým vokalickým a konsonantickým třídám segmentů a dále byl zkoumán i vliv přízvuku a hranice promluvového úseku.

V řeči českých mluvčích byly pozorovány následující tendence. Jednak se českým mluvčím nedařilo zcela postihnout rozdíly v trvání mezi jednotlivými třídami samohlásek a jejich dlouhé samohlásky byly tedy o něco kratší než u britských mluvčích. Dále samohláskám českých mluvčích částečně chyběl potřebný kontrast daný rozdílem v trvání v přízvučných a nepřízvučných pozicích. A z hlediska vlivu hranice, jak samohlásky tak i souhlásky českých mluvčích nedosahovaly takového stupně prodloužení v poslední slabice promluvového úseku, ke kterému docházelo u mluvčích britských.

Teoretická část práce se zabývá dvěma hlavními tématy. Prvním je cizinecký přízvuk s ohledem na stupeň přízvuku, jeho objektivní i subjektivní srozumitelnost a také na jeho důsledky pro komunikaci a na faktory, které mohou ovlivnit jeho míru. Druhá část pojednává o temporální doméně řeči vzhledem k trvání segmentů v češtině a angličtině a oblasti rytmu. Zmiňuje se též o významnosti trvání v percepci a krátce pojednává o vlivech, které by měly segmentální trvání ovlivňovat. Práce neopomíná ani téma českého přízvuku v angličtině a důležitosti trvání v cizím jazyce.

Cílem této práce je přispět k ucelenějšímu popisu české angličtiny a pomoci identifikovat správnou oblast, na kterou by se měla výuka anglické výslovnosti v českém kontextu zaměřit.

Klíčová slova: trvání segmentů, cizinecký přízvuk, česká angličtina, britská angličtina, čtená řeč, přízvuk, hranice

TABLE OF CONTENTS

List of Appendices

LIST OF ABBREVIATIONS

ANOVA – analysis of variance

AOL – age of learning

BBC – British Broadcasting Corporation

CP – critical period

EFL – English as a foreign language

ESL – English as a second language

JND – just noticeable difference

L1 – first language (mother tongue)

L2 – second language (sometimes used also in the context of foreign language)

LOR – length of residence

NS – native speaker

NNS – non-native speaker

Pre-Bound 0 – other syllables than the ultimate or pre-ultimate one before the boundary

Pre-Bound 1 – ultimate syllable before a boundary

Pre-Bound 2 – pre-ultimate syllable before a boundary

PVI – Pairwise Variability Index

SLM – Speech Learning Model

SPL – sound pressure level

ΔC - standard deviation of the duration of consonantal intervals

ΔV - standard deviation of the duration of vocalic intervals

%V - proportion of vocalic intervals

1 INTRODUCTION

Speaking a second or even a third language has become a necessity for many people around the world. Yet the apparent fact concerning the grand majority of these people is that their non-native speech is foreign-accented. If one learns a foreign language, he or she undoubtedly wants to be understood. Intelligibility is thus an important factor in a foreign-accented speech.

Research suggests that accentedness and intelligibility are connected but partially independent problems (Derwing & Munro, 2009). It has been found that intelligibility of non-native speech depends on rhythmic and durational patterns among other factors (e.g. Tajima, Port, & Dalby, 1997; Boula de Mareüil and Vieru-Dimulescu, 2006; Holm, 2008, cited in Quené & van Delft, 2010). Therefore, it is possible that incorrect temporal proportions in non-native speech influence both intelligibility and the degree of accent.

The present study aims to map the variability in the time domain of speech of Czech and British speakers of English. It will focus on the variability in segmental durations in the read speech of three Czech speakers when compared to three reference British speakers. Specifically, it will observe durational variation with regard to individual categories of vowels and consonants and will also examine the possible influence of stress and intonation phrase boundary. The study would like to contribute to the debate about the nature of Czech English accent by mapping the possible tendencies in its temporal plan.

The theoretical introduction to the study will focus on two main areas connected to the present research – foreign accent and the time domain of speech. The first part will cover the topics of accentedness, intelligibility and comprehensibility. It will discuss the possible social implications of a foreign-accented speech. Further, it will consider several factors which are reported to influence the degree of foreign accent. The topic of L1 and L2 interaction will be also touched upon and finally, we will consider the problem of an appropriate focus of pronunciation instruction. After this general introduction of foreign accent, the study will focus on the position of English in the world and in particular, on one of its varieties – Czech English.

The second principal part will deal with the time domain of speech. Specifically, it will mention phonological length, segmental duration in English and Czech, factors influencing segmental duration and why duration is important for perception. The section will be concluded with the topic of rhythm. Finally, the introductory part will be terminated by linking the two main topics and focusing on the importance of duration in foreign language.

2 FOREIGN-ACCENTED SPEECH

2.1 Several notes on L2 learning terminology

The subject area of language learning is connected to some terminological vacillation. Thus the terms *acquisition, learning, second language* and *foreign language* will be briefly explained in this section. Also the essence of the terms *L1 transfer* (*L1 interference*) and *interlanguage* needs to be clarified in connection with foreign language learning.

Although the terminological distinction is not made by all researchers, most adhere to the notional difference between *second language acquisition* and *foreign language learning*. In the former case, the non-native speaker of the second language (L2) acquires it relatively naturally while being immersed in the target language speaking environment. He or she lives in the country where this language is spoken, which is often the case of immigrants. On the other hand, foreign language learning applies to languages being taught and learnt in an artificial (non-real life) setting, often a classroom setting, in a country where the language is not spoken on a day-to-day basis.

Nevertheless, the terminology is not always used in a straightforward way. Some researchers deem that the term *acquisition* should be used solely for the act of acquiring a first language (i.e. mother tongue) and the word *learning* for all further languages. Similarly, there is sometimes a discrepancy in the use of the term *second language*. There are three most common meanings (uses) of this term.

First, it is the one mentioned above, i.e. a language used in the community of the L2 native speakers (NS) where the non-native speaker (NNS) permanently resides. Second, it is used in countries where the official language or the language of instruction differs from the mother tongue(s) of its inhabitants. This is mostly due to large linguistic heterogeneity of the area and is common, for instance, in post-colonial countries. This official language is then said to be the inhabitants' second language. The situation differs from the preceding one in that that the NNS's use of the L2 is usually register-specific. The NNS uses the L2 only in some communicative situations and he learns the language most probably at school, often without the presence of NSs.

Last, the term second language may be used to refer to any language which is being learnt / acquired other than the mother tongue (i.e. first language). However, this last case is often called *foreign language* as well. It can be seen there is some instability in the terminology in this area. Consequently, in this paper if the context requires it, the terms

second language and *foreign language* will be used as defined at the beginning of this section – second language as being acquired (usually in a community of NSs) and foreign language as being learnt and taught in other contexts. However, if no distinction is needed, it will be referred to both simply as the *target language*.

When one learns a foreign language, it is quite likely that the language he or she produces is going to be influenced by their mother tongue, especially if the mother tongue and the target language differ to a great extent. This effect is called *L1 transfer* or *L1 interference* (L1 standing for first language, i.e. mother tongue) and may concern any level of language – from phonological to textual – and most of the time more levels are affected at the same time. The resulting linguistic hybrid which is thus established by the non-native speaker is called *interlanguage*.

In the *Dictionary of Linguistics and Phonetics*, Crystal defines interlanguage as follows:

> The linguistic system created by someone in the course of learning a foreign language, different from either the speaker's first language or the target language being acquired. It reflects the learner's evolving system of rules, and results from a variety of processes, including the influence of the first language ('transfer'), contrastive interference from the target language, and the overgeneralization of newly encountered rules. (Crystal, 2008, p. 249)

He thus proposes that interlanguage is a matter of not L1 transfer only but of other factors as well. However, the mother tongue is usually regarded to play a principal role in an interlanguage. As noted above, interlanguage often manifests itself on all levels of language and consequently, the speech, or the concrete phonetic realization of the target language, is often affected as well (for more information on phonological interference see Section 2.4.8).

The experimental part of present study will examine the speech of one such interlanguage, namely Czech English which will be described in more detail in Section 3.1. Nevertheless, before proceeding to the Czech English interlanguage, more theoretical background on foreign accent and all its aspects is needed.

2.2 Accentedness, intelligibility and comprehensibility

In the world there are hundreds of millions of people who can speak more languages than just their mother tongue. However fluent some of them are, one thing usually betrays the fact that they are not native speakers of the language they are using – their accent. Of course, both NSs and NNSs have an accent. NSs' accent is based on their region, class or register

used at the time of speaking. NNSs' accent is influenced mainly by their first language. The important thing many people agree upon is that no accent is good or bad - it just simply is.

Nevertheless, many foreign or second language speakers are concerned about their accent. Since the present study examines the case of Czech English as contrasted to British English, the word *accent* will be used here to refer to a foreign accent, i.e. to the ways in which the speech of NNSs of a target language differs from the speech of its NSs. Thus the present study deals particularly with the differences between the accent in Czech English speech and the southern dialect of British English.

Whether we incline towards the point of view that foreign accent does not matter, or whether we are of the opposite opinion, one thing seems to be certain. Foreign accent features can be very salient (Scovel, 1988, in Derwing & Munro, 2009), meaning that listeners are very sensitive to the presence of foreign accent and they can detect it easily. This was pointed out by many studies. For instance, it has been found that listeners can detect a heavy foreign accent on the basis of just 30 ms of speech (Flege, 1984) or even in backwards speech (Munro, Derwing, & Burgess, 2003) – both as reported in Derwing & Munro (2009). However, even if a NNS has a strong accent, it does not mean that his or her speech cannot be fully intelligible (see e.g. Munro & Derwing, 1995a, 1995b). Successful intercultural communication depends on many factors and having an accent is just one of them. *Accentedness* and *intelligibility* are thus related but partially independent.

It is necessary to specify what we mean by *intelligibility*. Munro & Derwing (1995a) define intelligibility as the "extent to which a speaker's message is actually understood by a listener" (p. 76). It is thus very important in communication. Yet Derwing & Munro (2009), point to the fact that intelligibilty is rather difficult to assess. They mention some assessment techniques which include comprehension questions, summaries of heard speech, sentence verification tasks, or counting the percentage of correctly transcribed words.

Accentedness is usually assessed by independent raters, most often using rating scales with one end of the scale marked as "native-like pronunciation" or "no foreign accent" and the other end with the other extreme – "heavy foreign accent" (Piske, MacKay, & Flege, 2001, p. 194). In previous research, 5-point scales have been used most often, but there have been also cases of 3-point, 4-point, 6-point or even 9-point scales. In several studies conducted by Flege and his colleagues (e.g. Flege, Munro, & McKay, 1995) a continuous scale has been used in order to find out how many scale values are adequate for capturing the listeners' sensitivity to the range of foreign accents (Piske et al., 2001, p. 194).

Besides accentedness and intelligibility, there is a third important concept associated with foreign speech - *comprehensibility*. Derwing & Munro (2009) define it as the "listener's perception of how easy or difficult it is to understand a given speech sample" (p. 478). It differs from intelligibility by being a judgment of difficulty and not how difficult it really is for the listeners, i.e. how much is eventually understood. Just like accentedness, it is measured using an evaluation scale. Munro & Derwing (1995b) found that sentences with low comprehensibility ratings take more time to process than ones which are judged as more comprehensible. Moreover, they point out that there was no evidence in their data that the degree of accent would be related to the processing time. Consequently, even if two speech samples are both fully intelligible, they may differ in comprehensibility. The listener has to concentrate more on the less comprehensible one, which prolongs the reaction time, and as many psycholinguistic experiments show, in the long run it also makes the listener more tired.

To summarize this section, accentedness, intelligibility and comprehensibility are all partially interdependent but distinct entities (see also Derwing & Munro, 1997). A nice simplified comparison of the terms is provided by Derwing & Munro (2009): "[A]ccent is about difference, comprehensibility is about the listener's effort and intelligibility is the end result: how much the listener actually understands" (p. 480). Consequently, all three are important dimensions of a foreign speech.

2.3 Social dimensions of foreign accent

There are both advantages and disadvantages to speaking with a foreign accent. When a NNS with a perceivable accent interacts with a NS, it is likely that the NS will adjust his or her speech so that the NNS would understand more easily. This kind of speech on the part of the NS is sometimes called *foreigner talk* and might involve, for example, slower speech rate and better enunciation than if the person was speaking to other NSs. This, obviously, may facilitate communication between people from various language backgrounds. Yet learners with a stronger accent who have otherwise no problems with other areas of the target language, like understanding NSs' colloquial speech, might feel slightly offended by such an approach. For some speakers there can be another advantage of having an accent. Some foreign accents might be considered desirable – NSs might think of them as cute or sophisticated. Derwing and Munro (2009) give an example of Maurice Chevalier who was allegedly asked to exaggerate his French accent because people considered it 'charming' (p. 484). However, this is probably not the case for the majority of foreign language speakers.

On the other hand, there are also disadvantages to a foreign accent. The first and perhaps the most prominent one is the possible loss of intelligibility. For a language learner, there is nothing as frustrating as not being understood - maybe just not being able to understand. Although, as was noted in Section 2.2, there is not a simple correlation between accentedness and intelligibility, having a strong accent (or some dimensions of it) may cause misunderstanding between the interlocutors. Naturally, this may be very uncomfortable for both the speaker and the listener.

Another disadvantage of accent is social evaluation on the part of the listener. Leather & James (1996) note that to have an accent acceptable for NSs does not necessarily mean to speak in a native-like fashion. The acceptability depends on many factors including which variety of the target language the accent approximates, and the degree of bias towards the accent. They add that having an "overperfect pronunciation" may be also viewed quite negatively by some NSs (p. 271). What is more, consciously or subconsciously, many people hold a lot of prejudice or deep-rooted stereotypes connected to accented speech. As Munro, Derwing and Sato (2006) report, minority accents can be viewed as "signs of ignorance" or "lack of sophistication" and media and the film industry often promote such stereotypes (p. 68). Consequently, people with accented speech can face discrimination at various spheres of their social life. The most notable is probably discrimination at workplace. For example, research has shown that NSs of English might judge NNSs as less suitable for high-status jobs than NSs applicants and that in some cases the strength of foreign accent might negatively correlate with income (Kalin & Raiko, 1978; Bohara & Saenz, 1993; both in Munro et al., 2006). Basically, if language learners are eager to work on their accent, these might be quite powerful reasons. Nevertheless, such a negative social evaluation is often due to intolerant listeners (even if this might be subconscious). As Derwing and Munro (2009) stress, "accents [as such] do not cause discrimination", the fault is with the listeners (p. 486).

There is another valid statement in their paper: both the speaker and the listener are responsible for successful communication (Derwing and Munro, 2009, p. 486). Consequently, both sides should try to do their best in order for their interaction to be effective. Yet even if the listeners have no bias towards accented speech and even if it is quite intelligible, they may have problems understanding this kind of speech simply because they are not used to communicating with NNSs. Thus they might not have enough confidence in such communicative situations and the communication may break down. It has been shown that familiarity instruction with accented speech improves comprehension and facilitates intercultural communication in such cases (Gas & Varonis, 1984; Derwing, Rossiter, &

Munro, 2002; both in Derwing & Munro, 2009). In conclusion, in our today's globalized world, stress should not be put solely on foreign language teaching with pronunciation training and listening comprehension as its indispensable parts, but also on familiarity instruction and listener education of NSs and some basics of intercultural communication for everybody.

2.4 Factors affecting the degree of foreign accent

When one learns a foreign or second language, there are many factors which seem to influence its successful attainment, and more specifically, in the area of pronunciation – the degree of foreign accent. However, it sometimes needs to be differentiated between *second language acquisition* and *foreign language learning* (as defined in Section 2.1) because the factors might be different in the two cases due to the divergent learning contexts.

In the following sections the factors which might affect the degree of foreign accent will be shortly discussed. The article by Piske et al (2001) will serve as the main source of information for this part since it contains a relatively extensive review of literature on this topic. However, the article deals solely with foreign accent in a *second* language, so the area of factors influencing foreign accent in a *foreign* language is not covered in their review. Where necessary, the present study will thus try to draw attention to areas where the factors apply only to one case, but not the other. The factors discussed below are (1) age of learning, (2) length of residence, (3) gender, (4) formal instruction, (5) motivation, (6) language learning aptitude, (7) language use, and (8) mother tongue.

Several of these factors – namely length of residence, formal instruction, language use, and partially also age of learning – could be subsumed under a more general category, i.e. 'quantity and quality of target language input'. The reason is that it is the input and its amount and quality that is implied in each of these factors. Researchers only attempt to quantify input by means of these factors (or rather its amount because the quality is very difficult to be quantified) and test them whether they are suitable predictors of the degree of foreign accent. Each of the above-mentioned factors will be now discussed one by one.

2.4.1 Age of target language learning

The factor entitled here as the age of target language learning signifies the age when the subject first became exposed to the target language. In the case the foreign language learning, this means the beginning of instruction in the target language. In the case of second language acquisition, this age factor is usually called the *age of learning* (shortened as AOL) and it signifies the age when the subject arrived into the L2 speaking country, i.e. since when he or she has been permanently exposed to this language. Taking into consideration the amount of input NNSs receive when acquiring the language in a foreign country in contrast to the amount of input they might get in a classroom setting in their home country, it follows that the two age factors will most likely affect the degree of foreign accent to a different extent.

However, what the two age factors have in common is the theoretical link to the *critical period (CP) hypothesis*. This hypothesis was originally proposed by Lenneberg in 1967 for first language (L1) acquisition and it was based on evidence of impaired L1 skills and the question whether it is possible to acquire L1 after a certain age (Abello-Contesse, 2009). Later the CP hypothesis was extended to the domain of second and foreign language learning and acquisition. The most restrictive version of the hypothesis would say that after the end of the supposed CP, complete attainment of the target language is not possible any more. It has been suggested that there might be more different CPs for different linguistic abilities and the capacity of native-like pronunciation is supposed to be the first one to be lost (see Piske et al., 2001). Thus people who began learning the target language before the end of the CP are supposed to have significantly better pronunciation than people who commenced after the end of the CP. The reasons behind CP effects are claimed to be the age-related loss of neural plasticity, some kind of neurofunctional reorganization, or the result of the interaction between a learner's L1 and L2 systems (Piske et al., 2001, p. 196).

However, it has been suggested that what is at play is more likely to be only a "sensitive" period rather than a "critical" one and that the decline in linguistic abilities is thus more gradual. The crucial question which the research on the effect of age-related factors on foreign accent is trying to answer is when exactly the sensitive period takes place and when the latest point is when a person should start acquiring the target language in order not to sound foreign. The results seem to be divergent with the ages of 6 and about 12 years cited most often, but as Piske et al. (2001) claim, "no study has as yet provided convincing evidence for the claim that L2 speech will automatically be accent-free if it is learned before the age of about 6 years and that it will definitely be foreign accented if learned after puberty"

(p. 197). The reason is that the matter gets further complicated by other factors affecting the degree of foreign accent and also by the fact that the factor of AOL is often confounded with other factors like length of residence or amount of L1 and L2 use. Nevertheless, AOL often accounts for a great amount of variance in foreign accent ratings. For example, in the study of Flege et al. (1995) it accounted for an average of 59% of variance in their ratings, which means that it was overwhelmingly the most influential factor considered in their study.

2.4.2 Length of residence

Another variable which has been reported to correlate with the degree of foreign accent to some extent is the *length of residence* (LOR). The LOR is a measure of the number of years spent in the target language speaking country. It thus applies only to the case of second language acquisition and plays no role in foreign language learning. In the context of foreign language learning, the number of years learning the language might be used as a similar measure, but again probably with differing effects.

With regard to second language acquisition, LOR effects have been found to be much smaller than in the case AOL and some studies even did not even found the effect to be significant, which might have been allegedly due to a too narrow range of LOR values according to Flege and Fletcher (1992, in Piske et al., 2001). However, some effect of LOR is often reported and it has been suggested that the degree of LOR effects depends on whether the learners are still in their early phase of learning or not (e.g. Flege 1988, in Piske et al., 2001).

2.4.3 Gender

There have been divergent findings as regards the influence of gender on the degree of foreign accent. Most studies have not found gender to have a significant effect, but Asher and García (1969, in Piske et al., 2001) and Flege and his colleagues (1995) reported some influence on the degree of foreign accent. On the whole, however, any strong conclusions cannot be reached on the basis of available evidence with respect to gender.

2.4.4 Formal instruction

Although some studies detected the effect of formal instruction on the degree of second language foreign accent (e.g. Flege & Fletcher, 1992, in Piske et al., 2001), many of them did not. This might be due to the lack or insufficiency of pronunciation instruction in most language classrooms. The effect of formal instruction, sometimes specified as the number of years of target language instruction, then might become greater if students receive special training in the target language pronunciation. This view has been supported by Bongaerts et al. (1997), Moyer (1999) and Missaglia (1999) - all cited in Piske et al., 2001. It seems that the instruction should include both suprasegmental and segmental training although the suprasegmental one might be more important (for more information see Section 2.5).

However, in the case of foreign language learning, hypothetically, the results regarding the effect of the number of years spent in the language classroom might differ to some extent from the ones reported above. Due to the lack of factors like AOL in this context, the extent of effects like the length of instruction might be more significant. Nevertheless, the conclusion that if pronunciation training is focused upon in language teaching, the results might be even better is probably also valid in this case.

An important point to bear in mind is, however, that not only quantity (length) of instruction but also its quality is involved. If the instruction is long, but not effective (good), there will not be much effect on improving one's accent. The opposite might hold as well. If the instruction is good, but lasts too short, then the effect might also be limited. This would need to be confirmed in a larger number of longitudinal studies which are, however, still quite scarce. On top of that quality of input might be very hard to assess. Quantity and quality of instruction should not thus be confounded even though the degree of accent probably depends on both.

2.4.5 Motivation

Generally, it is believed that motivation plays a role in foreign and second language learning. Consequently, its influence on the degree of foreign accent has been examined in several studies. Although there have also been studies which have showed no significant effect of motivation on foreign accent, most studies did find it to be significant. Motivation might be quantified as "integrative motivation", "concern for L2 pronunciation", or also

"professional motivation" if the accuracy of pronunciation is seen to be important for the subjects' occupation. These factors have been found to correlate with the degree of foreign accent although sometimes they account for only a small portion of the variance. Piske et al. (2001) conclude that some influence of these factors have been found but that high level of motivation does not automatically mean that the speech would be accent-free. They also add that the motivation factors need to be quantified more precisely in future research.

2.4.6 Language learning aptitude

Many people believe that whereas for some people learning languages is easy, for others it is almost impossible however hard they try. To examine whether this belief is true or whether it is unfounded, some accent studies have tried to examine aptitude factors as well. The biggest problem is, however, how to measure the aptitude for learning languages. What is the factor or factors which should account for such an aptitude? Is it musical ability and if yes, does it influence the degree of foreign accent? Or is it mimicry ability? Not too many studies have dealt with this topic, but a few have.

Until recently, it has seemed that musical ability does not significantly affect the degree of foreign accent. However, in 2007, Sheppard, Haysashi, and Ohmori examined the degree of foreign accent in the speech of Japanese foreign language learners and found that self-proclaimed musical ability accounted for 32% of variance in their data. However, this has been an isolated study which has showed such a strong correlation. More research is needed to confirm any stronger conclusions and a stricter quantification of the musical ability would probably be better.

The ability to mimic unfamiliar speech sounds has been found significant by more studies. Although it usually accounted for only small amount of variance (only around 2 to 5 percent), all studies except of one examined by Piske et al. (2001) have identified mimicry ability as a significant predictor of the degree of foreign accent. The question connected to both musical ability and mimicry ability, as well as language learning aptitude as such is whether these are inborn abilities or whether one can learn them or at least improve them during one's lifetime. Consequently, more research is still needed in this area.

2.4.7 Language use

Also language use factors have been examined in order to find out whether and how much they affect the degree of foreign accent. By language use factors it is usually meant the amount and frequency of L2 use (either at work, at home, social use, or everywhere) and sometimes also the amount of L1 (mother tongue) use. Again not all studies have found a significant effect of language use factors but some have. For example, Flege et al. (1995) found that L2 language use factors were the second most important predictor of the degree of foreign accent, right after AOL, accounting for as much as 15% of the variance in their data. The amount of L1 use also seems to be an important factor in second language acquisition data. In the study of Piske et al. (2001) the self-reported use of L1 was found to affect the degree of L2 foreign accent significantly. Also here it was the second most influential factor after AOL.

2.4.8 Mother tongue – L1 and L2 interaction

It is quite likely that the character of the NNSs' mother tongue might also have an influence on the degree of foreign accent. People with different L1s seem to learn different target languages with various amount of success and the same might hold also for the degree of foreign accent. If this is so, there seem to be two (mutually interdependent) factors which might account for the divergent influence of L1s on the target language. The first one is the size of L1 phonological system and the second one involves the degree of similarity between the L1 and the target language, often simplified as the phonological systems similarity.

Although Iverson and Evans's (2007) study did not deal with the degree of foreign accent as such but with recognition of target language (English) vowels, it could be hypothesized that their findings might apply to foreign accent as well. They found that subjects with larger and more complex L1 vowel systems (German and Norwegian) were more successful in recognizing English vowels than subjects with smaller L1 systems (Spanish and French). A very tentative conclusion thus may be proposed on the basis of this study - having an L1 with a large phonemic system might be beneficial for learning foreign languages.

Closely connected to the size of the L1 phonological system is the degree of its similarity to the target language (L2) phonological system. It has been proposed that the L1 and L2 systems interact in a common phonological space of each language learner (Flege,

1995, in Flege, 2007) and the nature of these interactions is going to be the topic of the remaining part of this section.

When one learns a foreign language, it seems that for most people some sounds of the target language are easier to pronounce than others. A lay view of many learners would probably be that if the target language segment is present also in the L1, then it will not pose any problems to the learner; however, if the segment is not part of the L1, then its production might be more difficult. However, recent theoretical accounts of the problem have suggested that the picture might be more complicated. The theoretical models of Best (1995, in Strange, 1999) and Flege (1995, in Flege, 2007), which both deal with the phonological aspect L2 learning, will be shortly described below.

Best's Perceptual Assimilation model (1995) is succinctly summarized in Strange (1999). As its name suggests, the model deals mainly with perception of non-native segments, but it is relevant also for production. It is well-known that the areas of production and perception are interconnected and that a learner will most probably not be able to produce a segment correctly unless he or she is able to perceive it correctly. Thus Best's model, if valid, explains partially why it is so difficult to pronounce some L2 sounds.

