# The Voicing Contrast in Serbian Stops

Mirjana Sokolović-Perović

Thesis Submitted for the Degree of Doctor of Philosophy

School of Education, Communication and Language Sciences
Newcastle University

September 2012

# Abstract

This study investigates aspects of the phonetic realisation of the voicing contrast in Serbian stops. It determines the basic set of acoustic correlates of the voicing contrast, and examines the effect of several linguistic and speaker factors on these correlates. The thesis explores fine details of the phonetic realisation of the voicing contrast that are specific to Serbian, and evaluates the existing theoretical models of the voicing contrast in relation to Serbian data.

Twelve native Serbian speakers produced stops in a range of positions in isolated words and in a sentence frame. Acoustic analysis revealed that the voicing contrast is robust in Serbian in all word positions and for each speaker. Utterance-initially Serbian contrasts prevoiced stops and stops with short to intermediate positive VOT values. In word-initial intervocalic position the relevant correlates are duration of voicing in the closure and closure duration; in word-medial and final position the correlates are duration of voicing in the closure, closure duration, and preceding vowel duration. The following factors affect the realisation of the voicing contrast: the place of stop articulation, the vowel environment, gender, age, and place of birth of speakers. This variability is only partly attributable to universal constraints, and is mostly specific to Serbian.

The results suggest that the existing models cannot account for the type of realisation of the voicing contrast found in Serbian, in particular for the status of intermediate VOTs and the role of closure duration and preceding vowel duration. Some of the main assumptions of these models should be re-assessed in order to include these findings. Further, these models are unable to account for non-universal and non-contrastive variability found in Serbian and other languages. Advantages and difficulties associated with the integration of the existing models with elements of an exemplar-based model of phonological knowledge are discussed.

# Acknowledgments

I would like to thank my supervisor Gerry Docherty for his support, suggestions, and feedback, for his patience and his scientific rigour, and for encouraging me, little by little, to step outside my comfort zone. He is a great inspiration, and I feel very privileged to have had him as my supervisor.

I would also like to express my gratitude to Ghada Khattab, Katya Samoylova, and Jalal Al-Tamimi for their positive comments and suggestions at my annual review panels. I am also particularly grateful to Gerry, Ghada, and Katya for giving me a teaching opportunity, which was a great experience, and brought me immense satisfaction. Many thanks to Jalal for his help in detangling some statistical issues, and to Katya for her practical advice and for our chats. And thank you all for your efforts in running the Phonetics Research Group, which provided a lively and stimulating research atmosphere for my work.

The participants in this study deserve special mention, as do people who helped me recruit them. This study would not have been possible without their time and effort, and I cannot thank them enough.

A big thank you to my fellow students and friends from office 2.05 and from Phonetics Research Group, who made my time in Newcastle very special. I will never forget our chats, our food parties, and our outings.

Another big thank you to my friends in Serbia, who supported me from the distance, and who believed in me at every step. Special thanks go to Danijela for always being there, through good and bad times, and for all those last-minute reference checks; and to Jordana for helping me stay in touch with the latest research in Serbia.

Last, but not least, a special thank you to my family for their support and love, for their patience and for their belief in me on this journey.

# Table of contents

# List of figures

# List of tables

# Introduction

One of the main questions of phonetic theory has been the relationship between phonological representations and the continuous activity of the vocal tract leading to their realisations. The voicing contrast has been in the centre of this research for many decades.

Many languages have a contrast in obstruents that is traditionally described as the voicing contrast, although there is an on-going debate about the nature of phonological and phonetic categories of this contrast. What is usually described as the voicing contrast seems straightforward, but when it is examined in depth, the relevant phonetic dimensions and details of its phonetic realisation appear to be more complex than the straightforward descriptions would suggest. For stops, the traditional intuitive view that on the phonetic level the contrast is simply realised as the contrast between vocal fold vibration during the closure and its absence is not supported by the evidence from languages that in stops do not use voicing contrastively, but instead use aspiration. For a number of years the phonetic aspect of the voicing contrast in stops was related to the notion of Voice Onset Time (VOT), and based on very few languages. However, the range of the phenomena that need to be documented and accounted for is much greater, including a number of other acoustic correlates and a range of factors that can affect their realisation. It was only recently that Cho and Ladefoged (1999) brought to our attention the fact that there is much more variability in the VOT than previously thought, something which we can expect to find in other languages, and in other correlates. The aim of the present study is to look at the issue of phonetic realisation of the voicing contrast in stops in a language that has not been studied before – Serbian.

The present study is organised as follows. Chapter 1 gives an overview of the previous research on acoustic correlates of the voicing contrast in stops. It introduces the most important and best-researched correlates in a number of languages, as well as several factors that have been found to influence the realisation of these correlates, and demonstrates the complexity of the phonetic realisation of the voicing contrast. I highlight the fact that there is a lack of research on languages that use voicing contrastively, and a lack of systematic research on correlates other than VOT and on non-initial word positions.

Chapter 2 presents a review of several theoretical models of the voicing contrast. I point out that despite fundamental differences in how they envisage the relationship between phonological representations and their phonetic realisations, these models have in common that they are biased towards the acoustic correlates that are more relevant for languages that use aspiration contrastively, rather than voicing, and they are unable to account for the complex patterns of phonetic realisation reviewed in Chapter 1. In the second part of the chapter, predictions that these models make for Serbian are analysed, and the aims of the present study outlined. Finally, Chapter 2 gives a review of the existing literature and of the features of Serbian sound system that are relevant for the present study.

Chapter 3 describes the methodology used in the present study, including details about subjects, data collection, acoustic analysis, and statistical analysis.

Chapters 4 to 7 present experimental results for the following acoustic correlates: VOT, closure duration, voicing in the closure, and preceding vowel duration, respectively. The significance of each correlate for the voicing contrast in Serbian is examined, both in the pooled data and for each subject, and for relevant word positions. Several factors that can affect the realisation of these correlates are also investigated, and the findings discussed.

Chapter 8 provides a summary of results, outlining the realisation of the voicing contrast in Serbian in each word position, and the nature and the extent of the variability induced by the factors examined. These results are further discussed in relation to several of the theoretical models described in Chapter 2, especially with respect to their ability to account for the type of phonetic realisation found in Serbian, and for the observed variability, and some implications for the models are discussed. Finally, limitations of the present study and an outline of future work are presented.

# Chapter 1 Research on acoustic correlates of the voicing contrast in stops

In this chapter I review the existing literature about the most important acoustic correlates of the voicing contrast in stops. Because our knowledge about the phonetics of the voicing contrast is mainly based on acoustic analysis, this literature review primarily focuses on acoustic studies. Perceptual and articulatory studies are not discussed in great detail, unless it is necessary for a particular topic.

The following acoustic correlates are reviewed: Voice Onset Time, closure duration, voicing in the closure, properties of the release burst, preceding vowel duration, and frequency of the first formant and fundamental frequency at voicing onset and offset. The first four correlates are properties of the consonant itself, while remaining correlates are found in the preceding and the following vowel. In addition to this, a number of linguistic and speaker factors that have been found to affect the phonetic realisation of the voicing contrast are also discussed in this chapter, as well as differences in phonetic realisation of the voicing contrast across languages.

In the present study the terms *phonologically voiced stops* and *phonologically voiceless stops* are used to refer to phonological categories. The terms *voiced* and *voiceless* are used to refer to phonetic voicing, that is, to periods with or without vocal fold vibration. The term *stop* is used to refer to oral stops.

## 1.1 Voice Onset Time

### 1.1.1 Categories of Voice Onset Time

Lisker and Abramson (1964, 1965) proposed that Voice Onset Time (VOT), defined as the time interval between the release of a stop and the onset of laryngeal vibration, could be used to distinguish word-initial aspirated and unaspirated stops, as well as voiced and voiceless stops in a number of languages. The motivation behind this proposal was to find an underlying phonetic dimension that would bring together phonetic characteristics of voicing, aspiration and force of articulation in the description of the voicing contrast in stops.

Their analysis of eleven languages suggested that VOT values tend to group into three ranges they called voicing lead, short lag and long lag, with the median VOTs of -100 ms, 10 ms, and 75 ms, respectively. In this sample they found that languages with a two-way contrast use either the voicing lead and short lag categories (e.g. Dutch, Spanish, Hungarian and Tamil), or short and long lag categories (e.g. English and Cantonese), while languages with a three-way contrast, such as Thai and Eastern Armenian, use all three categories (Lisker & Abramson, 1964). Languages with a three-way or four-way contrast will not be discussed here.

Lisker and Abramson also carried out perceptual studies with synthetic speech stimuli, where VOT was varied across the range observed in production, and found that the three VOT categories were perceptually important. In identification tasks native listeners of American English, Latin American Spanish and Thai categorised stimuli into categories that broadly matched the categories observed in production for each language (Abramson & Lisker, 1973; Lisker & Abramson, 1970). In addition to this, discrimination tended to sharpen at the phoneme boundaries, specific for each language, which suggested that to some extent the speakers' ability to discriminate between categories is shaped by their own language experience (Abramson & Lisker, 1970, 1973). Subsequent studies reinforced these findings (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Williams, 1977).

Since the initial research by Lisker and Abramson, numerous studies dealing with various aspects of VOT realisation and perception in a number of languages have added support to the finding that languages with a two-way voicing contrast mainly belong to either the group that contrasts voicing lead with short lag VOT (also called *voicing languages* or *true voice languages*), or to the group that contrasts short and long lag VOT (also called *aspirating languages*[1]). Examples of languages from the former group include Romance and Slavic languages, for example French (Abdelli-Beruh, 2004, 2009; Caramazza & Yeni-Komshian, 1974), Portuguese (Lousada, Jesus, & Hall, 2010), Spanish (Poch-Olivé, 1987; Rosner, López-Bascuas, García-Albea, & Fahey, 2000; Williams, 1977), Polish (Keating, 1980; Rojczyk, 2009), and Russian (Ringen & Kulikov, fc); also Hungarian (Gósy, 2001; Gósy & Ringen, 2009) and Arabic (Flege & Port, 1981; Yeni-Komshian, Caramazza, & Preston, 1977), to mention but a few. Some

[1] In the present study I adopt this terminology: I use the term *voicing languages* for languages that in utterance-initial position contrast prevoiced and zero to short lag VOT stops, and the term *aspirating languages* for languages that contrast zero to short lag VOT stops and long lag VOT stops (after Jansen, 2004).

languages from the Germanic family also belong to this group: Dutch (Slis & Cohen, 1969a; van Alphen & Smits, 2004), Afrikaans, Frisian, Yiddish, Scottish English and Rhineland German (Jansen, 2004, p. 41). To the latter group belong, among others, languages from the Germanic family, such as English (Docherty, 1992; Flege, 1982; Lisker & Abramson, 1964, 1967; Smith, 1978), German (Jessen, 1998), Danish (Fischer-Jørgensen, 1954), Icelandic, Norwegian, and Faroese (see Jansen, 2004, p. 41 and references there); also Cantonese (Lisker & Abramson, 1964), Mandarin (Jessen, 1998, p. 236) and the Turkic languages (Jansen, 2004, p. 41). The number of studies and the depth of research vary greatly from language to language, but in recent years there has been a renewed interest in the issues related to the voicing contrast, and in particular VOT.

An exception to this division is Swedish, which contrasts prevoiced and long lag stops (Beckman, Helgason, McMurray, & Ringen, 2011; Helgason & Ringen, 2008). The number of languages with this type of contrasts could be higher, potentially including, among others, Turkish, Norwegian, Farsi, Swahili, and some dialects of Armenian (see Helgason & Ringen, 2008, and references there).

In addition to this, there are languages in which VOT does not seem to be a relevant dimension. In these languages the opposition between the two stop classes is expressed through closure duration, for example in Zapotec, Jawon, Rembarrnga, and Swiss German, or through burst amplitude, f0 onset, and breathy phonation, such as in Musey (Jessen, 1998, p. 275).

It has been suggested that this may apply to Canadian French as well, for which it was reported that VOT values for the two voicing categories overlap in production and that VOT is not utilised in perception (Caramazza & Yeni-Komshian, 1974; Caramazza, et al., 1973). The authors hypothesised that this could be because of the influence of Canadian English (but cf. Jacques, 1987; Ryalls, Cliché, Fortier-blanc, Coulombe, & Prud'hommeaux, 1997, who reported little or no overlap between the voicing categories). Canadian French seems to pattern with some other languages that exhibit a similar overlap between phonetic VOT categories, possibly because of influence from another language with a different type of contrast, as has been suggested for Dutch (van Alphen & Smits, 2004), and Fenno-Swedish (Ringen & Suomi, 2012).

For a number of years the three phonetic categories of VOT established by Lisker and Abramson were regarded as universal, and any exceptions to this taxonomy were unlikely to be acknowledged. These three phonetic categories were even used as a basis for Keating’s (1984a) phonological model of the voicing contrast (discussed in Chapter 2). However, a growing body of evidence has suggested that situation is much more complex, and a number of authors questioned the universal and categorical nature of this division (Cho & Ladefoged, 1999; Docherty, 1992; Raphael et al., 1995; Scobbie, 2005). Exceptions to the proposed VOT taxonomy fall into three broad categories.

First, there is the evidence of bimodal VOT distribution in the realisation of /b, d, g/ in English and some other languages. English is usually said to contrast unaspirated and aspirated (short lag and long lag) stops, based on Lisker and Abramson’s (1964) result and some consequent studies, but even Lisker and Abramson had instances of prevoicing instead of short lag VOT values (about 20% of tokens), produced mainly by one of their four speakers. Other studies also found tokens of /b, d, g/ realised with prevoicing (Caramazza, et al., 1973; Docherty, 1992; Flege, 1982; Ryalls, Simon, & Thomason, 2004; Ryalls, Zipprer, & Baldauff, 1997; Smith, 1978), some of them as many as 59% of tokens (Flege, 1982), but they also reported between- and within-subject variability in the number of prevoiced tokens.

English is not unique in this respect. Varying percentages of phonologically voiced stops realised with prevoicing instead of short lag VOT were also reported for Turkish (Kallestinova, 2004) and Persian (Bijankhan & Nourbakhsh, 2009; Heselwood & Mahmoodzade, 2007), with similar between-speaker variation.

The second type of evidence against the universality of phonetic VOT categories comes from the absence of any clear (and universal) boundary between unaspirated and aspirated (or short lag and long lag) stops. Although Lisker and Abramson considered short lag stops to have VOT values between 0 and 25 ms, and long lag stops to have VOT values above 60 ms, in a number of languages with a two-way contrast between voicing lead and voicing lag, VOT values fall in the area between 25 and 60 ms, or so called *intermediate* values of VOT (Raphael, et al., 1995; Riney, Takagi, Ota, & Uchida, 2007). Examples of languages from this group include Hebrew (Obler, 1982), Hungarian (Gósy, 2001), Japanese (Riney, et al., 2007; Shimizu, 1989), and Polish (Keating, Mikos, & Ganong III, 1981), to mention but a few. Several of the endangered languages from Cho and Ladefoged’s (1999) study also belong to this category. If studies on bilingual populations are included, the list is much longer. It is the velar /k/ that is most often realised with intermediate VOT values, but not exclusively, because

/p/ and /t/ can also be produced in this way. Intermediate values of VOT will be discussed in detail in Section 8.2.1.

These findings have received little attention. As Scobbie (2005) pointed out, if a language was considered to have only one category of voicing lag, it was automatically assumed that this category was short lag. Any variation that did not fit the established categories of short and long lag VOT was likely to be dismissed as irrelevant (cf. Keating, 1984a). This was reinforced by perceptual work that seemed to suggest that there might be a psychoacoustic, non-linguistic basis for the perception of the contrast between short lag and long lag VOT stimuli (Keating, 1984a).

Finally, a related problem is the question of how many phonetic VOT categories there are at all. Cho and Ladefoged (1999) examined stops realised with positive VOTs (irrespective of the nature of the voicing contrast) in eighteen languages. For the velar /k/ they found a continuum of VOT values across languages, rather than clear-cut categories of short lag and long lag stops and concluded that:

> it is not at all clear that there are just two phonetic categories from which languages can choose. … it would certainly be plausible to say that there are four phonetic categories, one around 30 ms representing unaspirated stops, another around 50 ms for slightly aspirated stops, a third for aspirated stops at around 90 ms, and a fourth for the highly aspirated stops of Tlingit and Navajo (Cho & Ladefoged, 1999, p. 223).

They observed that there is no phonological reason why there should be four categories, because they do not correspond to the number of categories that these languages have. Cho and Ladefoged's conclusion is that there are no discrete VOT categories, but "at best modal values within the continua formed by the physical scales" (1999, p. 225).

Cho & Ladefoged's findings are very important because they show, on data from a large number of languages, that there is no clear-cut boundary between the categories of short and long lag VOT, which is further confirmed by the evidence that there are intermediate lag VOT values in some languages with a two-way contrast. These results are also in line with earlier findings for British English, where the extent of within-category variability was such that the binary division into unaspirated and aspirated stops was questioned (Docherty, 1992). Together with the fact that that in some languages (for some speakers) the short lag category has a bimodal distribution, these findings all suggest that there are no three clearly separated universal VOT categories of voicing lead, and short and long lag VOT.

In addition to the abovementioned issues, VOT is influenced by a number of linguistic and speaker-related factors. Below is a review of the most important factors.

### 1.1.2 Effect of place of articulation on VOT

Lisker and Abramson (1964) observed that in stops realised with positive VOT values VOT shows sensitivity to place of articulation and that there is a tendency for velars to have longer VOTs than bilabials and apicals. A large number of consequent studies confirmed this finding. Results for phonologically voiceless stops are shown in Table 1.1, and results for phonologically voiced stops in Table 1.2.

It can be observed from Table 1.1 that, in general, as the place of articulation moves further back in the oral cavity, VOT increases. However, this tendency is realised somewhat differently in different languages. If we look at the studies that applied statistical tests on their data, some studies found that VOT results are statistically significant for all three pairwise comparisons, i.e. /p/-/t/, /t/-/k/ and /p/-/k/. This is the case with French, European and Canadian (Abdelli-Beruh, 2009; Ryalls, Cliché, et al., 1997), Canadian English (Nearey & Rochet, 1994), Swedish (Helgason & Ringen, 2008), and for both series of German stops, short lag and long lag, in utterance-initial position (Jessen, 1998). The same sort of relationship was observed in a number of other studies, but no statistical analysis was performed, for example in Dutch (Lisker & Abramson, 1964), English (Caramazza, et al., 1973; Lisker & Abramson, 1964, 1967), Hebrew (Obler, 1982), Hungarian (Lisker & Abramson, 1964), Portuguese (Lousada, et al., 2010). Another pattern is that VOT for the velar is significantly longer than VOT for the other two stops, such as in Hungarian (Gósy, 2001; Gósy & Ringen, 2009), Spanish (Rosner, et al., 2000), and Japanese (Riney, et al., 2007). Docherty (1992), on the other hand, found that in British English, it was the bilabial stop that had significantly shorter VOT values than the other two stops.

Cho and Ladefoged (1999) also found that VOT increases with the more back place of articulation in the eighteen languages they investigated. They reported that, with one exception, in languages that do not have uvular stops, velars had the longest VOT. In languages that have uvulars, either velars or uvulars had the longest VOT. Differences between bilabials and coronals (dentals/alveolars) were not statistically significant.

| Language | /p/ | /t/ | /ṭ/ | /k/ | /q/ | stat. sign. |
|---|---|---|---|---|---|---|
| Arabic, Leb (Yeni-Komsh. et al. 1977) $ ui |  | 25 | 23 | 28 | 30 |  |
| Danish (Fischer-Jørgensen 1954) iv | 60-70 | 80 |  | 70 |  |  |
| Dutch (Lisker & Abramson 1964) ui | 10 | 15 |  | 25 |  |  |
| Dutch (van Alphen & Smits 2004) Exp 2 ui | 19 | 31 |  |  |  | * /p/</t/ |
| English, Am (Lisker & Abramson 1964) ui | 58 | 70 |  | 80 |  |  |
| English, Am (Lisker & Abramson 1967) iv | 34 | 45 |  | 53 |  |  |
| English, Am (Klatt 1975) iv | 47 | 65 |  | 70 |  |  |
| English, Am (Zue 1976) iv | 58 | 71 |  | 73 |  |  |
| English, Br (Docherty 1992) total | 42 | 63 |  | 63 |  | */p/</t/,/p/</k/ |
| ui | 46 | 67 |  | 66 |  |  |
| iv | 42 | 65 |  | 62 |  |  |
| English, Ca (Carramazza et al. 1973) ui | 62 | 70 |  | 90 |  |  |
| English, Ca (Nearey & Rochet 1994) iv | 67 | 74 |  | 79 |  | * /p/</t/</k/ |
| French (Yeni-Komshian et al. 1977) $ | 20 | 32 |  | 40 |  |  |
| French (Nearey & Rochet 1994) $ iv | 32 | 35 |  | 46 |  | * /p/</t/</k/ |
| French (Abdelli-Beruh 2009) iv | 15 | 23 |  | 32 |  | * /p/</t/</k/ |
| French, Ca (Carramazza et al. 1973) ui | 18 | 23 |  | 32 |  |  |
| French, Ca (Jacques 1987) ui | 10 | 35 |  | 33 |  |  |
| French, Ca (Ryalls, Cl et al 1997) y/old ui | 41/34 | 58/41 |  | 73/62 |  | * /p/</t/</k/ |
| German (Jessen 1998) ui | 63 | 75 |  | 83 |  | * /p/</t/</k/ |
| iv | 49 | 72 |  | 58 |  | */p/</k/</t/ |
| Hebrew (Obler 1982) ui | 26 | 34 |  | 64 |  |  |
| Hungarian (Lisker & Abramson 1964) ui | 2 | 16 |  | 29 |  |  |
| Hungarian (Gósy 2001) ui | 25 | 23 |  | 50 |  | */p/</k/, /t/</k/ |
| Hungarian (Gósy & Ringen 2009) ui | 10 | 16 |  | 38 |  | * /p/</t/ |
| iv | 18 | 20 |  | 43 |  | */p/</k/, /t/</k/ |
| Japanese (Shimizu 1989) ui | 44 | 27 |  | 68 |  |  |
| Japanese (Riney et al. 2007) ui | 30 | 29 |  | 57 |  | */p/</k/, /t/</k/ |
| Persian (Bijankhan & Nourbakhsh 2009) ui | 69 | 80 |  | 98 |  | */p/</t/</k/ |
| iv | 45 | 54 |  | 51 |  | ns |
| Polish (Keating et al. 1981) ui | 22 | 28 |  | 53 |  |  |
| Polish (Kopzyńsky1970 in Rojczyk2009) ui | 38 | 33 |  | 49 |  |  |
| Portuguese (Lousada et al. 2010) iv | 20 | 28 |  | 51 |  |  |
| Russian (Ringen & Kulikov fc) ui | 18 | 20 |  | 38 |  |  |
| iv | 18 | 18 |  | 35 |  |  |

| | | | | | |
|---|---|---|---|---|---|
| Spanish (Poch-Olive 1987) | iv | 17 | 20 | 30 | |
| Spanish (Rosner et al. 2000) Castilian | ui | 13 | 14 | 27 | */p/</k/, /t/</k/ |
| Spanish (Lisker & Abr. 1964) PuertoR | ui | 4 | 9 | 29 | |
| Spanish (Williams 1977) Vene/Peru/Gua | ui | 14/15/10 | 21/16/10 | 33/30/26 | *overall |
| Swedish (Helgason & Ringen 2008) | ui | 49 | 65 | 78 | */p/</t/</k/ |
| Tamil (Lisker & Abramson 1964) | | 12 | 8 | 24 | |

Table 1.1 VOT (ms) for phonologically voiceless stops reported in some previous studies

Note. Results were rounded to the nearest millisecond. Results marked with $ were calculated from original papers and represent the mean of mean VOTs before different vowels. The following abbreviations were used: ui = utterance-initial position, iv = intervocalic position.

When phonologically voiced stops in English were realised with positive VOTs, the VOT value increased from front to back place of articulation, as in phonologically voiceless stops (Docherty, 1992; Klatt, 1975; Lisker & Abramson, 1964; Smith, 1978). The same was found in Dutch /b/ and /d/ tokens realised as voiceless unaspirated (van Alphen & Smits, 2004), and in French /b, d, g/ realised with interrupted voicing in intervocalic position (Abdelli-Beruh, 2009). In English, Smith (1978) reported statistically significant differences between all three stops, but Docherty (1992) found that only at the bilabial place of articulation VOT was significantly shorter than at the other two places (the same as for /p, t, k/).

| Language | | /b/ | /d/ | /ḍ/ | /g/ | /ɢ/ | stat. sign. [b] |
|---|---|---|---|---|---|---|---|
| Arabic, Leb (Yeni-Komsh. et al. 1977)$ | ui | -65 | -57 | -60 | | | |
| Danish (Fischer-Jørgensen 1954) | iv | 15 | 20 | | 25 | | |
| Dutch (Lisker & Abramson 1964) | ui | -85 | -80 | | | | |
| Dutch (van Alphen & Smits 2004) | Exp1 ui | -113 | -104 | | | | ns |
| | Exp2 ui[a] | -83/12 | -71/19 | | | | ns |
| English, Am (Lisker & Abrams 1964) | ui [a] | -101/1 | -102/5 | | -88/21 | | |
| English, Am (Klatt 1975) | iv | 11 | 17 | | 27 | | |
| English, Am (Smith 1978) | ui [a] | -74/11 | -71/18 | | -65/26 | | */b/>/d/>/g/ |
| English, Am (Zue 1976) | iv | 13 | 19 | | 30 | | |
| English, Br (Docherty 1992) | total | 18 | 26 | | 31 | | */b/</d/,/b/</g/ |
| | ui | 25 | 33 | | 40 | | |
| | iv | 15 | 21 | | 27 | | |

| | | | | | |
|---|---|---|---|---|---|
| French (Abdelli-Beruh 2009) | iv | 8 | 14 | 19 | */b/</d/</g/ |
| French, Ca (Yeni-Komshian et al. 1977) | $ | -77 | -63 | -70 | |
| French, Ca (Jacques 1987) | ui | -60 | -40 | -51 | |
| French, Ca (Ryalls, C et al 1997) y/old | ui | -131/-112 | -122/-108 | -120/-108 | ns |
| German (Jessen 1998) | ui | 15 | 21 | 26 | */b/</d/</g/ |
| | iv | 16 | 20 | 28 | */b/</d/</g/ |
| Hebrew (Obler 1982) | ui | -111 | -96 | -101 | |
| Hungarian (Lisker & Abramson 1964) | ui | -90 | -87 | -58 | |
| Hungarian (Gósy & Ringen 2009) | ui | -95 | -95 | -90 | ns |
| Japanese (Shimizu 1989) | ui | -72 | -58 | -64 | |
| Persian (Bijankhan& Nourbakhsh 2009) | ui[a] | -34/3 | -43/7 | -40/15 | |
| | | | | -15/8 | |
| Polish (Keating et al 1981) | ui | -88 | -90 | -66 | |
| Polish (Kopzynsky1970 in Rojczyk2009) | ui | -78 | -72 | -61 | |
| Portuguese (Lousada et al. 2010) | iv | 28 | 16 | 17 | |
| Russian (Ringen & Kulikov fc) | ui | -70 | -75 | -78 | |
| Spanish (Lisker & Abr 1964) PuertoR | ui | -138 | -110 | -108 | |
| Spanish (Williams 1977) Venezuelan | ui | -95 | -79 | -64 | *overall |
| Peruvian | ui | -102 | -110 | -98 | |
| Guatemalan | ui | -120 | -109 | -101 | |
| Spanish (Rosner et al 2000) Castilian | ui | -92 | -92 | -74 | */b/>/g/, /d>/g/ |
| Swedish (Helgason & Ringen 2008) | ui | -96 | -90 | -61 | */b/>/g/, /d>/g/ |
| Tamil (Lisker & Abramson 1964) | ui | -74 | -78 | -62 | |

Table 1.2 VOT (ms) for phonologically voiced stops reported in some previous studies

Note. Results were rounded to the nearest millisecond. Results marked with $ were calculated from original papers and represent the mean of mean VOTs before different vowels. The following abbreviations were used: ui = utterance-initial position, iv = intervocalic position.

[a] The first number is for phonologically voiced stops realised with prevoicing, the second number for the same stops realised as voiceless unaspirated.

[b] Ordered by absolute VOT values, i.e. for negative VOTs according to duration of prevoicing.

There are fewer studies about the effect of stop place of articulation on prevoicing, but from their results it seems that this effect could be twofold: place of articulation affects the proportion of /b, d, g/ tokens that are realised as prevoiced, and it affects the duration of prevoicing. For example, Smith (1978) found that in English the more forward the place of articulation the higher frequency of occurrence of prevoiced stops, and the same was observed in Persian (Bijankhan & Nourbakhsh, 2009) and

Dutch (van Alphen & Smits, 2004). On the other hand, Caramazza et al. (1974) found that bilabials were least likely to be realised as prevoiced in French (both European and Canadian).

The duration of prevoicing has generally been found to decrease from front to back place of articulation, but there is also a lot of variation between languages (Table 1.2). In English, Smith (1978) found statistically significant differences in prevoicing duration at all three places of articulation, while in Spanish (Rosner, et al., 2000) and Swedish (Helgason & Ringen, 2008) /g/ had significantly shorter prevoicing than /b/ and /d/. In contrast, van Alphen and Smits (2004) for Dutch, and Gósy and Ringen (2009) for Hungarian reported no place-related significant differences in the duration of prevoicing.

According to the myoelastic-aerodynamic theory of voice production (van den Berg, 1958), the following conditions have to be satisfied to produce vocal cord vibration during the closure of a stop. First, the vocal folds need to be adducted and tensed for voicing, and second, there has to exist an appropriate difference in pressure between the subglottal and supraglottal cavity, so that resulting airflow can first initiate (if necessary) and then sustain vibration of the vocal folds (details about the pressures needed to start and sustain voicing can be found in Jansen, 2004; Keating, 1984b). However, because in stop production there is a complete closure in the oral cavity and no air is allowed to leave while the closure is held, the air coming from the lungs through the glottis accumulates in the oral cavity and increases the supraglottal pressure, thus reducing the pressure differential. This makes it difficult to initiate voicing after a pause. Alternatively, if voicing is already present, at some point the pressure differential falls to zero, and voicing stops.

Voicing can be prolonged if the volume of the supraglottal cavity is increased so that oral pressure increases at a slower rate, sustaining the pressure differential and vocal fold vibration for longer. It has been proposed that there are two ways to expand the supraglottal cavity and increase its volume. It can be expanded actively, by lowering the larynx, raising the soft palate, advancing the tongue root or by moving the tongue root and blade down (Westbury, 1983), or by expanding pharyngeal cavity through lateral movement of pharyngeal walls, coupled with tongue root advancement (Ohala, 2011). It can also be expanded passively, if vocal tract walls are lax and yield due to the increasing pressure. A computer simulation of the vocal tract confirmed that voicing is sustained for longer if the walls are lax (Keating, 1984b). Another active manoeuvre, which reduces oral pressure and thus maintains the trans-glottal pressure difference, but does not involve cavity enlargement, is nasal or oral leakage, which releases airflow through an incomplete velopharyngeal or oral closure (Ohala, 2011; Solé, 2011; Solé & Sprouse, 2011; Westbury, 1983). These mechanisms for vocal tract expansion are difficult to separate, but they are all considered to play a role in production of voicing utterance-initially.

Explanations that have been proposed in the literature for the effect of place of articulation on prevoicing duration have only concentrated on aerodynamic and physiological factors.

According to Smith (1978), place-related differences in prevoicing duration are directly linked to the size of supraglottal cavity volume in stop production. The bigger the cavity volume, the longer it takes for the trans-glottal pressure difference to fall below the threshold necessary to maintain voicing, and therefore stops produced with larger cavity volume, such as labials, are expected to have longer prevoicing than stops produced with smaller cavity volume, such as velars. However, Ohala and Riordan (1979) and Ohala (1983) argued that cavity volume difference in itself is not sufficient to explain this effect. Instead, they suggested that passive expansion of vocal tract through tissue compliance was more likely to prolong voicing duration. Because stops with more forward place of articulation have bigger cavity, they also have bigger surface area, which can expand to sustain voicing for longer. For example, in velars, the available surfaces are the pharyngeal walls and parts of the soft palate; for alveolars, in addition to these surfaces, there are the surface of the soft palate and a big part of the tongue; for bilabials, all these surfaces, plus enlarged pharyngeal cavity (Ohala, 1983).

Both proposals were supported by experimental evidence. In a modelling experiment, Keating (1984b) found that in an intervocalic stop, all else being equal, the duration of voicing in the closure of a velar could be 30% shorter than that of a labial, due to the difference in the compliant surface area. For intervocalic stops as well, Keating (1984b) found that properties of surface area have the biggest effect on voicing duration, so that stops produced at more forward places, with larger surface area, have more voicing.

Different scope for passive cavity enlargement across places of articulation has also been used to explain the finding that labials are more frequently produced with prevoicing than stops at more posterior places of articulation (Bijankhan & Nourbakhsh, 2009; van Alphen & Smits, 2004), although these two phenomena are not connected in an obvious way. If a stop is produced without any voicing at all, then passive cavity enlargement is not necessary and does not play any role whatsoever, since there is no voicing during the closure to be maintained. As Smith (1978) pointed out, although "at a more general level, one of the tendencies observed in these data seem to be that those conditions which facilitate the greatest prevoicing *duration* also seem to result in more *frequent* occurrence of that phenomenon" (p. 171), it is not clear why prevoicing is more frequent for labials than for alveolars and velars, since "for the production of any given voiced stop, a speaker may employ *either* the prevoicing mode *or* the short lag mode" (p. 171). Despite this contradiction, Smith's (1978) conclusion that speakers learn to produce certain articulatory events (or sequences of events) more frequently because they are less demanding but achieve desirable acoustic result could help explain this phenomenon:

> It is possible that as speakers acquire a language, they may learn what timing relationships between glottal and supraglottal events are most conducive for particular aerodynamic events and then produce them more frequently. ... It seems for many aspects of speech that speakers employ production models which are "preferable" because they are in some way physiologically less complex or demand less of the speech production system (p. 171).

The aerodynamic explanation can, however, go some way to explaining the higher number of partially devoiced velars, where voicing subsides at some point before the release, not only in utterance-initial position, but also in intervocalic position. For example, it has been argued that stops in intervocalic position in aspirated languages, such as German and English, are passively voiced (Beckman, Jessen, & Ringen, fc; Jansen, 2004; Jessen, 1998), in which case this explanation could be very relevant. In languages that have active voicing, this factor could play a role if active manoeuvres are insufficient to sustain voicing for long enough, but this issue is under-researched.

For stops produced with positive VOTs, Cho and Ladefoged (1999) summarised possible explanations for place-related VOT differences in stops, which are physiological or aerodynamic in nature (see references therein):

1. The volume of the cavity in front of the constriction. Bigger front cavity in velar stops means more air and greater obstruction to the air accumulated behind the constriction, and consequently longer period until trans-glottal pressure drop reaches the level for the start of voicing.

2. The volume of the cavity behind the constriction. Smaller back cavity in velars results in more pressure build-up during constriction, which takes longer to be reduced to the level appropriate to start voicing.

3. Velocity of articulators. Labials and alveolars are expected to have shorter VOTs than velars because lips and the tip of the tongue are more mobile than the dorsum, have faster release, and less time is needed to reduce pressure behind the constriction and achieve the trans-glottal pressure drop necessary for voicing.

4. The extent of contact area. Because the contact area is greater in velars and in laminal dentals, it takes longer for the constriction to be released and for the appropriate pressure difference to be achieved.

5. Change of glottal opening area in voiceless aspirated stops, which is reduced more slowly for velars, because they have slower drop in intraoral pressure than other stops.

6. Temporal adjustment between VOT and closure duration, so that the voiceless period remains uniform.

According to Cho and Ladefoged, the first four factors better explain processes in unaspirated or slightly aspirated stops, sixth in both, and fifth only in aspirated stops. Nevertheless, physiological and aerodynamic factors cannot account for all place-related variation in VOT. In Cho and Ladefoged’s (1999) sample this was especially true for aspirated stops, and they concluded that the grammar might be supplying this value for each place of articulation.

The evidence for the sixth factor (temporal adjustment between closure duration and VOT) is unconvincing. Weismer (1980) proposed that in voiceless stops there is a relatively constant period during which vocal folds do not vibrate. He argued that this so-called *devoicing gesture*, expressed through *the voiceless interval* (closure duration + VOT), is independent of place of articulation and pre-programmed. This further implies that place-related VOT differences are simply a consequence of place-related differences in closure duration, and that the two measures are inversely related. Weismer found that voiceless interval was indeed fairly constant in his data from American English, as did Abdelli-Beruh (2009) for French. However, the prediction that closure duration and VOT would be inversely correlated was confirmed neither by Abdelli-Beruh (2009) nor by Docherty (1992) for British English. These results suggest that VOT variations cannot simply be explained by variations in closure duration, especially taking into account the inconsistency in findings about place-related differences in closure duration (which are discussed in Section 1.2).

A view similar to that of Cho and Ladefoged (1999) was expressed by Klatt (1975), who suggested that place-related VOT differences in short lag English stops (/b, d, g/) are due to physiological constraints, but that in long lag stops (/p, t, k/) they are implemented by a phonological rule for perceptual reasons. In British English, Docherty (1992) found that labials had the shortest VOTs, and there was very little difference between alveolars and velars. He suggested that "these findings reflect an aspect of the systematic language-specific micro-variability of this accent of English" (p. 139), and that there could exist a rule which can partially override the aerodynamic and/or physiological processes. Similarly, Jessen (1998) found that while for German /b, d, g/ VOT increases significantly from the bilabial to alveolar to velar in all contexts under investigation, for /p, t, k/ this was only true in absolute initial position (in intervocalic position the order was /t/ > /k/ > /p/). He attributed this effect of place of articulation on VOT in /b, d, g/ to the passive, aerodynamic processes, arguing that this would explain the fact that they are uniform across contexts. For /p, t, k/ he suggested that observed differences in aspiration duration are "actively controlled by oral-laryngeal coordination" (Jessen, 1998, p. 323), and agreed with Docherty's view that they could be under the control of the speaker.

In sum, place of articulation is one of the most researched factors that can induce variability in VOT, and has been found in the majority of languages that were investigated. Overall, if positive VOT values in a language increase, and duration of prevoicing decreases as the place of articulation moves from front to back, they are considered to be caused by the passive aerodynamic processes. As this is rarely the case for both stop classes and in all environments in a language, some of the place-related VOT differences are likely to be actively controlled. It is, therefore, important to gather data about other, under-researched languages, such as Serbian, to gain better understanding of VOT patterning related to place of articulation.

### 1.1.3 Effect of the quality of the following vowel on VOT

Although early studies by Lisker and Abramson (1967) and Zue (1976) found no effect of the following vowel on VOT, other researchers have found the quality of the following vowel to influence VOT values in both prevoiced and lag stops.

For English, Smith (1978) found that the percentage of /b, d, g/ tokens realised as prevoiced is higher before high vowels than before low vowels (54% and 43% respectively), and that prevoicing is longer before high vowels than before low vowels (with the mean difference of 11 ms). In Castilian Spanish, Rosner et al. (2000) reported longer prevoicing in /b/ and /g/ tokens before /o/ than before /a/ (with mean differences of about 16 ms and 10 ms respectively). Neither of these effects was found in Dutch, French, and Latin American Spanish (van Alphen & Smits, 2004; Williams, 1977; Yeni-Komshian, et al., 1977).

According to Smith (1978), longer prevoicing before high vowels is a result of bigger cavity volume in high vowels compared to low vowels. On the other hand, Ohala and Riordan (1979) and Ohala (1983) argued that this effect is due to the passive expansion of vocal tract through tissue compliance: since high vowels have bigger cavity volume, they also have bigger area that can be expanded through this mechanism, which directly allows longer voicing. Findings by Keating (1984b) and Westbury and Keating (1986) support their arguments (as discussed in relation to the place of articulation effect on VOT).

There is more research about the effect of the following vowel on lag stops.

In English, Smith (1978) and Docherty (1992) found that in /b, d, g/ tokens realised with positive VOTs, VOT was significantly longer before high vowels than before low vowels (by 5 ms and 3 ms on average). Other studies found not only an overall effect of vowel height, but also that its exact realisation varied with the stop place of articulation (Morris, McCrea, & Herring, 2008; Nearey & Rochet, 1994).

VOT values of phonologically voiceless stops also tend to be longer before high vowels than before long vowels. In English aspirated stops, Klatt (1975) and Docherty (1992) found a significant difference (12 and 5 ms respectively), as did Nearey and Rochet (1994). For German, Jessen (1998) reported that aspiration duration of /p, t, k/ was significantly longer before tense than before lax vowels. A similar vowel effect was observed in phonologically voiceless stops in a number of voicing languages, including Lebanese Arabic, French, Hungarian, Italian, Portuguese and Spanish (Esposito, 2002;

Gósy, 2001; Lousada, et al., 2010; Rosner, et al., 2000; Yeni-Komshian, et al., 1977). An interaction of this effect with place of articulation was found in French (Nearey & Rochet, 1994) and English (Morris et al., 2008).

Two explanations have been suggested for the finding that lag VOT values tend to be higher before high vowels than before low vowels. The first is greater resistance to the airflow in high vowels after the release of a stop. As a result, it takes longer to establish the appropriate trans-glottal drop in pressure necessary to start voicing, which results in longer VOT (Ohala, 1981). Chang et al. (1999) measured oral pressure decay after /t/ in /atɪ/ and /ata/ environments, and found that there indeed exists a correlation between oral pressure decay and VOT, which led them to conclude that there is a mechanical link between VOT and vowel height. A somewhat different account was put forward by Smith (1978), who attributes this delay to the fact that high vowels have bigger cavity volume than low vowels, and the voicing cannot commence until the pressure build-up in supraglottal cavity is removed.

The second explanation is a possible pull on the larynx resulting from raising the tongue for high vowels, which in turn increases the tension and the resistance in the area of glottis, making it more difficult for voicing to start (Docherty, 1992; Morris, et al., 2008).

To sum up, although a number of studies have found an effect of the quality of the following vowel on VOT, there is a lot of between-language variability in the realisation of this effect, as well as within-language variability, which often comes from the interaction between this effect and place of articulation. These results suggest that, in addition to the aerodynamic and anatomical factors proposed above, some language-specific active processes are involved, as was the case with the effect of place of articulation on VOT. It is noteworthy that studies vary in the number of vowels investigated, and in the magnitude of the effect observed. For stops with positive VOTs, differences induced by the following vowel tend to be fairly small. There are fewer studies on this effect in prevoiced stops, but their results are conflicting.

### 1.1.4 Effect of other linguistic factors on VOT

Several other linguistic factors have been found to affect VOT.

It has been suggested that VOT tends to be shorter in disyllables than in monosyllables. In English, Klatt (1975) found this difference to be small, about 8% for /p, t, k/, while Lisker and Abramson (1967) reported a 19 ms difference for word-initial /k/ in stressed monosyllables vs. stressed disyllables.

The condition in which word tokens with stops are produced, whether it is in isolation, in a sentence frame, or in spontaneous speech, also affect VOT values and distributions. Lisker and Abramson (1967) found that stops produced in isolated words had longer VOTs than when the same words were embedded in a sentence. In isolation, there was clear separation between the cognate stop pairs along the VOT dimension, but in the sentence condition there was some overlap, especially in unstressed position. Similarly, Docherty (1992) found a tendency for shorter VOT and some overlap in the sentence condition, as opposed to no overlap in isolation. A shortening effect was found in Hungarian /p/ and /k/ (not /t/), where VOT values were smaller in spontaneous speech than in syllables or words spoken in isolation (Gósy, 2001).

Stress is another factor that has received some attention. Lisker and Abramson (1967) found that stressed /p, t, k/ (i.e. at the beginning of a stressed syllable) tend to be produced with longer VOTs than unstressed ones, with the difference being larger in isolated words (29 ms) than in sentences (6 ms). This finding was supported by Klatt (1975). In Dutch, the effect is in the opposite direction for /t/ (Cho & McQueen, 2005). In French, Jacques (1987) found a mixed effect of stress on VOT: in stressed syllables /b/ and /g/ had longer prevoicing (there was no change for /d/), and /t/ and /k/ longer positive VOTs, while for /p/ it was reduced.

Certain phonetic environments have been found to affect VOT. For CVC(C) words with initial /p, t, k/, Port and Rotunno (1979) found that the nature of the final consonant or cluster affects VOT in initial stops: if the word ended in voiceless cluster /pt/, VOT was on average 20% shorter than if the final consonant was the nasal /n/. Immediately following phoneme was found to affect VOT in English and Dutch. In English, all stops except /b/ had longer VOTs in stop-sonorant sequences than in stop-vowel sequences (Docherty 1992), while Dutch stops /b, d/ were more often produced as prevoiced if followed by a vowel than if followed by a consonant (van Alphen and Smits 2004).

In some instances, the preceding environment can exhibit an influence on VOT, such as in /s/+/p, t, k/ sequences in English, where VOT is reduced in comparison to single stops (Docherty, 1992; Klatt, 1975). Docherty (1992) also found a tendency for shortening of VOT in word-initial English stops preceded by a voiceless context vs. a voiced context, while Abdelli-Beruh (2009) found no such effect in French stops with positive VOTs. However, Abdelli-Beruh (2004) found that when a CVC word with initial stops was embedded between two voiceless fricatives, lag VOTs for both /p, t, k/ and /b, d, g/ were significantly longer than between two vowels.

### 1.1.5 Effect of speaking rate on VOT

Speaking rate has been found to affect both production and perception of VOT. In production, at slower speaking rates, positive VOT was found to increase in English (Kessinger & Blumstein, 1998; Miller, Green, & Reeves, 1986; Nagao & de Jong, 2007; Volaitis & Miller, 1992) and in Icelandic (Pind, 1995). The effect of rate was much greater on aspirated than on unaspirated stops (Miller, et al., 1986; Nagao & de Jong, 2007; Pind, 1995; Volaitis & Miller, 1992). In all of these studies, the VOT value that most effectively separated the two voicing categories also increased as speaking rate decreased. Pind (1995) proposed that this asymmetry comes from the fact that there is a limit on the unaspirated VOT category: it cannot stretch a great deal without the risk of overlap, while aspirated category can.