Strange's (1999, p. 2513) account of Best's model is the following. The model defines phonetic categories in terms of "gestural coordinative structures". These underlie both perception and production. When we acquire our L1, these gestural structures are strengthened over time as the L1 phonological system develops. Consequently, adult speakers process their L1 by the means of these structures. Only such acoustic variation which is phonologically distinctive in the L1 is considered significant by the listener while processing this stream of speech. However, the core of the model predicts how non-native (L2) categories will be perceived. It says that beginning L2 learners will experience relative difficulty in discriminating the contrasting L2 segments based on how well these L2 categories "fit" the L1 categories (which is quite close to the lay view expressed above). Yet the model proposes more than this.

The Perceptual Assimilation model offers different possibilities which can occur while discriminating L2 sounds. The most difficult situation for the L2 learners is when two contrasting L2 segments are perceived as equally "good" instances of the same L1 phonological category ("the single category pattern"). If the segments are both assimilated to the same category, but one is viewed to be a better exemplar of the category than the other, discrimination will be slightly less difficult ("category-goodness pattern"). The easiest case for successful discrimination is when the two L2 segments are judged to belong to different

L1 categories ("two-category pattern"). Best thus describes perception of L2 sounds as constrained by the interference of L1 phonological categories (cited in Strange, 1999, p. 2513).

Flege's Speech Learning Model (1995, in Flege, 2007) is similar to Best's model in trying to describe the interaction of L1 and L2 phonological systems and in using the phonetic similarity of categories as one of its principal concepts. In contrast to Best, Flege suggests that the systems interaction is bi-directional because they co-exist in a "common phonological space" (p. 366) – more specifically, that it is not only the L1 system which influences the L2 system, but later the L2 system can also affect the mother tongue.

The Speech Learning Model (SLM) (Flege, 1995, in Flege, 2007) tries to account for changes in speech learning over time. Flege claims that the nature of L1-L2 interactions changes with age and thus might account for some of the age effects on speech learning. The SLM predicts two mechanisms of interaction of the L1 and L2 systems: category assimilation and category dissimilation. Which of the two mechanisms is at work is said to depend on whether category formation has or has not occurred (and on the degree of similarity between the L1 and L2 categories).

Phonetic category assimilation takes place when L2 learners are unable to establish a new category for a perceptually distinct L2 sound because either the sound is perceptually incorporated into an existing L1 category or because it is considered to be too similar to an L1 sound (or both). The assimilated category may later evolve into one which differs from both the original (monolingual) L1 and L2 categories which came to be assimilated. However, this "merged" category may be closer to one of these original categories.

The second case proposed by SLM – category dissimilation – takes place if a new category is formed for a speech sound which is found in the L2 but not the L1. Flege claims that if such a new category is formed, the phonetic space becomes more crowded and this may cause the new category to "deflect away" from the closest L1 category in order to maintain phonetic contrast. In his 2007 paper, Flege supports his model by examples of both category assimilation and category dissimilation by bilinguals.

Strange (1999) comments on the two models summarized above by proposing that they might involve the problem of circularity. She claims that the concept of phonetic similarity needs to be defined by other measures than the ones used to assess perceptual difficulty if this concept is to be used as its predictor (p. 2513). Besides proposing direct methods to measure phonetic similarity, she also reasons that its cross-language comparison on the basis of phoneme inventories is "far too abstract" and suggests an intermediate level of

abstraction which is able to capture "listeners' knowledge of systematic allophonic, phonotactic, and prosodic variation" (pp. 2513-2514).

Moreover, Best (1995, in Strange, 1999) and Flege (1995, in Flege, 2007) deal with L1 transfer only on the phonological (segmental) level, but it has been hypothesized that cross-language interference functions also on the suprasegmental level and that it might be even more significant for the strength of foreign accent. Volín & Skarnitzl (2010a) claim that in general, research has been biased toward the segmental dimension of foreign accent (p. 271).

To conclude, it seems that the influence of the mother tongue on the target language is a concept that is not doubted any more. However, as Flege suggests, it is possible that this interaction is mutual and changes over time. In line with Strange, more research into this area is still needed. Direct methods of phonetic similarity examination as well as trying to pinpoint the most appropriate level of abstraction could be useful for cross-language comparison and for detecting the principles of target speech learning.

2.5 Teaching pronunciation – where to put the focus

As was hinted at in the previous sections, there are many aspects to speaking a foreign or second language. What should the learners concentrate on then if they want to work on their pronunciation? Or from the teachers' viewpoint, what aspects of pronunciation should they teach so that their students would become successful communicators in the target language? Since in most language classes there is only limited time one can devote to pronunciation, it is necessary to try to focus on the most important aspects of pronunciation. But what are these?

Although with the rise of the communicative language teaching not much attention has been given to pronunciation teaching and there have even been some doubts whether teaching pronunciation can be effective, nowadays not too many people believe that teaching pronunciation has no effect at all and that it is only a waste of time. Also recent research has shown that pronunciation training can be beneficial for the learners. However, as Derwing and Munro (2009) point out, merely knowing that it can be effective is insufficient (p. 482). As suggested in the previous paragraph, we need to know where to put the focus.

An indispensable part of trying to answer this question is the need to realize what we want the learners to improve. Is it the accentedness that bothers the students, or is it rather that they are not being understood? Or is it that their interlocutors find it difficult to understand

them? Thus, as explained in Section 3.2, accentedness, intelligibility and comprehensibility are three different aspects of foreign speech and need to be kept apart. For most students intelligibility will probably be the most important goal. However, an advanced learner who is intelligible enough might like to work on accentedness to get rid of the possible negative social evaluation by native speakers. Of course, the aim of ideal pronunciation training would be improving all three. However, it seems that what works for accentedness does not need to work for intelligibility and comprehensibility, and so on.

To resolve the question which particular aspects of pronunciation should be focused on, research of pronunciation training needs to be consulted. Although research in this area has grown in the last decade or two, there have not been any strong conclusions reached yet. The reason for this might be that more longitudinal classroom-based research is still needed. Also the effect of pronunciation training on solely foreign accent or on solely intelligibility has often been examined and thus more comprehensive studies which would include all three aspects of pronunciation could be beneficial.

Nevertheless, there have been some interesting preliminary findings which should be taken into account when considering the right focus of pronunciation instruction. With regard to intelligibility, Zielinski (2008) found that English native listeners had problems understanding foreign accented speech if the speech samples included non-standard syllable stress patterns (non-standard number and pattern of strong and week syllables) and non-standard segments in strong syllables. It thus seems that rhythmical properties of accented speech as well as the segmental make-up of strong syllables might contribute to intelligibility in English, at least for native listeners. Another example of a study which examined intelligibility would be Hahn (2004, in Derwing & Munro, 2005) who measured the effect of nuclear stress (sentence stress) manipulation on intelligibility. She found that the speech samples with unmanipulated nuclear stress were much more intelligible to the lay listeners than the manipulated ones. It thus seems that it might be predominantly suprasegmental categories which are important for being understood, but a role may also be played by certain segments. In this respect, Munro & Derwing (2006, in Derwing & Munro, 2009) propose that functional load of segmentals seems to affect comprehensibility ratings and that they should thus also be considered in future research. If pronunciation teaching is to focus also on individual segments, it is possible that these should be the ones carrying a high functional load. This, however, as Munro & Derwing note, is still a very tentative conclusion.

One of the very few studies that examined the effect of pronunciation training on more variables, namely comprehensibility, accentedness and fluency, was Derwing, Munro,

& Wiebe (1998). In their longitudinal study of ESL [1] learners, two types of instruction (focus on segmental accuracy vs. global speaking habits with emphasis on prosodic factors) were examined in order to find out whether they have any influence on the above-mentioned variables. There was also a control group which received no specific pronunciation instruction. After a 12-week course both groups that were instructed in pronunciation showed improvement in both comprehensibility and accentedness in read sentences. With regard to the sentences, the segmental group's improvement on accentedness was significantly greater than that of the other two groups. However, only the global group showed improvement in comprehensibility and fluency in an extemporaneously produced narrative. Nevertheless, Derwing and his colleagues do not dismiss segmental instruction completely. They stress that ESL students can benefit from both global and segmental instruction.

Similar results, but with regards to accentedness, were found in Missaglia's study of native Italian learners of German (1999, in Piske et al., 2001). Those students who received prosody-centered training improved their pronunciation of German significantly more than the ones who received segment-centered training. What is interesting is that the prosody-centered group improved in both suprasegmental and segmental production. Thus although all the findings are still quite preliminary and much more research is needed, the following conclusion by Derwing & Munro (2005) could be supported. They write:

> Evidence is accumulating that what's important are the macroscopic things, including general speaking habits, volume, stress, rhythm, syllable structure and segmentals with a high functional load.
>
> Derwing & Munro (2005, in Derwing & Munro, 2009)

Time and more research will show whether these conclusions were in the right direction. One important note must be made here, however. All findings reported in this section (with the exception of Missaglia (1999)) concerned English pronunciation. It is possible that the results for other languages might be relatively disparate, especially for the ones which significantly differ from English. Consequently, these conclusions, though tentative, should not be generalized to teaching of other languages. Future research might show that the preliminary conclusions hinted at above might well apply also to learning of other languages, but at this point such generalization is premature.

[1] ESL is an acronym for English as a second language

3 ENGLISH AROUND THE WORLD

English has become a part of people's lives in many parts of the world today and its influence is growing every day. There are many supporters of this development as well as its opponents; however, for most people it has stopped being a matter of preference and it is rather a matter of necessity. English has become the number one 'global language' or a 'global lingua franca', be it for international communication in business, education, science, or just staying in touch with one's friends in another part of the world. As Crystal writes in his *English as a Global Language*, "There has never been a language so widely spread or spoken by so many people as English (Crystal, 2003, p. 189)."

The importance of English is thus rising. Many non-native English speakers want to speak it in order to be part of this global English-speaking world. However, what English are they supposed to learn? Nowadays, there is a considerable plurality in the field of Englishes. There is not a single idea of English as there used to be for most people still at the beginning of the twentieth century; rather the plural terms "the Englishes", "world Englishes" and "the English languages" are in use at present (McArthur, 2006, p. 361). Many varieties of English are distinguished today. They differ according to where they are spoken, by whom, to what purposes and in what circumstances.

McArthur distinguishes five terminological areas of English varieties:
(1) based on their "geographical location" (e.g. British English, New York English, Hong Kong English), (2) on "linguistic and ethnic association" (e.g. Bengali English, Chinese English), (3) on "activities such as commerce, technology, education, culture and social life" (e.g. legal English, medical English), (4) on "combinations of location and activity" (e.g. American legal English, Australian Standard English) and last, but not least, (5) "fusions of English with the names of other languages" (e.g. Frenglish, Chinglish, or Japlish) (McArthur, 2006, pp. 362-363). It is the last group of Englishes that is the concern of the present study. More specifically, a variety called Czech English, sometimes dubbed Czenglish in accordance with the hybrid names above, and even more specifically – only its sound.

3.1 The case of Czech English

Crystal (2006) points out, however, that labels such as Spanglish, or Czenglish in our case, should be used with caution because they have often been used to describe many different language situations and because they often have stereotypical connotations. In this

study, the terms Czech English or Czenglish will be used in the same way as in Králová's thesis *Czenglish: A Basic Outline of an EFL Variety* (2010). Her definition of *Czenglish* is the following: "an English-based interlanguage used by Czech speakers containing linguistic features transferred from Czech" (Králová, 2010, p. 11) – for the term "interlanguage" go back to Section 2.1. Czenglish approximates the target language, English, to a varying degree depending on many factors including the learner's linguistic abilities, age of learning and the amount and quality of instruction.

An outline of Czenglish as a variety of English is given in Králová's (2010) above-mentioned thesis. She attempts to give a succinct account of this interlanguage on all levels of linguistic description giving some examples of encountered Czenglish. Literature which deals with this variety, for example Sparling's *English or Czenglish* (1989), is also discussed there. It is, however, by no means the aim of the present study to compete with her general description. This study will describe only one aspect of this variety, namely the phonetic level, more specifically it will focus on temporal properties of read Czech English.

3.1.1 Czech accent in English

Step by step, research in the field of accentedness of Czech English has tried to uncover the nature of Czech accent. Although much still remains to be examined or confirmed, the most important findings about the phonetic level of Czech English will be summarized in this section.

What seems to be quite unambiguous is the possible salience of the Czech accent. In Skarnitzl, Volín, & Drenková (2005), two groups of listeners - ten native speakers of English and ten Czech proficient speakers of English - evaluated the degree of accentedness of Czech English extracts on a 5-point scale (see Section 2.2 for more information on accentedness evaluation). It has been found that there was a considerable inter-rater agreement between both groups. In a study examining the strength of Czech accent under adverse listening conditions, Volín & Skarnitzl (2010b) point out that strong accent can be very salient even in degraded signal (which resembles real life conditions), but that in certain kind of noise the differences in the degree of accentedness diminish. They found that in such conditions listeners are not able to distinguish the speakers with just a slight foreign accent from heavily-accented speakers to the same extent as in a 'clean' signal conditions.

Yet it is much less clear what exactly constitutes the nature of a strong Czech accent. As noted in Section 3.1, it is likely that L1 interference plays a relatively large role in the

Czech accent in English. Several suggestions have been made with regard to possible candidate phenomena that might influence the salience of Czech accent. The proposed features of Czech English have been the following: rhythmic differences as compared to native speakers of English, general lack of durational reduction in polysyllabic words, differences in duration and prominence of un/stressed syllables, different intonation patterns and several types of segmental errors.

There seems to be a rhythmic interference of Czech in Czech English. This hypothesis has been supported by means of PVI measurements in Volín & Skarnitzl (2010a) (for PVI see Section 4.7.3). The same authors have also found that less variation in the PVI of vocalic intervals means more accented speech and vice versa (Volín & Skarnitzl, 2010b, p. 1019). In a similar vein, Volín (2005) has found a general lack of durational reduction in polysyllabic words in Czech English. Yet he points out that the specific rhythmic patterns of individual words are affected by not only the speakers' L1 but also by factors from the target language (p. 291).

Connected to the rhythmic differences between Czech and British English are also the findings of Volín & Poesová (2008). They state that in their Czech English data the stressed syllables were either shorter than or of the same duration as stressed syllables in the British English data while the opposite held for unstressed syllables (Volín & Poesová, 2008, p. 26). They say that this again might be an example of interference from Czech because in Czech there is a "moderate tendency towards shortening stressed vowels", as reported by Janota & Palková, (1974, in Volín & Poesová, 2008, p. 26). All these studies basically report the same feature of Czech-accented English – 'lack' of rhythmicity or rather divergent rhythmical properties from native English, which might cause difficulties to the listener (see Section 4.7.1).

Also the sound pressure level (SPL) can reveal something about the issue of stress and rhythm. By measuring the SPL, Volín & Skarnitzl (2010b) confirmed the observation that "Czech speakers of English are often unable to convey the proper prominence of stressed syllables and suppress the prominence of unstressed syllables in terms of their relative loudness" (p. 1012). Duration together with SPL might be then quite good indicators of the divergent rhythm of Czech English.

Last suprasegmental phenomenon to be mentioned is intonation as represented by the F0 contour. It has been found that the contours of Czech English are "smaller" or "less extreme" than those of native English speakers, which would account for the alleged remarks of native speakers of English that Czech English often sounds "disinterested" or "bored"

(Volín & Skarnitzl, 2010b, p. 1012). A relatively surprising finding of the same study was that in less accented Czech English there seem to be greater downtrends.

In respect of segmental errors, there are also several which are reported to be typical of Czech English. There appears to be a lack of not only durational, but also spectral reduction in Czech English weak syllables (Volín & Poesová, 2008). What is typical not only for Czech English is the difficulty of Czech speakers to pronounce the English interdental fricatives. Also the velar nasal causes problems and the '-ing' ending is often pronounced with a [k] or [g] sound at the end (see e.g. Volín, 2000). Czech speakers further experience difficulties with the English open front vowel /æ/ which is not part of their native vowel system. The Czech English pronunciation of this vowel tends to be more closed and fronted towards the Czech /e/ (Šturm, 2011, p. 33). There are definitely many more segmental errors which occur with Czech speakers, but the ones mentioned above are probably the most conspicuous ones. A question is how much these errors contribute to lowered intelligibility of Czech accented English.

To the best of the author's knowledge, no study of Czech English has attempted to measure the relative contributions of the individual proposed accentedness phenomena on intelligibility and comprehensibility yet. However, for preliminary analysis with regard to accentedness, see Volín & Skarnitzl (2010a, 2010b). To conclude, although several candidates for factors influencing the degree of accent in Czech English have been proposed, these need to be confirmed in future research. Also the relative contributions of these factors towards accentedness, intelligibility and comprehensibility need to be ascertained in order to find an appropriate focus of pronunciation instruction in the Czech context.

4 TIME DOMAIN OF SPEECH

Like any other human activity, speech takes place in the domain of time. When we speak, we produce utterances, words and speech sounds along the time axis and consequently, any speech segment can be measured for duration. Segmental duration seems to play a significant role in speech rhythm and the temporal plan of speech in general. It is duration of these smallest speech units which builds up into durations of larger units – syllables, feet, intonation phrases and utterances. Consequently, the duration of even such small units as segments can influence a lot in the communication process. Since the topic of the present paper is concerned with durational properties of Czech English and British English segments, some theoretical background about the time domain is needed at this point.

In the history of phonetics there has been relatively extensive research devoted to the topic of duration, especially since the 1970s. This research has been motivated by various reasons, ranging from purely scientific ones, when investigators have tried to uncover the reality of linguistic units, to applications oriented research with the speech technologies in the foreground (Noteboom, 1997, pp. 664-665). Whatever the motivation, the research up to date has brought many interesting findings on duration which will be summarized below.

The topics of the next sections are segmental duration in English and Czech (4.2 and 4.3), factors influencing segmental duration (4.5), the use of durational cues in speech perception (4.6), and finally the issue of speech rhythm (4.7). However, before proceeding to the topic of duration as such, the relationship between duration and length needs to be clarified.

4.1 A short note on phonological length

Duration, a physical, more specifically acoustic characteristics, should not be confounded with *length*, which is considered to be a phonological feature. Phonological *length* has to do with the language system of individual languages. Some languages distinguish phonologically long and short vowels (e.g. Czech) and some languages distinguish long and short consonants (e.g. Icelandic) even though rarely certain languages are said to distinguish even three degrees of length. However, since the subject of the present study does not concern itself with the question of phonological length, this topic will not be gone into detail at this point. For more information on length across languages see e.g. Laddefoged &

Maddieson (1996) or Duběda (2005), and for information on the specific languages' phonological inventories see e.g. Campbell (1995).

In the case of the two languages which play a part in the present study of Czech English, namely Czech and English, there are both similarities and differences regarding length. What the two languages have in common is that both have long and short vowels but no short and long consonants. However, Czech differs from English in the conception of vowel length. With the exception of the /ɪ/x /iː/ contrast (see Podlipský, Volín, & Skarnitzl (2009) for discussion), the Czech corresponding long and short vowels differ in their length only. On the other hand, in English the difference is not only in their length but also in the spectral quality of the respective vowels, which is sometimes said to be of larger perceptual importance for native speakers of English. Consequently, length in English is not a truly a phonological feature as such since it is not the only distinguishing feature of the corresponding long and short vowels (with the exception of [ɜː] and [ə]). Therefore, instead of long and short, the English monophthongs are often labelled as tense and lax, which is said to describe the difference more appropriately.

Although phonologically long vowels or consonants are reported to have in general larger mean durations than their short counterparts, it is by no means a rule that all single realizations of phonologically long segments would be acoustically longer than the phonologically short ones. Thus the relationship between phonological *length* and actual segmental *duration* is not straightforward and is most probably language-specific. However, as noted above, it is not *length* but *duration* which is the focus of the present study, so let us proceed to the topic of *segmental duration*.

4.2 Segmental duration in English

As a starting point to the topic of segmental durations, the phonological systems of English and Czech will be shortly described together with durational values for both languages as measured in previous research. Since the British English and Czech English accents will be examined in this study, a general idea of both systems is needed because some interference of the Czech durational system has been hypothesized in Czech English (as noted in Section 3.1.1). First, English durations will be covered followed by Czech durations in the next section (4.3).

The British English vowel system is differentiated into monophthongs and diphthongs. The monophthongs include both short (lax) vowels /ɪ ɛ æ ʌ ə ɒ ʊ/ and long (tense)

vowels /iː ɜː ɑː ɔː uː/, at least according to the traditional account of the English vowel system.

A point worth mentioning is that the vowel /æ/ tends to be longer than the other short vowels and some researchers thus include it among the long vowels even though it is not marked by the length symbolising colon mark ː as the other ones are. And finally, to conclude the list of vowels, the British English diphthongs include the following: /eɪ aɪ ɔɪ aʊ əʊ ɪə eə ʊə/.

The British English consonant system includes plosives (sometimes also called stops) /p b t d k g/, fricatives /f v θ ð s z ʃ ʒ h/, affricates /tʃ dʒ/, nasals /m n ŋ/, and approximants /r l j w/. The approximants, are sometimes further subdivided into liquids /l r/ and glides /j w/.

To have a more detailed idea of English segmental durational characteristics, two sources have been chosen – van Santen (1992) and Crystal and House (1988). Although both studies include useful overviews of their measurements, (which will be cited below), both make potential comparisons with our data difficult. The reason is that not British English, but American English was analysed in both studies. Consequently, these measurements are included in the present study only in order to have some idea of the temporal properties of English phones and classes of phones, but any comparisons must be very careful since two different varieties of English are concerned.

Van Santen's (1992) measurements of American English vowels can be seen in Table 1. Raw (uncorrected) mean durations of vowels under different stress conditions are given for vowels in utterance-medial position. The number in the parentheses after each durational value is a half of the 95% confidence interval. For example, the 95% confidence interval for unstressed /ɛ/ is 77-101 ms. Since van Santen (1992) uses slightly different phonetic symbols, the second column with the usual IPA transcription was added by the author of the present study in order to make the table more comprehensible. The vowel /æ/ is subsumed under long vowels by van Santen, and since American English, and not British English, is concerned, the monophthong /ɒ/ and the centering diphthongs /ɪə eə ʊə/ are not present. The data were obtained from sentences read in isolation by a male American speaker. There were a total of 18 046 vowel segments analysed.

Segment – original	*Segment - IPA*	*Unstressed*	*Primary stress*	*Secondary stress*
ə	ə	68 (2)	-	-
ɪ	ɪ	74 (2)	90 (2)	96 (10)
ʊ	ʊ	87 (28)	104 (4)	92 (8)
ʌ	ʌ	77 (26)	116 (6)	123 (23)
ɛ	e	89 (12)	118 (4)	110 (18)
i	iː	101 (4)	137 (4)	121 (14)
u	uː	105 (12)	136 (8)	-
ə~	ɜː	98 (2)	154 (6)	140 (22)
e^{j}	eɪ	139 (14)	162 (4)	147 (12)
o	oʊ	142 (18)	162 (6)	145 (8)
ɑ	ɑː	151 (22)	176 (6)	168 (8)
a^{j}	aɪ	172 (18)	171 (4)	186 (22)
æ	æ	148 (20)	173 (4)	161 (16)
ɔ	ɔː	-	189 (6)	-
a^{w}	aʊ	-	203 (12)	-
$ɔ^{j}$	ɔɪ	-	222 (24)	-

Table 1 Raw vowel durations (with 95% confidence interval halves in parentheses), in ms. Vowels were restricted to utterance-medial positions in accented words. The second column was added by the author of the present study. (adapted from van Santen, 1992, p. 523)

When exploring the values in Table 1, it can be seen that the ratio of the longest vowel to the shortest one within the primary stressed conditions is slightly larger than 2:1 (van Santen, 1992, p. 523).

While van Santen (1992) concentrated on vowel durations only, Crystal and House (1988) analysed both vowels and consonants. Their measurements can be seen in Table 2. They divided their results into separate speech-sound categories, similar to the ones reported above. However, again, especially with regard to the vowel categories, we must bear in mind that the variety analysed in their study was American and not British English. The numbers in parentheses after each category name signify the number of types followed by the number of tokens for each category. Crystal and House (1988) analysed connected read speech of six speakers, out of whom three were chosen as fast speakers and three as slower speakers based on the total reading time (for details see their earlier paper, Crystal & House, 1982). In Table 2, the mean durations of these two tempo groups are given in milliseconds for each speech-segment category, as well as the overall mean durations for all six speakers. On the whole, the data consisted of about 3600 words and approximately 10 300 measured speech segments.

	Talkers					
	Slow		**Fast**		**All**	
Category	Mn	s.d.	Mn	s.d.	Mn	s.d.
All phones (47, 10 303)	88	52	76	43	82	48
Vowels (18, 3850)	108	65	95	52	102	59
Monophthongs (15, 3480)	98	58	86	47	92	53
Long vowels (7, 1328)	140	58	119	47	129	54
Short vowels (4, 1446)	75	36	68	32	71	34
Diphthongs (3, 369)	188	70	163	50	175	63
Consonants (29, 6453)	75	37	64	30	70	34
All stops (6, 1891)	76	36	66	30	71	33
Complete stops (6, 1119)	89	34	81	27	85	31
Voiceless (3, 744)	97	35	87	27	92	32
Voiced (3, 375)	74	27	70	23	72	25
Hold-only stops (6, 705)	57	27	49	22	53	24
Voiceless (3, 366)	60	28	53	23	56	26
Voiced (3, 339)	53	24	46	20	49	22
Flapped *t* (1, 161)	33	11	26	9	49	11
Fricatives (7, 1833)	80	42	68	33	74	39
Voiceless (4, 925)	107	36	87	31	97	35
Voiced (3, 908)	53	29	48	22	50	26
Affricates (2, 134)	123	42	105	31	114	38
Complete (2, 122)	129	39	109	29	118	36
Voiceless (1, 77)	136	41	116	27	126	36
Voiced (1, 45)	115	32	96	27	105	31
Nonvocalic sonorants (7, 2174)	72	31	61	24	67	28
Nasals (3, 1030)	76	33	60	23	68	29
Liquids (2, 726)	72	29	64	24	68	27
Glides (3, 421)	63	29	58	24	60	27

Table 2 Mean durations (Mn) and standard derivations (s.d.) of speech-sound categories. All values in milliseconds. Six talkers (three SLOW, three FAST); two complete scripts. Numbers in parentheses indicate total *types* followed by total *tokens* for category. (from Crystal & House, 1988, p. 1555)

Looking at Table 2, it may be concluded that the average durations of the slow speakers for each category are larger than in the group of the fast speakers. Out of the consonant categories, the affricates are the longest, followed by fricatives and stops, which are about the same, and not far behind are the nonvocalic sonorants. However, since there are not too many affricates in the sample (only 134 tokens), the data with regard to the affricates might not be quite representative. Also it can be seen that the voiceless consonants are always longer than their voiced counterparts, which seems to be a general tendency in all languages. With regard to vowels, Crystal & House's long vowels are almost twice as long as the short vowels and diphthongs are the longest of all, outnumbering the long vowels by about 40 ms.

In order to compare the values for vowels as given by van Santen and by Crystal & House, van Santen's mean values of individual vowels were converted into mean values for short monophthongs, long monophthongs and diphthongs. In line with van Santen's view, the vowel /æ/ was counted among long vowels in these results. All values are given Table 3. Like in Table 2, Crystal and House's values are given for the two groups of speakers – slow and fast – and the mean values for all speakers. Reflecting Table 1, the values of van Santen are given for different stress conditions – unstressed, primary stress, secondary stress and a total mean for all the conditions was counted as well. The mean values counted for all stress conditions approximate Crystal and House's slow speaker values the most, only with the exception of diphthongs which are slightly shorter in van Santen's data. This is not too surprising because van Santen's speaker's pronunciation was allegedly very careful (1992, p. 515) and thus we may suppose it was also relatively slow.

	Crystal & House			**van Santen**			
	Slow	Fast	**All speakers**	Unstres.	Primary	Second.	**All conditions**
Short vowels	75	68	**71**	79	107	105	**96**
Long vowels	140	119	**129**	120	160	148	**144**
Diphthongs	188	163	**175**	151	184	160	**168**

Table 3 Comparison of the mean values as given by Crystal & House (1988) and means counted from van Santen's (1992) individual vowels. Crystal & House's values are of slow, fast, and all speakers. Van Santen's values are of different stress conditions – unstressed, primary stress, secondary stress, and all stress conditions. (adapted from Crystal & House (1988) and van Santen (1992))

A question may be asked whether the data of the present study is going to approximate the values of van Santen and Crystal & House as given in Tables 1, 2 and 3, but it may be hypothesized that the individual phonetic categories should probably have similar relations to each other.

4.3 Segmental duration in Czech

To have some comparison of English and Czech, the Czech phonological system will now be shortly discussed along with some approximate durational values. The Czech vowels include short vowels /ɪ ɛ a ɔ u/, long vowels /i: ɛ: a: (ɔ:) u:/ and diphthongs /(eu) (au) ɔu/. Segments /ɔ:/, /eu/, and /au/ are given in parentheses because they entered the Czech vowel system in loan words and are not perceived as completely domestic (Palková, 997, p. 170). The Czech short and long vowels are reported to differ in the

respective pairs in their duration only, with one exception – the /ɪ/-/iː/pair. Podlipský, et al. (2009) report that in the /ɪ/- /iː/ contrast both spectral and durational cues are utilized. The Czech consonant system includes plosives /p b t d c ɟ k g/, fricatives /f v s z ʃ ʒ x ɦ/, affricates /ʦ ʧ/, nasals /m n ɲ/, approximants /j l/, and two types of trills /r r̝/.

With regard to the Czech durational values, Palková (1997) will be cited as a source here because she gives overviews of measurements of both Czech vowels and consonants found in studies up to 1997. We can see the durational data of Czech vowels given in milliseconds in Table 4. Palková reports the measurements of Chlumský (1928), Kaiserová & Janota (1964) (= Janota), Mluvnice češtiny (1986) (= MČ) and Borovičková & Maláč (1967) (= B-M). Chlumský gives both mean durations and the range, Janota and B-M give the mean durations, and MČ gives solely the range for all short vowels and for all long vowels. As for the data types, Palková claims that Chlumský used connected speech, as probably MČ did, Janota used isolated words and B-M worked with nonsense words (Palková, 1997, p. 179) Thus we should bear these different conditions in mind when comparing and contrasting these data.

Segment	**Authors**				
	Chlumský		***Janota***	***MČ***	***B-M***
	Range	**Mn**	**Mn**	**Range**	**Mn**
ɪ	50-100	80	100	40-160	182
iː	140-200	170	200	for all	-
ɛ	60-120	90	120	short	184
ɛː	160-230	190	210	vowels	-
a	90-160	120	120		185
aː	190-300	240	240	80-320	-
ɔ	70-130	100	120	for all	185
ɔː	160-250	200	200	long	-
u	60-120	90	100	vowels	186
uː	120-240	180	-		-

Table 4 Mean durations and durational range of Czech vowels in ms. Data from 4 sources are cited: Chlumský (1928), Kaiserová & Janota (1964) (= Janota), Mluvnice češtiny (1986) (= MČ) and Borovičková & Maláč (1967) (= B-M). (adapted from Palková, 1997, p. 179)

Palková further claims that traditionally Czech long vowels should be around twice as long as their short counterparts. However, more recent measurements have shown that this ratio is smaller, at least nowadays. Podlipský et al. (2009) report that the smallest ratio is for the /ɪ/-/iː/pair – the long vowel is only 1.29 times longer than the short one, which is

probably also due to their spectral differentiation. They further report that in the case of the other Czech vowel pairs the long vowels are 1.60 to 1.79 longer than their short counterparts. So especially for the /ɪ/- /iː/pair, the difference seems to be much smaller today than in the data reported in Table 4.

Further, Chlumský's mean duration would agree with the measurements of Skarnitzl and Machač (2007) who claim that the high vowels are the shortest and low vowels are the longest. What is more, this finding is in accordance with the hypothesis about inherent phonological duration of vowels reported below in Section 4.5.5.

Table 5 reports the mean durations of Czech consonants as given by Palková (1997). This time three original sources are reported: MČ, B-M and Chlumský (1911, 1928) as cited in Hála (1962) (= Ch-H). Again all measurements are in ms. For the fricatives, trills and approximants, MČ and B-M give their data for the intervocalic and initial positions separately; however, for the plosives and affricates, they do not distinguish between the word positions. Also Ch-H gives only one mean value for all the positions for each of the consonantal segments. As for the data types, Palková does not specify these this time, but we may hypothesize that they are the same or similar as for the vowel data of the respective authors.

Segment	Authors			Segment	Authors				
	MČ	*B-M*	*Ch-H*		*MČ*		*B-M*		*Ch-H*
					Intervoc.	Initial	Intervoc.	Initial	
p	200	211	180						
b	150	143	120	f	240	110	240	109	200
m	150	137	120	v	120	110	123	124	100
t	220	216	160	s	260	180	259	176	230
d	130	121	110	z	170	180	172	177	120
n	140	130	100	ʃ	240	180	235	182	220
c	200	213	170	ʒ	160	170	158	166	110
ɟ	130	154	120	x	240	160	239	161	200
ɲ	140	134	110	ɦ	140	120	135	123	110
k	210	207	190	r̝	170	160	158	167	120
g	140	142	140	r	100	120	85	125	30-70
ts	210	269	220	l	100	120	95	122	70
tʃ	220	250	210	j	110	120	109	128	90

Table 5 Mean durations of Czech consonants in ms. Durations in initial and intervocalic positions are distinguished by MČ and B-M for some segments. If not specified, mean values for all positions in a word are given. Data from 3 sources are cited: Mluvnice češtiny (1986) (= MČ), Borovičková & Maláč (1967) (= B-M), and Chlumský as cited in Hála (1962) (= Ch-H). (adapted from Palková, p. 221)

If we look at Table 5, we may draw some tentative conclusions about the separate consonant categories. As in English, the voiceless consonants are longer than their voiced

counterparts; however, with fricatives this difference is to be seen in the initial position only. The affricates are not the longest consonants in Czech although they definitely belong to the longer ones. It seems that the fricatives are longer than plosives although this difference might hold only for voiceless fricatives and maybe only for the initial position. Also the data shows that the sonorants /r l j/ are the shortest among the consonant categories (Palková, 1997, p. 222). If we compare these Czech consonantal data to the American English ones reported above, many relations seem similar, but definitely not all of them. Moreover, we do not know how representative the data are because we do not have much information about them.

4.4 Multifunctional nature of segmental duration

After the English and Czech systems and their durations were recapitulated, the present study will progress to more general topics with regard to the temporal domain. In the following two chapters, factors influencing segmental duration and the utility of durational cues in speech perception will be discussed. The topic of time domain of speech will be concluded by the issue of rhythm.

Although the durational properties of languages other than English have been investigated as well, the majority of durational research has focused on English, which might be connected to the dominance of English touched upon in Section 3, or also to the greater funding of research in English-speaking countries. One of the early articles which deal with duration in English is Klatt's (1976) well-known study. He focuses on two main areas. These two areas are aptly summarized by Nooteboom who stresses that the biggest difficulty with speech sound durations is their multifunctional nature because they are "*affected by* a great many very divergent factors in production, and *affect* a great many very divergent perceived aspects of speech" (Nooteboom, 1997, p. 660; emphasis added). Thus the areas of production and perception – or, in other words, what durations are influenced by and what they have influence on - are the two areas Klatt distinguishes in his study and they are going to be covered in the next sections of the present paper as well.

4.5 Factors influencing segmental duration

Klatt points out that "speech timing is specified or modified at many different levels in the sentence generation process" (1976, p. 1209). This means the duration of words as well as individual speech segments is the outcome of interplay of many factors in the utterance

production process. According to Klatt, these include extralinguistic, discourse level, semantic, syntactic, word-level, phonological or phonetic, as well as physiological factors (1976). For an overview of these factors and their examples based on research findings, see Table 6.

Extralinguistic	Psychological and physical state	*(Williams & Stevens, 1972)*
	Speaking rate	*(Huggins, 1964; Goldman-Eisler, 1968)*
Discourse level	Position within a paragraph	*(Lehiste, 1975b)*
Semantic	Emphasis and semantic novelty	*(Coker et al, 1973)*
Syntactic	Phrase-structure lengthening	*(Martin, 1973; Klatt, 1975)*
Word level	Word-final lengthening	*(Lehiste, 1972; Oller, 1973)*
Phonological/Phonetic	Inherent phonological duration for a segment	*(Peterson & Lehiste, 1960)*
	Effect of linguistic stress	*(Parmenter & Trevino, 1936)*
	Effect of postvocalic consonant	*(House & Fairbanks, 1953)*
	Segmental interactions, e.g. consonant clusters	*(Klatt, 1973b; Haggard, 1973)*
Physiological	Incompressibility	*(Klatt, 1973a)*

Table 6 Factors that influence the durational structure of a sentence (from Klatt, 1976, p. 1210)

Although it would be certainly interesting and beneficial to examine all the factors presented in Table 6 in the present study, it was simply beyond its scope. Consequently, only three of the factors introduced below were taken into account when examining the durational variation of speech segments in Czech and British English: (1) lexical stress, (2) intonation phrase boundary, and (3) articulation rate.

The effect of lexical stress was explored with regard to vowel duration and boundary was examined with regard to both vowels and consonants. Although the effect of articulation rate as such was not analysed, it was also partially taken into account because a normalization method was applied to the present study data to remove the interspeaker articulation rate differences.

At present, all important factors mentioned by Klatt will be shortly commented upon in the following sections in order to obtain a comprehensive idea of the extent of possible durational variation.

4.5.1 Speaking rate

Klatt subsumes speaking rate under extralinguistic factors. However, the word *extralinguistic* is not used nowadays any more. Maybe *paralinguistic* would be a more suitable word in the context of affective states (Klatt also includes a psychological and physical state in this category), but speaking rate belong really belong to this groups of

factors. Affective states can influence speaking rate but speaking rate as such should rather be included under prosodic factors. Be it as it may, speaking rate and its changes can have many linguistic implications. One of them is the interaction with duration.

Speaking rate has been found to influence the durational patterns of a sentence in rather complex ways. For instance, when speaking rate increases, higher occurrence of phonological and phonetic simplifications has been found as well as differential shortening of vowels and consonants (Klatt, 1976). Lehiste claims that speech tempo affects unstressed syllables more than stressed ones, at least in English (Lehiste, 1970, in Nooteboom, 1997). However, it seems that the influence of speaking rate on durational phenomena is language-specific and that different languages show different effects of speaking rate manipulation (Vaane, 1980, in Crystal & House, 1982).

4.5.2 Emphasis, contrastive stress and semantic novelty

Semantic factors also play a role in the durational behaviour of a sentence. Emphasis and contrastive stress has been found to lengthen the duration of a word by at least 10-20% (Coker et al., 1973, in Klatt 1976). Similarly, semantically new information exerts an influence on duration. When an unusual word is mentioned for the first time, it is longer than when it is mentioned again later (Umeda, 1975, in Klatt, 1976).

4.5.3 Phrase-structure lengthening

There are also several lengthening phenomena connected to "syntactic" matters according to Klatt. However, again, by *syntactic* factors Klatt means rather factors connected to *prosodic structure*. Prosodic structure can correspond to syntactic structure, but not necessarily, and it is prosodic structure and not syntactic which is vital in speech and consequently also in the temporal domain. With regard to the lengthening caused by prosodic structure, Klatt (1976) calls the principal lengthening effect 'prepausal lengthening' although, as he rightly points out, it occurs not only before silent pauses. Thus more accurately it should probably be called 'pre-boundary lengthening', a term used by van Santen (1992).

The effect of the presence of a pause was also examined in Ondrušková's (2011) study of Czech monosyllabic words. She discovered that if a word preceded a pause its duration tended to be longer than if the word was found after a pause or if no pause was present in the vicinity of the word. Also according to van Santen, a question may be asked which type of

boundaries cause pre-boundary lengthening. The possibilities he suggests are minor phrase boundaries, major phrase boundaries and utterance boundaries (van Santen, 1992, p. 535). Nevertheless, in his experiment he investigates utterance boundaries only and in these he confirms the lengthening effect (van Santen, 1992, pp. 538).

Examples of detected phrase-final lengthening may be, however, found for instance in Martin (1970), Klatt (1975) or Cooper (1975), all cited in Klatt (1976), or Cummins (1999). Wightman and his colleagues also investigated phrase-final lengthening and came to a conclusion that not only that lengthening can be detected in the rhyme of the last syllable before a phrase boundary, but also that four different types of boundaries can be distinguished on its basis (Wightman et al., 1992). The reasons for pre-boundary lengthening are unclear. It is interesting what Klatt points out in this respect:

> It is not known whether a speaker learns to lengthen segments at the end of phrase boundaries in order to help the listener decode the message, or if there is simply a natural tendency to slow down at the ends of all motor sequences or planning units. Since utterance-final lengthening often extends over several syllables, it is probably related to the general deceleration of motor activity at the ends of speaking acts. This is in contrast to the lengthening seen at sentence-internal phrase boundaries which is usually localized to the phrase final syllable. (Klatt, 1976, p. 1212)

If the lengthening was due to perceptual reasons, this would nicely illustrate the interconnectedness of the production and perception viewpoints on segmental duration. Phrase boundaries cause the phrase-final syllables to lengthen and at the same time we lengthen the phrase-final syllables so that the listener would understand us better, which makes it a kind of a hen-or-egg problem.

The present study will examine the influence of an intonation phrase boundary on the final two syllables in order to established the extent of the potential lengthening with regard to the distance from the boundary.

4.5.4 Effect of position within a word

Even at word boundaries there is a slight tendency for segments to be somewhat longer than segments within words (Nooteboom, 1997). It has been found that this occurs even in non-phrase-final positions (Oller, 1973; Klatt, 1975, both in Klatt, 1976). Nevertheless, not all investigators have observed word-final lengthening (Harris and Umeda, 1974, in Klatt, 1976; Turk & Shattuck-Hufnagel, 2000). However small the difference in duration may be, Quené (1989) found that it can help word boundary detection with up to 80

per cent accuracy rate (as cited in Nooteboom, 1997). On the other hand, word-final position does not necessarily have to be the "longest". With consonants the case is somewhat more complicated. They are longest in word-initial position, about 10-30 ms shorter when placed word-finally and even shorter in the middle of a word (Oller, 1973; Klatt, 1974; Umeda, 1975b; all in Klatt 1976). Yet if the word-final consonants are also phrase-final, then they have the longest duration of all positions.

4.5.5 Inherent phonological duration

Out of the phonological and phonetic factors, we must first name the inherent or intrinsic phonological duration. Each segment has its own inherent duration which is given mainly by the amount of articulation effort needed for its production. It is thus not only a phonological factor, but also a physiological one. For example, when we pronounce the vowel /æ/, the jaw has to make a bigger movement than during the production of the /ɪ/ vowel and thus the inherent duration of /æ/ is going to be longer than the duration of /ɪ/. Also Peterson and Lehiste (1960) claim that /ɪ, ɛ, ʌ, ʊ/ are shorter in duration than other vowels in English (as cited in Klatt, 1976), voiceless fricatives tend to be about 40 ms longer than their respective voiced counterparts and small differences may be observed also regarding the place of articulation of consonants: bilabial stops are usually somewhat longer than velars and alveolars (Klatt, 1976).

4.5.6 Effect of lexical stress

Another phonetic factor which has influence on duration is lexical stress. However, stress, to actualize Klatt's grouping again, should be better placed under prosodic factors. In English, stressed vowels have longer durations than unstressed ones, which is a finding that has long been known to the phonetic community (documented also by Parmenter and Treviso, 1936; or Lehiste 1975a, both in Klatt, 1976). Klatt further points out that when the unstressed vowel is schwa /ə/ the duration is even shorter.

However, it is not only vowels that are influenced by stress. The effect of stress has also been observed with consonant durations. Klatt reports studies (Oller, 1973; Klatt, 1974; Umeda, 1975b; all in Klatt, 1976) which found that prestressed consonants had slightly longer durations than other consonants. However, the term "prestressed" is not in use nowadays

anymore. What Klatt means by a “prestressed consonant” is probably a consonant in the onset of a stressed syllable.

Yet the present study will examine the effect of stress only with regard to vowel duration because since they constitute the core of the syllable they are considered to be the main carriers of prominence.

4.5.7 Effect of postvocalic consonant

In languages like English (so-called coda languages), there is a tendency for vowels to be slightly shorter before voiceless consonants than if before their voiced counterparts (Delattre, 1962, in Klatt, 1976). An example of this tendency would be the difference in the duration of the vowel /æ/ in “bat” as opposed to “bad”, or the duration of the diphthong /aɪ/ in “life” /laɪf/ as opposed to the adjective “live” /laɪv/. House and Fairbanks (1953) report quite a significant durational difference of about 50-100 ms in the case of this postvocalic voiceless-voiced distinction in phrase-final syllables in English (as cited in Klatt, 1976). Crystal and House (1988) came to a similar conclusion; however, only in pre-pausal word-final contexts. They have observed no significant shortening before non-prepausal word-final consonants. This is in accordance with the findings reported in Klatt (1976) (e.g. Lisker, 1974; Klatt, 1975) that in non-phrase-final syllables this difference is much smaller, about 10-20 ms only.

Van Santen (1992) also detected the effect of postvocalic voicing, but on top of that he discovered that the manner of articulation mattered as well. For instance, he found that voiced fricatives produced longer durations of the preceding vowel than voiced stops and voiceless fricatives than voiceless stops. According to that, /eɪ/ should have longer duration in “maize” than in “maid” and /ʊ/ should be longer in “push” than in “put”. Another related postvocalic effect reported in past research (Crystal & House, 1988; Chen, 1970, in Crystal & House, 1988; van Santen, 1992) is that this vowel lengthening due to the voicing of a postvocalic consonant functions even across an intervening sonorant, as in “bend” and “bent”. Thus the /e/ in “bent” is going to be shorter than the one in “bend”. Consequently, the context of a vowel matters in English, at least the right-hand one does.

In Czech, however, the situation seems to be different. Machač and Skarnitzl (2007) point out that compensation tendencies in CV sequences are stronger in Czech than in VC sequences, which means that it is not the postvocalic consonant that affects its duration but

the prevocalic one. They found that vowels are shorter after voiceless plosives than after voiced ones, but that a similar tendency with a postvocalic consonant is much weaker. They suggest this might be due to the tautosyllabicity of CV sequences, which is not the case with VC sequences.

4.5.8 Consonant clusters

The behaviour of consonants in clusters is also important from the durational perspective. It has been found that there is a tendency for consonants to be shortened in most clusters (Klatt, 1973a; Haggard, 1973, both in Klatt, 1976; Crystal & House, 1988). Klatt (1976) speculates that this might be due to physical constraints of coordination of different articulators, to some perceptual reasons, or to keep the duration of words nearly constant, the last of which he regards as quite unlikely. Yet something like compensation tendencies might be at work in this case. However, this would need to be subjected to future research.

4.5.9 Incompressibility

The last factor Klatt (1976) mentions regarding the durational profiles of segments is incompressibility, a physiological factor. He claims that stressed vowels have certain minimal duration which is the lowest possible duration for a vowel in order not to be perceived as reduced. Consequently, if there are some shortening effects causing a vowel to compress, the vowel cannot be shortened beyond this minimal duration level (Klatt, 1973b, in Klatt, 1976).

To conclude, we can see that there are a great many factors which can affect segmental durations and most probably not even all have been mentioned. Consequently, the picture of temporal planning is quite complex. There are simply many variables which a researcher who studies durational characteristics has to take into consideration. Nevertheless, the picture gets even more complicated when we realize that there are strong quantitative interactions between these factors (Noteboom, 1997). Many factors interact with each other in complex ways and researchers need to be aware of these interactions.

4.6 Durational cues

As hinted in the previous sections, segmental durations are not random and can serve as a cue to a number of linguistic distinctions (e.g. Quené & van Delft, 2010; Klatt, 1976). Out of these distinctions, let us mention the following:

(1) phonologically long and short vowels

(2) distinction between voiced and voiceless consonants

(3) location of word boundaries

(4) location of phrase boundaries

(5) lexical and phrasal stress

However, before we proceed to the individual linguistic distinctions, we must make a short digression. For a cue (be it a durational or any other cue) to be perceptually relevant, one concept is essential, namely a 'just-noticeable difference'.

4.6.1 Just-noticeable differences in segmental duration

The *just-noticeable difference* (JND), also known as the *difference limen* or the *differential threshold*, is the smallest difference between two stimuli which is observable by human cognition. If the change between the two stimuli is smaller than one JND, then we are not able to recognize the stimuli as different because the change is too small for our cognition to be perceived. Consequently, for a contrast in segmental durations to be perceptually relevant, it must be larger than about one JND, no matter how regular and thus potentially important this difference is (Klatt, 1976, p. 1218).

In the following section, the just-noticeable differences in segmental durations found in previous research are summed up. Huggins (1972a, in Klatt, 1976) reports a JND of about 20 ms for single phonetic segments in an experiment in which he measured how much listeners are able to detect a change in duration of individual segments embedded in a sentence. Burghard (1973 a, 1973b, both in Nooteboom, 1997) comes with a similar finding that filled intervals shorter than 20 ms seem to have no subjective duration at all. Moreover, he also states that intervals shorter than about 40 ms and longer than 250 ms are less accurately perceived than intervals which lie in between these two values.

However, in an experiment with two-syllable words spoken in isolation or in a carrier sentence, Fujisaki et al. (1975, in Klatt 1976) found a JND of 10 ms for a segment of about 100 ms, i.e. of about 10 %. Similar conclusion was reached in two other studies (Ruhm et al.,

1966; Abel, 1972a, both in Nooteboom, 1997) which claim that the JND should be approximately 5 to 15 % for speech sounds with duration in between the mentioned 40 to 250 ms band.

Both Klatt (1976) and Nooteboom (1997) point out a deficiency in these perceptual experiments: the stimuli (be it a segment, word, or sentence) are played over and over again, so the participants might build a fixed temporal reference pattern which they might use in judging the differences. It is thus not quite clear whether the same principles work in connected speech and whether there is a similar kind of JND for segmental duration there although there is some evidence that it might be not too different (Nooteboom and Doodeman, 1980, in Nootemboom, 1997).

Klatt and Cooper (1975, in Klatt, 1976) attempted to overcome this deficiency by randomizing a set of seven sentences in relatively few experimental trials and they found a JND of at least 25 ms. Consequently, the JND in segmental duration probably depends on the type of task in which it is measured. Nevertheless, we may suppose, it would be at least around 10 to 15 % of the duration of the individual segment or maybe more. If the difference in the duration is smaller, the listener will most probably not notice any change and consequently this difference will not serve as a linguistically relevant perceptual cue.

4.6.2 Phonologically long and short vowels

Let us return to the list of potential linguistic distinctions made on the basis of durational cues now. Each of the distinctions will be shortly commented upon in Sections 4.6.2 to 4.6.5. The inherent phonological duration for vowels can be distinguished by durational cues. In 1973 in his synthetic speech study of Dutch long and short vowels (as cited in Klatt, 1976) Nooteboom found out that the durational difference between these two types of segments is perceptually important in their distinguishing. Klatt (1976) mentions a similar unpublished study by Noteboom in which he examined the English “bed” vs. “bad” contrast and concludes that since the difference in duration between the two vowels is about 40 %, duration will probably play a perceptual role in the identification of qualitatively similar vowel pairs. However, Podlipský, et al. (2009) mention studies of Finnish, German, Hungarian and Icelandic (Kirmse et al., 2008; Mády & Reichel, 2007; Pind, 1996) and state that in quantity languages “the role of segmental duration may vary depending on whether short and long categories are distinguished by duration alone or not.” Consequently, although duration often plays a relatively significant role in perception of the short-long vowel

contrasts, vowel quality should not be obliterated when establishing the perceptual importance of vowel duration (cf. Podlipský et al., 2009).

4.6.3 Distinction between voiced and voiceless consonants

It has been found that segmental durations may serve as a cue when distinguishing voiced and voiceless consonants (e.g. Slis & Cohen, 1969, in Quené & van Delft, 2010). A similar finding is reported by Cole & Cooper (1975, in Klatt, 1976) who state that for changing a voiceless fricative into a voiced one, duration is a sufficient perceptual cue. Thus we may see that duration is important in distinguishing not only vowels but also consonants, or at least fricatives.

A connected phenomenon is vowel duration as a cue to the voicing of the following consonant. It has been reported (Denes, 1995; Raphael, 1972, both in Klatt, 1976; Kluender, 1988, in Machač & Skarnitzl, 2007) that the duration of a preceding vowel may serve as a cue in recognition of the voicing feature of a consonant. This is based on the finding reported above in Section 4.5.7 that in many languages vowels are shorter before voiceless consonants than before their voiced counterparts. As can be seen, this can be used also as a perceptual cue. Klatt (1976) notes, however, that it probably serves as a primary perceptual cue only in phrase-final positions where the durational difference is the largest.