Conversely, at faster speaking rates VOT of English stops decreased (Diehl, Souther, & Convis, 1980; Kessinger & Blumstein, 1997). Kessinger and Blumstein (1997) further found that this change in speaking rate affects VOT categories in Thai, French and English in an asymmetrical way: at the fast speaking rate the prevoiced categories in Thai and French and the long lag categories in Thai and English were significantly smaller than at the normal speaking rate, and they shifted towards the range for the short lag category. On the other hand, the short lag categories in all three languages did not change significantly at the faster speaking rate. There was no overlap between prevoiced and short lag categories in Thai and French, and little overlap between the short lag and long lag categories in Thai and English. Despite observed changes, the voicing categories remained distinct at the fast speaking rate in all three languages. This finding was reinforced by Beckman et al.’s (2011) results for Swedish.

They found that both prevoiced and long lag category were significantly shorter at fast speaking rate, and this was accompanied by the appropriate changes in VOT distributions. There was no overlap between the two categories (however, because Swedish has a contrast between prevoiced and long lag stops, changes would need to be considerable for the two categories to overlap).

The pattern of results found in Kessinger and Blumstein's (1997) study cannot simply be explained by the need to preserve the voicing contrast at faster speaking rates. While in Thai and English this could be the case, because of the potential overlap between short lag and long lag categories, the short lag category in French is potentially free to change with the speaking rate, but it does not happen. Kessinger and Blumstein suggest that the articulatory gestures used to produce prevoiced and short lag stops are different and this acts as a natural boundary between the two categories in French. There is no overlap between the two because that would mean a change of articulatory gesture from that for a prevoiced stop to that for an unaspirated stop. They further argue that, in the same way, different articulatory gestures are employed in the production of unaspirated and aspirated stops. However, this still does not explain why the short lag category in French did not change at faster rates, given the fact that voiceless stops in French were realised as unaspirated to slightly aspirated. There is also a lack of research on the prevoiced category, especially at slower rates of speech. It is unclear what happens with the prevoiced category at slower rates, and if there is an asymmetry between this category and short lag category. All these questions warrant further investigation.

Perceptual studies have generally found the same effect of speaking rate on VOT in English: as speaking rate was changed from slow to fast (i.e. syllable duration or vowel duration of stimuli decreased), the perceptual VOT boundary between the voicing categories moved to smaller VOT values (Miller & Volaitis, 1989; Nagao & de Jong, 2007; Summerfield, 1981; Volaitis & Miller, 1992). A similar effect was found in Icelandic (Pind, 1995, 1996), but the shift in the perceptual VOT boundary as the rate changed from slow to fast was very small, and smaller than that found in English (Miller and Volaitis, 1989; Volaitis and Miller 1992). Miller and Volaitis (1989) and Volaitis and Miller (1992) further found that this effect of speaking rate occurred throughout the ranges for /p/ and /k/, changing the range of stimuli that listeners identified as belonging to a particular category. What is more, the internal perceptual structure of a phonetic category, where some members are perceived as better exemplars of that category than others, changed in accordance with the speaking rate.

### 1.1.6 Effect of individual differences between speakers on VOT

Individual differences between speakers are another source of variability in the production of VOT in /p, t, k/ in English (Allen, Miller, & DeSteno, 2003; Theodore, Miller, & DeSteno, 2009). Allen et al. (2003) suggested that 8-15% of variability can be due to this factor alone. These differences were not simply a consequence of individual differences in speaking rate, since they were present even when the effect of speaking rate on VOT was controlled for (Allen, et al., 2003; Theodore, et al., 2009). The effect of speaking rate on VOT was also speaker-specific, so that the magnitude of change in VOT values for the same change in speaking rate varied with the speaker (Theodore, et al., 2009). On the other hand, the effect of place of articulation on VOT did not differ between speakers, and the speaker-specific effect of rate on VOT was stable across the three places of articulation (Theodore, et al., 2009).

### 1.1.7 Effect of gender on VOT

Speaker gender[2] has been found to affect VOT values and the frequency of prevoicing in some languages. For example, Smith (1978) reported that the number of prevoiced /b, d, g/ tokens was higher for male than for female speakers of English (56% vs. 40%), as did van Alphen and Smits (2004) for Dutch (86% for males vs. 65% for females). However, in Hungarian, there was no difference, because both male and female subjects realised all tokens with prevoicing (Gósy & Ringen, 2009).

The majority of the studies further found that males produced prevoicing of longer duration than females: in English (Smith, 1978), Dutch (van Alphen & Smits, 2004), and Swedish (Helgason & Ringen, 2008), where the difference was statistically significant. Ryalls, Zipprer et al. (1997) reported that younger male speakers of English

[2] When discussing male-female differences in production, it is important to dissociate between biological differences in the size and properties of the vocal tract and learned differences, which might be sociophonetic. In this thesis the term *gender* is used throughout, but when I am discussing male-female differences I make it clear whether I am discussing one or the other.

had significantly longer mean prevoicing than females, but since they reported pooled results for /b, d, g/ tokens realised with prevoicing and tokens realised with positive VOT, it is difficult to know the exact extent and duration of prevoicing in this sample. This phenomenon is usually explained by differences in length and size of the vocal tract in men and women: because men have larger vocal tracts and larger supraglottal cavity volume, supraglottal pressure increases more slowly in men than in women, making it easier to produce voicing (Helgason & Ringen, 2008; Smith, 1978; van Alphen & Smits, 2004). Smith (1978) mentioned other factors that could be contributing to this effect, such as vocal fold length, airflow rate, and subglottal pressure, but did not discuss them further.

On the other hand, Karlsson et al. (2004) found that in Swedish prevoicing was significantly longer in females than in males, as did Gósy and Ringen (2009) in Hungarian. This could not be explained by differences in vocal tract size, so other possible reasons were considered, such as differences in speech tempo between males and females, and a tendency for females to use clear speech (Gósy & Ringen, 2009; Helgason & Ringen, 2008).

For English /b, d, g/ realised with short lag VOT, Sweeting and Baken (1982) and Morris, et al. (2008) found small, non-significant gender-related differences. Smith (1978), however, reported significantly longer VOTs for male subjects than for female subjects (although the difference in means was only 4 ms).

In English aspirated stops females tend to produce longer VOTs than males. For /p, t, k/ reported differences in means were 10 - 13 ms for younger speakers (Ryalls, Zipprer, et al., 1997), and 5 - 11 ms for old speakers (Ryalls, et al., 2004). Morris et al. (2008) found a difference of about 5 ms for /p, t, k/, and Sweeting & Baken (1982) a difference of about 8 ms for /p/. Only in Ryalls, Zipprer et al.'s (1997) study these differences reached statistical significance.

For other languages the influence of gender on lag VOT tends to be in the opposite direction. In Hungarian, males produced longer VOTs for voiceless unaspirated stops than females; these differences were small, around 2-3 ms, but significant for /p/ and /t/ (Gósy & Ringen, 2009). In Swedish aspirated stops, Helgason & Ringen (2008) found that male subjects produced slightly longer VOTs than female subjects. Differences were significant in intervocalic position, but not in absolute initial position, with mean differences of 2 ms and 3 ms respectively (but cf. Karlsson, et al., 2004, who found no statistically significant differences in Swedish). Oh (2011) also found that males produced significantly longer VOTs than females in Korean aspirated stops, and mean differences were bigger than in other languages: 13 ms in isolation and 19 ms in the sentence frame. For the other two Korean stop categories, although there was the same tendency, it was not consistently present in both conditions, and differences were small and non-significant.

Longer lag VOTs in males than in females were explained by Karlsson et al. (2004) as resulting from differences in oral airflow during stop release. They argue that males have a bigger build-up of pressure during the closure and a bigger airflow during the release, which acts as an obstacle to the air coming from the lungs. This makes it more difficult for voicing to start soon after the release. Females, on the other hand, have relatively weak airflow at the stop release, which makes it more likely for voicing to start sooner after the release. However, Subtelny, Worth and Sakuda (1966) measured significantly higher amplitudes of oral pressure in females (and children) than in males in stop production for /p, b, t, d/, while Koenig (2000) found no significant differences in the peak pressure for /p/. Both studies contradict Karlsson et al.'s (2004) explanation.

Lung volume has also been investigated as a possible source of variation in VOT. Hoit, Solomon and Hixon (1993) found that VOT tends to be longer at higher lung volumes and shorter at low lung volumes. They hypothesized that the first finding could be explained by a "tracheal tug" (p. 519): "the diaphragm usually flattens and pulls the trachea and larynx caudally, exerting a force that tends to abduct the vocal folds" (p. 516), which delays VOT. The second finding, where shorter VOT values were associated with low lung volumes, was explained as a need to conserve air during stop production. Stathopoulos and Sapienza (1997) found that several lung volume measures were significantly different for women and for men. They found that "women initiated speech at higher lung volumes, and ended speech utterances at lower lung, rib cage and abdominal volumes" (p. 607), which can potentially affect their VOT production.

Koenig (2000) argued that particularities of laryngeal setting and subglottal and supraglottal pressure levels induce VOT differences between men and women. Koenig investigated patterns of intervocalic /h/ and /p, t/ production in English and found that men are more likely to voice /h/ tokens than women. Since there was a significant correlation between /p, t/ production and /h/ production, she proposed that men are likely to have shorter VOTs than women (i.e. to voice more), caused by physical differences at the glottis. She speculated that greater thickness and smaller stiffness in the vocal folds and smaller glottal convergence angles in men could cause these differences. Koenig reported VOT values for /b, p, d, t/ that were in line with this prediction, although the difference between men and women was non-significant.

Whiteside, Henry, and Dobbin (2004) used similar reasoning to explain longer VOT values in 13-year old girls compared to 13-year old boys. One possible explanation is differences between male and female larynx: female larynx has a higher level of tissue stiffness, which results in a higher level of glottal resistance, which, in turn, leads to longer VOT values. The other explanation is found in male-female differences in supraglottal area: females have smaller supraglottal vocal tracts and vocal tract constrictions, which lead to higher airway resistance and longer VOT.

Oh (2011), however, argued that gender-related differences in VOT production cannot be caused only by anatomical or physiological differences. If the cause was biological, it would be universal, and this is in contradiction with the results reported so far. In Korean aspirated stops, differences between male and female VOT values were not caused by differences in speaking rate either. Instead, Oh proposed that gender differences in VOT production “index sociophonetic gender variations” (p. 65), and these patterns vary with language or dialect. Even when there is a biological base for these differences, at least in some cases “a sociophonetic factor can override the physiological factor, and these sociophonetic contents need to be adjusted in the process of language acquisition” (p. 66). In support of this argument, Oh discussed developmental data from Whiteside and Marshall (2001), where boys and girls seem to change their VOT production between the ages of 9 and 11 to make it adult-like, with girls achieving larger separation between the categories. This, according to Oh, suggests that innate VOT differences caused by gender are changed for sociophonetic reasons until they reach adult-like values.

In summary, gender-related differences in VOT production are documented in a number of studies, but the direction and the extent of differences and the reasons behind them are far from clear. It seems that generally male speakers prevoice more often, but it is not clear if they prevoice for longer periods. There is also a fairly consistent body of evidence that females produced aspirated stops in English with longer VOTs, while this is not the case with /p, t, k/ in other languages, whether realised as aspirated or not. Explanations for these differences tend to focus on biological differences between males and females. They are often based on some type of model of the vocal tract, but one of the striking features of this research is that relatively few of them have been tested (partly due to the fact that they are difficult to test). Finally, although some of the reasons for gender differences in VOT could be biological, the diversity of findings suggests that other factors are likely to play a role, such as sociophonetic factors, speaking rate, and a tendency for clear speech in females.

### 1.1.8 Effect of age on VOT

There are a number of changes associated with normal ageing that can have an effect on speech production, for example anatomical and physiological changes in the respiratory and supralaryngeal systems, and changes in the larynx, such as reduction of mobility, decrease of muscle strength and bulk, reduced speed of neural impulse transmission, and slower motoric movements (Neiman, Klich, & Shuey, 1983; Ryalls, Cliché, et al., 1997; Sweeting & Baken, 1982, and references there). It has also been hypothesised that because VOT as a measure reflects the timing of glottal and supraglottal events in stop production, gradual loss of coordination that occurs with ageing, in addition to the factors mentioned above, can be expected to lengthen VOT values and result in more variability in VOT production (Neiman, et al., 1983; Sweeting & Baken, 1982).

Results from acoustic studies suggested that normal ageing affects VOT in several ways. For English, Sweeting and Baken (1982) found that, although lag VOT values for /p/ and /b/ did not change significantly in their older subjects (over 75 years), standard deviations did, so that they had more variability in VOT production than the control group (25-39 years). In addition to this, minimal separation between /b/ and /p/ was significantly smaller in the older subjects than in the control group and in the intermediate age group (65-74 years). This change was a result of shortening of VOTs for /p/. Sweeting and Baken argued that these changes could be caused by a loss of precision of fine motor coordination required to control laryngeal-supralaryngeal timing in stop production. In order to explain why it is VOT production of /p/ that is less stable, the authors suggested that long lag stops require more careful timing in the innervation of the articulators and more complex muscular movement in adducting the vocal folds, both of which can diminish in older subjects.

Neiman et al. (1983), like Sweeting and Baken (1982), did not find statistically significant differences in the production of /p/-/b/ and /k/-/g/ between two groups of women: the control group (20-30 years) and the older group (70-80 years). They reasoned that either the changes in the laryngeal musculature caused by ageing had little effect on the timing of voicing in stop production, or that the older women used a different way of controlling it.

On the other hand, Ryalls et al. (2004) found statistically significant differences in VOT production between younger (20-30 years, reported in Ryalls, Zipprer, et al., 1997) and older (50-70 years) speakers of English. Phonologically voiceless stops were produced with shorter VOTs by the older group, while phonologically voiced stops had on average more prevoicing in the older group (pooled means for /b, d, g/ were negative, which means that a proportion of tokens was realised as prevoiced[3]). For shorter mean VOT in /p, t, k/ the authors proposed that a smaller lung volume in older speakers can be responsible for the difference (following Hoit, et al., 1993), but for /b, d, g/ the reason was unclear.

For English spoken along the English-Scottish border, Docherty, Watt, Llamas, Hall, and Nycz (2011) found that older subjects (57 years or older) were more likely to produce prevoiced stops, and they had significantly shorter VOTs for phonologically voiceless stops, which is a result very similar to that of Ryalls et al. (2004). However, the situation was further complicated by the interaction of age with social factors, such as country of origin (England/Scotland), and the position on the coast (East/West).

Ryalls, Cliché et al. (1997) investigated VOT production of /p, t, k/ and /b, d, g/ for two groups of speakers of Canadian French (with the mean ages of 24 and 67 years). They found that the older subjects produced /b, d, g/ with shorter prevoicing than the younger subjects (average difference 14 ms), and /p, t, k/ with shorter positive VOTs (average difference 12 ms). Consequently, the average difference between the two

[3] There is a methodological difference between Sweeting and Baken (1982) and Neiman et al. (1983), on the one side, and Ryalls et al. (2004), on the other side. In the first two papers stops were measured in intervocalic position (in words in a sentence frame), and consequently all VOT values were positive. In Ryalls et al. (2004) words were produced in isolation, and negative and positive VOT values were pooled in the means reported for /b, d, g/ and in the statistical analysis. For older speakers, average VOTs are quite long for English (-87 ms for /b/, -90 ms for /d/, -76 ms for /g/, which suggests that a high proportion of /b, d, g/ tokens was prevoiced and that this prevoicing was relatively long). This may have contributed to the result being statistically significant. A problem with this way of presenting results is that we cannot determine the proportion of prevoiced tokens in the sample, and a comparison with data obtained in intervocalic position is difficult.

voicing categories was smaller in the older subjects than in the younger subjects. Standard deviations for VOT were different in the two groups: for voiced stops, standard deviations in the older subjects were significantly larger than those in the younger subjects, while for voiceless stops they were smaller in the older subjects (although, on the whole, production of the younger speakers was more uniform for both voicing categories). This finding is in disagreement with Sweeting and Baken (1982), who found that their older subjects were more variable in VOT production of both short lag and long lag stops in English.

According to Ryalls, Cliché et al. (1997), smaller positive VOTs and shorter prevoicing could be explained by smaller lung volumes in the older speakers (Hoit, et al., 1993). They further hypothesised (similarly to Sweeting and Baken, 1982, for English long lag stops) that prevoiced stops are more difficult to produce and may become more variable with age, while (simpler) voiceless stops may become less variable with age. This would explain differences in precision. Further, they noted that because in French there is more separation between the two voicing categories, there is a greater margin for change and the effect of ageing on VOT production may be bigger in French than in English.

To sum up, this literature review suggests that normal ageing can affect VOT production in a number of ways. It is not only VOT duration that can change with ageing, but variability of production changes in older speakers, which might not be in the same direction for each voicing category. There is also a possible effect of language (or the type of the voicing contrast), with some languages potentially more affected by age-related changes in VOT production than others. However, this summary is based on a small number of studies and almost all of them were on English (with one exception). This topic is largely under-researched, and in order to be able to make any generalisations, much larger body of data is needed.

Linguistic factors aside, the reasons behind this effect have mainly been sought in the anatomical and physiological changes that result from ageing. Because these hypotheses are generally difficult to test, and because there is a lack of research that is aimed directly at establishing the relationship between biological aspects of ageing and VOT, these considerations remain speculative at most part.

### 1.1.9 Summary

This literature review suggests that the three VOT categories, although undoubtedly very useful as a descriptive and classificatory tool for the voicing contrast, cannot be considered universal, nor are they restricted to certain ranges on the VOT continuum. As has been shown in this section, there is a lot of research that demonstrates that VOT categories are not universal: bimodal distribution instead of expected short lag category in some languages, VOT values intermediate between short lag and long lag category and the absence of a clear boundary between them, as well as variability in VOT realisation due to a number of factors. There exists not only between-language variability, but also a lot of within-language variability, caused by linguistic and non-linguistic factors, including some speaker-specific factors. In some cases there is an interaction between some of these factors, which can also be language-specific. What is more, despite attempts to find universal, biological or aerodynamic explanations for within-language variability (for example for the effect of place of articulation, the following vowel, and gender and age of speaker), it seems that part of the observed variability cannot be explained by universal factors and is language-specific.

This poses a problem for some of the models reviewed in Chapter 2, such as that proposed by Keating (1984a), which is based on three universal VOT categories, and gives little consideration to language-specific variation, apart from that related to the choice of the phonetic VOT categories for any particular language. Language-specific variation of the kind reviewed in this section has not been properly elaborated in the model by Cho and Ladefoged (1999) either, despite the fact that it acknowledges the non-discrete nature of the VOT categories.

It should also be mentioned that this research suffers from certain methodological problems. Many studies were based on a small number of subjects and on a small data set, and some studies did not report statistical test results. In addition to this, many studies report small differences in VOT of only a few milliseconds, which are probably within the measurement error, irrespective of the statistical significance of the results. Without some kind of effect size measure, evaluation and comparison of these studies is often difficult.

## 1.2 Closure duration

It is often suggested that closure duration of phonologically voiced stops is shorter than that of phonologically voiceless stops. Closure duration as a correlate of the voicing contrast is most often associated with word-medial and word-final stops.

In isolated words in English, in word-medial position, Lisker (1957) found that on average the closure of /b/ was 45 ms shorter than the closure of /p/. For words in a sentence frame, Stathopoulos and Weismer (1983) found these differences between phonologically voiceless and voiced stops to be 4 - 6 ms in stressed position, and 3 - 21 ms in unstressed position. Edwards (1981), on the other hand, found the opposite effect in stressed syllables, while for continuous speech Umeda (1977) reported inconsistent results. The same effect was reported for final stops in isolated words (Chen, 1970; Wolf, 1978) and in a sentence frame (Luce & Charles-Luce, 1985; Stathopoulos & Weismer, 1983), with mean difference of 52 ms in Chen's study, and up to 24 ms in sentence condition; and also for final stops in continuous speech (Umeda, 1977). Suen and Beddoes (1974) reported a 33 ms difference in pooled results for medial and final stops.

In languages other than English, in word-medial stops, closures were mostly shorter in phonologically voiced than voiceless stops. Mean difference in isolated words was between 28 ms in Dutch (Slis & Cohen, 1969a) and 38 ms in Polish (Keating, 1980). In the sentence condition, mean difference was between 5 - 8 ms in French (Jacques, 1987) and 46 ms in Portuguese (Lousada, et al., 2010). The same relationship was found in word-final position in the sentence frame, with a mean difference of 47 ms in Portuguese (Lousada, et al., 2010) and a smaller, but significant difference of 21 ms in French (Abdelli-Beruh, 2004).

Results for German are inconsistent. Word-medially closures of phonologically voiced stops were found to be significantly longer by Jessen (1998), while Fuchs (2005) reported the same for the pair /d/-/t/ in post-stressed, but not in stressed position. Word-finally, Fuchs (2005) and Smith, Hayes-Harb, Bruss, and Harker (2009) found no significant differences, although majority of Brunner's (2005) speakers produced /g/ closures that were significantly shorter than /t/ closures.

Voicing-related differences in closure duration were observed in word-initial (intervocalic) stops as well, but results were not uniform across conditions and languages. In French, phonologically voiced stops were found to have significantly shorter closures than phonologically voiceless stops, on average by 22 ms (Abdelli-Beruh, 2004). The same was reported for Portuguese, with the mean difference of 55 ms (Lousada, et al., 2010), and for Arabic, with mean difference of about 10 ms (Flege & Port, 1981).

On the other hand, for initial stops in English, Docherty (1992) found shorter closure for /b/ compared to /p/ and /g/ compared to /k/, but the opposite for the pair /d/-/t/. Stathopoulos and Weismer (1983) reported that closures of phonologically voiced stops were longer in word-initial stressed position, and equal or shorter in word-initial unstressed position. In Danish, Fischer-Jørgensen (1954) found that phonologically voiced stops had significantly longer closures (with mean differences of 26 - 45 ms).

In continuous speech in English, Crystal and House (1988a) found that word-initially phonologically voiced stops had longer closures than their voiceless counterparts, while Umeda (1977) found a mixed pattern, dependent on word stress. However, pooled results for stops in all word positions in continuous speech suggest that the overall difference in closure duration is rather small or disappears: Byrd (1993) found that closures of phonologically voiced stops were 7 ms shorter (a significant result), while Crystal and House (1988a) found a negligible difference in closure duration of their complete stops.

Following these production results, several perceptual studies found that synthetic and edited natural stimuli with silent closures were perceived as containing voiceless stops if closures were longer, and as containing voiced stops if closures were shorter, for example in English (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Lisker, 1957) and Dutch (Slis & Cohen, 1969a). Perception of medial stops in English was sensitive to changes of speaking tempo in the preceding carrier sentence so that, as speaking rate increased, less silence was needed for a voiceless percept (Port, 1979).

In perception, closure duration can be traded with the presence or absence of vocal fold vibration during the closure interval. Presence of voicing during the closure increases the number of voiced percepts (Kingston & Diehl, 1995; Kingston, Diehl, Kluender, & Parker, 1990; Parker, Diehl, & Kluender, 1986). In a study with edited natural speech, Raphael (1981) concluded that for final stops the role of closure duration as a cue depends on the extent of closure voicing: if the closure is voiced, assigning it a duration appropriate to its voiceless cognate does not affect the perception of voicing; only if closure is silent or near silent does the number of voiceless percepts increase.