4.6.4 Location of word and phrase boundaries

There is also some evidence that segmental durations can contribute to the location of word boundaries in connected speech (Quené, 1992; Shatzman & McQueen, 2006, both in Quené & van Delft, 2010). Furthermore, the lengthening at phrase boundaries seems to be a cue to the distinguishing of syntactic units. Duration as a perceptual cue at phrase boundaries has been reported, for instance, by Van Santen (1994, in Quené & van Delft, 2010). Lehiste has discovered that if the last metric foot of an utterance is not longer than the preceding feet, listeners will perceive the last foot as too short (Lehiste, 1973, in Klatt, 1976). Klatt and Cooper (1975, in Klatt, 1976) report a similar lengthening expectancy at both internal and utterance-final phrase boundaries. Klatt (1976) further proposes that since the lengthening of the phrase-final syllables is quite large, it probably serves as one of the primary cues in decoding the utterance structure. To sum up, it is quite likely that speakers lengthen at phrase

boundaries in order to communicate their message as clearly as possible by enabling the listener to process the information more easily.

4.6.5 Lexical and phrasal stress

Influence of duration on the perception of stress, both lexical and phrasal, has also been hypothesized (Eefting & Nooteboom, 1993, Sluijter & Van Heuven, 1995, both in Quené & van Delft, 2010). In a study of two-syllable noun-verb pairs, it has been found that a change in the relative duration can change the perceived stress pattern and therefore also the word class (Fry, 1958, in Klatt, 1976). It is known that for the perception of English stress, or rather prominence to be more precise, four factors are important: loudness, length, pitch and vowel quality, out of which pitch and length are the most significant ones (Roach, 2000, pp. 94-95). Duration is then not the only perceptual cue for the perception of stress, but together with the other factors, especially pitch, it is one of the means of distinguishing a stressed syllable. Cross-linguistically, the problem of stress is much more complex, however, and the factors contributing to its perception, as well as the extent to which they do, depend on the individual languages.

4.7 Rhythm

Another important area of the time domain of speech connected to segmental duration is that of rhythm. The question of rhythm is also a quite challenging one, maybe even more than the one of stress (although these two areas are interconnected in some respects) and many researchers have tried, often in vain, to penetrate the nature of rhythm. There are still many unanswered questions considering rhythm nowadays, but one thing seems to be relatively clear – its importance.

Let us begin with a very general definition of rhythm. If something repeats regularly, it is said to be rhythmical. Such regularity can be found in speech as well although it is not such a strict regularity as in dance music, for instance. Consequently, the rhythm of speech is described as quasi-regular rather than regular. What it is that repeats itself in speech and makes it rhythmical is a rather complicated matter and we will try to tackle the problem below in Sections 4.7.2 and 4.7.3, but first it needs to be discussed why researchers consider the rhythm of speech so important.

4.7.1 The importance of rhythm

Rhythm of speech seems to be primarily a perceptual phenomenon (Lehiste, 1977). It is thus the listener who is bound to benefit the most from a regular rhythm of speech. It seems that listening to predictable rhythmic configurations in speech is less demanding than listening to arrhythmic ones (Volín, 2010, pp. 297-298). It probably takes less time and less energy to cognitively process such rhythmical speech (e.g. Quené & Port, 2005; Volín, 2005). If the speech is arrhythmic, the listener needs to concentrate on it more, which is, in the long run, also more tiring. It appears thus that the biggest advantage of rhythmical speech is its communicative value because it enables easier speech processing.

4.7.2 Isochrony and rhythmical classification of languages

Yet even if the importance of rhythm in speech is recognized by most researchers nowadays, one thing still seems not to be satisfyingly accounted for – its nature. As has been mentioned, rhythm must be based on some kind of regularity, which has been proposed to be *isochrony*, the regular repetition of some events in time, which was originally thought to be of physical, acoustic origin.

For a long time, languages were classified into *stress-timed languages* and *syllable-timed languages* according to what researchers thought to be the basis of their rhythm. The rhythmic beats in the *stress-timed languages* were supposed to be carried by the stressed syllables and the alternation of stressed and unstressed syllables thus should have contributed to the rhythmic regularity. The individual feet (or interstress intervals) were thus supposed to have equal durations. On the other hand, *syllable-timed languages* were supposed to have the syllable, not the foot, as the smallest isochronous unit and all syllable durations were to be the same. Later, a third grouping was made of *mora-timed languages* having the mora as their basic rhythmic unit.

There are several problems with this account of rhythm, however. First, as pointed out above in Section 4.7.1, rhythm appears to be primarily a matter of perception, not of production (Lehiste, 1977). Consequently, it has proven to be very difficult to find something tangible about the regularity of speech. Researchers have tried to find the places in the speech signal which are equidistant and thus supposedly make speech rhythmical. For instance, in English, an example of a stress-timed language, they have measured distances between the onsets of stressed syllables, between the onsets of stressed vowels, and elsewhere, but without

much success. The key problem is probably, as Volín points out in accordance with Lehiste, that "rhythm is not a property of the acoustic signal but a perceptual phenomenon" (Volín, 2010, p. 298).

Nevertheless, the discovery of the so-called P-centres has given this stream of research some hope. P-centres (perception centres) were discovered in word-manipulation experiments (for a more detailed account see Volín, 2005, p. 280). The participants were asked to adjust the distances between monosyllabic words in a sound editor, so that the resultant row of words would sound rhythmical. It was found that after such a manipulation, certain points (later called P-centres) near the beginning of the vowel onset in each word were distributed equidistantly. Yet the exact position of P-centres depends on a number of factors, e.g. the consonants in the vicinity of the nucleus as well as the quality of the vowel in the nucleus itself. Volín comments on the discovery of P-centres:

> [The] exact algorithm for calculating the position of the P-centre is not simple, but the issue confirms that the perception of rhythm is induced by the interplay of various acoustic properties of the structure, rather than just simple time intervals between the beginnings of the elements.
>
> (Volín, 2005, p. 280)

As can be seen, some points which are regularly spaced in time and which probably make speech rhythmical seem to have been found. However, the issue of rhythm is even more complex.

Second problem with the classification of languages into different groups according to their supposed rhythm is that this classification has never been satisfyingly proven. Kohler points out that the author of the terms "stressed-timed" and "syllable-timed", Kenneth L. Pike, had in mind that "languages contain both types of rhythmical structuring, but differ in favouring more the one or the other", which was later forgotten for some time and twisted into trying to prove that languages belong to either this or that category (Kohler, 2009, p. 30). In contrast, in accordance with Pike, it has been shown that languages form a continuum with stress-timed languages, syllable-timed languages and mora-timed languages as permeable categories, rather than separate, self-contained ones (e.g. Low, Grabe, & Nolan, 2000; Ramus, Nespor, & Mehler, 1999). Nevertheless, even for languages that are supposed to belong predominantly to one category, e.g. English as an example of a stress-timed language or Spanish as a syllable-timed language, the distinction has often been hazy.

4.7.3 Global temporal patterns

After realizing that interstress intervals or syllables durations are an insufficient measure for distinguishing different types of rhythm, measures which would better account for rhythm variability have been searched for. It has been proposed that metrics of global temporal patterns of speech might be the key measure. Two well-known examples of such research would be Ramus et al. (1999) and Low et al. (2000). Both groups of researchers used rhythm metrics "based solely on durational measures of vocalic and consonantal intervals" (Volín & Skarnitzl, 2010b, p. 1012). The former group (Ramus et al., 1999) attempted to distinguish the rhythm types of 8 languages by measuring vocalic and consonantal intervals in an utterance and counted three measures for each language - %V (proportion of vocalic intervals), ΔC (standard deviation of the duration of consonantal intervals) and ΔV (standard deviation of the duration of vocalic intervals). According to them, the combination of the first two measures (%V and ΔC) can successfully distinguish between the individual rhythm types (i.e. stress-timed l., syllable-timed l. and mora-timed l.). To sum up, Ramus and his colleagues (1999) claim that the phonotactical constraints of a language are related in some way to its rhythm type.

In a study of rhythm in British and Singapore English, the latter group of researchers (Low et al., 2000) presented another measure which they claim should cover the cross-language (and cross-varietal) rhythm variability even more precisely – *Pairwise Variability Index* (PVI). This index expresses the mean absolute difference between successive pairs of vowels in an utterance and is normalized for speaking rate. They claim that this measure better reflects the alternation of longer and shorter vowels which is characteristic for stress-timed languages like British English. By the means of PVI and differences in spectral vowel reduction they established a distinction between the British English and Singapore English, the former a stress-timed variety and the latter allegedly a syllable-timed variety. In a later study by the same authors (Grabe & Low, 2002), they applied this index to 18 languages which they plotted on a continuum with the end points of stress-timed and syllable-timed languages (in Kohler, 2009).

Consequently, these above described rhythm metrics have been found useful in the research of prosody and it seems they can be differentiate not only between different languages, but also between different regional accents of one language (White et al., 2007, in Volín & Skarnitzl, 2010b, p. 1012). However, Kohler (2009) claims that neither of these measures (neither the combination of %V and ΔC, nor PVI) are truly global parameters

because they deal solely with local segmental durational distinctions. He proposes that the global variables, which are indispensable for capturing the rhythmic differences, should be "patterns of syllabic timing, of fundamental frequency and of energy, recurrent with some degree of regularity over time" (Kohler, 2009, p. 34). So again like in the analysis of stress above (Section 4.6.5), there seem to be more factors involved and segmental duration is probably only one of them. Nevertheless, there are still many unanswered questions in the area of rhythm research and thus the research community will probably have to wait some more time for a further clarification of the nature of rhythm.

Yet it is not the aim of this study to penetrate the question of rhythm. The goal of this section was to indicate the connection of segmental duration and rhythm and to point out that languages (and most likely also regional accents of the same language) differ in their use of rhythm, or more exactly, in the nature of their rhythm. What is important in connection with the present study is that it has been hypothesized (e.g. Volín, 2005; Volín & Poesová, 2008; Volín & Skarnitzl, 2010b) that the rhythm of Czech English is different from the rhythm of native British English speakers and measurements by means of PVI seem to support this hypothesis (Volín & Skarnitzl, 2010a, p. 277).

It has been proposed that the reason for this is the rhythmic interference of Czech in Czech English. Czech is usually classified more at the syllable-timed end of the rhythmic continuum, whereas British English, as was already pointed out, is considered to be a stress-timed language. Thus, we hypothesise that such differences might be detectable also in the present-study data, especially in the different realisation of stressed and unstressed syllables by Czech English speakers when compared to British speakers.

5 IMPORTANCE OF DURATION IN A FOREIGN LANGUAGE

From the preceding sections, it is more than clear that duration most likely plays a role not only in a mother tongue, but also in a foreign language, and research seems to confirm this assumption. This section will mention several studies which deal with the importance of duration for perception or production of a foreign language, or both. Before proceeding to the studies themselves, it must be stressed, however, that it is never duration alone that serves as a perceptual cue in the various tasks of the below-mentioned studies. There are always more factors at play and this should be kept in mind when reading the following paragraphs.

In a prosody transplantation study, Boula de Mareüil and Vieru-Dimulescu (2006) found out that prosody (by which they mean specifically timing and melody) plays an important role in foreign accent identification, specifically, in the identification of Spanish-accented Italian and Italian-accented Spanish. In Iverson & Evans's (2007) study of vowel identification already mentioned in Section 2.4.8, the subjects from four L1 backgrounds (Spanish, French, German and Norwegian) used durational cues (besides spectral ones) in order to identify English vowels. Similarly, in a cross-language assimilation task with goodness rating and a discrimination task, Lengeris (2009) discovered that his Greek and Japanese listeners were sensitive to durational cues, but that they *temporally* assimilated L2 vowels to L1 categories. However, "temporal information was available in discrimination only when the listeners' L1 duration category/categories did not interfere with the target duration categories" (Lengeris, 2009, p. 169). Lengeris concludes that durational cues do not have any special position in L2 vowel perception when compared to spectral cues (p. 184). It is thus both spectral and durational cues that are important in vowel identification. Yet durational cues should not be overused. Cebrian (2006) reports overreliance of Catalan listeners on duration when distinguishing the English /ɪ/ - /i/ contrast although (or maybe because) Catalan has no temporal contrasts. He also states that this overreliance did not correlate with the amount of experience of the Catalan speakers of English, specifically with the LOR (length of residence) factor. Consequently, durational cues may be important in vowel identification, but they should not be relied on too much in case of some vowel contrasts.

The last two studies reported in this section (Quené & van Delft, 2010; Tajima et al., 1997) will be speech manipulation studies in which duration patterns were mutually transplanted between a NS's and a NNS's versions of speech material. Quené and van Delft's (2010) study was a Speech Reception Threshold study where Polish-accented and native-

Dutch versions of Dutch sentences were manipulated. The study aimed to "to quantify the relative contributions of non-native durational patterns and non-native speech sounds to intelligibility" (Quené & van Delft, 2010, p. 911). In order to do that, differences in speaking rate and pitch contours had to be removed so that they would not act as confounding factors. Intelligibility of the sentences was measured by asking participants to repeat sentences and then evaluating the repetitions as right or wrong. Quené & van Delft report clear effects of both the durational patterns and the speech segments, but the durational effects were found to be relatively small when compared to segmental effects (2010, p. 917). Consequently, they conclude that "the major part of the difference in intelligibility must be ascribed to segmental errors, either phonemic or sub-phonemic" but that attention to both native-like speech sounds and durational patterns might be desirable when learning to speak a target language (p. 917). Moreover, the material of the study was a read speech and, as the authors hypothesize, as for intelligibility in spontaneous conversations, durational patterns might play a larger role because the content is usually more predictable (p. 917).

Tajima et al.'s (1997) experiment was partly similar to the one by Quené & van Delft (2010) but intelligibility was measured by a forced choice identification task and the speech reception threshold was not used. Sentences spoken by a Chinese speaker of English and a native English speaker were also subjected to mutual durational transplantation and the results were very similar to the ones reported by Quené & van Delft:

> Intelligibility of the unmodified Chinese-accented phrases was poor (39% correct), but improved significantly (to 58%) after temporal correction. Performance on the native productions was high (94%), but declined significantly (to 83%) after temporal distortion according to the Chinese speaker's timing. (Tajima et al., 1997, p. 1)

Although the studies by Quené & van Delft (2010) and Tajima et al. (1997) report clear although small durational effects on intelligibility, we must be careful when coming to any more general conclusions because there was only one native and non-native speaker involved in each study - even though the non-native speakers might well have been representative of other non-native speakers. Future research should examine more subjects and more different languages in a similar manner to establish whether these conclusions are valid cross-linguistically.

To conclude this section, it seems that durational patterns play a significant role in production and perception of a foreign language. Although this role might be relatively small when compared to other factors involved – e.g. spectral cues for vowel identification or

segmental make-up in general for intelligibility as such – it is worth to research it in more detail. As pointed out above, durational cues serve as means of distinguishing several features including phonological length, stress and rhythm and might be thus significant for the degree of accent, intelligibility as well as comprehensibility. This, however, needs further research. The present production study is trying to contribute to this durational debate by providing data from read continuous speech by Czech and British speakers of English whose durational patterns will be mapped and where appropriate also compared and contrasted.

6 AIMS OF THE PRESENT STUDY

Before proceeding to the second part of the present study, its aim should be reminded - it is to compare and contrast the variability in the time domain of speech of Czech and British speakers of English. Specifically, it will investigate the variability of segmental durations in both groups of speakers. The study would like to contribute to the debate about the nature of Czech English accent by providing grounds for analyses of its temporal plan. Since studies which have examined the timing in Czech English have concentrated either on its rhythmical properties or on the duration of specific single segments, the present study aims to provide a more complete picture of the time domain of Czech English by clarifying the relationships among mean durations of vocalic and consonantal speech sound classes and to contrast these to the durational patterns in native, British, English.

As noted in Section 4.5, there are many factors that affect segmental duration. Out of these, the influence of two factors will be examined in the present study - first, lexical stress, and second, intonation phrase boundary (the effect of pre-boundary lengthening). Stress will be examined with regard to its influence on vowel duration and boundary with respect to both vowel and consonantal durations. The study will examine the robustness of the influence of these two factors on segmental duration and whether the extent of these effects differs for Czech and British speakers of English.

The study is designed as exploratory in nature and consequently, no specific working hypotheses are going to be tested. Yet a null hypothesis might be coined that there is no difference in the durational profiles of the Czech and British speakers. In case that the null hypothesis is falsified, the study will attempt to point out the most significant differences between the Czech and British speakers which might be then considered for testing in future research.

7 METHOD

7.1 *Subjects*

In order to be able to examine the temporal properties of Czech English, a referential group of native speakers was needed. British English was chosen as our referential native English variety because it is the variety most commonly taught in Europe.

The participants of the present study were 6 women – 3 native British English speakers and 3 Czech speakers of English. All were undergraduate students of a university degree in English and volunteered to take part in the research. All participants were in their early twentieth. None of the subjects had any self-reported hearing or speech impairments.

The British speakers all grew up in southern England and were thus native speakers of a southern British accent (see e.g. Wells, 1982). The Czech speakers were advanced speakers of English. This fact is due to the very demanding entrance examinations to the English studies degree at the Faculty of Arts of the Charles University which is hardly ever passed by students below this level. However, all three Czech speakers were assessed as having a relatively strong Czech English accent. They were selected on the basis of a study of accentedness evaluation (Skarnitzl et al., 2005).

7.2 Recording

The recording of the Czech ad British speakers took place in two different recording conditions. Nevertheless, the quality of the recordings was sufficient for the given purpose in both cases. The Czech speakers were recorded in a sound-attenuated booth of the Institute of Phonetics of the Charles University in Prague with a studio electret microphone IMG ECM 2000 and digitalized at the sampling rate of 32,000 Hz with a 16-bit resolution. The recording of the British speakers took place in a normal, not sound-treated room in Canterbury, UK. The recording was done with a portable high-resolution digital recorder Roland Edirol R-09HR which was set on a tabletop tripod laid on a table. The original 48-kHz sampling rate of the recordings was later resampled to 32 kHz with a 16-bit resolution.

All speakers read a regular BBC news bulletin. The participants had enough time before the recording to get acquainted with the text. They were asked to repeat a sentence if larger dysfluences occurred during the recording, which was, however, unfortunately not

followed by all the speakers. It took the individual speakers from about 3.5 to 4.5 minutes to read the text during the recording.

7.3 Text and its characteristics

All participants read a news bulletin which was originally broadcast in 2002 by the BBC World Service. The news bulletin was divided into 7 paragraphs reporting individual news items. On top of that there were three sentence-long BBC signal calls present in the text – an initial, medial and final one. On the whole, the text consisted of 25 sentences and 478 word-tokens. The British speakers read the text on average in 40 breath groups; whereas, the Czech speakers needed on average 57 breath intakes. The text is enclosed in Appendix 1.

The relative frequency of occurrence of the individual speech segments is given in Table 7. The segments are in order of their relative frequency which is given in percentages.

Overall rank	Speech segment	Relative frequency (%)	Overall rank	Speech segment	Relative frequency (%)
1	ə	12.3	23	aɪ	1.3
2	ɪ	8.9	24	g	1.2
3	t	8.7	24	ŋ	1.2
4	n	7.6	24	w	1.2
5	s	5.8	24	j	1.2
6	l	4.4	28	ʌ	1.1
7	d	4.2	29	ʃ	0.9
8	r	4.0	30	ɒ	0.8
9	z	3.0	30	əʊ	0.8
10	m	2.9	32	ɔː	0.6
11	e	2.8	33	ɜː	0.5
12	k	2.6	33	ɑː	0.5
13	ð	2.5	35	dʒ	0.4
13	p	2.5	35	aʊ	0.4
15	v	2.0	37	eə	0.3
15	f	2.0	37	tʃ	0.3
17	b	1.9	37	ʊ	0.3
18	eɪ	1.8	40	θ	0.2
18	æ	1.8	41	ɪə	0.1
20	iː	1.7	41	ʊə	0.1
20	h	1.7	43	ɔɪ	0.0
22	uː	1.3	43	ʒ	0.0

Table 7 Relative frequency of occurrence of individual speech segments in the text (in %). Relative frequency counted from the canonical number of phonemes.

The textual relative frequency was counted from the canonical number of phonemes. The most frequent speech segment was [ə]. At the other end of the scale of frequencies, not a

single instance of [ʒ] was found in the text and only one instance of the diphthong [ɔɪ] in the read text. Also the following segments were quite rare: [tʃ dʒ au eə ɪə uə u u θ] - there were fewer than 10 occurrences of each of these sounds in the text. Despite these minor deficiencies, the text can be regarded as quite representative with respect to the relative frequencies of the individual speech sounds. For the total number of occurrences of each of these sounds, see Table 9, and for information how these were distributed among the six speakers, see Appendix 3.

7.4 Preparation of the material for analysis

This part will describe the process of the preparation of the recorded material for the subsequent analysis. The main part of this process was a segmentation procedure which, due to the nature of the study, was a significant and relatively demanding part of the whole endeavour. Therefore, the segmentation method will be described in greater depth.

First, the boundaries of individual breath groups were found in each recording with the help of the Praat computer program (Boersma & Weenink, 2011) which was the principal program for the whole material preparation process. The recordings were then cut into the separate breath groups using a Praat script. They were orthographically transcribed into a Praat textgrid tier and then other two tiers were created - one for word segmentation and another one for phoneme segmentation.

7.4.1 Segmentation

Segmentation was done manually by the author of the present study as consistently as possible in accordance with general segmentation criteria. Specifically, the suggested guidelines for boundary placement as given in the segmentation handbook by Machač & Skarnitzl (2009) were followed where possible.

In general, a canonical version of pronunciation was used for the transcription and segmentation of both British English and Czech English material. This was done in order to be able to compare the two varieties. This procedure is based on a presupposition that both groups of speakers have the same (or at least very similar) mental representation of a correct English pronunciation which they are trying to materialize in their speech.

To this end a pronunciation dictionary (Roach et al., 2003) was consulted. The British variants were used for this canonical segmentation allowing for the potential rhoticity of the

Czech speakers. If there were more canonical variants in the British part of the dictionary entry, the one closest to the actual pronunciation of the individual speaker was chosen. Consequently, minor divergences have been created between the transcriptions of the six speakers examined by the present study. All speech sounds based on the canonical transcription were marked in the recordings in the interval phoneme tier. Those phonemes that were not realized, i.e. the ones that were elided, in the real speech sample were marked as only five-millisecond-long stretches of speech and were later excluded from the durational analysis. Also stretches of speech that were significantly altered, contained hesitations or dysfluencies were marked as special cases and discarded from the analysis. Similarly, foreign sounding proper names were not analysed since the Czech and sometimes also the British speakers were not sure about the pronunciation.

As to the segmentation itself, boundaries of a given speech sound were determined in the signal studying a spectrogram and a waveform in the Praat program (Boersma & Weenink, 2011) while listening to the signal in high-quality earphones. General segmentation conventions were used when trying to pinpoint the segment boundaries (see Machač & Skarnitzl, 2009). Among these were, for instance, marking the boundaries of a vocalic onset and offset with the first and last visible full formant structure respectively. However, the ending of a final vowel or sonorant was marked at the elbow in the dynamic envelope of the soundwave and not with the last formant structure. Relatively unequivocal was marking of vowel-nasal or nasal-vowel boundaries because the nasal antiformants were mostly well visible. As the voiceless plosives in English stressed syllables are usually heavily aspirated, the aspiration covers part of the following vowel. The aspiration was not, however, included into the vowel duration and the vowel onset was marked again with the formant structure (see also van Santen, 1992, p. 517). Also the beginnings of initial voiceless and devoiced plosives had to be marked arbitrarily - 40 ms of “silence” before the plosion was marked as the beginning of the hold (closure) phase. All boundaries were placed at nearest zero crossings. Pauses were marked as pauses if they were at least 80 ms long; otherwise, they were split in half and included into the neighbouring segments.

As boundary placement is partially based on arbitrary decisions, instances have occurred in which the best point to place a boundary was not quite straightforward. These problematic points included the segmentation of approximants (especially an [r] sound) from vocalic environment, the segmentation of vowel-vowel sequences, the segmentation of approximants (mainly [r], but sometimes also [l]) after initial voiceless plosives and the

segmentation of [ð] and [ə] in the definite article ‘the’ especially if placed utterance-initially. Each of these problematic cases will be shortly commented upon.

The case of vowel-vowel sequences and the one of vowel-approximant or approximant vowel have much in common. Since sonorant sounds are involved on both sides of the searched-for boundary, the transition between the two sounds may be very gradual and the boundary hard to identify. In such cases both the spectrogram display and the waveform have been studied thoroughly for any discontinuities (i.e. changes that would indicate the transition of one sonorant into the other). If no such point was found, then listening to the signal had to be taken as the principal criterion and the boundary was placed at a point where the two sonorants sounded as separate as possible. The most difficult case proved to be a transition between an [r] sound and a vowel because the rhotic timbre often penetrated deep into the neighbouring vowel. When segmenting the rhotic instances of the Czech speakers, the postvocalic [r] was separated from the preceding vowel if possible. However, some cases of [ɚ] and [ɜ˞ː] had to be included as special cases of allophonic variation and thus were marked by special symbols in cases where the [ə] and [r] or [ɜː] and [r] could not be distinguished as separate sounds.

Another difficult case was when approximants followed an initial voiceless plosive, or more exactly rather a voiceless plosive in a stressed position. Then this voiceless plosive often underwent partial fricativization and hid partially or entirely the following approximant. Generally, such cases were segmented like in the case of vowels, i.e. the onset of the approximant was marked at the point of visible formant structure. Only if the formants of the approximant were visible already during the friction, then this voiceless part of the approximant was included into the duration of the approximant itself. However, such cases were relatively rare.

The last case to be mentioned at this point is the one of the definite article ‘the’ and the segmentation of its two parts. Since the canonical parts [ð] and [ə] were often pronounced simultaneously and thus the features of the fricative and the vowel overlapped, their segmentation proved to be very demanding. Again, the spectrogram, waveform and listening had to be considered and a qualified guess had to be made to place a boundary between the two segments. If the definite article was in the phrase-initial position, [ð] was often partially devoiced, and then a 40-ms stretch of silence was included to the duration of the fricative like in the case of plosives since it has been shown that [ð] sometimes behaves partially like a plosive and also in our data something like a plosion could be seen in some

cases, even in the case of the British speakers. Nevertheless, in the follow-up analysis, the [ð] sound was included among the fricatives.

As can be seen from this section, phonetic segmentation is not always an easy and straightforward task. Rather it is one full of arbitrary decisions which should be, however, in accordance with previous research so that the results were comparable. By following the instructions of Machač & Skarnitzl (2009), the present study hopes to contribute such comparable data.

7.4.2 Stress and intonation phrase boundary assignment

Word boundaries were marked and the words labeled so that in the analysis each segment could be identified with the word it was pronounced in. As one of the aims of the present paper is to analyze the influence of the presence of lexical stress and of intonation phrase boundary, these were also marked in the material in separate tiers.

The stress in the material was marked according to the canonical stress placement in line with the pronunciation dictionary (Roach et al., 2003) with minor allowances regarding the text material. So, for instance, some cases of stress shift had to be taken into account. To mention just two examples, the word 'European' is usually stressed on the third syllable with a secondary stress on the first syllable [ˌjʊərəˈpiːən]. However, if it is in the phrase 'European Court', then the stress shifts to the first syllable and the third syllable loses this property altogether [ˌjʊərəpiːənˈkɔːt]. Consequently, the word 'European' was marked with the stress on the first syllable. Similarly, the word 'international' usually has the main stress on the third syllable [ˌɪntəˈnæʃᵊnᵊl], but in 'International Press Institute' it shifts again to the front and the word 'institute' loses its stress in favour of the stronger one on 'press' [ˌɪntənæʃᵊnᵊlˈpresɪnstɪtjuːt]. Thus the nature of the text had to be taken into account. However, this was not the case with regard to the idiosyncrasies of the individual speakers. If a speaker stressed a word on a different syllable than the canonical one (as was often the case with the Czech speakers), the stress marking in the Praat textgrid was placed still where it canonically belonged. This approach is again supposed to bring the comparability of the British and Czech data sets. It may also help determine whether the possible incorrect stress placement by Czech speakers has an influence on the durational properties of their speech material. The stress marks were placed in the middle of each stressed vowel. Consequently, since vowels are generally considered to be the principal

carriers of stress and to be affected more than consonants, stress was analysed only with regard to vowels and not consonants.

The intonation phrase boundaries were marked according to the actual phrasing of each speaker, and thus, in contrast to the marking of stress they truthfully displayed the speakers' idiosyncrasies. The major phrase boundaries and utterance boundaries were marked in the present study speech material (see e.g. van Santen, 1992, p. 535) in order to study the lengthening effect of a boundary on the preceding two syllables. The strength of boundary was not distinguished in the data. In the follow-up analysis the final and penultimate syllables before each boundary were marked manually in the Excel program (final as '1', penultimate as '2', all other syllables had a 'zero'). The aim was to find out whether the lengthening effect (if any) extended only to the last syllable before the boundary or even to the penultimate one (see Section 4.5.3 for pre-boundary lengthening).

7.5 Analysis

Once all material was prepared with the necessary information in the textgrids, the information for each speaker has been extracted with a Praat script and inserted into the Excel program which served as the main analysing tool. The information extracted for each segment was the following: the label and duration of the segment, the word in which it was contained, the name of textgrid in which it was found as well as the information whether the segment included a stress mark and whether it was the last segment before a boundary. As already mentioned in the preceding section, the segments of the last two syllables before the boundary had to be marked manually.

7.5.1 Articulation rate normalization

Since the speakers read the text with different speeds, the data had to be normalized for articulation rate so that the data would be better comparable. The normalization method was the following. The mean articulation rate for each speaker was counted in phones per second with a Praat script. The number of phones was based on the number of realized segments. The mean articulation rate of each speaker was then used to calculate the speaker's personal coefficient by means of the following formula:

$$C_i = AR_i / AR_{gm}$$

where:

C_i is the personal normalization coefficient of a given speaker to be calculated,

AR_i is the mean articulation rate of a given speaker,

AR_{gm} is the grand mean (i.e. mean articulation rate of all six speakers)

This coefficient was then used to multiply the durational value of each segment of the given speaker in order to gain the normalized duration of each segment. Thus if a speaker's mean articulation rate was greater than the grand mean, the coefficient was greater than 1 and consequently, all normalized durations of the given speaker were slightly longer than her raw durations. On the other hand, if the mean articulation rate was smaller than the grand mean,

	British speakers			**Czech speakers**			**AR_{gm}** (ph/s)
	BrE 1	**BrE 2**	**BrE 3**	**CzE 1**	**CzE 2**	**CzE 3**	
AR_i (ph/s)	13.97	14.83	13.01	12.76	13.04	11.48	13.185
C_i	1.06	1.12	0.99	0.97	0.99	0.87	

Table 8 Mean articulation rates of individual speakers. AR_i (in phones per second) - mean articulation rate of a given speaker, AR_{gm} (also in ph/s) - grand mean, i.e. the mean articulation rate of all six speakers, C_i - personal normalization coefficient of each speaker.

then all normalized durations of this speaker were shorter than the speaker's raw durations. The specific values of personal articulation rates and normalization coefficients are displayed in Table 8.

The three British English speakers are going to be referred to from now on as BrE 1, BrE 2 and BrE 3 and the Czech English speakers as CzE 1, CzE 2 and CzE 3. The British speaker BrE 2 was the fastest with 14.83 phones per second whereas the Czech speaker CzE 3 with 11.48 phones per second was the slowest. On the whole, the British speakers were faster than their Czech colleagues. This normalization procedure aims to reduce the interpersonal differences caused by divergent articulation rates and to enable easier comparison of the temporal properties of the given segments.

7.5.2 Data analysis

When the normalized durational values were added to the data, the analysis itself was undertaken. Two types of analysis were performed - first, a general analysis with regard to the individual segments and phonetic categories, and second, an analysis of the possible influence of the stress and boundary factors.

For the general analysis, the data were sorted according to the segment type, vowel or consonant category and speaker. The mean duration of each segment type for each speaker was counted together with the standard deviation from the mean. The means were always counted for both the raw and the normalized data and both will be given in the Results section since there are advantages but also disadvantages to each of them. The few cases of the Czech English rhotic [ɚ] which could not be further segmented into [ə] and [r], as noted in Section 7.3, were subsumed under the [ə] category. Also [ɝː] and [ɜː] were merged into one category in the end.

Mean durations for each phonetic category were counted for each speaker. The vowel categories distinguished were short vowels, long vowels and diphthongs. With regard to consonants, the present study differentiated plosives, fricatives, affricates, nasals, liquids and glides. Plosives and fricatives were further classified into voiced and voiceless. The means for each segment type for each group of speakers were also counted, e.g. a mean value of all three British speakers for the [p] segment and another mean for the Czech speakers. Finally, the average prototypical example of each phonetic category was counted for each group of speakers.

Finally, the data was searched for the presence of stress and for subsequent boundary and analysed with respect to vowel and consonant categories, the country of origin of the speaker. An analysis of variance (ANOVA) was performed on the data in order to test the statistical significance of the results. A maximum of three factors were taken into account. When needed, post hoc Tukey HSD test was also performed. Only the most general and significant results were chosen to be presented in the next sections.

7.6 Representativeness of the data: number of occurrences

In order to establish the representativeness of the data analysed, the number of occurrences need to be described. This section will summarize the numbers of segments in each phonetic category and of each type individually. Because of the analysis of effect of the two above-mentioned factors, the numbers of occurrences in the different pre-boundary positions need to be presented and with regard to vowels also the number of stressed and unstressed vocalic segments.

Table 9 displays numbers of occurrences of analysed segments in the individual phonetic categories for all six speakers. The speech data comprised of a total of 11 840 analysed speech segments - 4433 vowels and 7407 consonants. This number is comparable to the size of the corpus analysed by Crystal and House (1988) who had 10 300 measured phones. However, the number of vowels is relatively small when compared to van Santen's (1992) study which examined a much larger corpus of 18 046 vowel segments.

	Total
All vowels	4433
Short monophthongs	3341
Long monophthongs	554
Diphthongs	538
All consonants	7407
Plosives	2426
Voiceless	1584
Voiced	842
Fricatives	2120
Voiceless	1240
Voiced	880
Affricates	83
Nasals	1396
Liquids	1104
Glides	278

Table 9 Overall number of occurrences of analysed segments in individual phonetic categories out of a total of 11 840 speech segments.

In the present corpus, there were more than six times as many short vowels than long vowels or diphthongs. There were 2426 plosives and 2120 fricatives which were the two most numerable consonant categories. On the other hand, there were only 278 glides and mere 83 affricates analysed. Thus affricates were relatively underrepresented in comparison with the other categories. For an overview of number occurrences in the individual categories as realized by the individual speakers, see Appendix 2.

	VOWELS		Total
Monophthongs	*Short monophth.*		
		ɪ	1059
		e	354
		æ	216
		ʌ	125
		ə	1449
		ɒ	98
		ʊ	40
	Long monophth.		
		i:	201
		ɜ:	59
		ɑ:	62
		ɔ:	73
		u:	159
Diphthongs	*Diphthongs*		
		eɪ	212
		aɪ	149
		ɔɪ	6
		aʊ	46
		əʊ	93
		ɪə	8
		eə	17
		ʊə	7

CONSONANTS			Total	
Plosives				
	Voiceless	p	293	***Obstruents***
		t	983	
		k	308	
	Voiced	b	227	
		d	468	
		g	147	
Fricatives				
	Voiceless	f	246	
		θ	21	
		s	671	
		ʃ	109	
		h	193	
	Voiced	v	241	
		ð	277	
		z	362	
		ʒ	0	
Affricates				
		tʃ	36	
		dʒ	47	
Nasals				***Sonorants***
		m	344	
		n	906	
		ŋ	146	
Liquids				
		r	577	
		l	527	
Glides				
		j	139	
		w	139	

Table 10 Overall number of occurrences of analysed individual speech segments out of a total of 11 840 speech segments.

To provide an exhaustive description of the present study data, the overall numbers of occurrences of individual speech segments are given in Table 10. Since this is relatively detailed data, it will not be commented upon. For the number of occurrences of individual segments as realized by the individual speakers see Appendix 3.

The data need to be described also with respect to the number of occurrences of the individual phonetic categories with reference to the different positions as given by stress and position before a boundary. Table 11 shows the number of occurrences of the individual vowel categories with regard to stress conditions and L1 group. It may be seen that while the

majority of short vowels were unstressed, the major part of long vowels and diphthongs were stressed. There were only about 70 occurrences of both long vowels and diphthongs in the unstressed position.

L1 Group	*Stress*	*Short*	*Long*	*Diphthongs*	*All vowels*
BrE	Yes	403	208	212	823
CzE	Yes	427	202	181	810
BrE	No	1201	71	73	1345
CzE	No	1306	74	71	1451

Table 11 Number of occurrences of stressed (Yes) and unstressed (No) vowels in individual vowel categories for British (BrE) and Czech (CzE) speakers.

The number of occurrences of vowels in the individual vowel categories with regard to pre-boundary position is displayed in Table 12. The data is again distinguished for British and Czech speakers. Three situations with regard to position before boundary were distinguished: ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Since the overall number of long vowels and diphthongs was

L1 Group	*Pre-Bound*	*Short*	*Long*	*Diphthongs*	*All vowels*
BrE	0	1199	192	197	1588
CzE	0	1211	183	153	1547
BrE	1	181	50	61	292
CzE	1	237	50	68	355
BrE	2	224	37	27	288
CzE	2	285	43	31	359

Table 12 Number of occurrences of vowels in individual vowel categories in different positions with regard to boundary for British (BrE) and Czech (CzE) speakers. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0).

L1 Group	*Stress*	*Pre-boundary position*		
		0	1	2
BrE	Yes	585	113	125
CzE	Yes	538	131	141
BrE	No	1003	179	163
CzE	No	1009	224	218

Table 13 Number of occurrences of all vowels (without distinction of category) with regard to both stress and boundary conditions for British (BrE) and Czech (CzE) speakers. Stressed (Yes) and unstressed (No) vowels. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (1), penultimate syllable (2), and all other syllables (0).

relatively low, also the number of occurrences in the scarcer positions (Pre-Bound 1 and 2) was quite low. This should be considered when analyzing the data although the tests of statistical significance partially cater for differences in the amount of data.

Table 13 captures the number of occurrences for the combination of the above-mentioned factors – stress and syllable - this time only with regard to all vowels without distinction because the individual sub-classes would not be numerous enough.

Since the only factor examined for consonant duration was boundary, Table 14 is the only table displaying the number of occurrences for consonants. The data is shown again for the two groups of speakers based on their nationality. It may be observed that the numbers in some categories are really meager - especially in affricates and glides. This explains why the results for these two categories in Section 8.2.4 were by no means significant (even if taken for all six speakers) whereas in the other groups they were significant.

L1 Group	*Pre-Bound*	*Plosives*	*Fricatives*	*Affricates*	*Nasals*	*Liquids*	*Glides*	*All consonants*
BrE	0	793	774	38	518	334	107	2564
CzE	0	757	694	33	461	400	98	2443
BrE	1	359	171	1	126	75	22	754
CzE	1	322	213	3	163	133	27	861
BrE	2	143	113	3	58	70	9	396
CzE	2	152	156	4	69	92	14	487

Table 14 Number of occurrences of consonants in individual consonant categories in different positions with regard to boundary for British (BrE) and Czech (CzE) speakers. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0).

8 RESULTS

8.1 Vowels

The durational results as found by the present study are given in this and the following section. First, Section 8.1 will compare and contrast the durations for vowels, and then, in Section 8.2 consonants will be covered. First, mean durations for the general phonetic categories for the two L1-based groups of speakers will be always presented. Then a segmental "peculiarity" in which the Czech speakers differed from their British colleagues the most will be described. Finally, the following sections will cover the effect of the two examined factors on durational variation.

In the next section (Section 8.1.1) it will be revealed how Czech and British speakers realize individual vowel categories, i.e. short vowels, long vowels and diphthongs. In Section 8.1.2 the case of Czech English [æ] will be described. Section 8.1.3 will explore the influence of stress on vowel duration whereas Section 8.1.4 will examine the influence of boundary. The final section concerning vowel duration (Section 8.1.5) will then show what happens when these two factors are combined

8.1.1 *Vowel categories*

The most general results with regard to vowels are depicted in Figures 1 and 2. Figure 1 shows the raw mean durations for the categories of short vowels, long vowels and diphthongs for the two groups of speakers examined in our study – the British speakers of English and the Czech speakers of English. The normalized mean values for the same data are given in Figure 2.

The analysis of variance (ANOVA) found highly significant interaction between the L1 group and the type of vowel category in the raw data: $F(2, 4423) = 15.3$; $p < 0.001$. Figure 1 shows that the short vowels of the Czech speakers were slightly longer (about 5 ms) than their British English counterparts and that Czech English diphthongs were almost 20 ms longer than the British ones. The post hoc Tukey HSD test revealed that both these differences are highly significant ($p < 0.001$). The raw mean duration of Czech and British long vowels did not differ significantly.

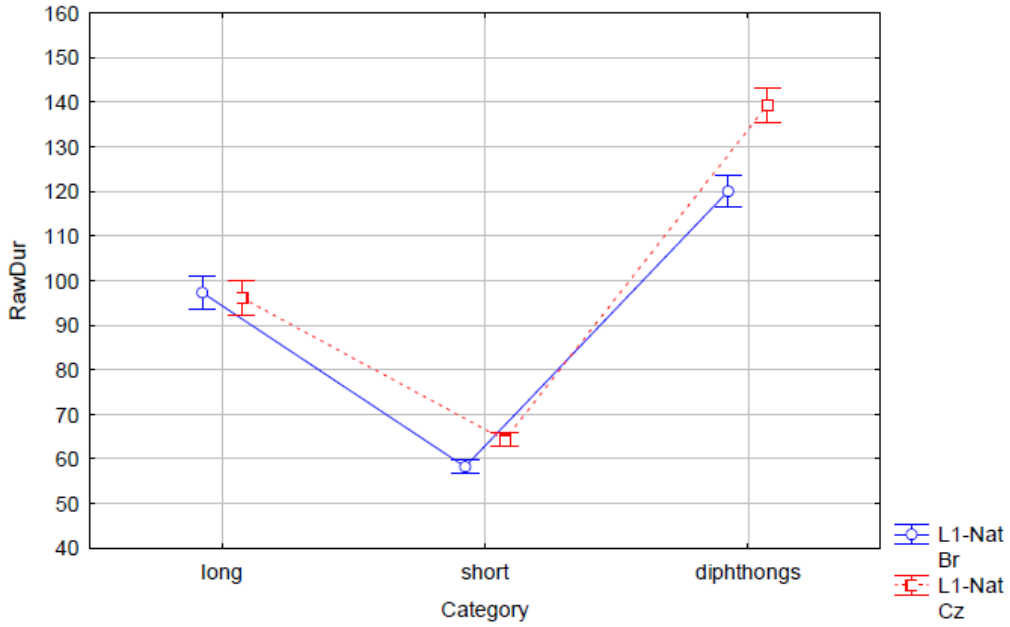

Fig. 1 Raw mean durations of vowel categories for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. (Error bars indicate 95% confidence intervals.)

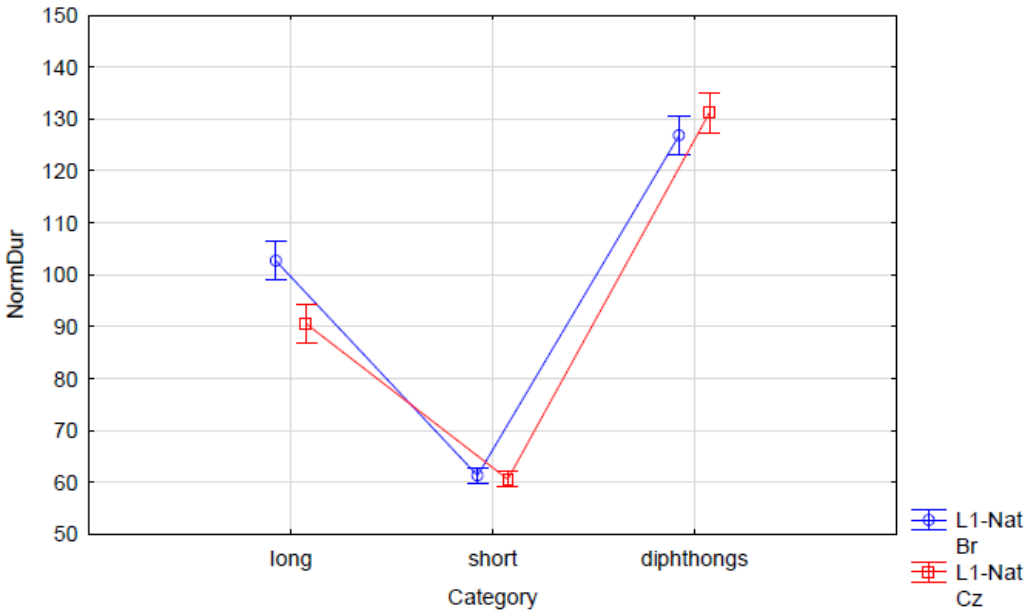

Fig. 2 Normalized mean durations of vowel categories for British and Czech speakers. NormDur - normalized duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. (Error bars indicate 95% confidence intervals.)

However, the normalized data in Figure 2 show a slightly different picture. In this data where the articulation rate differences should be removed, the short vowels and diphthongs of both groups had almost the same duration, but it was the long vowels that differ the most – the Czech English long vowels were 13 ms shorter than their British counterparts. The post hoc tests found this difference also highly significant ($p < 0.001$). It thus seems from these

group means that if there was something that the present data Czech speakers experienced problems with, it might have been the adequate duration of the long vowels.

The relationships among durations of the three vowel categories are also interesting. Table 15 gives the durational ratios between the individual categories. The Long to Short ratio, i.e. how many times are the long vowels longer than the short vowels, of the British speakers is 1.67 whereas the Czech one is only 1.5. It can thus be seen that the Czech English long vowels in our data were really shorter when compared to the British ones if the short vowels are taken as a reference measure. The ratios for the diphthongs (diphthongs to short vowels) are more similar. Both numbers are slightly larger than 2, which means that both Czech English and British English diphthongs were approximately twice as long as the respective short vowels in those two groups. Even though Czech English diphthongs were slightly longer than the British ones, this difference is not significant. These relationships are graphically portrayed in Figure 2.

	Long to Short Ratio	Diphthongs to Short Ratio
British speakers	1.67	2.06
Czech speakers	1.5	2.17

Table 15 The durational ratios between the vowel categories. The ratios of the duration of long vowels to short vowels and of diphthongs to short vowels are given for British and Czech speakers.

8.1.2 *Short vowels: The case of [æ]*

In the present study attention was not paid to the duration of the individual vowels unless some significant differences have been observed. This was the case of the duration of the Czech English [æ] as opposed to its British English equivalent. Before looking specifically at [æ], its durational standing among other English short vowels needs to be touched upon.

Figure 3 portrays the normalized durations for short vowels of British and Czech speakers. Looking at the British short vowels, the inherent vowel durations based on physiological constraints can be clearly observed. If we leave [ə] aside, the degree of vowel height (openness) correlates with vowel duration. Thus the close vowels [ɪ] and [ʊ] are the shortest out of the six short vowels. On the other hand, the open front vowel [æ] is the longest British short vowel. The other three vowels take their place in between as can be seen

in Figure 3. The central mid vowel [ə] is reported to be the shortest vowel in English since it occurs only in unstressed positions and this is also confirmed by our data. The vowel [ə] is really the shortest vowel of the present study British speakers although it is not much shorter than [ɪ] or [ʊ].

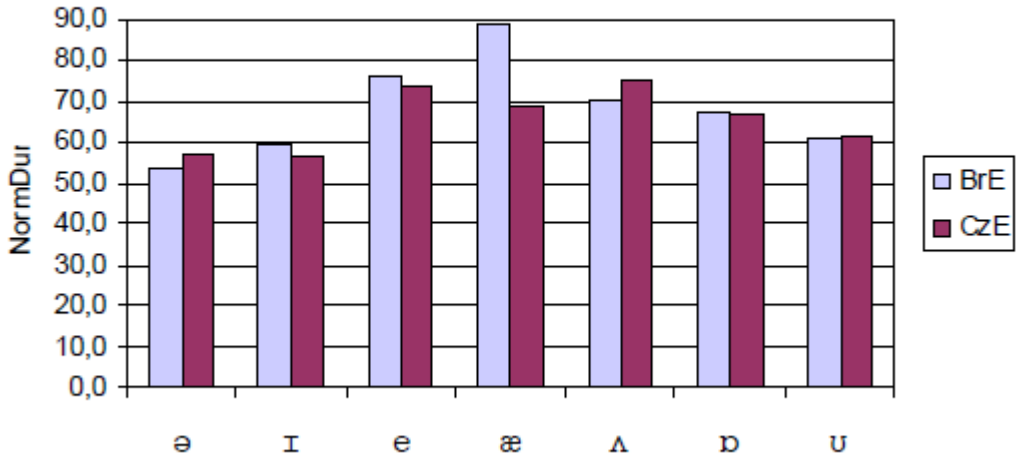

Fig. 3 Normalized mean durations of individual short vowels for British and Czech speakers. BrE – data for British speakers, CzE – data for Czech speakers. NormDur - normalized duration in milliseconds.

When the duration of British and Czech English short vowels are contrasted, the most conspicuous difference is the relative shortness of the vowel [æ]. The vowel [æ] as realized by the Czech speakers is much shorter than the one of their British colleagues. It is shorter by 20 ms, which has been found to be a highly significant difference ($p < 0.001$). The influence of the L1 group (British speakers versus Czech speakers) on the duration of [æ] was tested by a single-factor ANOVA and a highly significant interaction was found: $F(1, 214) = 38{,}5$; $p < 0.001$. The normalized duration of the vowel [æ] is displayed separately in Figure 4. The same interaction was found also in the raw data although to a lesser degree: $F(1, 214) = 13{,}8$; $p < 0.001$. Figure 5 shows the raw durations of [æ]. It may be noticed that the difference between the British and Czech speakers is smaller (about 12 ms) but still significant ($p < 0.001$).

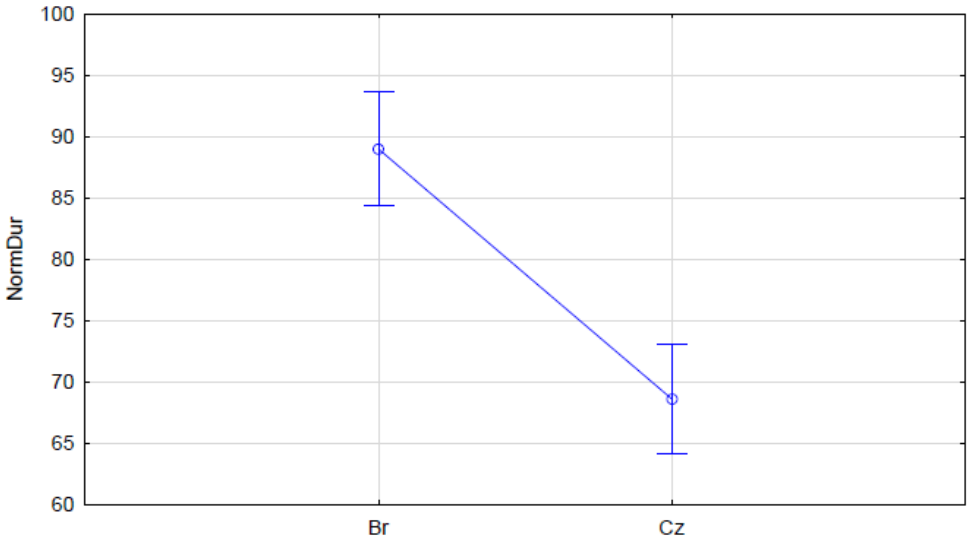

Fig. 4 Normalized mean duration of the [æ] vowel for British (Br) and Czech (Cz) speakers. NormDur - normalized duration in milliseconds. (Error bars indicate 95% confidence intervals.)