The effect of place of articulation on stop closure duration has been observed in a number of studies but the direction of this influence varies. In majority of studies there was a tendency for labials to have longer closures than alveolars (or dentals) and velars. Some studies found this relationship to be labial > alveolar > velar, for example Gósy and Ringen (2009) for medial and final /b, d, g/ in Hungarian. Other studies reported that the order was labial > velar > alveolar (Luce & Charles-Luce, 1985; Sharf, 1962; Suen & Beddoes, 1974 for English), or labial > alveolar = velar, for example Docherty (1992) for English and Jacques (1987) and Abdelli-Beruh (2009) for French. In studies that examined a number of environments, results tend to differ across environments. For example, Lousada et al. (2010) found that in initial stops in Portuguese the order was labial > alveolar > velar, while in word-medial and word-final stops there was a mixture of results. Stathopoulos and Weismer (1983) found the order to be labial > velar > alveolar in word-medial and final English stops, but in word-initial position there was little difference in closure duration between the velars and the alveolars. Esposito (2002) found different order for phonologically voiced and voiceless stops in Italian: /p/ = /t/ > /k/ but /b/ > /d/ = /g/. A statistically significant effect of place of articulation was reported by Luce & Charles-Luce (1985), Esposito (2002), and Abdelli-Beruh (2009) for phonologically voiceless stops only.

Studies on continuous speech also present mixed results, although only data for English is available. Byrd (1993) reported statistically significant effect of place, but slightly different for phonologically voiceless and voiced stops (/p/ > /k/ > /t/ and /b/ > /g/ = /d/). On the other hand, Crystal and House (1988a) found that place-related differences in closure duration were small and inconsistent, as did Umeda (1977) for initial and medial stops (but in final stops Umeda found larger differences in phonologically voiceless stops in the order /p/ > /k/ > /t/).

Although there is no lack of studies on this topic, it is difficult to compare and evaluate their results because of methodological differences. For example, target words were spoken either in isolation, in a sentence frame, or in a continuous speech sample. The position of the stop under investigation within the word varied (initial, medial, final), and whether the target syllable was stressed or unstressed. Statistical analysis was not supplied in some studies, and this, together with lack of reports on effect size measures, makes it difficult to assess the relevance of results.

Despite this, it could be said that closure duration seems to be a relevant correlate of the voicing distinction in word-final and word-medial position in a number of languages, both voicing and aspirating. The magnitude of closure duration differences between the two stop classes varies with language and with other factors, such as stress, utterance position etc., but differences found in English are comparable to those found in other languages. The same relationship is found in word-initial position, although to a larger extent in voicing languages than in English, for which results are less consistent between studies. In continuous speech differences in closure duration between the two stop classes are generally small or inconsistent. In addition to this, there are place-related differences in closure duration in both stop classes, but the exact order and magnitude varies between languages, and even within a language, depending on the stop class or on the position in the word.

## 1.3 Voicing in the closure

Voicing in the stop closure is considered to be an important correlate of the voicing contrast in voicing languages, while in aspirating languages its role is limited to certain contexts. In English, in word-medial post-stressed position (such as in *rapid* vs. *rabid*) phonologically voiced and voiceless stops were found to differ in amount of closure voicing. Lisker (1957) observed that in such word pairs voicing continues throughout the entire closure duration for the majority of /b/ tokens, and that in most /p/ tokens there is no voicing present. In the same context, Edwards (1981) measured the average duration of voicing in the closure of phonologically voiced stops to be three times longer than that of their voiceless cognates (78 ms vs. 25 ms), and about 45% of voiced stops had voicing throughout the closure interval.

Studies on English stops in word-final pre-pausal position found that some glottal pulsing was present during the closure in realisation of phonologically voiced stops, but not during the closure of phonologically voiceless stops (Hogan & Rozsypal, 1980; Revoile, Pickett, Holden, & Talkin, 1982; Wolf, 1978). Revoile et al. (1982) reported that on average the first 87% of the closure was occupied by voicing in phonologically voiced stops. In pooled results for intervocalic and utterance-final stops, Smith et al. (2009) found that native English speakers on average produced phonologically voiced stops with 74% voicing in the closure (36 ms) and phonologically voiceless stops with 10% voicing in the closure (7 ms).

For word-initial and word-final intervocalic positions in English, Docherty (1992) reported that in most cases in the realisation of both phonologically voiced and voiceless stops voicing continues during the closure, but that the former have significantly longer intervals of closure voicing than the latter. Word-initially, phonologically voiced stops had on average 52 - 67% of the closure voiced, while in phonologically voiceless stops it was 14 - 18%. However, in phonologically voiced stops voicing was interrupted in the majority of tokens. The result was similar in word-final intervocalic position: voicing continued for significantly longer periods in phonologically voiced stops (62 - 67%) than in phonologically voiceless stops (15 - 27%). Although some tokens of /b, d, g/ had fully voiced closures, 46% of all tokens did not have any voicing at all. The absence of voicing was more frequently observed in phonologically voiceless stops. In general, in post-vocalic environment they tended to have some amount of voicing carried over.

Results for German are similar to English ones in the pattern of realisation, but less consistent across subjects. In word-medial intervocalic position Jessen (1998) found three times longer duration of voicing in /b, d, g/ closures compared to /p, t, k/ closures (45 ms vs. 15 ms), for all his subjects. Brunner (2005) also found that voicing in the closure duration was longer for /g/ than for /k/, except for two subjects in one context. For final stops, Smith at al. (2009) found that although overall duration of voicing in the closure was significantly longer for phonologically voiced than voiceless stops (25% vs. 21%), in majority of their subjects differences were very small and within the range of measurement error.

The number of studies that deal with the extent of voicing in the closure in voicing languages is limited, and hardly any of them give a thorough overview of all word positions and all places of articulation for both stop classes. Gósy and Ringen (2009) reported that in medial intervocalic position in Hungarian, some 96% of /b, d, g/ closures were fully voiced, while in word-final pre-pausal condition on average 70 - 74% of closure was voiced. In Portuguese, the percentage of voicing in the closure is higher in phonologically voiced stops (Lousada, et al., 2010). In French, in word-initial but sentence-medial intervocalic position, 99% of /b, d, g/ tokens were phonated (defined as having 75% or more voicing in the closure), while there were no /p, t, k/ closures that were phonated (the amount of voicing was below 25%); a similar result was obtained for word-final intervocalic position (Abdelli-Beruh, 2004). Also for French, Snoeren et al. (2006) reported that in word-final position before a voiced sound, there was a statistically significant difference in the amount of voicing in the closure in voiced stops and in voiceless stops, 97%  and 30% respectively.

Similarly, for Swedish /b, d, g/ in intervocalic and pre-pausal position, Helgason and Ringen (2008) found that they were predominantly voiced for all subjects.

An effect of place of articulation on voicing in the closure was observed in several languages, so that duration of voicing was longer and percentage of devoiced tokens smaller at more forward places of articulation. For example, in the post-voiceless context in French, /b/ closures were significantly more voiced than /d/ or /g/ closures (Abdelli-Beruh, 2009), and the same tendency was found in Hungarian intervocalic and final voiced stops (Gósy & Ringen, 2009). In word-initial and word-medial position in Portuguese the percentage of devoiced tokens increased as the place of articulation moved further back in the oral cavity (Lousada, et al., 2010). A similar place effect on closure voicing in /b, d, g/ was observed in German (Jessen, 1998). Jessen found that there was significantly more voicing in the closure at the more forward place of articulation (labial >alveolar>velar). These results are in agreement with explanations which suggest that at the more forward place of articulation the area of compliant tissue is bigger, which makes it easier to sustain voicing (see above, Section 1.1.2). Docherty (1992), however, having found the opposite result in English initial and final stops, where labial stops /b/ and /p/ tended to have less voicing in the closure than corresponding alveolar and velar stops, suggested that speakers have active control in the production of closure voicing governed by auditory goals.

There is only limited data about the perceptual role of closure voicing. Perceptual studies have mainly used synthetic speech stimuli with silent closures, for the purpose of determining the perceptual role of other acoustic features, such as closure duration (Liberman, et al., 1961; Lisker, 1957; Port, 1979). In English, Lisker (1957) and Port (1979) found voicing in the closure to be of primary importance in medial intervocalic position: if glottal pulsing is maintained throughout the whole closure interval, the stop stimuli are perceived as voiced no matter how long the closure interval is. Only when the closure interval is silent, other factors, such as closure and vowel duration, can play a role in perception.

Some studies with edited natural speech concluded that voicing in the closure is important for the perception of English final stops as voiced (Revoile, et al., 1982; Wolf, 1978), while others concluded that a voiced closure interval is required for hearing a voiced stop neither by English (Hillenbrand, Ingrisano, Smith, & Fledge, 1984), nor by French listeners (Flege & Hillenbrand, 1987).

Perceptual studies on voicing languages are rare. Slis and Cohen (1969b) found that the presence or absence of the voice bar in V-stop-V stimuli changed the percept from voiced to voiceless in most cases for Dutch listeners. In an experiment with edited natural stimuli, Keating (1980) established a hierarchy of cues for Polish intervocalic /t/-/d/ contrast: the most important cue to a voiced percept was presence of at least some strong voicing in the closure, and moderate amounts of silence in the closure did not affect perception. For extreme durations of silence the percept was voiceless. Low-amplitude voicing in a /t/ closure was found to have the same effect as silence. In stimuli with moderate amounts of silence, the percept depended on the cues in preceding and following syllables (such as the nature of voicing offset and burst voicing).

In sum, production studies have found systematic differences in duration of voicing in the closure of phonologically voiced and voiceless stops. In aspirating languages, such as English and German, this was limited to certain word positions and contexts: in English in intervocalic and word-final pre-pausal position, in German only in word-medial intervocalic position, but this was also speaker-dependent. In voicing languages this correlate consistently separates the two stop classes in all word positions where the voicing contrast is present. There is also a place of articulation effect on duration of voicing in the closure, which, depending on the language, can be caused by universal constraints in speech production or is controlled by speakers.

## 1.4 Release burst

Although properties of the release burst have been studied less than other correlates of voicing, previous research has suggested that intensity and duration of the burst, as well as its spectral properties, can be relevant for the voicing contrast.

Phonologically voiceless stops have stronger release bursts than phonologically voiced stops in English and German (Halle, Hughes, & Radley, 1957; Hayward, 2000; Smith, et al., 2009). In Dutch, release bursts of phonologically voiceless stops are longer and higher in intensity, and also have higher spectral centre of gravity (SCG) than those of phonologically voiced stops (Slis & Cohen, 1969a; van Alphen & Smits, 2004). Lousada et al. (2010) found that releases of phonologically voiceless stops were longer in word-initial position in Portuguese, but the opposite was true in word-medial and word-final position.

That these properties have perceptual effect has been shown in experiments with synthetic speech. Slis and Cohen (1969a) found that longer noise bursts and higher intensity of noise both favour voiceless percepts in Dutch. A similar result was obtained by Repp (1979) for English syllable-initial stops. He varied the amplitude of aspiration noise relative to the following vowel and found that, as the amplitude of aspiration was increased (or the amplitude of the vowel decreased), the number of voiceless responses increased.

For word-final English stops experiments have been performed mainly with edited natural speech, where portions of consonant and vowel were progressively cut back and such stimuli presented to listeners for identification. Generally, released stops were better identified than unreleased stops (Wang, 1959). When release bursts were removed, the number of correct voicing identifications was significantly reduced for voiceless stops (Malecot, 1958; Revoile, et al., 1982; Wang, 1959; Wolf, 1978), but identification of voiced stops was not strongly affected (Flege & Hillenbrand, 1987; Hillenbrand, et al., 1984; Malecot, 1958; Raphael, 1981; Revoile, et al., 1982; Wang, 1959; Wolf, 1978). These results suggested that releases of voiceless stops are more important for the perception of the voicing contrast than releases of voiced stops.

It has been hypothesised that the importance of the release might be different in languages other than English. In a study by Flege and Hillenbrand (1987), removing release burst from English word-final /g/ affected the voicing judgments of native

French, but not of native English listeners. A possible explanation could lie in the fact that word-final stops are usually strongly released in French (Laeufer, 1992), which is not the case with English.

Several explanations have been offered for the relationship between the voicing contrast and the properties of the burst. One is that constrictions of voiceless stops are articulated with more force, and with higher pressure build-up, which result in longer and stronger release bursts (Halle, et al., 1957; van Alphen & Smits, 2004). The other is that more extended contact at the place of constriction in voiceless plosives has the same effect on burst duration and intensity (van Alphen & Smits, 2004). For SCG van Alphen and Smits (2004) suggested the following factors: higher air velocity caused by higher subglottal pressure in voiceless stops, which results in higher SCG; the presence of voicing in the burst of voiced stops, which shifts the energy and SCG toward lower frequencies; and one language-specific factor for Dutch - slightly more forward place of articulation for /t/ than for /d/, which, due to the smaller front cavity for /t/, results in higher SCG.

In sum, although limited, previous research has suggested that properties of the release burst and its presence or absence could be relevant for the voicing distinction in stops, but that implementation of this correlate is likely to be language-specific. Further research is needed, especially because of differences in approach and methodology in the existing studies.

## 1.5 Preceding vowel duration

Out of the correlates of the voicing contrast that are found in the segments surrounding the stop or obstruent in question, preceding vowel duration has received most attention. For word-final and syllable-final context, research has shown that vowels preceding phonologically voiced obstruents are longer than vowels preceding phonologically voiceless obstruents[4]. This durational difference is commonly expressed as a ratio: the duration of the vowel preceding a voiceless obstruent divided by the

[4] Some authors discuss this as shortening of vowel duration before /p, t, k/, for example Roach (2000, p. 35) and Wells (1990), who uses the term *pre-fortis clipping* for this phenomenon.

duration of the vowel preceding a voiced obstruent, for example 2:3 or 0.67, or as a percentage (67%). This phenomenon has been investigated in English in a number of environments (Chen, 1970; Cochrane, 1970; Edwards, 1981; Hogan & Rozsypal, 1980; House, 1961; House & Fairbanks, 1953; Klatt, 1973; Laeufer, 1992; Luce & Charles-Luce, 1985; Mack, 1982; Peterson & Lehiste, 1960; Sharf, 1962; Smith, et al., 2009). Depending on the condition and position in the word and sentence, differences vary from 28 ms (Klatt, 1973) to 140 ms (House, 1961), and ratios vary from 0.53 (Mack, 1982) to 0.82 (Laeufer, 1992).

The voicing effect has also been documented for a number of other languages, including French (Abdelli-Beruh, 2004; Chen, 1970; Laeufer, 1992; Mack, 1982), Italian (Esposito, 2002), Spanish (Zimmerman & Sapon, 1958), Portuguese (Lousada, et al., 2010), Russian (Chen, 1970), Dutch (Slis & Cohen, 1969a), German (Chen, 1970; Fuchs, 2005; Smith, et al., 2009), Korean, and Norwegian (Chen, 1970). Differences reported in these studies were between 13 ms (Abdelli-Beruh, 2004 for French) and 53 ms (Chen, 1970 for French), and ratios between 0.74 (Mack, 1982 for French) and 0.91 (Smith, et al., 2009 for German).

Cross-linguistic validity of this feature and its nature was investigated by Chen (1970), on data from four languages: French, Russian, Korean and English. All languages showed the same effect, with mean ratios: 0.61 for English, 0.87 for French, 0.82 for Russian, and 0.78 for Korean. The ratio varied from language to language, with English having notably larger differences (smaller ratios) than other languages. Taking into account his own results and the results from several previous studies for English, German, Spanish and Norwegian, Chen concluded that the variation in vowel duration depending on the voicing of the following consonant is a universal phenomenon, while its extent is language specific.

However, there are languages that do not exhibit this effect of obstruent voicing on the preceding vowel duration. Keating (1985) found that for both Polish and Czech differences in vowel duration before medial voiced vs. voiceless stops in disyllables are negligible, with average ratios of 0.99 and 0.98, respectively. Similar results were obtained for Saudi Arabian Arabic by Flege and Port (1981), who reported the mean difference of about 6 - 7 ms and the ratio of 0.97, and for Hungarian, for which Gráczi (2011) reported hardly any effect in intervocalic position.

These findings, however, need to be interpreted with caution. The above studies differ in many respects: the word material used for analysis, the position of obstruent within the word and sentence, the manner of articulation and place of articulation of obstruent in question, the identity of the vowel, speaking rate etc. English is by far the best-researched language, but even in English, large effects were observed mainly in isolated words and phrase-final position. The effect is much smaller in other positions, and can completely disappear in continuous speech. Some factors that were investigated as possible sources of variability in English are:

1. Number of syllables: the effect is bigger in monosyllabic words than in polysyllabic words (Klatt, 1973; Port, 1981; Sharf, 1962).

2. Position in the sentence: Luce and Charles-Luce (1985) found that differences in vowel duration were significantly bigger for phrase-final than for non-phrase-final position, while Klatt (1976) argued that vowel duration cue has primary importance only in phrase-final environments. In continuous speech, Umeda (1975) found the effect of the following stop and fricative voicing on vowel duration only in pre-pausal position.

3. Speaking rate: the effect is smaller at a fast speaking rate than at a slow speaking rate (Port, 1981).

4. Stress: the effect is inconsistent or absent in unstressed vowels. Davis and Van Summers (1989) found that the effect was clearly present in stressed vowels. There was a tendency for unstressed vowels to be longer before phonologically voiced obstruents, but not consistently, and the difference was not significant in all contexts. In continuous speech, Crystal and House (1988b) found this effect only in stressed vowels followed by word-final pre-pausal stops, but not in unstressed vowels.

5. Manner of articulation: the effect is larger in the context of fricatives than in the context of stops (Hogan & Rozsypal, 1980; House, 1961; House & Fairbanks, 1953; Laeufer, 1992; Peterson & Lehiste, 1960).

6. Vowel quality/quantity: the effect of consonants on the preceding vowel duration seems to be bigger for intrinsically long vowels than for intrinsically short vowels (or for tense than for lax vowels), as reported by several authors (Crystal & House, 1982; House, 1961; Luce & Charles-Luce, 1985; Peterson & Lehiste, 1960); but cf. Port (1981), who found no significant differences and Hogan and Rozsypal (1980), who reported the opposite result.

Fewer studies have investigated the perceptual importance of vowel duration for the voicing distinction in the following obstruents. These studies were concerned mainly with the word-final position in English. For fricatives, Deneš (1955) and Derr and Massaro (1980) found that for the minimal pair of synthetic stimuli /jus/-/juz/, where both vowel duration and fricative duration were varied, the number of voiced responses increased as vowel duration increased and as frication duration decreased. For stops, supporting evidence comes mainly from studies with synthetic speech stimuli by Raphael and his colleagues (Raphael, 1972; Raphael, Dorman, Freeman, & Tobin, 1975; Raphael, Dorman, & Liberman, 1980). Raphael (1972) reported that listeners perceive a final consonant (stop or fricative) as voiced if the preceding vowel was long, and as voiceless if the preceding vowel was short. However, the presence or absence of voicing during the consonant closure, although a secondary cue, was also used in perception. A similar experiment by Hogan and Rozsypal (1980) revealed that vowel duration as a cue was not always sufficient and was accompanied by several secondary cues: the duration of voicing in the closure, silent closure duration and burst/frication duration. The relative importance of these cues varied, depending on the vowels and consonants involved - vowel duration was more important for fricatives than for stops.

Experiments based on edited natural speech did not confirm the relevance of this cue. Wardrip-Fruin (1982) found that the presence or absence of voicing during closure was more relevant in perception than preceding vowel duration. Other studies reported that expanding vowel duration in words ending in voiceless stops did not increase significantly the number of voiced percepts (Hogan & Rozsypal, 1980; Revoile, et al., 1982), and that shortening vowel duration in words ending in voiced stops did not increase significantly the number of voiceless percepts (Hogan & Rozsypal, 1980; Raphael, 1981; Revoile, et al., 1982; Wardrip-Fruin, 1982).

There are hardly any perceptual studies on this subject for other languages. Slis and Cohen (1969a) asked subjects to adjust the vowel length in Dutch words containing voiced and voiceless stops and fricatives. As a result, vowels before voiced consonants were made 25 ms longer than vowels preceding voiceless ones, and there was no difference between stops and fricatives in this respect.

There is an inverse relationship between closure duration and the preceding vowel duration: duration of the vowel preceding a phonologically voiced stop is greater than duration of the vowel preceding a phonologically voiceless stop, while phonologically voiced stops have shorter closures than phonologically voiceless stops. This has been observed in a number of languages and is regarded as “nearly universal, though the magnitude of the effect varies from language to language” (Hayward, 2000, p. 196). Port’s (1981) analysis of natural speech proposed that there was a constant syllable duration (VC) in medial English stops, so that any changes in vowel duration caused by stop voicing were compensated for by appropriate changes in closure duration, but this was not supported by Chen (1970) for final stops.

Port (1981) also suggested that a duration ratio, defined as the ratio of stop closure duration and the preceding vowel duration (C/V), which was higher for /p/ than for /b/, could be invariant for each stop across contextual changes. The opposite conclusion was reached in a production study by Luce and Charles-Luce (1985) for word-final stops, where vowel duration difference distinguished the voicing categories in all instances, but the C/V ratio was influenced by contextual factors and failed to serve as a cue for some minimal pairs. In a perceptual study, Port and Dalby (1982) varied closure duration and preceding vowel duration in the synthetic words *dibber-dipper* and *digger-dicker,* and found that in both cases perceptual boundary values cluster around a certain value of the C/V ratio (0.35 for labials and 0.4 for velars), and that this value was fairly independent of speaking rate.

The C/V ratio has another meaning in the auditory enhancement theory by Kingston and Diehl (outlined in Section 2.2.5). They argue that the vowel duration cue has the function of perceptually enhancing the closure duration cue: a longer preceding vowel makes the following closure seem shorter, and therefore more voiced, and a shorter preceding vowel makes the following closure seem longer, and therefore less voiced (Diehl, Kluender, & Walsh, 1990; Kluender, Diehl, & Wright, 1988).

Several types of explanations have been put forward for voicing-conditioned vowel duration, which are either articulatory or auditory in nature. Articulatory explanations were sought out partly because of the assumed universality of this phenomenon. Chen (1970) discussed a number of possibilities, including compensatory temporal adjustment (to maintain VC dyad/syllable duration, cf. Port, 1981), and laryngeal adjustment (longer time is needed for fine laryngeal adjustment to achieve active voicing for the following voiced stop), but he concluded that the rate of closure transition is “the best, if partial explanation” (p. 152). Voiceless stops, argues Chen, are produced with greater articulatory force and the movement of articulators are faster, and the transition from the preceding vowel to the full closure is achieved in a shorter time, making the previous vowel shorter than that before a voiced stop. Klatt (1976) proposed that vowels before voiceless stops are shorter due to an early glottal opening gesture to prevent any voicing during the closure. Another alternative is the auditory oriented explanation proposed by Diehl et al. (1990) and Kluender et al. (1988), explained above, where vowel duration is under the control of the speaker.

Whatever the nature of the mechanism(s) behind this effect, there seems to be an agreement that English exploits this process more than other languages, which led to the suggestion that in English there exists a (low-level) phonological rule which requires vowel lengthening before phonologically voiced obstruents (Chomsky & Halle, 1968; Klatt, 1976). Laeufer (1992), however, argued that in all languages there exists a relatively uniform effect of voicing-conditioned vowel duration, which is physiological in nature. In some languages and in certain contexts, this effect can be enhanced, which is the case in English when compared to French, for example. This is due to languages-specific linguistic differences related to their prosodic systems, syllable structure, and the phonetic realisation of the voicing contrast, rather than presence or absence of a low-level phonological rule.

The issue of whether there is a similar effect of obstruent voicing on preceding vowel duration in Serbian is especially interesting, because Serbian is one of the few Slavonic languages with word-final voicing contrast in obstruents. Out of other Slavonic languages investigated, Russian, Czech, and Polish all have a neutralisation of voicing contrast in word-final position. The effect of consonant voicing on the preceding vowel duration in Russian was reported in older studies (Chen, 1970; Kozhevnikov & Chistovich, 1966), while recent studies found small, non-significant effects word-finally (Dmitrieva, Jongman, & Sereno, 2010; Shrager, 2005). Keating (1985) did not find this effect in Polish and Czech in medial position, although for Polish Slowiaczek and Dinnsen (1985) found a 10% difference in some speakers in word-final position. The position of Serbian in this spectrum remains to be demonstrated, as well as any potential perceptual role of this effect. My preliminary research suggested that this effect is present both in Standard Serbian and Southern Serbian (non-standard), with overall mean ratios between 0.82 and 0.84 (Sokolović-Perović, 2009; Sokolović, 2010).

## 1.6 Frequency of the first formant (F1)

Two further correlates of the voicing contrast are found in the vowels preceding or following the stop: frequency of the first formant and the fundamental frequency. I discuss frequency of the first formant in this section and fundamental frequency in Section 1.7.

Interest in the transitions of F1 of the following vowel and their relevance for the voicing contrast initially came from perceptual studies, mainly in English. Early pattern playback experiments suggested that initial /b, d, g/ and /p, t, k/ could be distinguished by the F1 transition in the following vowel: the rising transition of the F1 is characteristic of phonologically voiced stops, while the absence of transition of the F1 is the feature of phonologically voiceless stops (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952). Further experiments delayed the onset of the F1 relative to the F2 and F3 by progressively removing parts of the F1 transition (the F1 cutback) and found that longer delays favour voiceless percepts (Liberman, Delattre, & Cooper, 1958). However, the cutback procedure changes two parameters simultaneously: time delay of the F1 onset and the F1 onset frequency (followed possibly by a transition). When manipulated separately, both the F1 onset delay and the F1 onset frequency were found to be important for perception (Liberman, et al., 1958). The F1 onset delay (F1 cutback) corresponds to aspiration. The term *F1 cutback* was later abandoned in favour of *Voice Onset Time/VOT*, reflecting the shift in focus from speech perception to speech production (Lisker, 1975). VOT and various aspects of its realisation have dominated research on the voicing contrast since it was introduced. The F1 onset, on the other hand, received less attention. It has remained a topic of debate exactly which acoustic properties at the F1 onset are responsible for the observed perceptual effect. Proposals include the duration of the F1 transition (Stevens & Klatt, 1974), the F1 onset frequency (Kluender, 1991; Lisker, 1975; Summerfield & Haggard, 1977), or both (Benki, 2001; Slis & Cohen, 1969a).

F1 frequency of the following vowel will rise after a stop due to the movement of the articulators from the constriction for the stop to the more open articulatory configuration for the following vowel. In English, this is visible after /b, d, g/, but after aspirated stops this movement of articulators cannot be observed because it is completed before the vocal fold vibration for the vowel begins. In this case the onset F1 frequency varies with VOT: the longer the VOT value (i.e. more aspiration), the higher the onset

F1 frequency. However, in French, where voiceless stops are unaspirated, the F1 transition is usually present in both series of stops, and this cue is less important than in English. Watson (1990) found that there was a highly significant difference in F1 onset frequency in initial /b, d, g/ vs. /p, t, k/ in the production of British English speakers, but small and non-significant difference in French speakers. Further evidence comes from perceptual experiments by Simon and Fourcin (1978), who found that British and French children respond differently to synthetic stimuli with varying VOT and F1 transition. British children learned to use the F1 transition as a cue to voicing from the age of four, and at the age of 11-12 they reached adult-like performance. French children, on the other hand, never used this cue in the perception of the voicing contrast.

For word-final consonants, the F1 frequency has been examined at several points in the preceding vowel: the F1 onset frequency, F1 steady-state frequency, F1 final transition, and F1 offset (endpoint) frequency. Almost all studies are on English.

In production, final voiced stops are associated with lower F1 frequency at vowel onset (Summers, 1987), with lower F1 steady-state frequency (Summers, 1987; Wolf, 1978), with lower average F1 at the end of the vowel (Wolf, 1978), and lower F1 offset frequency (Crowther & Mann, 1992; Summers, 1987), compared to voiceless stops.

In perception, it has been found that both low F1 onset and low F1 steady-state frequencies produced more voiced responses for the final stop, but is unclear if either is more important (Castelman & Diehl, 1996; Mermelstein, 1978; Summers, 1987, 1988). It has also been suggested that the F1 offset transition slope does not have an effect on voicing perception (Fischer & Ohde, 1990; Summers, 1988), but that F1 offset frequency does, with lower F1 offset values favouring voiced judgments and higher F1 offset values favouring voiceless judgments (Castelman & Diehl, 1996; Crowther & Mann, 1992; Fischer & Ohde, 1990; Summers, 1988). However, this cue seems to be more effective for non-high vowels than for high vowels (Fischer & Ohde, 1990; Hillenbrand, et al., 1984).

Perceptual studies with edited natural speech have confirmed the importance of the final portion of a vowel for the perception of voicing of the following obstruent (Hillenbrand, et al., 1984; Raphael, 1981; Wardrip-Fruin, 1982; Wolf, 1978). In addition to this, studies with natural speech are consistent in the finding that vowel offset cues are more important for voiced stops than for voiceless ones (O' Kane, 1978; Revoile, et al., 1982; Slis & Cohen, 1969b; Walsh & Parker, 1981).

Very little is known about this phenomenon in languages other than English. Unlike in French (Watson, 1990), in Italian the F1 onset values are significantly lower when voiced stops precede non-high vowels (Esposito, 2002). However, this effect is not present at vowel midpoint. The same effect, but smaller, was observed on the F1 offset values. In perception, in contrast to Simon and Fourcin's (1978) study on French, Slis and Cohen (1969a, 1969b) found that in Dutch initial stops both duration of F1 transitions and F1 onset frequency play a role, although small, in the perception of the voicing contrast. They further found that prevocalic F1 transitions are more important than the postvocalic ones.

It has been hypothesised that voicing-related differences in steady-state F1 values and F1 transitions in VC sequences could be a consequence of articulatory gestures involved in the stop production. Several proposals have been made. Crowther and Mann (1994) and Thomas (2000) argue that articulatory manoeuvres involved in the production of voicing during voiced stops, such as lowering of the larynx and tongue-root advancement, both of which lower F1, could explain observed differences. On the other hand, vocal fold vibration stops earlier for voiceless stops, before or around the beginning of the stop closure, which results in higher F1 values (Hillenbrand, et al., 1984).

Wolf (1978) suggested that for initial and final stops the amount of low-frequency energy near the onset or offset of the vowel, which includes not only low F1, but also low f0 and voicing in the closure, serves as a cue for the perception of the voicing contrast. The same idea was elaborated in the auditory enhancement theory by Kingston and Diehl (Section 2.2.5). In contrast to articulatory accounts, which assume that the process is universal and automatic, this theory argues that articulations are under control of the speaker and aimed at perceptual enhancement of the contrast in question. An advantage of this theory is that it applicable to both prevocalic and postvocalic stops.

However, Moreton (2004) argues that low-frequency hypothesis cannot account for the effect observed in English diphthongs, where the opposite relationship was found: voiced stops are associated with higher F1 values (and lower F2 values), and voiceless stops with lower F1 values (and higher F2 values). In order to be able to account for both effects in English (for low monophthongs and for diphthongs), Moreton, following Thomas (2000), proposes the Pre-Voiceless Hyperarticulation

Hypothesis. The acoustic correlate of hyperarticulation before voiceless stops is peripheralisation of formants, which for low monophthongs means rising of the F1 frequency, and for diphthongs lowering of the F1 frequency (and raising of the F2). A problem with this theory is that it is based only on English data, and might not be applicable to other languages. It is also unclear why hyperarticulation before voiceless stops occurs at all. Research into hyperarticulation and its relation to the voicing contrast is at early stages, and the suggested explanations have not been fully tested.

To sum up, differences in the F1 frequency associated with phonologically voiced and voiceless stops have been reasonably well documented for English, both in production and perception, although certain questions need further clarification. This issue, however, remains to be examined in other languages, especially in voicing languages, which, due to the different phonetic realisation of the voicing contrast, i.e. due to lack of aspiration in voiceless stops, may not rely on this feature to a great extent. The research on this topic has been dominated by perceptual studies, so this imbalance should further be addressed by focusing on production data. It is unclear if this effect is present in high vowels at all. Finally, although several explanations have been put forward, they have not been fully tested, and we still do not know enough about the mechanism that is behind this phenomenon.

## 1.7 Fundamental frequency (f0)

Previous research has reported higher fundamental frequency (f0) in the vowel adjacent to a phonologically voiceless stop than in the vowel adjacent to a phonologically voiced stop. This effect, for which the term *f0 perturbation* is also used (Hombert, Ohala, & Ewan, 1979; Jessen, 1998), has been found in many languages. Vowels following phonologically voiceless stops tend to have higher onset f0 and higher average f0 than vowels following phonologically voiced stops. Vowels following phonologically voiceless stops are also said to have an f0 trajectory that starts at a higher value and then decreases, while vowels following phonologically voiced stops have f0 trajectory that rises from a lower onset.

In English, higher average f0 was observed in vowels after word-initial or syllable initial (intervocalic) voiceless stops (Edwards, 1981; House & Fairbanks, 1953;

Lehiste & Peterson, 1961), but mean differences were small, ranging from 4 Hz (House & Fairbanks, 1953, for /k/-/g/) to 13 Hz (Lehiste & Peterson, 1961, for /t/-/d/ and /k/-/g/). Slis and Cohen (1969a) found a 6 Hz difference in the maximum f0 in Dutch.

Significantly higher onset f0 after phonologically voiceless stops was found in English (Hombert, 1978; Ohde, 1984), Persian (Heselwood & Mahmoodzade, 2007), Italian (Esposito, 2002), and Japanese (Shimizu, 1989). However, Jessen (1998) for German and Haggard, Summerfield and Roberts (1981) for English, found this to be the case only for some of their subjects.

A falling f0 trajectory after phonologically voiceless stops and rising trajectory after phonologically voiced stops was found in word-initial and syllable-initial position in Dutch (van Alphen & Smits, 2004), Italian (Esposito, 2002), English (Hombert, 1978), German (Kohler, 1982) and French (Hombert, 1978).

The cue value of f0 perturbation for the voicing contrast has also been tested in a number of perceptual studies.

In the perception of CV syllables, different f0 trajectories (high falling vs. low rising) have been found to influence the listeners’ judgments, but mainly when the VOT of the stimuli was in the boundary region, and therefore ambiguous (Abramson & Lisker, 1985; Fujimura, 1971; Haggard, Ambler, & Callow, 1970; Haggard, et al., 1981). It is not entirely clear how much of an influence f0 perturbation has on stimuli with unambiguous VOTs, although Whalen, Abramson, Lisker and Mody (1993) found that f0 values that did not co-vary with given VOT values slowed down reaction times, both for unambiguous and ambiguous VOTs.

Some authors, however, argued that in syllable-initial position the domain of the f0 cue is restricted to the voicing onset. Haggard et al. (1981) tested the relative importance of the f0 onset, f0 trajectory and the average f0 in the vowel in CV sequences and concluded that it was the onset f0 value that was used by listeners to make voicing judgments. Diehl and Molis (1995) replicated this finding for VCV disyllables.

There is less evidence that voicing value of post-vocalic stops has an influence on f0 in the preceding vowel. Kohler (1982) observed this effect in production in German, but it was not found in English (Gruenenfelder & Pisoni, 1980; Lehiste & Peterson, 1961), Italian (Esposito, 2002), or French (Snoeren, et al., 2006).

In perception, for syllable-final (intervocalic) stops in English, it has been suggested that both low steady-state f0 and low f0 offset value in the preceding vowel cue voiced stops (Castelman & Diehl, 1996). For German stops in the same position, Kohler (1985) and Kohler and van Dommelen (1986) found that particularly important was f0 trajectory in the final 100 ms of the vowel.

Several hypotheses have been put forward to explain the basis of the f0 perturbation effect.

The *aerodynamic hypothesis* suggests that f0 perturbations are a result of general aerodynamic factors associated with stop production: voiceless aspirated stops have higher rate of airflow after the release, which leads to a strong Bernoulli effect, which, in turn, increases f0 in the following vowel (Hombert, et al., 1979). A problem with this explanation is that this effect is not expected to last long, but Ohde (1984) suggested that it can extend to around 100 ms into the vowel. In addition to this, voiced stops and voiceless unaspirated stops are expected to induce less f0 perturbation than aspirated stops, but some production studies on German (Jessen, 1998) and English (Ohde, 1984) do not support this aspect of the hypothesis.

The *vocal fold tension hypothesis* is based on physiological considerations, and comprises of two components, one relating to horizontal, and the other to vertical vocal fold tension (Hombert et al., 1979).

During the production of voiced stops vocal folds are considered to be slack, while during the production of voiceless stops (unaspirated and aspirated) they are considered to be stiff (Halle & Stevens, 1971). These differences in horizontal vocal fold tension affect f0 in adjacent vowels so that it is lower next to a voiced stop and higher next to a voiceless stop. This hypothesis predicts that both following and preceding vowels would be affected, but, as mentioned above, while some studies confirmed the former prediction, there is little evidence from production for the latter.

Vertical vocal fold tension is associated with the lowering of the larynx, which is one of the cavity enlargement mechanisms, performed in order to sustain the trans-glottal pressure necessary to maintain voicing during the closure. Since higher larynx position is thought to be related to higher f0 and lower larynx position to lower f0, this hypothesis predicts that the effect of voiced stops on f0 would be different from that of both voiceless unaspirated and aspirated stops, but that voiceless unaspirated and aspirated stops would exhibit similar effect. Differences in larynx height seem to be the largest at the end of the stop closure, and they continue far into the following vowel (Hombert et al., 1979). Supporting evidence comes from English, where f0 was significantly higher after voiceless unaspirated and voiceless aspirated stops than after voiced stops (Ohde, 1984), and from voicing languages, such as French, Dutch and Italian (Esposito, 2002; Hombert, 1978; Shimizu, 1989; Slis & Cohen, 1969a; van Alphen & Smits, 2004). This hypothesis is also reinforced by the finding that f0 differences can persist into the following vowel (Hombert, 1978; Ohde, 1984; Shimizu, 1989; van Alphen & Smits, 2004), and by studies which found no perturbation effect in the preceding vowel (Esposito, 2002; Gruenenfelder & Pisoni, 1980; Lehiste & Peterson, 1961). However, as Fuchs (2005) pointed out, this hypothesis alone is not sufficient to explain the type of contrast found in languages that contrast voiceless unaspirated and aspirates stops, such as Danish.

The third hypothesis is perceptual, and usually referred to as the *low-frequency hypothesis*. This hypothesis predicts that the presence or absence of low-frequency spectral energy or periodicity in or near the stop closure is a cue for phonologically voiced and voiceless stops, respectively (Stevens & Blumstein, 1981). Its presence is manifested as vocal fold vibration during the closure and lower f0 and F1 in the vicinity of the closure, which is a cue for hearing a voiced stop. The most elaborated version of this hypothesis comes from the auditory enhancement theory (Section 2.2.5). The perceptual role of f0 was tested by Castelman and Diehl (1994, 1996), and Diehl and Molis (1995). They propose that the domain of f0 as a cue depends on the position in the utterance or syllable. For stops in utterance- or syllable-initial position the f0 cue is present only at the onset of the vowel, while for stops in utterance- and syllable-final position both low steady-state and low f0 offset give rise to voiced percepts. They further argue that this pattern is parallel to that found for the F1 frequency in the same positions. However, results from production studies do not fully support this hypothesis either, as was discussed above.

In sum, the relationship between f0 perturbation and the voicing contrast remains controversial. It is unclear what the exact domain of influence of f0 perturbation is (is it mean f0, onset/offset f0, f0 transition trajectory, or a combination). It is also unclear to what extent it is present in syllable-final position, if at all. Although there is no lack of possible explanations of this effect, none of them seems to be supported by a large body of evidence. It can be concluded from production studies that the f0 perturbation effect is not universal, since it was not found in all speakers and in all contexts (see above, for example Jessen, 1998; Haggard et al., 1981; van Alphen & Smits, 2004). This is in contradiction with both the aerodynamic hypothesis and the vocal fold tension hypothesis, and suggests that this effect might be under speaker control. The low-frequency hypothesis is based on such premises, i.e. that speakers intentionally use certain articulations in order to enhance the voicing contrast, but it is not fully supported by production data either, and more data from various languages is needed to test each of these hypotheses.

## 1.8 Summary

The above literature review illustrates the complexity of the phonetic realisation of the voicing contrast within and across languages, and gives an account of the wide range of factors that induce variability in the realisation of the voicing contrast. Although previous research has been thorough in many respects, there are still some aspects of phonetics of the voicing contrast that have not been adequately explored. VOT is the correlate that has been most systematically researched so far, and as a consequence, word positions other than initial have not received the same attention. Most of the research has been on English and, to a smaller extent, on other aspirating languages, and a lot is unknown about the voicing contrast in voicing languages. This imbalance is also reflected in this literature review, in that VOT and English are better represented than other topics and languages. Further, because of lack of systematic research on correlates such as release burst, f0, and the F1 frequency, the role of these correlates in signalling the voicing contrast is still unclear. Methodological differences between studies also make it difficult to evaluate relevance of some correlates and factors across contexts and across languages.

In Chapter 2, I review the existing theoretical models of the voicing contrast.

# Chapter 2 Theoretical background

## 2.1 Approaches to modelling the voicing contrast in obstruents

As mentioned in the introduction, many languages have a contrast in obstruents that has traditionally been described as the voicing contrast, but the nature of phonological and phonetic categories of this contrast, and their relationship, are still a matter of debate. One of the key questions is how to relate patterns of phonetic realisation to phonological representation. Traditionally, in pre-generative linguistics, the voicing contrast was seen as an abstract voiced/voiceless opposition, which had different realisations across languages. After the introduction of distinctive feature theory (Jakobson, Fant, & Halle, 1969; Jakobson & Halle, 1956), and the publication of *The Sound Pattern of English* (Chomsky & Halle, 1968), modelling of the voicing contrast became more complex, reflecting the development of the feature theory and of phonological theory in general, as well as the advances in phonetic knowledge. For years it was dominated by an approach centred around the specification of phonological features (Chomsky & Halle, 1968; Halle & Stevens, 1971; Jakobson, et al., 1969; Jakobson & Halle, 1956; Jessen, 1998; Keating, 1984a; Ladefoged, 1989), but in recent years some different phonological units were proposed, unrelated to the concept of distinctive features, such as articulatory gestures (Browman & Goldstein, 1986, 1992a). Two central issues characterising this theoretical development are the question of what the right features for the voicing contrast are, and the question of whether features (or any other proposed units of lexical representation) have a basis that is acoustic, articulatory, or auditory. Some of the strands of this research will be reviewed in this section. In the following sections, some of the most elaborated models of the voicing contrast will be discussed in relation to these questions.

Jakobson and colleagues (Jakobson, et al., 1969; Jakobson & Halle, 1956; Jakobson & Waugh, 1987) proposed a model of the voicing contrast within the framework of distinctive features. They developed a minimal set of binary distinctive features as the basis of the phonological systems of all the languages in the world. In this set, two distinctive features are related to the voicing contrast: the feature [voice] and the feature [tense]. The definition of each distinctive feature has both an acoustic and an articulatory component[5], and consequently the same features are used at the phonetic level. Phonetically, each distinctive feature is realised through the common phonetic denominator, invariant across all sources of variability. The concept of invariance (relational invariance in Jakobson & Waugh, 1987) is understood in relative terms: in each context the two values of the distinctive feature should be realised phonetically in such a way that they are sufficiently different from each other in order to signal the contrast, but the actual values may vary across contexts, speakers and other factors.

Chomsky and Halle (1968) proposed a modular view of the relationship between the phonological and the phonetic representation. The output of the phonological component consists of underlying forms, which are converted by the phonological rules into phonetic forms. Phonetic forms are the input to a speech production module. Both phonological and phonetic forms consist of matrices, in which segments are represented in columns and features in rows. The same features are used on both levels, but on the phonological level they have binary values, whereas on the phonetic level they have scalar values. The phonetic rules convert the binary values into continuous phonetic values, and the resulting phonetic specification is the input to the universal phonetic component, which then converts these values into continuous movements of articulators.

Chomsky and Halle define phonological features in articulatory terms, with the focus on the state and configuration of the active articulator (although neither acoustic nor articulatory aspects are considered to be more important). They proposed four phonological features for the voicing contrast: [voice], [tense], [heightened subglottal pressure], and [glottal constriction], but these were not widely accepted and were later superseded by other features (more detailed accounts of the development of the feature theory can be found in Clark & Yallop, 1995; Jessen, 1998; Keating, 1988b; Ladefoged, 1989, 2004).

Halle and Stevens (1971) replaced the feature [glottal constriction] by [constricted glottis], and the feature [heightened subglottal pressure], when used for aspirated obstruents, by [spread glottis]. They also proposed the features [stiff vocal cords] and [slack vocal cords] to describe the glottal configuration that controls vocal

[5] However, the acoustic component is seen as more important: “Features are defined in acoustic terms: articulatory means are to be seen only in the light of their ends, namely their use to distinguish perceptually words which are different in meaning” (Jakobson & Waugh, 1987, p. 3).

fold vibration. Out of the four features, the features [spread glottis] and [constricted glottis] have remained in use for aspiration and glottalisation, but the features that refer to vocal fold stiffness were more problematic (Keating, 1988b). The features [slack vocal cords] and [stiff vocal cords] are based on the assumption that in voiced obstruents vocal folds are slack in order to facilitate voicing, while in voiceless obstruents they are stiff in order to prevent it. However, Keating (1988b) pointed out that although this is a possible mechanism, in production of voiceless obstruents glottal spreading is more often used than stiffening of the vocal folds. She also criticised the feature [spread glottis] for being unable to capture different timing of glottal gestures: because it refers to the moment of release, it is unable to separate voiceless aspirated stops from voiceless unaspirated stops produced with glottal spreading (Esling & Harris, 2005, propose the term "prephonation state" for the state of the glottis used in production of voiceless unaspirated stops, in which the glottis is partially open). According to Keating, both classes are produced with glottal spreading, but in unaspirated stops the glottis in closed sooner, and there is no aspiration.

As pointed out by Ladefoged (2006) and Lindau and Ladefoged (1986), the models described so far (by Jakobson and colleagues, Chomsky & Halle and Halle & Stevens) have in common that each phonological feature has a single phonetic correlate. They argued that there is no reason why this should be the case, especially in the light of the fact that phonetic research has not been able to confirm that there is an invariant acoustic property for each phonological (distinctive) feature. Another criticism of the SPE model and the model of Halle and Stevens comes from Keating (1984a). She pointed out that, because the same features were used for phonetic categories and phonological representation, and because these models wanted to account for phonetic differences between languages, this resulted in a large number of features. Keating argued that some of these features are redundant, and that such a system can distinguish contrasts that real languages never use. As an improvement of the SPE model, Keating proposed a model with only one phonological feature [±voice], but where an additional level is introduced after the phonological level, which is based on the temporal phonetic dimension of VOT (and hence only applicable to stops). The phonological level does not contain specific phonetic information, but organises classes of sounds for phonological rules, while lower levels deal with the specifics of the phonetic realisation. In this model phonological features still have phonetic content, but the two sets of features are different, and phonological features are realised as phonetic features in a language-specific way. Keating's model is described in more detail in Section 2.2.1.