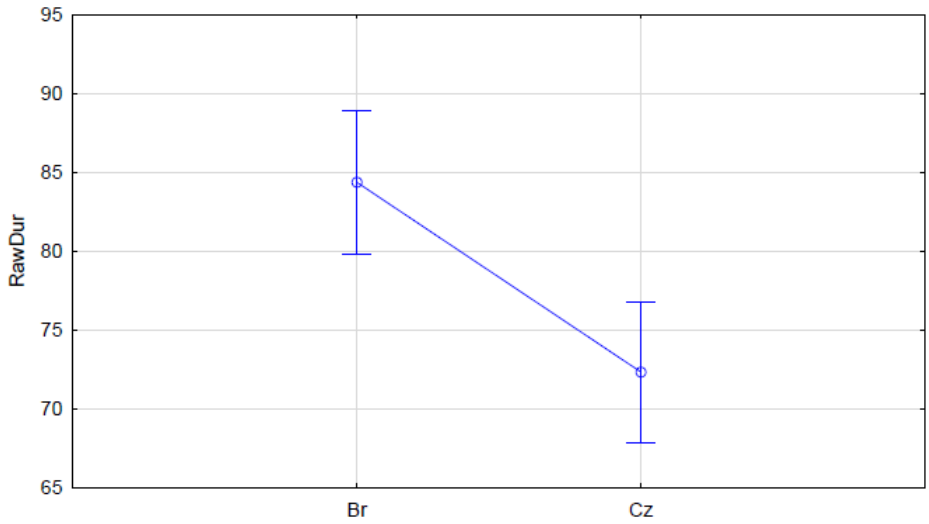

Fig. 5 Raw mean duration of the [æ] vowel for British (Br) and Czech (Cz) speakers. RawDur - raw duration in milliseconds. (Error bars indicate 95% confidence intervals.)

8.1.3 Influence of stress on vowel duration

One of the aims of the present study was to explore the influence of stress on vowel duration. The results will be presented in this section. As expected, the data confirm clear impact of stress on vowel duration. Stressed vowels are in general longer than unstressed ones ($F(1, 4425) = 751{,}9$; $p < 0.001$). We also wanted to find out whether Czech and British speakers realized stressed and unstressed vowels in the same way, so the interaction of stress

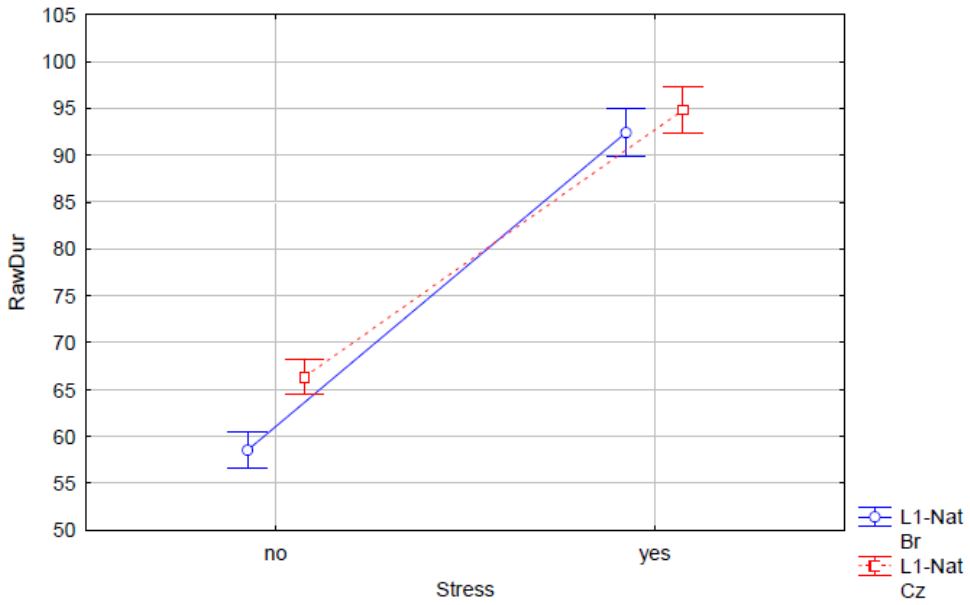

Fig. 6 Influence of stress on vowel duration for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Stress no – unstressed vowels, stress yes – stressed vowels. (Error bars indicate 95% confidence intervals.)

and L1 group was tested with a two-factor ANOVA and a significant interaction was confirmed: $F(1, 4425) = 5.6$; $p = 0.018$. The results are shown in Figure 6. It may be observed that in the raw data, the stressed vowels are approximately 30 ms longer than unstressed vowels. The duration of the British English and Czech English stressed vowels does not differ significantly. However, the Czech speakers' unstressed vowels are shorter than those of

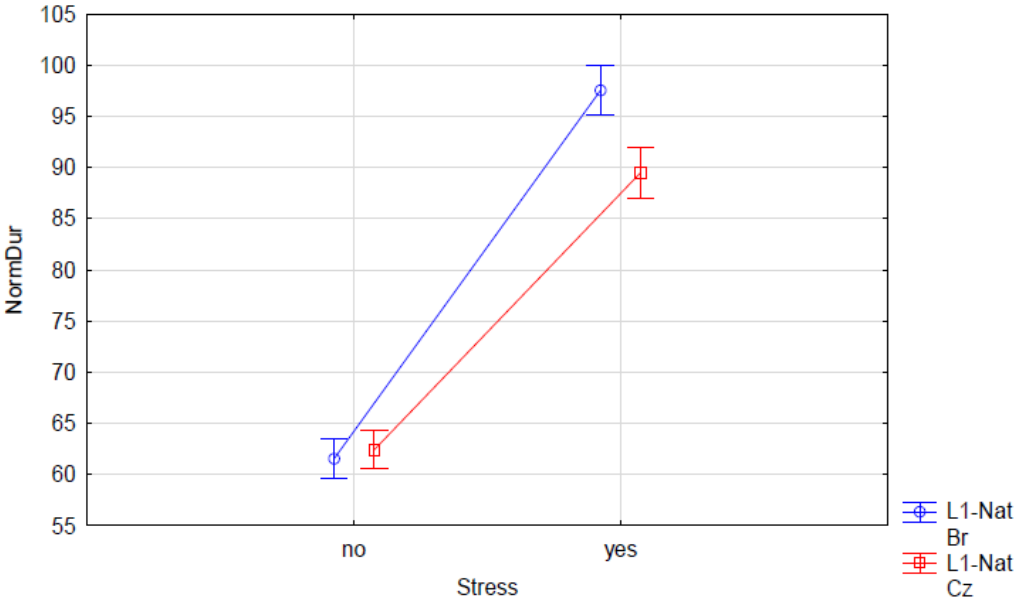

Fig. 7 Influence of stress on vowel duration for British and Czech speakers. NormDur - normalized duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Stress no – unstressed vowels, stress yes – stressed vowels. (Error bars indicate 95% confidence intervals.)

British speakers and this difference has been found highly significant in post hoc Tukey HSD test ($p < 0.001$). Thus although our Czech speakers' stressed vowels are as long as the British ones, the Czech speakers seem to temporally reduce the unstressed ones less. The difference between the stressed vowels and unstressed vowels is 28 ms for the Czech speakers and 34 ms for the British.

In the normalized data (Figure 7) the difference is even bigger (27 ms as against 36 ms) and an ANOVA showed a highly significant interaction this time: $F(1, 4425) = 15.5$; $p < 0.001$. After normalization the relationship gets reversed (like in the case of short vowels, long vowels and diphthongs) and it is the stressed vowels that differ.

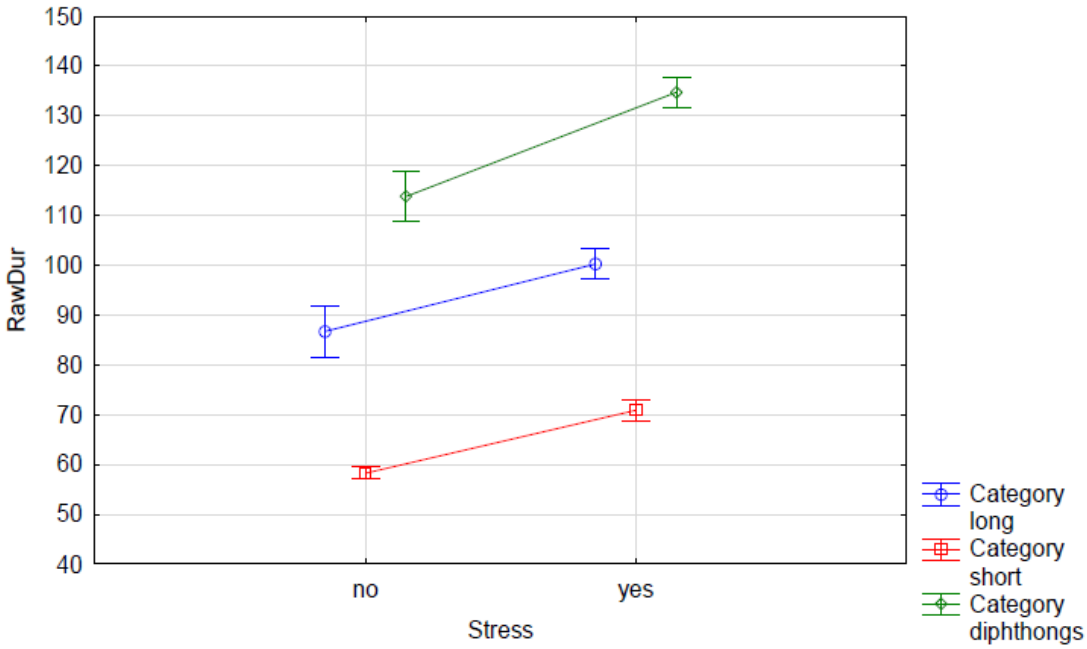

Fig. 8 Influence of stress on the duration of individual vowel categories. RawDur - raw duration in milliseconds. Stress no – unstressed vowels, stress yes – stressed vowels. (Error bars indicate 95% confidence intervals.)

Furthermore, we were interested whether different vowel categories behave in the same way when under different stress conditions. It was found that there was a small but still significant interaction between these two factors: $F(2, 4423) = 3.2$; $p = 0.04$. The behaviour of the three vowel categories can be seen in Figure 8. The short vowels and the long vowels seem to be lengthened to a similar degree in stressed positions (12 and 14 ms respectively). However, diphthongs tend to be lengthened slightly more (21 ms). The differences are not big, but still significant.

Last, it was explored whether British and Czech speakers lengthen different vowel categories differently under the two stress conditions. In a three-factor ANOVA, the interaction of all factors was found significant: $F(2, 4417) = 3.7$; $p = 0.025$. The raw mean

durations of different vowel categories as realized by the British and Czech speakers when stressed and when not stressed is displayed in Figure 9.

The left half of the graph portrays the realizations by the British speakers and the Czech means are on the right. The differences between the two groups of speakers are noticeable at first sight. The Czech English short vowels are significantly longer than the

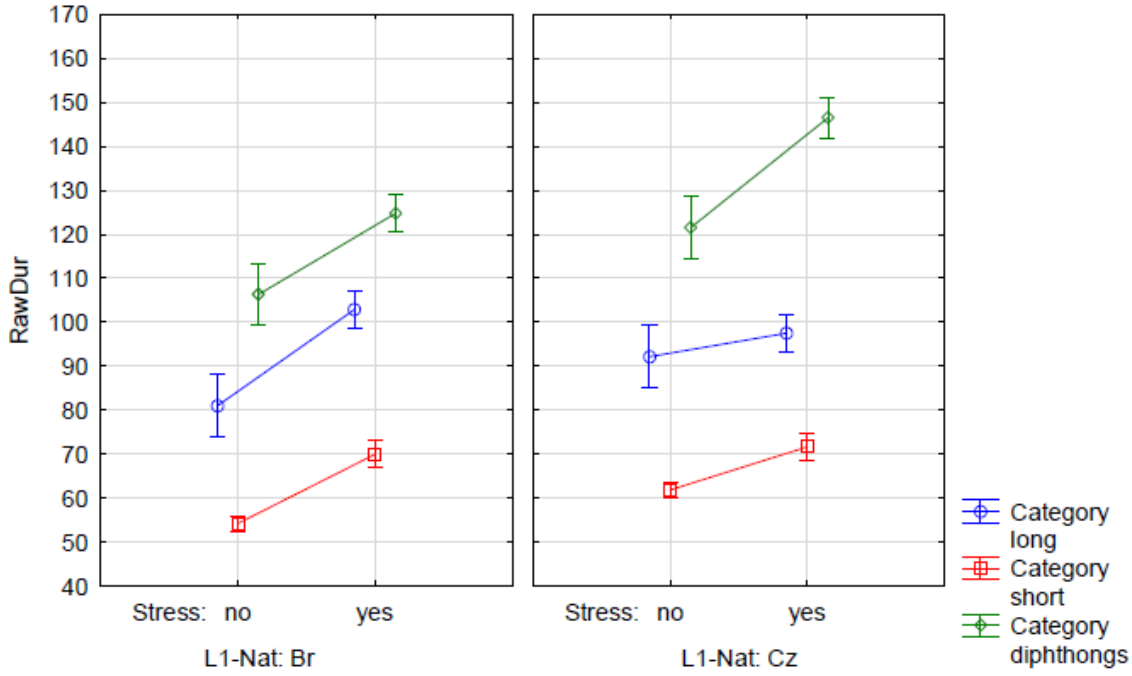

Fig. 9 Influence of stress on the duration of individual vowel categories for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Stress no – unstressed vowels, stress yes – stressed vowels. (Error bars indicate 95% confidence intervals.)

British ones when not stressed ($p < 0.001$). On the other hand, the Czech English diphthongs get lengthened more when stressed when compared to the British ones. However, the most notable difference lies in the long vowels. There is no significant difference ($p = 0.98$) between the stressed and unstressed long vowels as realized by Czech speakers. This situation sharply contrasts with the right-hand part of the graph which shows that the British speakers lengthen their long vowels 'adequately' when stressed.

8.1.4 Influence of boundary on vowel duration

The other main factor which was examined in connection to possible influence on segmental duration was the intonation phrase boundary. As described in Section 7.4.2, the segments in the two last syllables before the boundary were marked. Consequently, segments can be differentiated as belonging pre-boundary conditions: (1) segments in the ultimate

syllable before the boundary (Pre-Bound 1), (2) segments in the penultimate syllable (Pre-Bound 2), and (3) segments in all other syllables, i.e. being further than a two-syllable distance from the closest boundary (Pre-Bound 0).

As expected the position before the boundary has been found to have clear effect on vowel duration in English: $F(2, 4423) = 532.8$; $p < 0.001$. However, the two examined positions before the boundary did not behave in the same way. Whereas the last syllable before the boundary (Pre-Bound 1) was almost 50 ms longer than the one which was not in the vicinity of the boundary (Pre-Bound 0), the penultimate syllable (Pre-Bound 2) was only about 5 ms longer than the Pre-Bound 0 (although this was still a significant difference: $p = 0.001$).

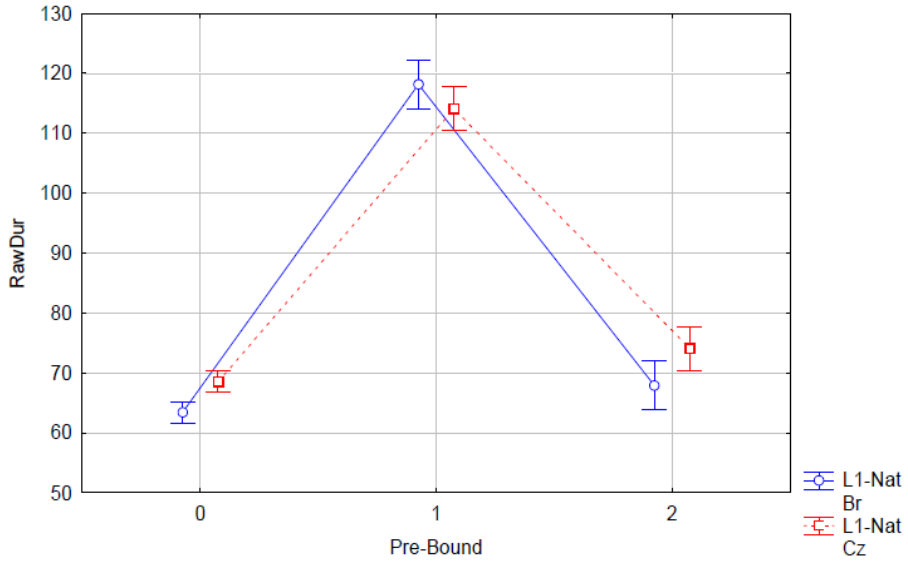

Fig. 10 Influence of boundary on vowel duration for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

Figure 10 shows the results of the interaction of the boundary and the L1-group affiliation in the raw data. This interaction has been found significant: $F(2, 4423) = 4,8$; $p = 0.008$. Although the difference between the Pre-Bound 1 segments of the British and Czech speakers does not differ significantly, the difference between their segments in Pre-Bound 0 is highly significant ($p < 0.001$). This means that the Czech speakers lengthen the last syllable less than the British speakers.

Even better observable is the lengthening difference in the normalized data in Figure 11. Here the Pre-Bound 0 and Pre-Bound 2 segments of both groups of speakers are

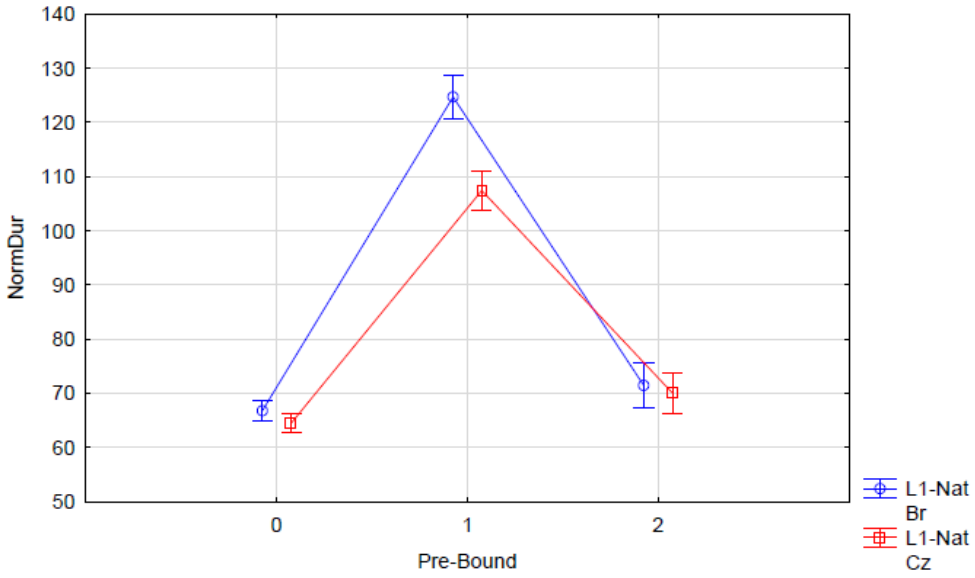

Fig. 11 Influence of boundary on vowel duration for British and Czech speakers. NormDur - normalized duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

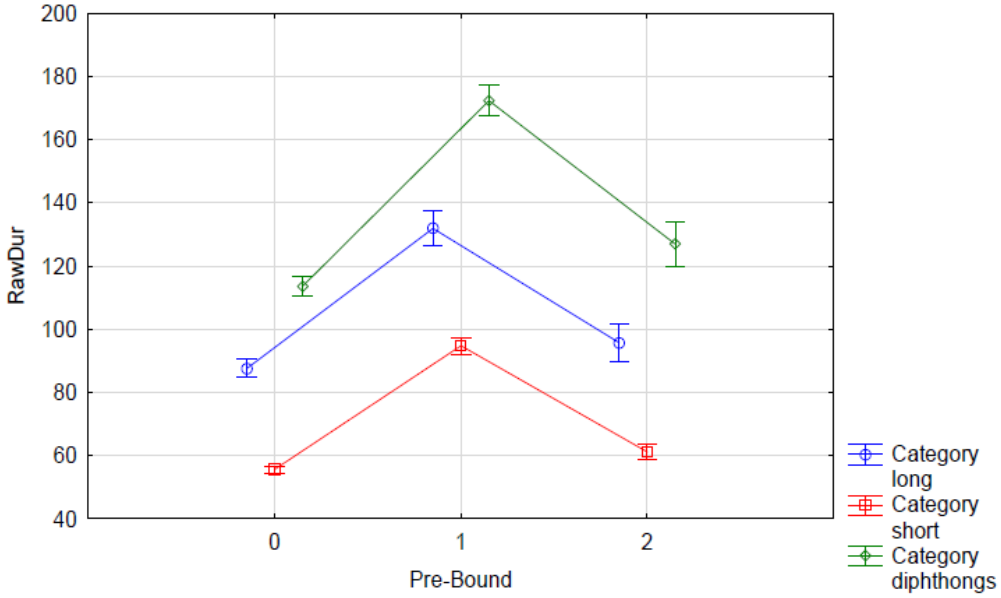

Fig. 12 Influence of boundary on the duration of individual vowel categories for British and Czech speakers. RawDur - raw duration in milliseconds. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

leveled and thus the differences in the normalized duration of the last syllable (Pre-Bound 1) are very conspicuous. The Czech speakers lengthen the segments in the last pre-boundary syllable for at least 15 ms less than their British colleagues and this difference has been found

highly significant ($p < 0.001$). The Pre-Bound 0 durations do not differ significantly from the durations in their respective Pre-Bound 2.

The final point of this section is the interference of pre-boundary lengthening and type of vowel category. This interference displayed in Figure 12 has been found also highly significant: $F(4, 4420) = 9.3$; $p < 0.001$. This is probably due to the larger lengthening in diphthongs as compared to short and long vowels. Regarding the potential differences between the L1 groups, the interaction among boundary, category and L1 group has been found insignificant ($p = 0.29$) and will not be thus described in this paper.

8.1.5 Interaction of stress and boundary in vowel duration

The last section describing the present data vowel duration will deal with the interaction of stress and boundary in vowels. The results with regard to this interaction are displayed in Figure 13. The graph shows how the stress and boundary together influence duration regardless of the speakers' provenience. An ANOVA found a highly significant interaction between the two factors: $F(2, 4423) = 24.8$; $p < 0.001$. The stressed vowels are lengthened in the last syllable before the boundary more than the unstressed ones.

Further, we were interested whether the British and Czech speakers differ in their realization of vowels under different stress and boundary conditions. The interaction among speakers' nationality, stress and boundary was been found significant by a three-factor ANOVA: $F(2, 4417) = 3.6$; $p = 0.03$. The differences between British and Czech speakers are captured in Figure 14. The left part of the graph shows the durations for the British speakers and the Czech values are on the right. The differences are interesting. It seems that the difference between the stressed and unstressed vowels in the ultimate syllable (as shown in Figure 13) is due predominantly to the Czech speakers. The British speakers display about the same amount of pre-boundary lengthening for stressed and unstressed vowels.

To conclude, there are a number of differences between the Czech and the British speakers with regard to vowel duration. The results for consonants will be reported in the following section.

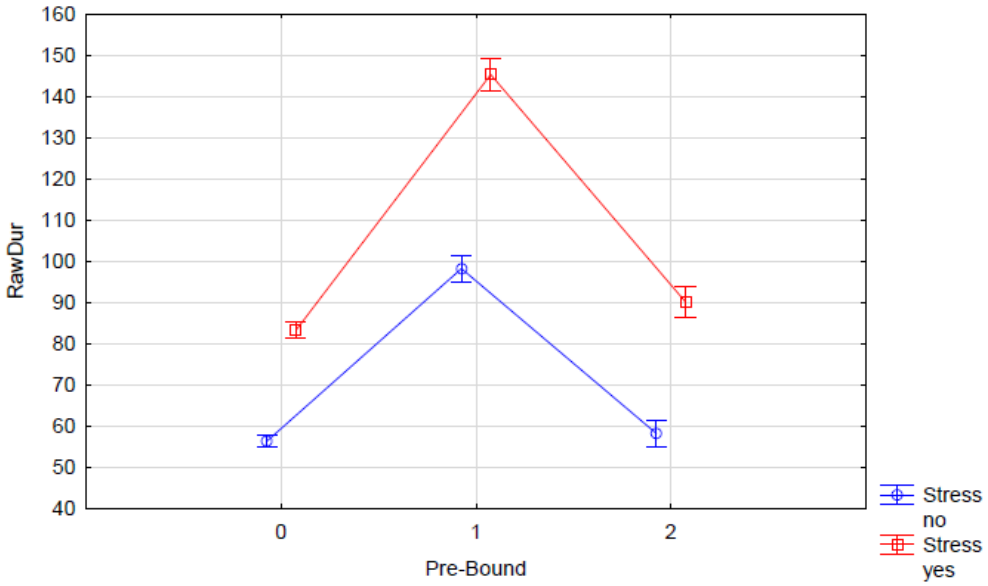

Fig. 13 The interaction of stress and boundary in their influence on vowel duration. RawDur - raw duration in milliseconds. Stress no – unstressed vowels, stress yes – stressed vowels. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

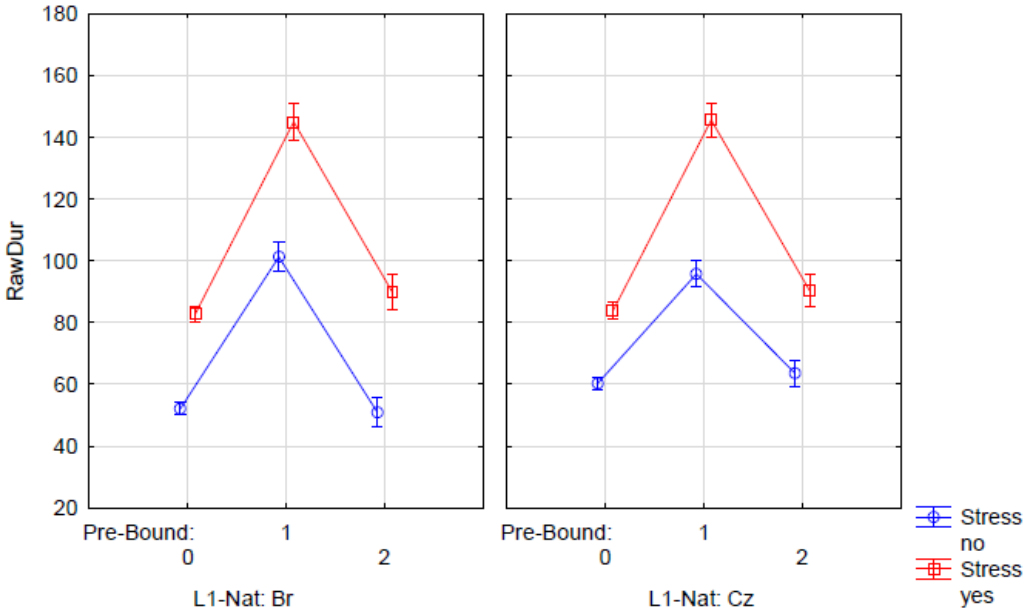

Fig. 14 Interaction of stress, boundary and L1 group in their influence on vowel duration. RawDur - raw duration in milliseconds. Stress no – unstressed vowels, stress yes – stressed vowels. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Error bars indicate 95% confidence intervals.

8.2 Consonants

The consonant durations will be reported in the following order. First, it will be shown how the consonant categories are realized by the British and Czech speakers (Section 8.2.1). Second, we will notice the durational difference between voiced and voiceless consonants as demonstrated by plosives and fricatives (Section 8.2.2). Third, the case of individual consonants will be explored, specifically the differences in the Czech and British realizations of [r] and [w] (Section 8.2.3). And last, the influence of boundary on consonant durations will be examined (Section 8.2.4).

8.2.1 Consonant categories

Like in the case of vowels, we were interested whether the British and Czech speakers realized the individual consonant categories in the same or different way from the durational viewpoint. The results for both groups are displayed in Figure 15. An ANOVA revealed that there was no significant difference between the ways the consonant categories were realized by the Czech and British speakers: $F(5, 7393) = 1.5$; $p = 0.17$. Although Figure 15 shows differences in the duration of categories of the British and Czech speakers, these departures were mostly due to the divergent articulation rates since the British speakers were in general faster and the Czech speakers slower. The similarity of both groups is thus even more easily observable in the normalized data of Figure 16 in which the articulation rate differences were removed. Consequently, the relationships among the consonant categories were almost the same for both groups of speakers.

The affricates were the longest consonant category which was confirmed despite the relatively small number of occurrences in the data. Fricatives and plosives were the second longest and were of nearly the same duration (they did not differ significantly). Then the sonorants ensued in this order - nasals, liquids and glides

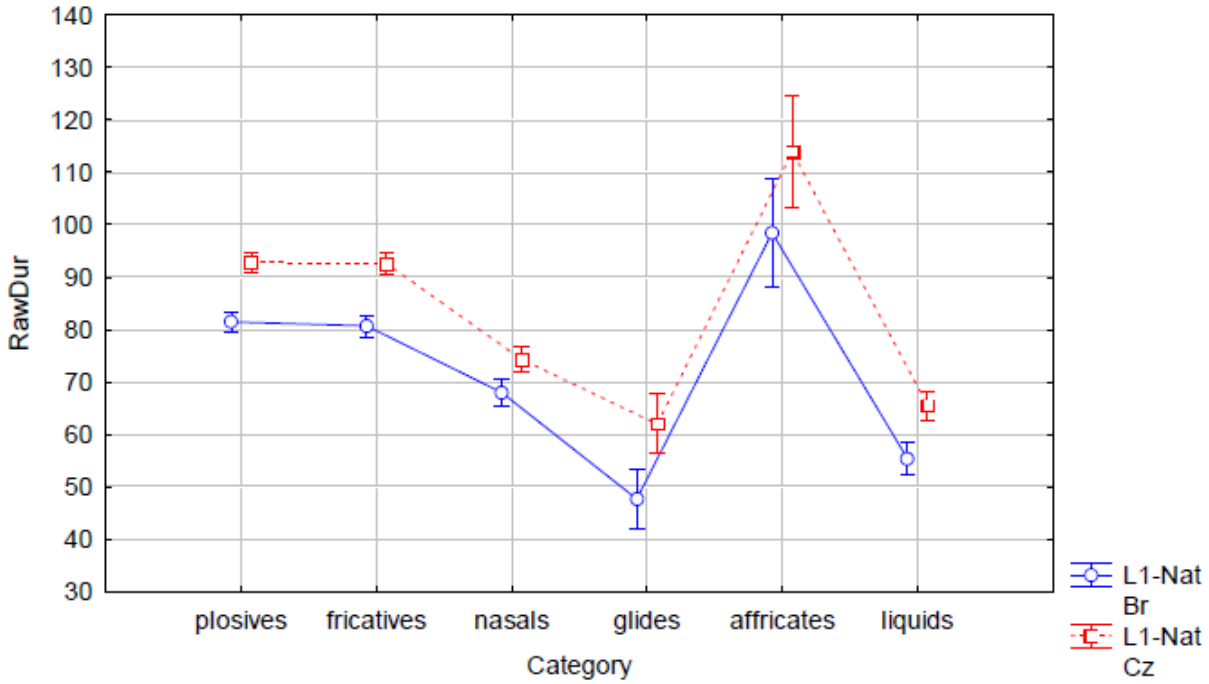

Fig. 15 Raw mean duration of consonant categories for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. (Error bars indicate 95% confidence intervals.)

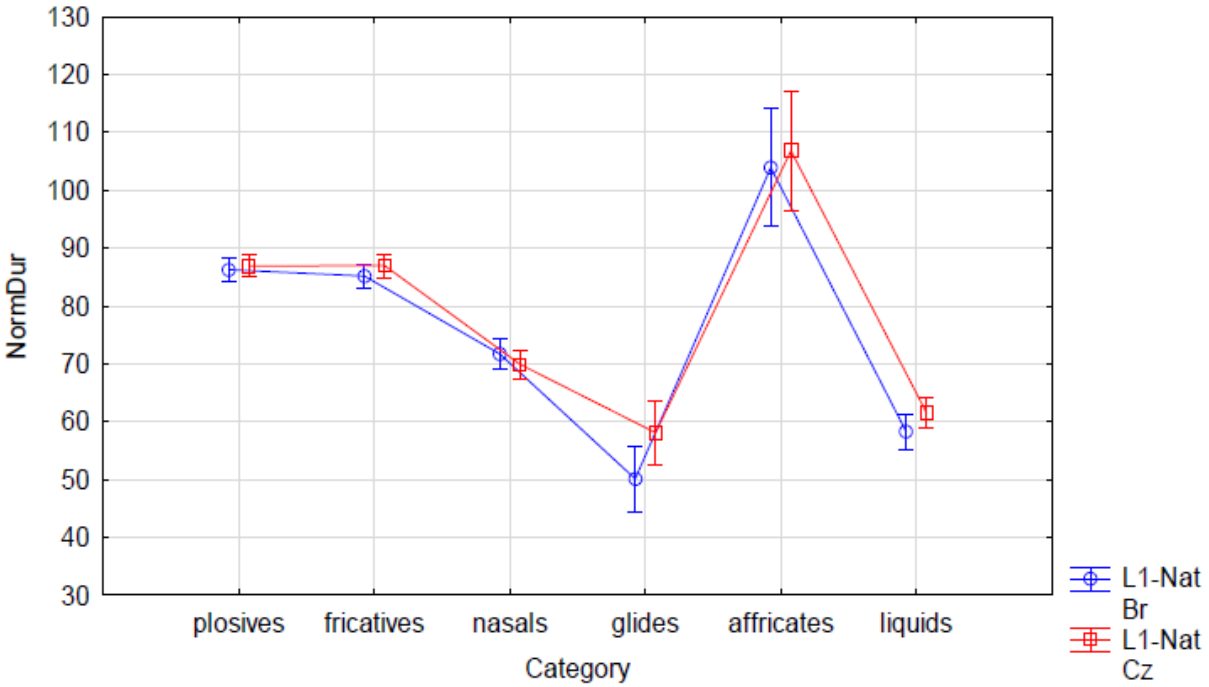

Fig. 16 Normalized mean duration of consonant categories for British and Czech speakers. NormDur - normalized duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. (Error bars indicate 95% confidence intervals.)

8.2.2 Voicing in plosives and fricatives

Further, we were interested whether, and if so, how much voiceless consonants differed from the voiced ones. This relationship was examined on plosives and fricatives since affricates were quite scarce in the data and would thus not form a representative category. The interaction between category type (plosives vs. fricatives) and voicing was found significant: $F(1, 4543) = 8.2$; $p = 0.004$.

Figure 17 captures the duration of the voiced and voiceless plosives and fricatives. It may be observed that the distinction between voiceless and voiced consonants (both plosives and fricatives) was highly significant ($p < 0.001$). It was 28 ms for fricatives and almost 35 ms for plosives. Whereas the difference between the duration of voiceless plosives and fricatives was not significant, the difference between their voiced counterparts was ($p = 0.008$). Thus the 7 milliseconds between 28 and 35 were a significant distinction. The British and Czech speakers were not found to realize the voicing contrast in a significantly different way.

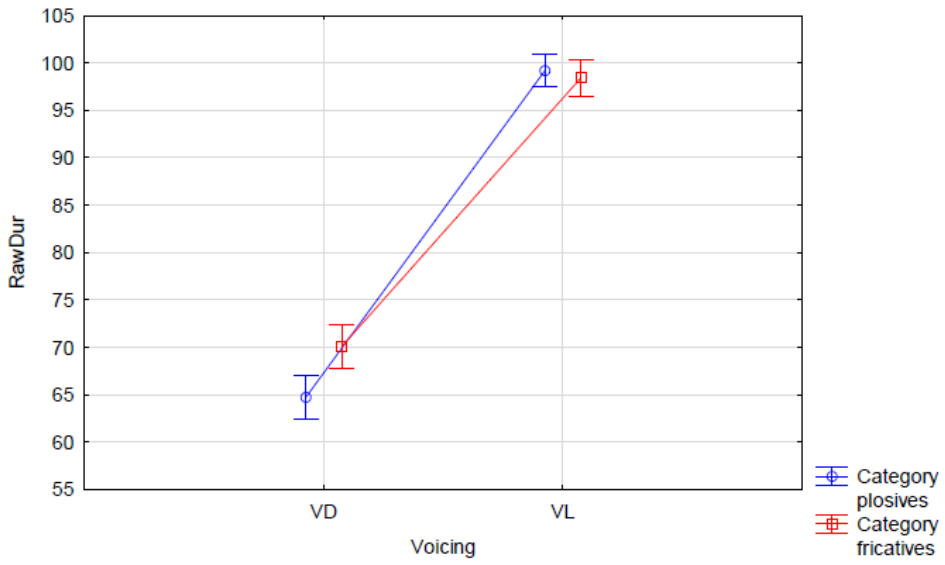

Fig. 17 Voicing distinction for plosives and fricatives. VD – voiced, VL – voiceless. RawDur - raw duration in milliseconds. (Error bars indicate 95% confidence intervals.)

8.2.3 Approximants: The case of [r] and [w]

Like in the case of vowels, the most conspicuous differences in the duration of individual segments will be noticed. The area that might deserve attention is that of approximants, specifically the case of [r] and [w]. Figure 18 portrays the raw mean

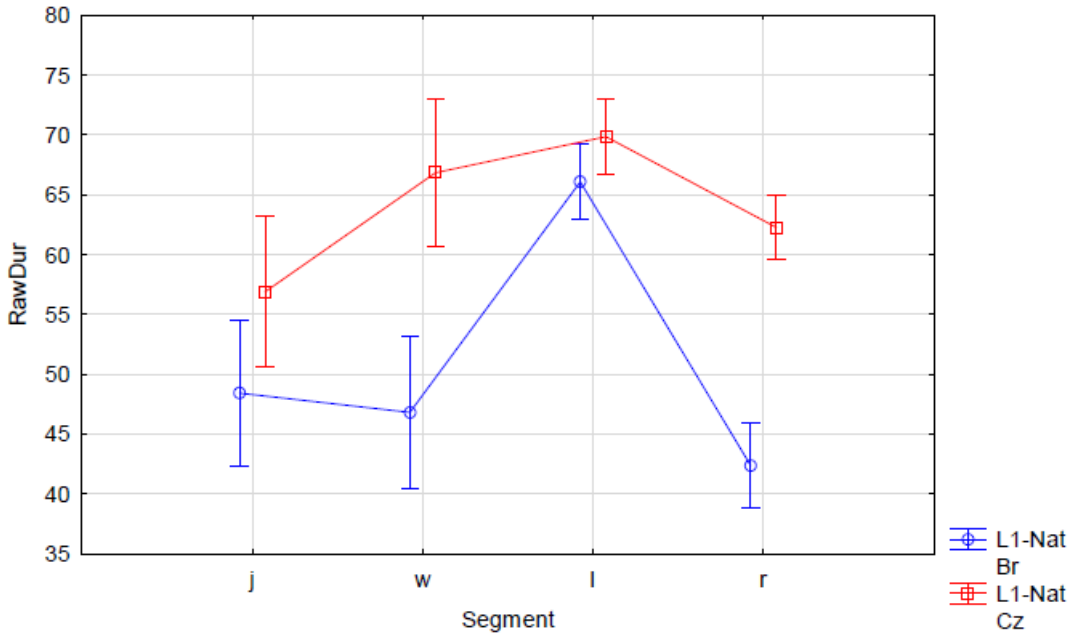

Fig. 18 Raw mean durations of individual approximants as realized by British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. (Error bars indicate 95% confidence intervals.)

durations for the individual approximants. A highly significant interaction was found between the type of segment and the L1 group: $F(3, 1373) = 9.6$; $p < 0.001$. The post hoc Tukey test revealed no difference between Czech and British cases of [j] or [l]. However, in the case of [w] and [r] highly significant differences ($p < 0.001$) were detected in the realizations of British and Czech speakers. The mean Czech English [w] was almost 20 ms longer than the British one and the same durational distinction was valid also for Czech English and British [r]. After normalization the differences got slightly smaller, but still stayed highly significant in the case of [r] ($p < 0.001$) and significant in [w] ($p = 0.37$).

8.2.4 Influence of boundary on consonant duration

The last section devoted to consonants will deal with the influence of the position before the boundary on consonant duration. First, the influence on all consonants without distinction will be examined with regard to differences between Czech and British speakers. The final part will then cover the influence of boundary on individual consonant categories.

Figure 19 depicts the differences between Czech and British raw mean consonant duration under different boundary conditions (for their description see Section 8.1.4). It may be observed that the situation in consonants is very similar to the one in vowels (Figure 10). The Czech speakers thus lengthened the consonants in the last syllable to a lesser degree when compared to the British speakers. Nevertheless, in the British speakers there was significant difference in the duration of the Pre-Bound 0 consonants and the Pre-Bound 2 consonants (p = 0.002). This contrasts both with Czech speakers and with the situation in vowels where neither group differed significantly in their Pre-Bound 0 and Pre-Bound 2 segments.

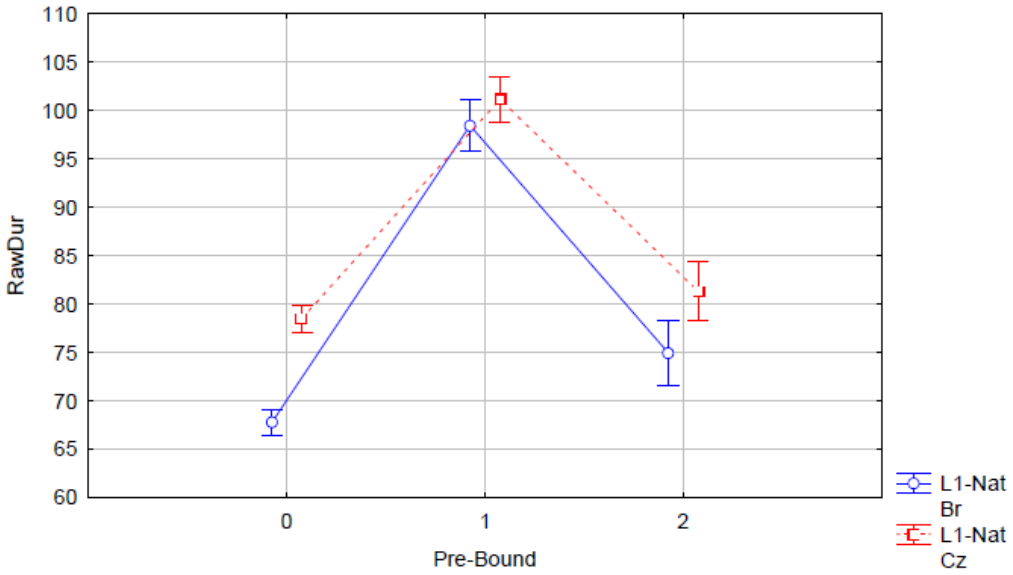

Fig. 19 Influence of boundary on consonant duration for British and Czech speakers. RawDur - raw duration in milliseconds. L1-Nat Br – data for British speakers, L1-Nat Cz – data for Czech speakers. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

Finally, the last results concern the influence of boundary on individual categories as realized by all speakers. There was a highly significant interaction between these two factors: $F(10, 7387) = 9.5$; $p < 0.001$. The individual consonant categories seem to behave differently in divergent boundary conditions. Only the Pre-Bound 0 and Pre-Bound 1 segments will be observed since the Pre-Bound 2 position has been shown as not too important.

The normalized durations displayed in Figure 20 reveal that the consonant category which got lengthened the most in the ultimate syllable was fricatives. The difference between the duration in the Pre-Bound 0 syllable and in the last syllable (Pre-Bound 1) for fricatives was 36 ms. Then the sonorants followed – nasals (30ms) and liquids (26ms). The plosives had the smallest difference (17 ms) among the significant ones. Consequently, the largest difference was between plosives and fricatives as can be observed in Figure 20. As noted in Section 7.6, thanks to the too small number of occurrences in the categories of glides and liquids, the differences for these categories have not been found significant.

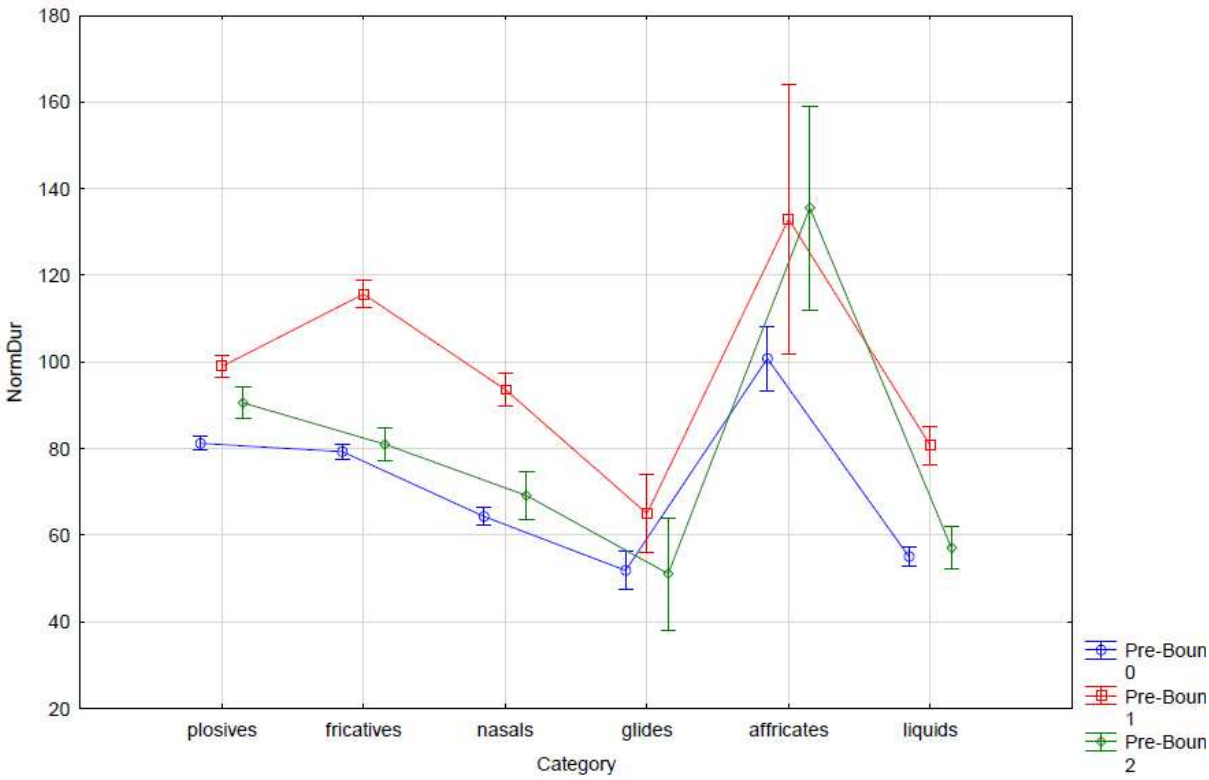

Fig. 20 Influence of boundary on consonant duration. NormDur - normalized duration in milliseconds. Three situations with regard to position before boundary were distinguished: Ultimate syllable before boundary (Pre-Bound 1), penultimate syllable (Pre-Bound 2), and all other syllables (Pre-Bound 0). Error bars indicate 95% confidence intervals.

9 DISCUSSION

9.1 Vowels

The data has shown that in both groups of speakers diphthongs had the largest durations among the vowel categories, then they were followed by the long vowels, and the short vowels, just as the category label implies, were the shortest. However, how do the durational values stand in comparison with other research?

Let us compare the present study vowel categories data with Table 3 from Section 4.2 which gave the mean values from the studies of Crystal and House (1988) and van Santen (1992). The data from Table 3 were reprinted here for convenience' sake and shown side by side with the present study raw data in Table 16. The raw, not normalized data were compared to the previous research since the normalization was done among the present study speakers only and comparing the normalized data would thus not make sense.

	Crystal & House (1988)			**van Santen (1992)**				**Present st. (2011)**	
	Slow	Fast	**All sp.**	Unstr.	Prim.	Secon.	**All c.**	**BrE**	**CzE**
Short vowels	75	68	**71**	79	107	105	**96**	**58**	**64**
Long vowels	140	119	**129**	120	160	148	**144**	**97**	**96**
Diphthongs	188	163	**175**	151	184	160	**168**	**120**	**139**

Table 16 Comparison of the present study data with the results of Crystal & House (1988) and van Santen (1992). All data given in ms. The mean values of Crystal & House are given for their slow speakers, fast speakers and all speakers without distinction. The mean values of van Santen are for different stress conditions – unstressed, primary stress, secondary stress and all stress conditions. The present study data is the raw data (BrE – British speakers, CzE – Czech speakers).

The relationship of the three categories – diphthongs the longest and short vowels the shortest - is confirmed in all cases of the previous research. However, otherwise the comparisons are not so straightforward. It can be seen in Table 16 that the present study data for both the British and Czech speakers approximate Crystal & House's fast speakers and van Santen's unstressed stress condition the most. This part of the present study data was not differentiated by stress condition and since the majority of the vowels are unstressed, it is natural that our values should approximate van Santen's unstressed (i.e. also the shortest) condition out of the three possible options. However, all the present study values are even shorter than in even those two cases named above. Nevertheless, this is only a very gross comparison and it should also not be forgotten that the data from both cited studies was of American English and not of British English which was our reference variety.

Yet there is a problem with the raw values that the articulation rate of the other speakers is not really known and consequently, not the raw values themselves but the ratios among the categories are more important and should be also compared. The ratios given in Section 8.1.1 are reprinted in Table 17 and compared to counted ratios from Crystal and House (1988) and van Santen (1992).

	Crystal & House			**van Santen**				**Present st.**	
	Slow	Fast	All sp.	Unstr.	Primary	Secon.	All c.	BrE	CzE
Long to Short	1.87	1.75	1.82	1.52	1.50	1.41	1.50	1.67	1.50
Diphth. to Short	2.51	2.40	2.46	1.91	1.72	1.52	1.75	2.07	2.17

Table 17 Ratio of long to short vowels and ratio of diphthongs to short vowels. Comparison of the present study data (BrE – British speakers, CzE – English speakers) to Crystal & House, 1988 (slow speakers, fast speakers and all speakers without distinction) and van Santen, 1992 (different stress conditions – unstressed, primary stress, secondary stress and all stress conditions).

It may be observed that Crystal and House's ratios for all their groups are much larger than van Santen's ratios. It is interesting that Czech speakers' long to short ratio is closer to all van Santen's values whereas that of out British speakers approaches Crystal and House's fast speakers the most. Were it not for the slow speakers of Crystal and House, we could hypothesize that van Santen's speaker spoke clearly and slowly and thus would be more alike to the rather slower Czech speakers. However, the data of Crystal and House do not fit into this hypothesis since their slow speakers' long to short ratio is even larger than the one of their fast speakers. The ratio of diphthongs to short vowels of our British speakers lies closest to the value in van Santen's unstressed position. The Czech speakers' value is in the middle between van Santen's unstressed and Crystal & House's fast group data.

Consequently, we may see that comparing the present study data to previous research is an uneasy task since the variability in the data seems quite extensive. It must not be forgotten that just like Czech English differs from British English, British English also differs from the American varieties and this difference is reported to be largest exactly in vowels. Then there might be many unknown factors that might play a role, e.g. the articulation rate, the context of the vowels, or exact segmentation procedure. Moreover, the present research presupposes the British variety to be the model for most Czech speakers including the ones who took part in our study. Thus although the comparison with previous research might seem to relativize our conclusions, the comparison of the Czech data with our British data is more important than data gathered in a different manner and different conditions.

It follows that, just as has been pointed out in Section 8.1.1, the present data Czech English speakers' proportions between the individual vowel categories are slightly distorted

when compared to their British colleagues. Either their short vowels and diphthongs are too long or their long vowels are too short, as has been revealed by the raw and normalized data respectively. Nevertheless, it can be argued that since length is not truly phonological in English, as noted in Section 4.1, the durational deficiency as observed in the Czech English long vowels might not be so important as long as the corresponding long and short vowels adequately differ in their spectral properties. However, the present study does not provide any data about vowel quality. Yet from observation it may be hypothesised that Czech English speakers experience problems not only with vowel quantity but also with the quality.

To proceed to the duration of individual short vowels, the British mean durations confirm Klatt's (1976) and Skarnitzl and Machač's (2007) general observations about the inherent duration of vowels, just as has been expected. The more open the vowel is, the longer its duration. This is based on physiological constraints and is to be seen in many languages. This general claim is closely related to the case of [æ] which the present study has indicated to be the most problematic short vowel for the Czech speakers. The Czech English [æ] was by more than 20 ms shorter than its British counterpart. Based on the openness-to-duration correlation, we could hypothesize that the [æ] as realized by Czech speakers is not open enough. This is exactly what was confirmed in Šturm (2011) in a study examining the spectral properties of Czech English [æ]. He reports that his Czech speakers pronounced this vowel as more closed and fronted towards Czech [ɛ]. His informal observation that some Czech speakers might lengthen their [æ] in order to compensate for the lack of openness (Šturm, 2011, p. 44) was, however, not confirmed in our data - rather to the contrary. The lack of openness in Czech English [æ] seems to correlate with its shorter duration.