Keating's (1984a) model, as well as models by Chomsky and Halle (1968) and Halle and Stevens (1971), are examples of a more general approach to the phonetics-phonology interface referred to as "extrinsic timing" models (Fowler, 1980), the key issue for which is how to map between the phonological categories and the complexities of phonetic realisation. They assume separate phonetic and phonological modules and an interface between them (also called "interface models", Jessen, 1998, p. 30). As pointed out by several authors, there are two general problems with this type of model: one is the problem of abstract, discrete, timeless (except for the linear ordering of segments) representations on one level, versus continuous phonetic realisations on the physical level, and the issue of the interface between the two levels; and the other is a small number of binary features versus multiple acoustic correlates on the physical level (Fowler, 1980; Fuchs, 2005; Jessen, 1998; Keating, 1988a; Pierrehumbert, Beckman, & Ladd, 2000). To overcome these limitations, other models of the voicing contrast were developed, drawing on earlier work by Fowler (1980, 1986), which assume that there is no division between phonetics and phonology and no need for translation (also called "integration models", Jessen, 1998, p. 30, or "intrinsic timing" models, Fowler, 1980). The best example of the models that take this approach is the model proposed by Browman and Goldstein (1986, 1992a), which does not assume the existence of features and uses articulatory gestures as units of lexical representation. Several other models still operate with binary features, but try to deal with the above-mentioned problems in different ways. In these models phonetic data serves as the basis for phonological generalisations, and there is a two-way communication between phonetic details and lexical representations. They differ in the number of features they propose, but also in the nature of postulated phonological categories. For example, the models proposed by Kohler (1984), and Kingston and Diehl (1994, 1995) assume the existence of only one binary feature, while the model proposed by Jessen (1998) assumes two binary features. They further differ in how they define the basis for the feature(s) that they propose. Kohler's (1984) feature [±fortis] is centred on differences in articulatory power between the two obstruent classes. Kingston and Diehl's (1994, 1995) feature [±voice] is auditory, based on the concept of auditory enhancement, which postulates that acoustic properties combine perceptually to enhance each other. Jessen's (1998) features [±voice] and [±tense] are defined in acoustic terms and based on generalisation of phonetic detail across different contexts. I discuss models by Kohler (1984), Jessen (1998), Kingston and Diehl (1994, 1995), and Browman and Goldstein (1986, 1992a) in more detail in Sections 2.2 and 2.3.

This discussion suggests that there is no agreement about the nature of the relationship between phonological representations and phonetic realisations of the voicing contrast. The models that have been proposed so far differ not only in this respect, but also in respect of the true nature (acoustic, articulatory or auditory) of the proposed features. Finally, the existing models also differ in the choice of the features for the voicing contrast. Some of them propose one general feature, whether [±voice], as in proposals by Keating (1984a), and Kingston & Diehl (1994, 1995), or [±fortis], as in Kohler (1984), which is at the phonetic level realised differently in voicing languages and aspirating languages. A more complex approach is to have two features, [±voice] and [±tense], as proposed by Jakobson and colleagues (Jakobson, et al., 1969; Jakobson & Halle, 1956; Jakobson & Waugh, 1987) and Jessen (1998), the first one to account for the way the voicing contrast is realised in voicing languages, the second for aspirating languages. Even more complex are models proposed by Chomsky and Halle (1968) and Halle and Stevens (1971), each with four binary features, or Ladefoged’s (1989) model with a number of features. On the other hand, Browman and Goldstein (1986, 1992) abandoned the concept of distinctive features completely, and based their model on the concept of articulatory gestures.

To sum up, there is a lot of uncertainty about the right phonological specification of the voicing contrast, which also depends on the author’s view of phonology and its relationship with phonetics. Furthermore, there is no agreement about the most appropriate set of features, or about the nature of the phonological features of the voicing contrast.

In addition to the above models, there are some important approaches to modelling the voicing contrast that are essentially phonetic, because they are not concerned with the relationship between the phonological representations and their phonetic realisations, but are focused only on the phonetic aspect of the voicing contrast (and as such they are relevant for the present study). The first one is the measure of Voice Onset Time proposed by Lisker and Abramson (1964) for stops, reviewed in Chapter 1. In contrast to the static features of Chomsky and Halle (1968) and Halle and Stevens (1971), a phonetic dimension of Voice Onset Time was based on the relative timing of glottal and supraglottal events in stop production. The concept of Voice Onset

Time was very influential in phonetic research of the voicing contrast for a number of years, although it did not have much influence on phonologists (Keating, 1988b).

The second is the model of timing of voicing in speech production proposed by Docherty (1992). The focus of his attention is the variation that is present in the phonetic realisation of the voicing contrast and how it can be modelled. He is particularly concerned with systematic, fine-grained variability, whether between- or within-language, especially below the level of segment, which cannot be captured by the feature-based models or by the gestural model of articulatory phonology. This model is discussed in Section 2.4. A growing body of research in recent years has added more data about variability in the phonetic realisation of the voicing contrast, in relation to a number of factors, such as linguistic, contextual factors, individual speaker characteristics or sociolinguistic factors (reviewed in Chapter 1). These findings present evidence that phonetic knowledge is the part of the grammar and raise the question of whether existing phonological models can account for these findings (see for example Docherty & Foulkes, 2000; Foulkes & Docherty, 2006; Pierrehumbert, et al., 2000).

In the remainder of this chapter I discuss in detail the most elaborated models of the relationship between phonological and phonetic categories of the voicing contrast, and of the phonetic realisation of the voicing contrast. I focus on the models that can describe the type of contrast found in voicing languages, with particular emphasis on stops. Other proposals, such as those that focus on the nature of the contrast in stops in aspirating languages, for example the feature [spread glottis] (Beckman, et al., fc; Jessen & Ringen, 2002), or proposals that focus on contrasts in languages that have a distinction with more than two categories, are not discussed.

## 2.2 Feature models

### 2.2.1 VOT-based feature [± voice] proposed by Keating (1984a)

The model proposed by Keating represents an extension of the SPE model. This model is based on VOT, and it concentrates only on stops. Keating's criticism of the generative models (Chomsky & Halle, 1968; Halle & Stevens, 1971) is mainly concerned with the fact that these models use "physical features describing specific articulatory states, both to represent phonetic categories and to serve as the basis for phonological representations" (1984a, p. 288). Keating, on the other hand, argues for a model in which the phonological level does not contain specific acoustic or articulatory details, but is able to "organize natural classes for phonological rules" (1984a, p. 290) and in which each level of representation would "characterize some aspect of sound systems" (1984a, p. 289). Keating's (1984a) model consists of three levels:

1. A phonological level with the phonological feature [±voice]. The number of feature values is determined by the number of natural classes in any given language (two in this case, since this model only deals with languages with a two-way contrast).

2. A phonetic level, with three major phonetic categories {voiced}, {voiceless unaspirated} and {voiceless aspirated}, based on traditional VOT categories of voicing lead, short lag VOT and long lag VOT in utterance-initial position. This is a fixed set of categories, provided by universal phonetics.

3. Pseudo-physical level of representation, which is "continuous in time and encompassing as many parameters as necessary for phonetic description" (p. 291).

This model differs from the SPE model in several respects: phonological and phonetic levels are separated, at both levels less phonetic detail is supplied, and representations at the phonetic level are also more abstract. Keating argues that it is necessary to separate phonological feature [±voice] from phonetic categories in order to account for the fact that there are rules that are equivalent across languages with different phonetic implementations of the voicing contrast. For example, the fact that in a number of languages vowels are longer before phonologically voiced stops than before phonologically voiceless stops, irrespective of their actual phonetic realisation, is taken as the evidence that there is a phonological feature [±voice] independent of the phonetic categories. Based on this and similar evidence, Keating argues that it is necessary to have separate phonetic and phonological representations, and to have phonological representations that are phonetically more abstract.

On the phonetic level, Keating introduces three major phonetic categories, {voiced}, {voiceless unaspirated} and {voiceless aspirated} which correspond to the traditional VOT categories of voicing lead and short and long lag VOT[6], but "they should be viewed as more abstract categories which include a number of acoustic correlates and articulatory mechanisms" (1984a, p. 290). However, these other acoustic correlates did not receive any further attention and the model was based only on VOT.

Keating argues that there are only three phonetic categories, which are discrete. She supports the first claim with the finding that languages contrast no more than three categories along the VOT dimension, and the research that suggests that these same three categories are used in different languages (Keating, Linker, & Huffman, 1983; Lisker & Abramson, 1964). Her own research supports the claim that the three categories are discrete: she found that categories {voiced} and {vl. unasp.} seem to be very well separated acoustically across languages – there is a gap in the area of low negative VOT values. She also offers evidence that values for the {vl. unasp.} category are usually constrained within a narrow area (short lag area), not only in languages with contrast between {vl. unasp.} and {vl. asp.}, but also in languages which contrast {voiced} and {vl. unasp.} categories, such as Spanish. In languages such as Spanish VOT values for {vl. unasp.} category could be expected to spread into the values for the {vl. asp.} category, according to the principle of *maximal dispersion* (Liljencrants & Lindblom, 1972), which proposes that languages keep phonetic categories maximally separated within the available perceptual space. Keating claims that "usually this does not happen" (1984a, p. 298), which supports the idea that these values are categorical.

Instead, Keating proposes a universal rule of polarisation of two adjacent categories. This rule ensures that the categories are maximally separated, and is similar to the dispersion theory of Liljencrants and Lindblom (1972), but here it operates on discrete categories. For example, the polarisation rule ensures that in Polish and English {vl. unasp.} category is maximally separated from the other category ({voiced} and {vl. asp.} respectively), although {vl. unasp.} category is realised with different VOT values in those two languages, and is slightly higher in Polish than in English. However, since

[6] Keating does not specify these categories in great detail, apart from noting that "positive VOT values to about 20-35 msec (depending on the place of articulation) are called 'short lag'; higher values are called 'long lag' " (1984a, p. 295).

polarisation principle cannot explain all variation found in languages, especially the ones with the contrast between {vl. unasp.} and {vl. asp.}, such as English and German, Keating calls for more research to test this principle. Should it turn out that this principle alone cannot account for all observed variation, she allows a possibility of introducing low-level language-specific phonetic rules.

Keating argues that there is also a perceptual basis for these three categories. Some studies found that boundaries between the three categories can be used in perception as extra discriminatory peaks (non-linguistic) by speakers who do not use them in their native language (Abramson & Lisker, 1973; Pisoni, 1977). In addition to this, perceptual experiments suggested that these three categories reflect non-linguistic division of the VOT continuum, resulting from common properties of the auditory system and are found not only in humans, but also in some animals (Kuhl & Miller, 1975; Waters & Wilson, 1976).

To sum up, on the level of lexical representation, there is one phonological feature [±voice] that is used for languages with different phonetic realisation of the voicing contrast, such as Polish and English. On the second level, the implementation of this phonological feature is different in different languages, but they still must choose from one of the three discrete categories {voiced}, {vl. unasp.}, and {vl. asp.}. In Polish and other voicing languages [+voice] stops are realised as {voiced} and [-voice] stops are realised as {vl. unasp.}, and Keating points out that there is little allophonic variation in this case. In English, phonological category [+voice] is realised as {voiced} or {vl. unasp.} and phonological category [-voice] is realised as {vl. asp.} or {vl. unasp.}. English shows more positional variation and more between-speaker variation. In addition to this, there are cross-linguistic differences in how allophonic variation is implemented in languages that use aspiration contrastively, such as English and German, and therefore they have different implementation rules.

At the level of phonetic output, the three major phonetic categories are realised through articulatory and acoustic parameters which are continuous in time. The relationship between major phonetic categories and their realisations is both universal (resulting from the definition of the three major phonetic categories) and language-specific. Since the three phonetic categories can be realised differently in physical terms, this must be specified for each language and for each context. A polarisation principle is suggested as a possible mechanism to deal with this, but there is also a possibility of introducing quantitative low-level phonetic rules which are language-specific.

Keating's proposal represents a very important theoretical model, which has its strengths and its weaknesses. Its strengths lie in the fact that it tries to account for a very persuasive phonological phenomenon, voicing contrast, which has a very complex phonetic realisation, based on what was known about VOT in different languages at the time. It is a powerful model which offers some clear ideas about how phonological representations can be mapped onto the level of phonetic implementation. Its weaknesses lie in the fact that in parts it is not explicit enough and is sometimes very challenging for the reader. Some of its premises are difficult to assess without more data, which is especially true for the level of phonetic realisation where the process of mapping of the three phonetic categories on their realisations remains unclear. If it is to be sustained in its existing form, it needs to accommodate for, for example, the possibility that phonetic categories are assigned different VOT values for different places of articulation, in cases where they do not result from physiological constraints. However, it is difficult to evaluate this piece of data within her model and to say if the model allows this or not.

In addition to this, without further elaboration the model is unable to capture a lot of variability in languages that occurs for reasons other than universal pressure. This includes language-specific variability, caused by either linguistic or non-linguistic factors. For example, Keating argues that in voicing languages, which contrast prevoiced and short lag stops, there is not much allophonic and between-speaker variation. Previous research on voicing languages, to the extent that it is available, shows that this is not necessarily the case. There is not only variation in the choice of VOT categories used (cf. for example, instances of overlap of the VOT categories in Canadian French, Dutch and other languages discussed in Section 1.1), or in the placement of a particular category (cf. evidence for intermediate values of VOT in a number of languages), but there is also variation due to factors such as place of articulation, the quality of the following vowel, stress, context, gender, and age, to the degree that is often comparable to that found in English and other aspirating languages, some of which is language-specific (for details see Section 1.1). More data from other voicing languages is needed to re-evaluate these aspects of Keating's proposal.

Keating's model was criticised by Docherty because of its focus on stops and emphasis on VOT, as well as lack of detail at the level of phonetic realisation, because it cannot explain "the fine-grained aspects of between and within-language variation" (1992, p. 83), and by Kohler (1984) because of its translationist nature (for overviews of Kohler's and Docherty's models see below, Sections 2.2.3 and 2.4).

### 2.2.2 Cho and Ladefoged's (1999) modification of Keating's model

Cho and Ladefoged agree in principle with Keating's approach, but argue that since there is a continuum of VOT values from which languages can choose, the three phonetic categories are not discrete, but represent "at best modal values within the continua formed by the physical scales – the parameters – that define each feature" (1999, p. 225).

The modal nature of the phonetic categories is derived from their research on VOT variations related to place of articulation across a large number of languages. They found that only some of the observed differences could be explained by physiological and aerodynamic factors, and that there are still differences that are language-specific and must be accounted for by the grammar of each language. Starting from these premises, they want to offer a model which would be able to account for both of these factors. Such a model should be able to explain phonological contrasts within each language and phonetic differences between languages. They offer a similar model based on VOT, but here VOT is conceived at a more abstract level as a phonological feature.

Cho and Ladefoged propose that at the phonological level there is a phonological feature VOT with three modal values [voiced], [vl. unasp.] and [vl. asp.]. The model is illustrated in Figure 2.1. In order to be able to establish VOT as a phonological feature, they redefined it in abstract terms as "the difference in time between the initiation of the articulatory gesture responsible for the release of a closure and the initiation of the laryngeal gesture responsible for vocal fold vibration" (1999, p. 225). Defined in this way, phonological feature VOT is not directly observable at this level, but its phonetic implementation is specified by the grammar of a particular language at lower levels. At the level of language-specific phonological rules, each language will choose between the appropriate modal VOT categories {voiced}, {vl. unasp.} and {vl. asp.}, and language-specific phonetic rules will then assign appropriate target values for timing of articulatory and laryngeal gestures. These language-specific rules, supplied by the grammar of the language, will be able to account for the way the voicing contrast is realised in that language, for allophonic variation, such as place of articulation differences in VOT that cannot be explained by physiological and aerodynamic constraints, but are language-specific (Abdelli-Beruh, 2009; Cho & Ladefoged, 1999; Docherty, 1992), and for cross-language differences. Up to this point the timing values are still abstract, and they are converted to real VOT values by the final level, universal phonetic implementation rules. These rules reflect universal physiological and aerodynamic processes that cause some of the variations in VOT values at different places of articulation.

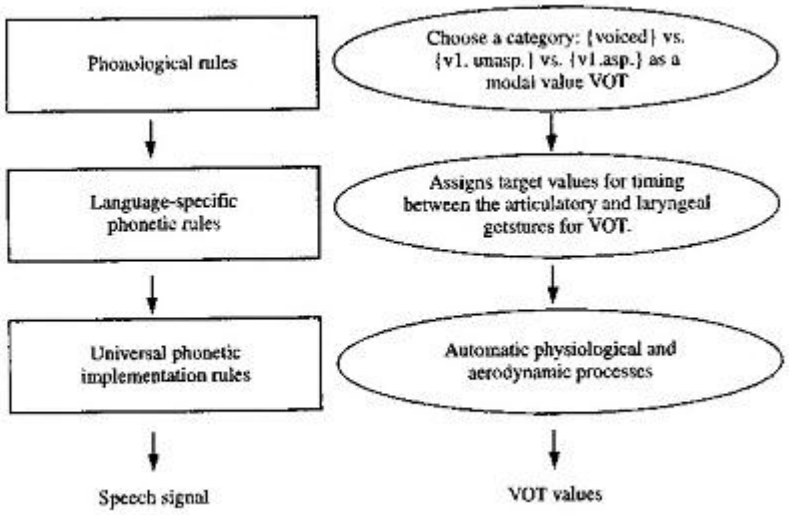
Phonological rules
Choose a category: {voiced} vs. {vl. unasp.} vs. {vl.asp.} as a modal value VOT
Language-specific phonetic rules
Assigns target values for timing between the articulatory and laryngeal getstures for VOT.
Universal phonetic implementation rules
Automatic physiological and aerodynamic processes
Speech signal
VOT values

Figure 2.1 Relationship between the phonological level and the physical output in Cho and Ladefoged's model (Cho & Ladefoged, 1999, p. 226)[7]

Cho & Ladefoged's model tries to fill a gap in Keating's model, namely the need to account for both language-specific (non-universal) variability and variability that can be explained by universal physiological and aerodynamic factors, by introducing into the model some ideas from articulatory phonology by Browman and Goldstein (1986, 1990, 1992a; for an account of this theory see Section 2.3). Cho and Ladefoged were able to postulate VOT at the level of phonological features by defining it in terms of abstract articulatory gestures, which are subjected to language-specific phonological and phonetic rules, and then finally to universal phonetic rules. The language-specific

[7] Reprinted from Journal of Phonetics, 27, Cho & Ladefoged, Variation and universals in VOT: evidence from 18 languages, p. 207-229, Copyright (1999), with permission from Elsevier.

phonetic rules are similar to gestural score in articulatory phonology, and universal phonetic rules have a role similar to that of task dynamics in the same model. However, Cho & Ladefoged's model suffers from the same problems as Keating's and Browman and Goldstein's models. Like Keating's model, it is not explicit enough and very difficult to evaluate. Despite the obvious advantage that it acknowledges the need to include both language-specific and universal factors that induce variability in VOT, it remains vague as to how this could be achieved for any particular language.

### 2.2.3 Feature [± fortis] proposed by Kohler (1984)

Kohler (1984) criticises the existing two-level models, in which phonological features are conceived as static and discrete and phonetic features as continuous and dynamic, and argues that translation models are inherently flawed because they do not incorporate the time dimension. He believes that this problem cannot be resolved by introducing a third level between phonological features and their physical manifestations, as Keating (1984a) does. Instead, he proposes a dynamic model of the voicing distinction in obstruents, which attempts to include the time dimension, and is based on the feature fortis/lenis or [±fortis].

Kohler's feature [±fortis] is based primarily on differences in articulatory power between fortis and lenis obstruents. He argues that these differences also have a functional role: fortis obstruents are auditorily more salient then lenis obstruents, because of the higher intensity at certain points in the acoustic signal. In this model the opposition between fortis and lenis obstruents is achieved by coordinating the actions of the three valves: oral, velopharyngeal, and glottal. Fortis obstruents are produced with tighter and more rapid stricture in the oral and the velopharyngeal valve, compared to lenis obstruents. The glottal valve action is different for fortis and lenis stops, and is manifested as aspiration, voicing, or glottalisation.

All three valves in this coordinative structure work together to achieve differences in intensity between the two categories, and their actions have different timing depending on the position within the utterance. It is by proposing this three-valve structure with components that can be coordinated in time that Kohler incorporates the time dimension in his model.

Kohler proposes that the feature [±fortis], conceptualised in this way, has two components:

1. Articulatory timing, representing the power and speed in supraglottal movements, and

2. Laryngeal power/tension, representing the action at the glottis, such as aspiration, voicing, or glottalisation[8].

The first component is considered by Kohler to be probably universal and the second to be language-specific. Since the first component is universal, this implies that in any particular language the opposition between fortis and lenis stops is achieved by choosing one of the three possibilities at the glottal valve – aspiration, voicing, or glottalisation, and by varying the timing of these events in relation to supraglottal gestures (Docherty, 1992).

Kohler proposes that the laryngeal component in fortis vs. lenis stops can be realised as the opposition between absence and presence of vocal fold vibration during the stop closure. This opposition can be present in all positions in an utterance, such as in French, or in non-final positions, such as in some Slavonic languages. Alternatively, the laryngeal component can be realised as the opposition between aspirated and unaspirated stops, either in all positions (such as in English), in non-final positions (such as in German), or in initial position only (such as in Danish). Another manifestation of the laryngeal component is glottalisation, which is present in final stops in English, for example.

The two components receive different weight in different utterance or word positions. In utterance-initial stops, the laryngeal component is more important than the articulatory component. The fortis/lenis distinction is centred on the release phase in languages that use contrastive aspiration. This is achieved by temporal coordination of the action of the two valves, oral and glottal. In languages that use closure voicing, the distinction is achieved by using active voicing during the closure.

In intervocalic stops, the two components are equally important.

In utterance-final stops, the articulatory component becomes more relevant, because in this position it is difficult to base the fortis/lenis distinction on the laryngeal

[8] A fourth parameter, f0 in the vowels preceding and following stops, was mentioned by Kohler, but not explicitly included as belonging to the correlates of laryngeal tension. Fortis stops are characterised by f0 in the following vowel that is falling from a higher value, while after lenis stops f0 is raising from a lower value. Kohler argues that these differences result from differences in vocal fold tension in fortis and lenis stops.

action - it is both difficult to maintain voicing and to perceive aspiration. In this case, the contrast is signalled by the power differences in the closing movement for the closure. Laryngeal features are considered to be secondary or to disappear in some languages. In acoustic terms, the articulatory component in fortis vs. lenis stops is realised through the duration of the closure and the duration of the vowel preceding the stop. Fortis stops are characterised by short preceding vowels and long closures, and lenis stops by long preceding vowels and short closures. Kohler claims that there is a tendency towards constant duration of VC sequence and reciprocal vowel and consonant lengths for fortis and lenis stops, which is probably a phonological universal.