The present study data has also confirmed the influence of stress on vowel duration. Both Czech and British speakers' stressed vowels were significantly longer than their unstressed ones. However, the difference between the stressed and unstressed vowels of the British speakers was significantly larger than that of the Czech speakers. The present study Czech speakers thus did not distinguish stressed and unstressed vowels quite adequately. As the raw data suggests they did not temporally reduce the unstressed vowels to the same degree as the British speakers. However, after the data were normalized, the results suggest that it was rather the stressed vowels which diverged - that the Czech English stressed vowels were too short in comparison with the British ones. The reality might be somewhere in between. It is possible that the Czech English unstressed vowels were slightly longer and their stressed vowels slightly shorter. After all, it is the difference between them that matters. The present

study thus confirms the findings by Volín (2005) and Volín & Poesová (2008) who discovered a very similar relationship. Consequently, smaller distinction between stressed and unstressed syllables might contribute towards the lack of rhythmicity in Czech English speech.

However, the present study introduces a new finding worth further examination - the three vowel categories might raise divergent difficulties to Czech speakers when trying to establish the stressed-unstressed contrast. The results show that it was the long vowels that constituted the greatest problem to the Czech speakers. They were not able to distinguish the long stressed and unstressed vowels to a significant degree. Even though the distinction in the short vowels was also smaller than at the British speakers, the long vowels are probably the ones that might be truly problematic. We may hypothesize that the shorter long stressed vowels contribute to the relative shortness of the Czech English vowels as a group. However, with regard to the influence of stress it needs to be reminded that the stress was marked canonically, i.e. in positions where it is realized by a typical native speaker. That means that the smaller variability in the present data Czech speakers might be influenced also simply by the fact that they might not stress all words correctly. Yet even if this was the case, the non-native durational patterns with regard to stress might be confusing for the native speakers and it also might reduce the non-native speakers' intelligibility.

The other examined factor was the influence of an intonation phrase boundary on the preceding two syllables. It has been found that the vowel in the penultimate syllable before the boundary did not get lengthened to a great degree. However, the vowel of the final syllable was affected by the pre-boundary lengthening truly significantly – it was almost 50 ms longer than in any other position in the phrase. Since the present study did not examine different strengths of boundary, we cannot confirm nor disprove Klatt's observation that at sentence-internal phrase boundaries only the last syllable gets lengthened whereas before utterance boundaries the lengthening might stretch over several syllables (1976, p.1212). However, it is most likely that in the present data there were significantly more sentence-internal boundaries than utterance-final ones. Consequently, our results might give some support to the claim that sentence-internal phrase-final lengthening affects only the final syllable before the boundary.

With regard to differences between the two groups of speakers, the Czechs showed significantly smaller pre-boundary lengthening of segments in the ultimate syllable than their British colleagues (more than 15 ms in the normalized data). This is harder to account for because pre-boundary lengthening is a feature reported also in Czech although the degree of

lengthening might differ in the two languages. Thus the smaller extent of pre-boundary lengthening is probably not due to L1 interference. It is possible that the differences might have been due to non-fluency on the part of some Czech speakers and their incorrect phrasing or their sudden breaking off which was then marked as a boundary. On the other hand, even one of the British speakers (BrE 3) had minor problems with fluency in certain parts of the text.

Finally, the last observation with regard to vowels is that boundary and stress seem to interact in their lengthening effect on vowels. Stressed phrase-final vowels were lengthened to a greater degree than unstressed phrase-final vowels when contrasted to the corresponding vowels in other positions. It is interesting that this difference was larger in the Czech speakers than in their British colleagues who lengthened both stressed and unstressed phrase-final vowels to a similar degree.

9.2 Consonants

In contrast to the vowel categories, no significant differences have been found between the Czech and British speakers in the duration of consonant categories. The durational relationships between the categories were also relatively standard. As seen in Table 18, Crystal and House's (1988) consonant categories correspond to the present study durations in the inter-category relationships. The affricates are overwhelmingly the longest, followed by fricatives and plosives. In the case of Crystal and House, fricatives are slightly longer than plosives for all their speakers. Then nasals follow for both the present study and their data. However, their nasals are much closer to plosives in their duration than in the case of both British and Czech speakers. Liquids are followed by glides which are the shortest consonant group.

	Crystal & House (1988)			Present study (2011)	
Category	**Slow**	**Fast**	**All**	**BrE**	**CzE**
Plosives	76	66	71	81	93
Fricatives	80	68	74	81	93
Affricates	123	105	114	98	114
Nasals	76	60	68	68	74
Liquids	72	64	68	55	65
Glides	63	58	60	48	62

Table 18 Comparison of the present study data with the results of Crystal & House (1988). All data given in ms. The mean values of Crystal & House are given for their slow speakers, fast speakers and all speakers without distinction. The present study data is the raw data (BrE – British speakers, CzE – Czech speakers).

Also as expected, a distinction has been found between voiced and voiceless consonants. The contrast was observed on plosives and fricatives and it slightly but significantly differed between the two categories. Voiceless fricatives were 28 ms longer than their voiced counterparts whereas the difference between voiceless and voiced plosives was almost 35 ms. Thus the distinction in fricatives in the present data was smaller than the 40 ms reported by Klatt in his section on inherent duration (Klatt, 1976, p. 1213).

Just like in the case of vowels, a most conspicuous digression of the Czech speakers in the mean duration of individual segments was chosen. It was the case of the liquid [r] and the glide [w]. Both consonants were significantly longer (by about 20 ms) when realized by Czech speakers than the ones of their British colleagues. Some possible reasons for this difference might be hypothesized. Neither of the consonants is native to the Czech speakers. Although there is a [r] in Czech, there it is a trill and not a liquid which is the most common realization in standard British English. In the case of [r] we might further hypothesize a possible influence of the Czech speakers' rhoticity. From observation it seems that rhotic, postvocalic [r] might be in general longer than prevocalic or intervocalic one. Moreover, since a rhotic [r] may occur also in the coda of a word-final syllable which is not the case in non-rhotic accents, it may thus be subject to pre-boundary lengthening if this word-final syllable is also phrase-final. Due to rhoticity each of the Czech speakers produced about 50 more [r]s than any of the British speakers (see Appendix 3) and consequently the influence of these postvocalic [r]s might be quite large.

With regard to pre-boundary lengthening in consonants, the situation was similar to that of vowels. Again there was significant lengthening observed in the final syllable and again Czechs lengthened the segment in this syllable slightly less than the British speakers. The only difference between vowels and consonants seems to be that consonants as realized by the British speakers were slightly, but significantly lengthened also in the penultimate syllable.

Finally, the mentioned pre-boundary lengthening in the ultimate syllable had divergent effects of on different consonant categories. The influence of the lengthening on fricative, liquids and nasals was larger than on plosives. This might be partly caused by the aerodynamic similarities of fricatives and liquids because both categories are continuants and thus they can be pronounced for a longer periodd of time than plosives. With nasals the explanation might partly lie in similar reasoning. During the hold phase, the air flows out through the nasal cavity and reduces thus the supra-glottal pressure in the oral cavity (which

is the reason for relatively short duration of plosives). Another explanation for the relatively large apparent lengthening of nasals might be that the word-final and thus potentially also phrase-final syllable is often occupied by first, a velar [ŋ] which is inherently a relatively long consonant and second, by a syllabic [n̩] which thanks to its syllabicity is also longer than in other positions. With regard to plosives there might be an unaccounted-for factor at play. The initial plosives are reported to be longer than final ones, especially in English due to aspiration and the segmentation conventions which mark the onset of the following vowel with the first visible formant structure. The influence of the position in a word was not tested in our data, but from the segmentation experience with the present material, it may be hypothesized that it is possible that initial plosives are longer than final ones. Initial plosives do not get to the final syllable before the boundary and thus oppose the lengthened phrase-final plosives by their also relatively long duration. Consequently, the difference between pre-boundary and non-pre-boundary plosives might not be so large as in the case of the other mentioned consonant categories.

9.3 Limitations and suggestions for future research

As noted in Section 4.5, there are many factors which seem to influence segmental duration. Out of these only the influence of stress on vowels and of boundary on both vowels and consonants has been taken into account. From this it follows that the study necessarily has many limitations. In future more extensive research, attention should be paid to more factors and their interference in influencing segmental duration. Out of these, segmental context and position within a word might be suitable candidates. However, also phrasal stress or contrastive stress might be worth the attention.

Yet even the influence of factors which have been analysed by the present study might be examined in greater depth. First, it might be also interesting to inspect the influence of stress on consonants or the possible influence of secondary stress. Second, different strengths of prosodic boundaries might be investigated. It would be definitely beneficial (for speech technologies for instance) to find out whether boundaries of different strength have divergent lengthening effects as suggested by Wightman et al.'s (1992) results.

Another area worth attention is the effect of articulation rate on segmental duration. Although a normalization procedure was used in order to remove the differences in articulation rate of the speakers, we may ask whether this normalization procedure was adequate. By multiplying all segments by a coefficient based on the speaker's articulation

rate, all vowels and consonants were shortened or lengthened to the same degree. Yet evidence accumulates that different categories of vowels and consonants are affected by changes in articulation rate to a different degree. Speaking faster or slower does not mean just compressing or stretching the soundwave. If this is done, we know from experience that speech might turn incomprehensible or, at the very least, unnatural. Thus it is a question whether normalized or raw, non-normalized values are more accurate or more valid. The disadvantage of raw values is that divergent articulation rates can constitute a significant unaccounted for factor in segmental duration. However, with normalization the values might be distorted by the normalization itself. Therefore, it would be valuable to try to establish a different normalization method which would better account for ways in which changing articulation rate changes the individual types of segments.

Finally, it might be also useful to gain more controlled data in which the numbers of occurrences of individual segmental categories, of the factors and their combination would be more balanced. Since this might be relatively challenging, our preliminary conclusions should be at least tested on a much larger data sample. Also controlled data might not provide an adequate description of every-day speech. Consequently, both types of research are needed: experiments with careful speech which would control for the factors which might influence segmental duration as well as examination of large databases of spontaneous or semi-spontaneous speech to gain a more specific idea of the real-life data.

10 CONCLUSION

As noted in the theoretical introduction, English is becoming the means of communication of an increasing number of people around the world and the Czech Republic is no exception. Also here many people learn English as a foreign language and strive to be successful in communication with both native speakers of English and speakers of other languages who have English also only as their second or foreign language. To communicate effectively, one needs to understand and be understood. Since speech still remains the main tool of interpersonal communication, the degree to which a language learner masters the spoken medium is vital. However, many language learners speak their target language with a foreign accent which has been found to affect the communicative process. The consequences of a strong foreign accent might be various - from reduced intelligibility to divergent social evaluation on the part of the listener. Therefore, most language learners want to do their best to learn to speak without a foreign accent.

Although the highest level of accentedness - native-like accent - has been regarded to be unattainable for the majority of adult learners by many specialists, it is by no means fruitless to try to learn "correct" pronunciation of the target language. It has been proposed that if pronunciation instruction focuses on appropriate areas, the degree of accent and the level of intelligibility might be significantly improved. Yet the difficulty is that researchers and teachers have not yet settled on a definite and satisfactory answer to the question where exactly the focus should be. Consequently, more research is still needed concerning the nature of foreign accent and also with respect to what features cause reduced intelligibility and how to be more comprehensible as a foreign language speaker. The present study aims to contribute to such a research.

This exploratory study has sought to map a sample of Czech accented English in the area of segmental duration by comparing and contrasting it to a reference variety – British English. Segmental duration was examined in the read speech of three Czech and three native British speakers. The time domain of speech with the associated areas of stress and rhythm seems to play a significant role in speech intelligibility and accentedness. The temporal plan of speech is constructed of individual segmental durations which together compose into duration of syllables and larger units. Therefore, we hypothesise that accentedness might partially lie also in such miniature units as segmental durations. Consequently, by exploring the durational behaviour of the individual segments and phonetic classes of speech sounds in

Czech English, the present study hopes to bring the debate about the appropriate focus of pronunciation instruction a small step closer to its denouement.

The present study inspected the behaviour of vowel and consonant classes of speech segments as realised by the six Czech and British speakers. The most notable instances of divergence between the Czech and British speakers in individual segment duration were also observed, namely the case of [æ] among the vowels and [r] and [w] from the group of consonants. Finally, segmental duration was also examined with regard to two potential factors which have been reported to influence it – stress and intonation phrase boundary.

The data revealed that the Czech speakers had slight difficulties with the inter-group proportions among the vowel categories. Their long vowels seemed to be a little too short (or their short vowels and diphthongs too long) in comparison with the British speakers. Also they did not manage to differentiate the stressed and unstressed vowels to the same degree as the native speakers. Specifically, the biggest difference was again in their long vowels which did not differ significantly in the duration in stressed and unstressed positions.

However, there was no significant difference in the Czech and British English durations of the consonant categories. The Czechs were found to produce parallel proportions between the consonant groups like their British colleagues. In respect of the intonation phrase boundary, it influenced the phrase-final syllable with high significance in the present data but had much smaller (if any) effect on the penultimate syllable. This was true for both vowels and consonants in the positions before the boundary. Moreover, it was found that pre-boundary lengthening affects different vowel and consonant categories to a disparate degree. The consonant category that was affected the most was the fricatives whereas in vowels it was the diphthongs. Like in the case of stress, Czechs were found to lengthen segments in the phrase-final syllable to a smaller degree compared to the British speakers. This again applies to both vowels and consonants. With regard to the above-mentioned individual most differing segments, the Czech [æ] was too short while their [r] and [w] were too long in comparison with the mean durations of the native speakers.

In conclusion, it might be hypothesized that the detected durational features of the present study Czech speakers might contribute to their spoken English being assessed as strongly accented although there are certainly also many other features which add to their accent. Consequently, the detected features should be tested on a greater variety of Czech accented English in order to establish whether they significantly contribute to strong Czech accent and whether they should be part of the focus of structured pronunciation instruction.

REFERENCES

Abello-Contesse, C. (2009). Age and the critical period hypothesis. *ELT Journal, 63*(2), 170-172.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.28. Retrieved from http://www.praat.org/ on 28 June 2011.

Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of a foreign accent. *Phonetica, 63*, 247-267.

Campbell, G. L. (1995). *Concise compendium of the world's languages*. London: Routledge.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*, 372-387.

Crystal, D. (2003). *English as a global language* (2nd ed.). New York: Cambridge University Press.

Crystal, D. (2006). Into the twenty-first century. In L. Mugglestone (Ed.), *The Oxford history of English* (pp. 394-413). New York: Oxford University Press.

Crystal, D. (2008). *Dictionary of linguistics and phonetics* (6th ed). Oxford: Blackwell Publishing.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *J. Acoust. Soc. Amer., 72*(3), 705-716.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *J. Acoust. Soc. Amer., 83*(4), 1553-1573.

Cummins, F. (1999). Some lengthening factors in English speech combine additively at most rates. *J. Acoust. Soc. Amer., 105*(1), 476-480.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition, 19*, 1-16.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*, 379-397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Lang. Teach., 42*(4), 476-490.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1998). Evidence in favor of a broad pronunciation instruction. *Language learning, 48*(3), 393-410.

Duběda, T. (2005). *Jazyky a jejich zvuky: univerzálie a typologie ve fonetice a fonologii*. Praha: Karolinum.

Flege, J. E., Munro, M. J., & McKay, I. R. A. (1995) Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Amer., 97*(5), 3125-3134.

Iverson, P. & Evans, B.G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *J. Acoust. Soc. Amer., 122*(5), 2842-2854.

Kohler, K. J. (2009). Rhythm in speech and language: A new research paradigm. *Phonetica, 66*, 29-45.

Králová, K. (2010). Czenglish: A basic outline of an EFL variety. Unpublished MA thesis, ÚAJD, UK FF.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Amer., 59*, 1208-1221.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishing.

Leather, J., & James, A. (1996). Second language speech. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 269-316). San Diego: Academic Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253-263.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica, 66*, 169-187.

Low, L. E., Grabe, E., & Nolan, F. (2000). *Language and Speech, 43*(4), 377-401.

Machač, P., & Skarnitzl, R. (2007). Temporal compensation in Czech? In *Proceedings of the 16th ICPhS* (pp. 537-540). Saarbrücken: Organizing Committee.

Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Praha: Nakladatelství Epocha.

McArthur, T. (2006) English word-wide in the twentieth century. In L. Mugglestone (Ed.), *The Oxford history of English* (pp. 360-393). New York: Oxford University Press.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech, 38*, 289-306.

Munro, M. J., Derwing, T. M., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect, 21*(1), 67-79.

Noteboom, S. (1997). The prosody of speech: melody and rhythm. In W. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 640-673). Oxford: Blackwell Publishing.

Ondrušková, L. (2011). Zvukové vlastnosti jednoslabičných slov v semispontánním dialogu a hlasitém čtení (Acoustic properties of monosyllabic words in semi-spontaneous dialogues and read speech). Unpublished MA thesis, FU, UK FF.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29*, 191-215.

Podlipský, V. J., Skarnitzl, R., & Volín, J. (2009). High front vowels in Czech: a contrast in quantity or quality? In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)* (pp. 132-135). Brighton: ISCA.

Port, F. P. (1981). Linguistic timing factors in combination. *J. Acoust. Soc. Amer., 69*(1), 262-274.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

Quené, H., & van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication, 52*, 911-918.

Quené, H., & Port, R.F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*, 1-13.

Roach, P. (2000). *English phonetics and phonology* (3rd ed.). Cambridge: Cambridge University Press.

Roach, P., Hartmann, J., & Setter J. (Eds.) (2003). *English Pronouncing Dictionary* [CD-ROM]. Cambridge University Press.

Sheppard, C., Hayashi, C., & Ohmori, A. (2007). Factors accounting for attainment in foreign language phonological competence. In *Proceedings of the 16th ICPhS* (pp. 1597-1600). Saarbrücken: Organizing Committee.

Strange, W. (1999). Levels of abstraction in characterizing cross-language phonetic similarity. *Proceedings of the XIVth International Congress of Phonetic Sciences, 3* (pp. 2513-2519). San Franscisco.

Šturm, P. (2011). Akustická analýza vokálu /æ/ v české výslovnosti angličtiny (Acoustic analysis of the vowel /æ/ in the Czech pronunciation of English). Unpublished BA thesis, FU, UK FF.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25,* 1-24.

Turk, A. E. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28,* 397-440.

van Santen, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication, 11*, 513-546.

Volín, J. (2000). Proč je /ŋ/ obtížnou hláskou pro české studenty angličtiny. *ATE Newsletter*, 11/1, 51-57.

Volín, J. (2005). Rhythmical properties of polysyllabic words in British and Czech English. In J. Čermák, A. Klégr, M. Malá, & P. Šaldová (Eds.), *Patterns: A Festschrift for Libuše Dušková* (pp. 279-292). Praha: Klub moderních filologů.

Volín, J. (2010). On the significance of the temporal structuring of speech. In M. Malá, & P. Šaldová (Eds.), *... for thy speech bewrayeth thee: A festschrift for Libuše Dušková* (pp. 289-305). Praha: UK FF.

Volín, J., & Poesová, K. (2008). Temporal and spectral reduction of vowels in English weak syllables. In A. Grmelová, L. Dušková, M. Farrel, & R. Pípalová (Eds.), *Plurality and diversity in English studies* (pp. 18-27). Praha: UK PedF.

Volín, J., & Skarnitzl, R. (2010a). Suprasegmental acoustic cues of foreignness in Czech English. In *Proceedings of Odyssey: The Speaker and Language Recognition Workshop Odyssey 2010* (pp. 271-278). Brno: VUT.

Volín, J., & Skarnitzl, R. (2010b). The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication, 52,* 1010-1021.

Wells, J. C. (1982). *Accents of English 2: The British Isles.* Cambridge: CUP.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Amer., 91*(3), 1707-1717.

Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System, 36,* 69-84.

LIST OF APPENDICES

APPENDIX 1: Text

BBC news with Jackie Leonard

The former United States president, Jimmy Carter, is in the Cuban capital Havana for a five-day visit. He's the first acting or former president to visit Cuba since the country's communist leader, Fidel Castro, came to power in 1959. Daniel Schweimler reports from Havana.

The main political party in the Israeli coalition government, Likud, is discussing whether it should block any future attempts to declare an independent Palestinian state. But the Israeli prime minister Ariel Sharon has urged members of his party not to vote on the resolution. He said it would be against Israel's interests to rule out any future settlement which included the creation of a Palestinian state. From Jerusalem, Michael Voss reports.

A terminally ill British woman, who lost a high-profile legal battle to allow her husband to help her commit suicide, has died. Diane Pretty, who was forty-three, had been suffering with motor-neural disease for several years. The family says she began experiencing breathing difficulties ten days ago and died at a hospice on Saturday. Diane Pretty took her case all the way to the European Court of Human Rights in an attempt to gain permission for her husband to help end her life.

You're listening to the news from the BBC in London.

There have been outbreaks of ethnic violence in Madagascar as the political deadlock continues between the newly declared president Mark Ravallo Manana and his rival the long-standing president Didier Ratsirack, who's refusing to step down. A human rights group says six people have been killed in a town in the west of Madagascar, from where Alistair Leathead reports.

The Russian government has sent a specialist civil emergency team to the Bajkonur Space Centre in Kazakhstan to reach some eight people trapped after part of the building collapsed. They were repairing the roof of one of the hangars used for assembling and testing space vehicles when part of it crashed eighty meters to the ground. The space centre dates from the nineteen fifties and was the place where the Soviet Union launched the first man-made satellite Sputnik.

The International Press Institute has criticized governments around the world for limiting civil liberties in the name of fighting terrorism. Delegates meeting in Slovenia issued a statement saying it was dangerous to limit civil liberties under the pretext of combating terrorism. The statement also said the struggle against international terrorism had left governments seeking dangerous controls over the free flow of information and freedom of expression.

Delegates at a conference in Bangladesh aimed at preserving one of the world's largest mangrove forests, the Sunderburns, have agreed to cooperate with conservation efforts. The Sunderburns, home to the royal Bengal tiger, is described as one of the last great coastal wetlands, but it's seriously threatened by pollution and human encroachment. The forest straddles the border between India and Bangladesh.

BBC news

APPENDIX 2: Number of analysed occurrences of segments in individual phonetic categories for individual speakers

	British speakers			Czech speakers			Total of analysed
	BrE 1	**BrE 2**	**BrE 3**	**CzE 1**	**CzE 2**	**CzE 3**	
All vowels	747	707	717	757	754	751	4433
Short vowels	557	520	529	582	578	575	3341
Long vowels	92	90	97	93	92	90	554
Diphthongs	98	97	91	82	84	86	538
All consonants	1230	1203	1179	1271	1261	1263	7407
Plosives	408	407	378	412	408	413	2426
Voiceless	268	267	242	266	270	271	1584
Voiced	140	140	136	146	138	142	842
Fricatives	359	346	351	351	353	360	2120
Voiceless	212	205	203	202	206	212	1240
Voiced	147	141	148	149	147	148	880
Affricates	14	14	14	13	14	14	83
Nasals	238	234	231	231	231	231	1396
Liquids	162	157	160	218	208	199	1104
Glides	49	45	45	46	47	46	278

BrE 1, 2, 3 – individual British speakers
CzE 1, 2, 3 – individual Czech speakers.

APPENDIX 3: Number of occurrences of individual speech segments for individual speakers

PART A – Vowels

VOWELS		British speakers						Czech speakers					
		BrE 1		BrE 2		BrE 3		CzE 1		CzE 2		CzE 3	
		A	N	A	N	A	N	A	N	A	N	A	N
Monophthongs	*Short vowels*												
	ɪ	174	4	164	15	171	13	180	3	186	2	184	1
	e	56		57		56	1	63		61		61	
	æ	37		35	1	33	2	39		37		35	1
	ʌ	21		23		21		21		18		21	
	ə	247	7	218	28	229	17	255	8	252	7	248	7
	ɒ	17		15	1	16	1	18		16	1	16	1
	ʊ	5		8		3	1	6	1	8		10	
	Long vowels												
	iː	33		34	1	34	2	33	1	34		33	1
	ɜː	10		10		11		10		8		10	
	ɑː	9		9		12	1	11		12		9	
	ɔː	13		12		13		12		11		12	
	uː	27		25		27	2	27		27		26	
Diphthongs	*Diphthongs*												
	eɪ	36		35	1	35	2	34	3	35	3	37	2
	aɪ	25		26		23	2	25		25		25	
	ɔɪ	1		1		1		1		1		1	
	aʊ	8		8		7		7	1	8		8	
	əʊ	16		16		16		15		15		15	1
	ɪə	4		3		1	1	0		0		0	
	eə	6		6		5	1	0		0		0	
	ʊə	2		2		3		0		0		0	

BrE 1, 2, 3 – individual British speakers
CzE 1, 2, 3 – individual Czech speakers

A - Number of analysed speech segments for each speaker
N - Number of not analysed speech segments for each speaker. These were either canonical segments that were not realized by the speaker (they were elided) or segments that were discarded from the analysis because of alterations, dysfluences or hesitations.

PART B – Consonants

CONSONANTS			British speakers						Czech speakers					
			BrE 1		BrE 2		BrE 3		CzE 1		CzE 2		CzE 3	
			A	N	A	N	A	N	A	N	A	N	A	N
Obstruents	*Plosives*													
	Voiceless	p	49		50		48	2	48	2	50		48	1
		t	167	8	166	9	143	32	166	8	168	6	173	2
		k	52		51	1	51	1	52	13	52	2	50	3
	Voiced	b	38		38		38		38		37		38	
		d	78	5	79	5	73	12	83		76	5	79	2
		g	24		23		25	2	25		25	1	25	1
	Fricatives													
	Voiceless	f	41		41		39		41		42		42	
		θ	4		4		4		3		3		3	
		s	115	1	111	4	111	10	108	7	111	3	115	2
		ʃ	19		18	1	19		18		17	2	18	1
		h	33	1	31	3	30	6	32	3	33	2	34	
	Voiced	v	40	1	40	2	41		40	2	40	1	40	1
		ð	43	7	42	8	48	3	49	3	47	2	48	
		z	64		59	1	59	1	60	3	60		60	1
		ʒ	0		0		0		0		0		0	
	Affricates													
		tʃ	6		6		6		6		6		6	
		dʒ	8		8		8		7	1	8		8	
Sonorants	*Nasals*													
		m	58		58		58	1	56	1	57	1	57	2
		n	155	1	152	3	148	4	150	1	150	4	151	3
		ŋ	25		24		25		25		24		23	1
	Liquids													
		r	72	11	69	11	75	7	130	4	119	1	112	3
		l	90		88	1	85	5	88	2	89		87	1
	Glides													
		j	24		24		23		22	2	23		23	1
		w	25		21	3	22	2	24		24	1	23	

BrE 1, 2, 3 – individual British speakers
CzE 1, 2, 3 – individual Czech speakers

A ... Number of analysed speech segments for each speaker
N ... Number of not analysed speech segments for each speaker. These were either canonical segments that were not realized by the speaker (they were elided) or segments that were discarded from the analysis because