Kohler's model is an ambitious attempt to account for the realisation of the voicing contrast within and across languages, and to overcome some inherent problems of the previous models. In order to be able to achieve this task, the proposed feature [±fortis] needs to account for a number of articulatory events and their acoustic consequences. Since the feature [±fortis] is based on one dimension only, namely differences in articulatory power, it was necessary to include in the model the coordinative structure of the three valves, oral, velopharyngeal and glottal, and the option of coordinating the work of the three valves in time, in order to achieve the separation between the two categories. Kohler's model also dispenses with the need for translation from the phonological level to the level of phonetic representations, i.e. the need for an interface. However, it remains relatively abstract, especially at the level of phonetic realisation. Kohler acknowledges himself that phonetic variability in the realisation of the feature [±fortis] "has to be accounted for in an adequate phonological description, over and above the specification as [±fortis]. The latter gives a general phonetic classification of elements within phonological obstruent systems by referring them to greater/smaller power and tension" (1984, p. 169). However, this description is not part of the model, as was pointed out by Docherty (1992) as well. In addition to this, fricatives remain somewhat less specified in this model than stops.

Despite its relative abstractness, Kohler's model makes reference to a number of acoustic (in addition to articulatory) correlates of the voicing contrast, and thus allows for certain predictions to be made about a particular language. It also allows for new data to be assessed against the model.

### 2.2.4 Features [±voice] and [±tense] proposed by Jessen (1998)

Jessen (1998) argues for reintroduction of the Jakobsonian feature [±tense] in the feature theory and examines the relationship between the feature [±tense] and the feature [±voice]. He starts from the feature theory proposed by Jakobson and his colleagues, in which both the feature [±tense] and the feature [±voice] belong to the universal set of distinctive features. According to this theory, distinctive features are defined phonetically on two levels.

On the general level, Jessen defines distinctive features through *the phonetic invariant* (Jakobson's *common phonetic denominator*) - a phonetic property which is invariant across all contexts, speakers and other sources of variability. Based on statistical analysis, Jessen searches for a correlate for which in all relevant contexts all subjects have a statistically significant difference between the measures taken for the two obstruent classes. On the specific level, distinctive features are defined through a number of phonetic correlates that are relevant in particular conditions in which the opposition in question occurs, and are specific to different contexts, languages, speakers or other factors.

Jessen further distinguishes two types of correlates of distinctive features: *basic correlate(s)* and *non-basic correlates*. Basic correlates occur in most conditions, while non-basic correlates are limited to certain contexts. A correlate is considered to be non-basic if: 1) it appears in a limited number of contexts, 2) its effect is not statistically significant (present more as a tendency), 3) it has limited importance in the perception of the opposition in question, 4) it is caused by the basic correlate or by its underlying production mechanism.

Based on his analysis of German, Jessen argues that the relevant feature for German stops is [±tense], and proposes the following account of the distinctive features [±tense] and [±voice] and their correlates.

On the general level, duration is the phonetic invariant for the feature [±tense], and voicing for the feature [±voice]. Duration is defined as the duration of the obstruent in question that has this particular feature, as well as durations of the surrounding segments, in particular the preceding vowel. The correlates of duration are: aspiration duration, closure duration and preceding vowel duration for stops, and preceding vowel duration and total duration for fricatives. For stops, aspiration duration is the basic correlate of the feature [±tense] in German, since it is relevant in most contexts and conditions. Jessen proposes that it is also the basic correlate in other languages which express the opposition between tense and lax obstruents in a similar way (such as English).

For German stops, Jessen proposes that non-basic correlates are closure duration and preceding vowel duration, f0 perturbation, breathy phonation, burst amplitude and the F1 onset frequency, since they are found only in certain contexts.

Non-basic correlates are further classified in two groups: *substitute correlates*, which are contextually more limited than the basic correlate, but in some contexts can replace the basic correlate, and *concomitant correlates*, which appear in the same contexts as the basic correlate, but cannot replace it in any of these contexts.

Closure duration and preceding vowel duration are substitute correlates in German, since they are only relevant in word-medial position, but can replace aspiration duration in signalling the contrast in question. The remaining four correlates are concomitant correlates: f0 perturbation, breathy phonation, burst amplitude and the F1 onset frequency. They appear in the same contexts as aspiration, since they are basically caused by underlying physiological factors necessary for producing aspiration, but taken alone are not sufficient to signal the tense/lax opposition.

This model is also used for defining the feature [±voice] in stops. Jessen proposes that voicing is the basic correlate of [±voice]. In parallel to the definition of aspiration duration as the basic correlate of the feature [±tense], voicing must be present in most contexts in a language, if it is to be considered as having the distinctive feature [±voice].

Non-basic correlates of the feature [±voice] are the same correlates that are non-basic correlates of the feature [±tense], but in the feature [±voice] they are used with the opposite polarity; for example, longer closure duration is a correlate of [+tense], while shorter closure duration is a correlate of [+voice], and vice versa. The only exception is breathy phonation, which is present in both the feature [±tense] and the feature [±voice]. Closure duration and preceding vowel duration are substitute correlates of the feature [±voice], and f0 onset, F1 onset, burst amplitude, and breathy phonation are concomitant correlates of the feature [±voice]. This model is represented in Figure 2.2.

In support of this model Jessen cites acoustical evidence from studies on other languages that employ the feature [±voice], such as Japanese, Russian, and Arabic. Perceptual relevance of these correlates is explained in the model proposed by Kingston and Diehl (1994, 1995), which will be discussed separately in Section 2.2.5.

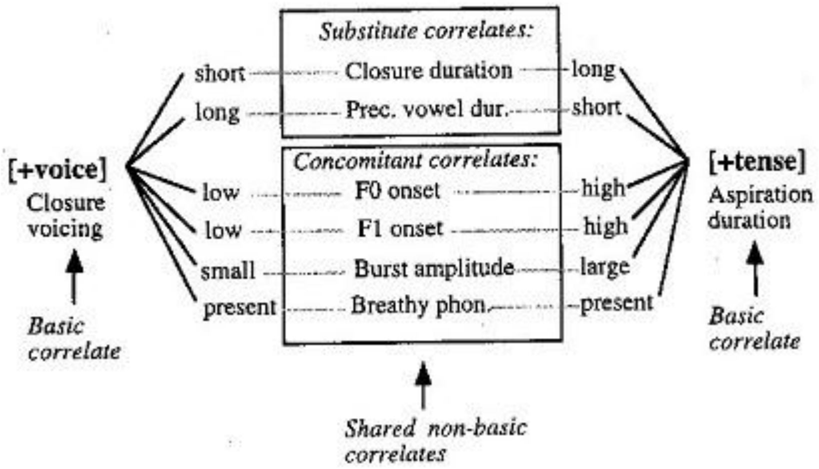

Figure 2.2 An illustration of basic and shared non-basic correlates of [±tense] and [±voice] in stops in Jessen's model (reproduced from Jessen, 1998, p. 270)

For German fricatives Jessen proposes that they employ both the feature [±tense] and the feature [±voice], and that therefore in German there exists a feature syncretism between [±tense] and [±voice]. For the feature [±tense], duration is the phonetic invariant in fricatives as well. The correlates of duration in fricatives are preceding vowel duration and total duration. Other correlates of [±tense] are breathy phonation and, to a smaller extent, f0 perturbation. Correlates of [±voice] are voicing duration, presence or absence of voicing, and f0 perturbation. It was not further specified which correlates are basic and which are non-basic (if any) in fricatives, and which are substitute or concomitant correlates. In addition to this, Jessen suggests that the two features [±tense] and [±voice] are used in fricatives in other languages as well, including Russian and Spanish. However, relevance of particular correlates may depend on whether in stops a language employs the feature [±tense] or the feature [±voice]. If it employs the feature [±tense] in stops, then in fricatives duration may be more important than voicing, and vice versa.

It should be noted that there is a discrepancy between the initial definition of the phonetic invariant, where a property has to signal distinction in question in all relevant contexts, and the definition of duration of aspiration and voicing as the basic correlates of [±tense] and [±voice] respectively, where these basic correlates are required to be relevant in most, but not in all contexts. Jessen acknowledges that there are two possibilities when performing the invariance analysis: the property has to be present in all contexts, or it has to be present in the majority of contexts. Only the first case represents true invariance. Since aspiration is not present in all contexts in German, duration is proposed by Jessen as the phonetic invariant of the feature [±tense] in stops, and aspiration as the basic correlate, not only in German, but in other languages (supported by the fact that word-finally in English it is not aspiration duration but preceding vowel duration that signals the opposition between tense and lax stops). Jessen's definition is then extended to voicing as the basic correlate of the feature [±voice] in stops, which is required to be distinctive in majority of relevant contexts. In this case, it is not clear what would be the phonetic invariant for the feature [±voice]. It is possible that among the languages that use the feature [±voice], there are languages that satisfy the strong version of the principle of contextual stability, and languages that satisfy the weak version of the same principle (Russian, Spanish).

The main advantage of Jessen's proposal is that it brings together the feature [±tense] and the feature [±voice] in the same model. By doing this, it overcomes problems of the models in which only the feature [±voice] is used for both voicing and aspirating languages. In addition to this, it is based on phonetic evidence, and consequently it incorporates a number of acoustic correlates of the voicing distinction that have been found to be relevant in many languages, which is one of its strong points. Time dimension is also incorporated, through several temporal correlates. The invariance analysis is detailed and explicit, and it could be applied to any language in order to establish the basic and non-basic correlates.

It should be mentioned that although the method for arriving at the phonetic invariant is based on statistical analysis, the definition of the relevant contexts is open to interpretation, and can lead to different conclusions depending on the exact application of the invariance analysis. It also is important that, when a decision about the phonetic invariant is based on statistical analysis, the number of tokens is taken into account.

On the other hand, Jessen's model suffers from similar problems as the previous models. First of all, although in principle it allows for the allophonic and other language- or speaker-specific variation to be incorporated in the model by using the invariance analysis, it cannot be used in the opposite direction, to make predictions about a particular context, environment, etc. Once the basic correlate is established for a language, non-basic correlates and their relationships are determined by the model, and it is unclear how language-specific information can be included. Further, it does not offer a way of expressing between-language differences for languages that use the same feature, either [±voice] or [±tense]. In a similar manner, although different word positions are taken into account during the invariance analysis for a particular language, the relative importance of each correlate in different word positions cannot be specified in the model. The same is true for gradient effects below the level of the segment (if present) such as, for example, different degrees of phonetic voicing in word-final stops in different environments. In other words, the specific strength of this model lies in the procedure of generalisation from the data for a particular language to the decision about the basic correlate, but problems arise in the opposite direction, from the model to the details of phonetic realisation. In this respect, Jessen's model is incomplete, as are other feature-based models. While it does offer a more explicit account of differences between languages that use the feature [±voice] and languages that use the feature [±tense], it is less specific when it comes to languages that use the same feature, and even less specific in describing the pattern of phonetic realisation in a particular language. Unlike some other models (by Keating, Kohler, and Browman and Goldstein), which were criticised by Docherty (1992) for not including systematic non-universal micro-variability in the timing of voicing, while at the same time acknowledging some of the variation coming from universal constraints in speech production, Jessen's model does not make any reference to either universal or non-universal factors in the realisation of the voicing contrast, except that concomitant correlates are considered to be an automatic consequence of the basic correlates, and therefore universal (although relevance of some of these correlates has not been fully established, as discussed in Chapter 1). Even well documented sources of variability, such as place of articulation effect on VOT, have not been included in the model. The model needs to be developed further so that it incorporates both types of variation.

### 2.2.5 Auditory-based feature [± voice] proposed by Kingston and Diehl

Auditory enhancement hypothesis is a contemporary model of speech perception that argues in favour of an auditory base of speech perception and production. The main points of this view are explained in several studies by Kingston, Diehl and their colleagues (Diehl, et al., 1990; Kingston & Diehl, 1994, 1995; Kingston, Diehl, Kirk, & Castelman, 2008).

Crucial to the theory is the notion of auditory enhancement. It is argued that speakers have a high degree of independent control in speech production (within constraints of physics and physiology), and that speech communities choose certain articulations so that phonological distinctions in a particular language are perceptually enhanced. Contrary to the position of the motor theory that some acoustic properties co-vary perceptually because they are results of the same articulatory gesture, Kingston and Diehl argue that "speakers covary articulation precisely because their acoustic consequences are auditorily similar enough to be integrated into more comprehensive perceptual properties, intermediate between the acoustic properties and distinctive feature values" (1995, p. 7). They also argue that speech perception does not depend on a specialized module (as proposed by the motor theory), but on general auditory processes, calling on the evidence of parallelism between the perception of speech and nonspeech sounds, and parallelism between human and nonhuman speech perception.

In the process of mapping acoustic properties to distinctive feature values, the authors introduce an additional level, *intermediate perceptual properties or IPPs* (Diehl & Molis, 1995; Kingston & Diehl, 1995). Several IPPs combine to specify distinctive feature values. Each IPP can be analysed into several subproperties, which have a mutually enhancing auditory effect. Some subproperties can contribute to more than one IPP.

Most of the work on the auditory enhancement theory has been concerned with the voicing distinction in stops, for which the authors propose the phonological feature [±voice]. In this model (Figure 2.3), the most important IPPs that contribute to the voicing distinction in stops are C/V duration ratio, low-frequency property, and aspiration. Their research has mainly been concerned with the first two properties, the C/V duration ratio and the low-frequency property, mostly in the intervocalic context.

To establish the role of the C/V duration ratio, Kingston, Diehl and colleagues have carried out a number of perceptual experiments with synthetic speech stimuli and non-speech stimuli in which acoustic correlates under investigation were varied independently. They found that within this IPP the following subproperties integrate perceptually: stop closure duration and preceding vowel duration (Kluender, et al., 1988), closure voicing and closure duration (Parker, et al., 1986), and low F1 frequency in the surrounding vowels and closure voicing (Kingston, et al., 1990). It was suggested that the vowel-duration cue has the function of perceptually enhancing the closure-duration cue: a longer preceding vowel makes the following consonant closure seem shorter (which favours a voiced percept), and a shorter preceding vowel makes the following consonant closure seem longer (which favours a voiceless percept). This is due to a general auditory effect, for which they use the term *durational contrast* (Diehl, et al., 1990). Similarly, the presence of glottal pulsing makes the perceived closure duration seem shorter and therefore shifts perception towards more voiced responses (Parker, et al., 1986). However, this process only takes place if there is a spectral continuity between the glottal pulsing and the surrounding segments provided by falling and rising F1 (Kingston & Diehl, 1995).

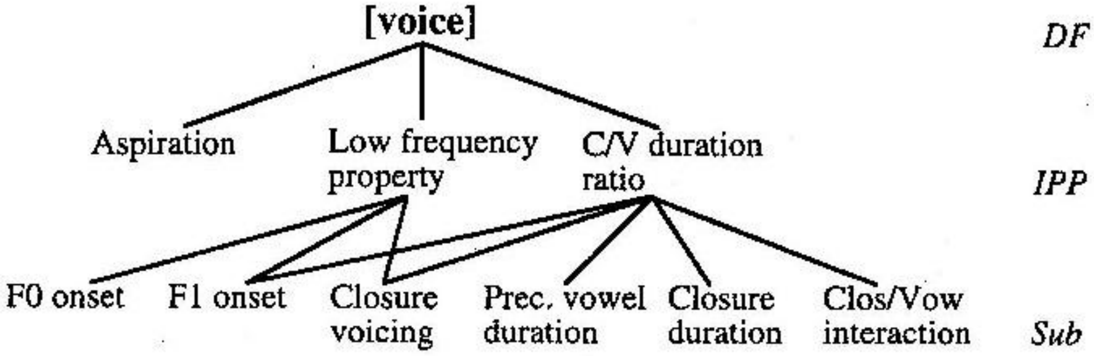

Figure 2.3 An illustration of the model by Kingston and Diehl (reproduced from Jessen, 1998, p. 266)

The low-frequency property is an IPP that can be analysed into at least three subproperties, which have mutually enhancing auditory effect: voicing during the constriction interval, a low f0 in the vicinity of the constriction interval and a low F1 frequency in the same interval (Kingston & Diehl, 1994, 1995; Kingston, et al., 2008). The role of low f0 was largely confirmed by Diehl & Molis (1995), Castelman & Diehl (1994, 1996), and the role of low F1 frequency by Castelman & Diehl (1996).

The low-frequency hypothesis predicts that these three subproperties integrate perceptually. Kingston et al (2008) confirmed that F1 and f0 each integrate with closure voicing. However, f0 and F1 did not integrate with each other, which suggested that it is not the amount of low-frequency energy that is perceptually important, but the continuation of low frequency energy from the vowel into the stop, i.e. low-frequency spectral continuity. This means that if voicing in the closure is present, either low/falling F1 or f0 can independently enhance the percept of the low-frequency property. However, if voicing in the closure is absent, neither F1, f0 nor both can create the low-frequency spectral continuity.

The role of the third proposed IPP, aspiration, was not elaborated in the theory. It seems that the authors include this IPP as a property relevant in initial position, in languages such as English, which suggests that relevant IPPs for word-initial position are aspiration and the low-frequency property, while for intervocalic position it is the low-frequency property and the C/V duration ratio (Kingston & Diehl, 1994). The model does not make any reference to VOT, although Diehl et al. (1990) and Kluender (1991) offer an auditory explanation for the trading relationship between VOT and F1 onset frequency word-initially.

The role of IPPs in word-final position was usually discussed in conjunction with medial position (Castelman & Diehl, 1996; Diehl & Molis, 1995), and it seems that the authors consider the two IPPs that are relevant for the medial/intervocalic position (the low-frequency property and the C/V duration ratio) to be used word-finally as well.

The theory of auditory enhancement is different from other models of the mapping of the phonological features onto phonetic representations, because it is based on the general auditory processes. The theory extends our knowledge about perception of acoustic correlates of voicing in intervocalic position, which has received less attention than initial position, and offers a detailed account of how these acoustic correlates integrate in perception. It accounts for the fact that the same distinctive feature can be signalled by different acoustic correlates or their combinations, thus allowing contextual and allophonic variation. In addition to this, as was pointed out by Hawkins (1999), it explicitly integrates information from different segments and syllables into intermediate perceptual properties, which are then directly mapped onto the features.

However, as it stands at the moment, this model is incomplete in several respects. Because it does not address word positions other than intervocalic/medial, it is not entirely clear if the model should be taken at face value, and extended to all word positions, as was done by Jessen (1998), for example. In that case, while it may be able to account for the voicing contrast in voicing languages, it cannot do so for aspirating languages. In such case, a modification, such as one proposed by Jessen (1998, p. 268), is necessary, which would need to be confirmed by a series of perceptual experiments to establish perceptual integration of aspiration with other subproperties. In addition to this, because the model is based on universal auditory processes, it does not discuss cross-linguistic differences in the realisation of the proposed feature [±voice], and cross-linguistic variation remains unspecified.

An inherent problem with a theory like this is the evaluation of the model for a particular language, since it is predominantly based on perceptual experiments with synthetic speech (and nonspeech) stimuli. Production data can only establish relevant subproperties, but cannot test for any integration of acoustic properties into IPPs. It is possible to use synthetic speech stimuli based on natural speech for a particular language, and present to native listeners, but since the theory is based on presumed universal auditory processes, and between-language variation is restricted to the choice of IPPs in a particular word position, it is not clear if listeners should be expected to respond using universal auditory processes or language-specific strategies.

On a more general level, the auditory enhancement theory was criticised by Nearey (1995) as being too strong an auditory theory. Nearey points out that speech perception may access properties below the level of IPPs, and his own experiments indicate that the effect of some subproperties, such as F1 and closure voicing, is essentially additive, not integrated. He also highlights that some claims of the theory can be explained in simpler ways. For example, lower F1 in the vowel preceding a stop does not have to be produced intentionally to achieve auditory enhancement. It is known that when a closure is made for a stop, the frequency of F1 decreases, and this can also be explained by gestural theories, which argue that acoustic and auditory effects are a consequence of articulatory gestures, not an aim themselves (Fowler, 1986; Liberman, 1996).

## 2.3 Articulatory phonology by Browman and Goldstein

Another model that attempts to overcome problems of previous models has been developed by Browman and Goldstein (Browman & Goldstein, 1986, 1990, 1992a, 2010; Goldstein & Browman, 1986). They propose a model in which articulatory gestures serve as the units of phonological representation. Gestures are typical classes of movements of articulators in space and in time. Each gesture represents a cluster of movements of articulators which can achieve the same goal (lip closure, for example) under a range of conditions, which vary with the linguistic context, speaking rate and speaker (Browman & Goldstein, 1986). Although they are the units of phonological representation, gestures do not necessarily correspond to either features or segments in traditional sense, and can spread across higher units, such as syllables.

Phonological contrast between two lexical items can be expressed in the following ways: a gesture can be present or absent, such as bilabial closure in *add* vs. *bad*; contrasting gestures can involve different sets of articulator and tract variables, such as lip closure vs. tongue closure in *bad* vs. *dad*; contrasting gestures can have different values of dynamical parameters, such as degree of the constriction, e.g. complete closure vs. turbulence generation (Browman & Goldstein, 1992a).

The coordination of different articulators is described using the task dynamic approach which models the movements of tract variables, not individual articulators. Each gesture is thus specified using a set of five related, but relatively independent vocal tract variables: lips (lip protrusion and lip aperture), tongue tip (constriction location and constriction degree), tongue body (constriction location and constriction degree), velum, and glottis.

All gestures involved in production of an utterance are coordinated to form a larger structure, represented by a gestural score (Browman & Goldstein, 2010). The gestural score corresponds to the phonological structure of that utterance. It is organised as a tiered structure where each row or tier represents one of the five vocal tract variables, and the horizontal axis represents time. The more important the gesture, the closer it is to the top of the gestural score. In this view, vowel gestures are the most important since they carry the rhythm and stress of speech, and velic gestures are at the bottom as the least important (Browman & Goldstein, 1986).

The time dimension is included in a gestural score not as real time, but as defined by the vowel gestures. Two vowel gestures define the full circle (360°) in production, and consonantal gestures are defined in relation to them at a quarter cycles or 90°, 180° and 270°. Thus, two articulatory events are considered to be simultaneous if they are occurring at the same quarter-cycle phase relative to the vowel cycle (Browman & Goldstein, 1986).

In a later version of the model (Browman & Goldstein, 1992a), schematic gestural scores display the duration of individual gestures and overlap between gestures, but not the explicit phasing relative to the vowel. In this model each gesture is represented by a box, whose horizontal dimension represents the interval of time in which this gesture is active. If there is overlap between articulatory gestures, it means that more than one gesture is activated at that particular time. In each box the parameter of that gesture is also given as the constriction degree and constriction location. An example is given in Figure 2.4 for word *pan* (TB = tongue body, TT = tongue tip, VEL = velic aperture, GLO = glottal apert).

The time during which vocal tract variables are activated and their overlap is controlled by the gestural score. In this model coarticulation is included as a consequence of gestural overlap. The gestural score itself is the input to the task dynamic model (for an overview see Hawkins, 1992), which then calculates the exact movements of articulators.

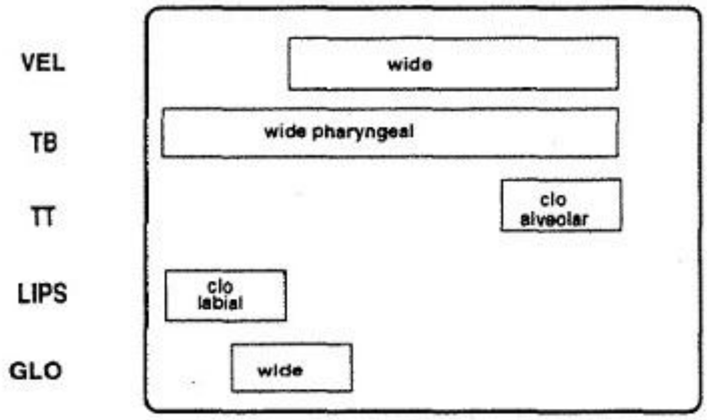

Figure 2.4 An illustration of a schematic gestural score for word *pan* in the model of articulatory phonology (reproduced from Browman & Goldstein, 1992a, p. 25)

Browman and Goldstein (1990) include a rhythmic tier to specify information about stress, as well as two functional tiers, one consonantal and one vocalic, to represent the articulatory overlap between vowels and consonants within a syllable.

The authors claim that this model can explain and capture both cross-linguistic differences and within-language contrasts that result from gestural overlap and differences in gestural timing. For example, they propose that difference between word-initial aspirated stops and unaspirated stops in /s/+stop sequences in English results from specific gestural organisation: there is only a single glottal gesture present word-initially, and it is synchronised with the release of a closure gesture for single stops, and with the middle of any fricative gestures present (Browman & Goldstein, 1986).

The voicing contrast has received little attention in the articulatory phonology model. Goldstein and Browman (1986) propose that differences between phonologically voiced and phonologically voiceless stops are based on the presence or absence of a glottal opening-and-closing gesture. In this view, voiceless stops consist of two gestures: an oral constriction gesture and a glottal opening-and-closing gesture, while voiced stops consist of a single oral constriction gesture. Differences between voiceless unaspirated and aspirated stops thus arise from different timing between the two gestures and also from different size of the glottal gesture, but neither timing differences nor size differences were further elaborated.

According to Goldstein and Browman (1986), in utterance-medial position phonologically voiceless stops usually have the glottal opening-and-closing gesture, while phonologically voiced stops do not. This is true both in languages such as English and French, although the timing and the size of the glottal gesture differ. In absolute initial position the opening part of the glottal opening-and-closing gesture cannot be observed since the glottis is already open for breathing, so the contrast is signalled by the closing part of the gesture. In this case phonologically voiced stops in English and French have glottal closing well before stop release, and in phonologically voiceless stops it occurs later.

Goldstein and Browman (1986) argue that by using articulatory gestures as the basis for the voicing contrast, phenomena such as voicing-conditioned vowel duration and differences in f0 onset values in the following vowel can easily be explained for different languages, irrespective of the exact phonetic realisation of the two stop categories. However, as it stands, this model of the voicing contrast is incomplete. It does not offer any detail about the way in which differences in realisation of the voicing contrast in languages such as French and English are achieved. It is unclear how timing and size differences are specified in the gestural score for different languages, and what the role of the task dynamic model is in the realisation of this contrast.

Articulatory phonology represents a valuable attempt to overcome limitations of previous phonological models. It is one of few models that offer explicit account of speech timing. By defining gestures as units of phonological representation and units of speech production, it removes the need for translation from the level of phonological representation of an utterance to its articulation, and narrows the gap between the two levels of representation. As pointed out by Hawkins, when coupled with the task dynamic model, it "unifies the traditional issues of coarticulation, speech rate and speech style into a single framework" (1992, p. 23).

However, certain problems arise from the fact that the task dynamic model has been developed to model skilled (non-speech) movement control and is based on general physiological and physical principles. As such, the task dynamic model is universal, and cannot accommodate for speech variability, either between- or within-language. All language-specific phonetic and phonological information is to be found in the gestural score (Browman & Goldstein, 1992b). On the other hand, the gestural score itself is unable to account for language-specific allophonic variability, and this cannot be resolved without either introducing another set of language-specific implementation rules after the gestural score, as proposed by Docherty (1992), or without allowing for some of this information to be modelled within the task dynamic model, as suggested by Hawkins (1992).

The mechanism for timing in the Browman and Goldstein's model was criticised by Byrd (1996) as being too constraining and unable to account for variation in timing due to various linguistic and extralinguistic factors, such as rate, stress or register. Byrd proposes a phase window framework, which would allow more variability in intergestural timing, as a function of linguistic and extralinguistic factors. There are two main concepts in this framework: phase windows and influencers. A phase window specifies the boundaries within which two gestures can overlap. The amount of overlap (or phase) between the two gestures is constrained by biological limits and language-specific limits. Biological limits restrict the amount of variability induced by language-specific limits. Byrd proposes only a small number of phase windows, one each for consonant-to-vowel, vowel-to-consonant, consonant-to-consonant and vowel-to-vowel type of gestures. Influencers are utterance-specific (task-specific) factors that further induce variability in gestural phasing. They can be linguistic or non-linguistic, and each of them contributes to the final phasing score. Their contributions are assessed probabilistically and each factor's contribution is weighed to achieve the final phasing relationship. This weighing procedure determines where within the phase window that defines permissible phase relationships is a particular token likely to be realised. Implemented in this way, phase window concept allows for additional variability in gestural overlap to be introduced in the model, but restricts the effect of this variability so that it cannot go beyond a certain (language-specific) limit.

## 2.4 Model of timing of voicing in speech production by Docherty (1992)

In his evaluation of how previous models of speech production deal with the voicing contrast, especially in relation to how they model the timing of voicing, Docherty points out that these models describe “essentially a level of contrast” (1992, p. 202), while variation, whether between- or within-language, received little attention or none at all. He argues that these models have in common the assumption that variation can essentially be explained by two factors: either by the underlying phonological contrast or by universal phonetic processes at the speech production stage (the motor programming and the execution stages). While it is true that some of the variability can be considered universal, language-specific variability has to be accounted for in the phonetic representation. However, the phonetic representation in the majority of the existing models was unable to fulfil this task.

Several aspects of these models were criticised by Docherty. For example, in the feature-based type of representation, developed within the framework of generative phonology, an utterance is represented as a string of phonemes, each with a corresponding set of (binary) distinctive features. It is possible to represent some allophonic variation in this representation, but not variation below the level of the segment. Since timing in these models is based on units of the size of a segment, they are unable to capture observed complexity in the realisation of the voicing contrast in the time domain. This variability is too fine for the coarse (segmental) description framework. For the same reason, they cannot capture gradience in the realisation of certain voicing correlates and the resulting allophony. While they are able to specify the contrast, they cannot specify variability that is non-contrastive. A similar problem is present in the gestural model developed within articulatory phonology (Browman & Goldstein, 1986, 1992a; Goldstein & Browman, 1986). Here universal variability results from the processes in the task dynamic model, but the gestural score is unable to account for language-specific variability (as discussed above).

In sum, these representations are unable to offer an account of “micro-variability”, especially the temporal aspects of phonetic realisation of the voicing contrast. The phenomena that need to be accounted for, in Docherty’s view, are “systematic, fine-grained patterns of phonetic variability” (1992, p. 210), e.g. between-language variability and “within-language context-determined variability (not capable of explanation on other grounds)” (1992, p. 208), as well as differences in strength of certain constraints on timing patterns, both between and within languages.

As a way of overcoming these problems, Docherty introduces a parametrically based framework for phonetic description of the timing of voicing. It is conceived as purely phonetic, descriptive supplementary tool, which would be capable of providing information about fine aspects of the timing of voicing below the level of segment, with better temporal resolution. This would enable it to capture differences in the timing of voicing both between languages and within a language. The proposed framework was based on the research of voicing patterns in obstruents in Southern British English.

The framework proposes three types of templates for the timing of voicing – one each for the onset phase, for the medial phase and the offset phase in obstruent production. For each phase, it establishes possible templates of the timing of voicing in terms of whether the voicing is present or not and, if it is present, it specifies its timing (its beginning and its end, if there was a delay in voicing onset or if there was an incursion of voicing from the previous sound). Second, it establishes all possible combinations of medial and transitional templates for each segment in a number of contexts.

Each segment in each context can be matched to the appropriate set of templates using simple binary assignment (“+” if a particular template was observed, and “-“ if it was not observed). The degree of detail on the time scale can be coarser or finer, e.g. each phase in obstruent production can be divided further, and scalar values can be used to describe the timing of voicing. The scalar specification has the advantage over the binary specification because it is capable of capturing gradient phenomena, such as varying degrees of aspiration in different languages or different contexts within a language, or varying degrees of carry-over voicing etc.

This descriptive framework is further incorporated in the phonetic representation module within the model of speech production.

Within phonetic representation, Docherty introduces a *voicing timing space*, based partly on the window model of coarticulation proposed by Keating (1990). This voicing timing space consists of a number of temporal windows. Each window represents an auditory parameter relevant for the voicing contrast, such as VOT, voicing in the closure etc., and its articulatory correlates. A window defines a set of acceptable values or timing between laryngeal and supralaryngeal gestures. These sets of acceptable values are specific for a particular language.

The width of a window represents the amount of variability that is allowed for each parameter and is also specific for each language. It is assumed to be negatively correlated with the perceptual importance of that particular parameter, i.e. if a window is wide, the parameter shows more variability, which corresponds to a small perceptual importance of that parameter, and vice versa. For example, the window for voicing in the closure would be expected to be narrow and thus allowing less variability in French than in English, reflecting the fact that this cue is more important in the former than in the latter.

In addition to this, the choice of different window lengths can potentially account for the fact that within the same language cues can have different importance depending on the context, such as, for example, word-initial and word-final position in English. What is more, a combination of windows with different widths that co-vary can account for the occurrence of trade-off between cues in a particular word position.

Another important consideration is the distribution of cases within each window. Each case is assumed to be functionally equivalent in a language, i.e. each case results in the same percept. The shape of distribution can potentially depend on at least two factors. First, it can reflect gestural constraints, so that the most cost-effective options would have the biggest probability. For example, since it is difficult to maintain voicing in fricatives, the most likely constellation is one that represents lack of voicing in fricatives. In stops, on the other hand, voicing can be maintained in the closure for some time, so the distribution of cases within the window should reflect that fact. Second, while the choice of a window and its width are language-specific, the distribution within a window could reflect within-speaker differences. An example of this is the fact that in utterance-initial position in English some speakers produce phonologically voiced stops both with short lag VOT values and with negative VOT values, while other speakers do not. Speakers from the latter group would have a distribution with one peak corresponding to short lag VOT realisation, whereas speakers who sometimes prevoice would have a bi-modal distribution.

As far as fine-grained variability is concerned, Docherty proposes two possibilities. One is to have a separate window for each separate context-dependent timing pattern, such as different VOT values at different places of articulation, which would lead to overlap of the respective windows. The other possibility is to have one level of rules when defining windows for each category (e.g. voiced vs. voiceless) and then another level of rules within each category to reflect contextual (place) variation.

Docherty's model of the timing of voicing, and the parametric framework it is based on, represent a valuable contribution to the area of speech production modelling. It achieves exactly what it sets out to do: it includes the time dimension in the model of speech production in relation to the voicing contrast. It offers a way of incorporating sub-segmental variation in the model, as well as context-dependent variability, and consequently a way of modelling non-universal between-language and within-language variability. Unlike some other models, it is explicit enough to allow its application (and further development) on new data and other languages, and is potentially compatible with a number of other models, for example with Byrd's (1996) phase window concept and with exemplar models of representation. However, its main shortcoming is that it is limited to the phonetic representation module and is non-committal regarding other levels of representation of the voicing contrast, and about possible categories there. As such, it runs a risk of being a model that simply re-states the facts about the realisation of the voicing contrast, but is somewhat detached from the associated level of phonological representation.

## 2.5 Evaluation of models

In this chapter I have discussed the most elaborated models of the voicing contrast in obstruents, and in particular in stops, in relation to three main issues: the proposed relationship between phonological representation and phonetic realisation, and the choice and nature of the features (or other units of phonological representation) used to represent the voicing contrast. The models that were reviewed differ in all three respects.

In relation to the first question, they either assume two (or more) separate levels of representation, such as models by Keating, Cho and Ladefoged, and Kingston and Diehl, or they propose a more integrated approach where phonological representation is directly related to phonetic realisation, such as models by Kohler, Jessen, and Browman and Goldstein. There is no consensus about the most appropriate feature either. While Keating, Cho and Ladefoged, and Kingston and Diehl, propose the feature [±voice], Kohler proposes the feature [±tense], Jessen both [±voice] and [±tense], and Browman and Goldstein no binary defined features at all. The basis of features is seen as articulatory, as in articulatory phonology model by Browman and Goldstein and in the models by Kohler and Cho and Ladefoged; as auditory, such as in auditory enhancement hypothesis by Kingston and Diehl; or as acoustic, as in the models by Keating and Jessen.

On the other hand, what these models have in common is that they attempt to overcome shortcomings of early segment-based feature models with one-to-one mapping between the phonological features and phonetic realisations, which were essentially static and did not include the time dimension. In doing this, they offer a valuable contribution to understanding many aspects of the voicing contrast. However, they also suffer from similar problems. Some of these problems were outlined by Docherty (1992) in relation to the three of these models (by Keating, Kohler, and Browman and Goldstein), but they also hold for the later models by Jessen and Kingston and Diehl. In short, Docherty points out that they all concentrate on modelling a level of contrast, both within a particular language or between languages, rather than on modelling the realisation of this contrast. While these models acknowledge some of the variability in the realisation of the voicing contrast, mostly that coming from universal constraints in speech production (or, in the case of auditory enhancement hypothesis, general auditory processes), they fail short of modelling systematic variability that is not universal, and thus they have limited scope. They also tend to focus on stops, rather than on all obstruents.

The literature review in Chapter 1 further illustrates the complexity in the phonetic realisation of the voicing contrast, by focusing on the research which was carried out mostly independently of the modelling theoretical work described in this chapter. The research on acoustic correlates has highlighted a number of linguistic and non-linguistic sources of variability in the realisation of the voicing contrast. It complements the research on modelling of the voicing contrast by identifying what is lacking from the models in terms of the details of phonetic realisation of this contrast.

Moreover, from a point of view of someone investigating the voicing contrast in a lesser researched language, the theoretical models discussed in this chapter lack predictive power. Once the choice is made at the highest level of representation, they are quite rigid in the choice of phonetic means to realise the contrast, which is the case with the models proposed by Keating, Kohler, Jessen, and Kingston and Diehl. At the same time, some of the models are vague about the most appropriate acoustic correlates for different word positions, such as Jessen's model and, to some extent, Kingston and Diehl's model, or they deal only with one word position, such as the models by Keating and Cho and Ladefoged. The model of the voicing contrast proposed within the framework of articulatory phonology has generally paid very little attention to the voicing contrast, which makes it difficult to make any specific predictions.

Regardless of their differences, some of the models have in common that they operate with a similar set of acoustic correlates of voicing, irrespective of the phonological feature they propose and its basis, which is the case with models by Kohler, Jessen, and Kingston and Diehl (exceptions are models by Keating and Cho and Ladefoged, which are based on one correlate only, and the model of articulatory phonology, which does not include acoustic correlates).

Another problem with the existing models is that they were mainly based on research about English and/or other aspirating languages, such as German, and only include sporadic phonetic evidence from voicing languages, even in the cases where the proposed feature is [±voice]. In modelling the voicing contrast, in particular in stops, it has been assumed that in voicing languages the voicing contrast is rather uncomplicated and manifested as simply presence vs. absence of vocal fold vibration. The importance and relevance of other correlates, such as preceding vowel duration, closure duration, F1 onset etc., was included in the model based on research on aspirating languages, mostly English, and non-systematic research, if any existed, on a small number of voicing languages (the exception is Keating's (1980) research on Polish, but this is restricted to one stop pair in non-final word positions, because of word-final neutralization of the voicing contrast in Polish). The existing models are incomplete in this respect as well, and likely to be biased towards a small number of languages.

One of the aims of the present study is to fill this gap by providing data about acoustic correlates in a voicing language that has the voicing contrast in stops in all word positions, and to evaluate them in the light of the existing models. Acoustic-phonetic research of the voicing contrast in Serbian obstruents is sparse, but it is usually mentioned in textbooks as a contrast between the presence of vocal fold vibration during the constriction interval and its absence. In addition to this, Serbian belongs to the group of Slavonic languages, which have a type of contrast that is best described by some kind of feature [±voice] for stops (if a featural approach is used), which suggests that Serbian is indeed a voicing language. In the remainder of this section I will compare predictions that can be made by some of the models about the phonetic realisation of the voicing contrast in a language of this type.

Keating's model predicts that realisation of the voicing contrast in stops in a voicing language will be similar to that in Polish: in absolute initial position [+voice] stops will be realised as {voiced}, and [-voice] stops as {vl. unasp.} and there will be little allophonic variation, and little between-speaker variation. The two phonetic categories {voiced} and {vl. unasp.} should be well separated in their phonetic realisation, due to the universal rule of polarisation, and the {vl. unasp} category concentrated within a narrow area of short lag VOT values. Although there is little reference to other word positions, it seems that word-medially and finally {voiced} stops are considered as having a certain amount of voicing during closure, or fully voiced closures, and {vl. unasp.} as having mostly voiceless closures and a short voiceless period after the release.

In contrast to Keating's model, the model by Cho and Ladefoged explicitly allows for allophonic variation, such as that coming from the effect of place of articulation, in addition to the variation caused by universal physiological and aerodynamic factors. This is defined by language-specific rules after the choice has been made between the three modal categories {voiced}, {vl. unasp.}, and {vl. asp.}, and before universal phonetic implementation rules. However, since at this point language-specific rules for Serbian are unknown, it is difficult to make any predictions at all. The model does not make reference to any other correlates of voicing, and, for a voicing language, does not consider any other word position apart from word-initial. To some extent, both models can be seen as simply re-stating what they observe (in the VOT research available at that point in time), and consequently they have very limited predictive power.

According to Kohler's model, a voicing language would be expected to choose voicing in the closure over aspiration word-initially. In intervocalic position it would give equal weight to the preceding vowel duration and closure duration, as the correlates of the articulatory timing component, and voicing in the closure as the correlate of the laryngeal component. Word-finally, it has a choice of either neutralising the contrast or expressing it through differences in preceding vowel duration and closure duration. Voicing in the closure is optional, since it is seen as secondary to the articulatory component and as variable (or even absent). In addition to this, since more weight is assigned to the universal articulatory timing component, this may reinforce the reciprocal timing relationship in VC sequence word-finally (as is the case in English).

Jessen’s model is even more explicit in terms of the acoustic correlates that are used for the voicing distinction in a voicing language. The basic correlate for the feature [±voice] in stops is closure voicing (although the phonetic invariant for this feature is not specified). Substitute correlates are closure duration and preceding vowel duration, shorter closures and longer preceding vowels for [+voice] stops, and the opposite for [-voice] stops. Concomitant correlates for [+voice] stops are low F1 and f0 onset, low burst amplitude and the presence of breathy phonation. The model does not explicitly state which correlates are used in which position within the word, although it does predict that when the basic correlate is absent, substitute correlates will take on its role.

The auditory enhancement hypothesis also makes a prediction of how the voicing contrast would be realised in a voicing language. For word-medial intervocalic position and for final position it predicts that relevant intermediate perceptual properties (IPPs) are low-frequency property and C/V duration ratio, and their subcorrelates (closure voicing, F1 onset and f0 onset, and closure duration and preceding vowel duration), and for absolute initial position it is presumably VOT.

It is not possible to make any predictions about acoustic correlates from articulatory phonology model by Browman and Goldstein. Finally, Docherty’s model can only be used as a descriptive tool, once the data for Serbian is available.

The above predictions are all incomplete. The first two, based on VOT, do not include other correlates of voicing and do not include other word positions, except initial. The remaining three models do include a set of acoustic correlates, (and essentially operate with the same set of acoustic correlates), but are vague as to which ones are relevant in which position. Kohler’s model is not explicit enough about the intervocalic position, where languages are expected to use several correlates in equal measure. Jessen’s model does not specify in which conditions the basic correlate is replaced by the substitute correlates for voicing languages. The auditory enhancement hypothesis predicts that in word-medial intervocalic and word-final position, all five correlates are relevant: closure voicing, F1 and f0 onset/offset, closure duration, and preceding vowel duration. As mentioned before, these predictions also offer a somewhat simplified picture of the variability of acoustic realisation of the voicing contrast, especially in the light of the findings reviewed in Chapter 1.

## 2.6 Motivation for the present study

As has been shown in Chapter 1 and in the above literature review, there are several problems with the existing research about the voicing contrast. First, despite some valuable attempts at modelling the relationship between phonological representations and their phonetic realisations, this relationship is still poorly understood, especially at the level of phonetic realisation. There are fundamental differences between the existing models in how they view the relationship between phonological and phonetic level, and about the nature of phonological representations. Furthermore, existing models have failed to include many aspects of variability on the phonetic level, especially those that are not caused by universal constraints, and those that are non-distinctive. Second, because of the focus on English, and, to a smaller extent, some other aspirating languages, and because of a lack of systematic research on voicing languages, existing models have focused on acoustic correlates and word positions that are more relevant for the former group of languages, and many aspects of the voicing contrast in voicing languages are assumed from English, rather than being based on substantial body of research. Third, research about the acoustic correlates of voicing that was carried out mostly independently of the theoretical models has mainly been dominated by English and VOT, and for a number of years motivated by perceptual research. There is an obvious lack of detailed acoustic-phonetic studies, in particular on voicing languages, and on correlates other than VOT.

To fill this gap further systematic research on voicing languages in needed. Serbian is an ideal candidate, because it has a voicing contrast in obstruents in all word positions. In this study I investigate several acoustic correlates of voicing in Serbian stops in relation to the above-mentioned issues.

In view of the lack of previous studies of the voicing contrast in Serbian, the decision which correlates to investigate in the present study was based mainly on the research on other languages, and on some characteristics of Serbian. The following correlates were chosen: VOT, closure duration, voicing in the closure, and preceding vowel duration. Properties of the release burst, and F1 frequency and f0 in the preceding and the following vowel were not included in the present study because of space limitations. They remain a possible topic for further research. For each of the chosen correlates, stops in an appropriate context were included. VOT was measured in absolute initial position. Closure duration and voicing in the closure were measured in word-initial and word-final stops. In word-initial stops they were measured in intervocalic position, and in word-final stops in both intervocalic and pre-pausal position, in order to establish if there is any phonetic devoicing word-finally in different contexts, and its extent (if it is present). Preceding vowel duration was investigated in two word positions in which it is considered to be relevant, that is in word-medial and word-final position, in words in isolation and in a sentence frame. In addition to this, several factors that have been found to have an effect on the phonetic realisation of the voicing contrast were included: place of articulation of the stop, the quality of the following vowel, condition (isolation vs. sentence frame), and several speaker factors: gender and age of speakers, and place of birth and living. Details about the design of the study are presented in Chapter 3.

The present study will provide a detailed account of the phonetic realisation of the voicing contrast in Serbian. It has the following aims:

1. To provide a quantitative account of a number of acoustic correlates of stop voicing in a range of environments,

2. To examine the effects of a number of factors that have been found to induce variability in the phonetic realisation of the voicing contrast in other languages, especially English,

3. To establish fine details of the phonetic realisation of this contrast that are language-specific, and to examine these results in relation to data from existing studies on other languages, especially voicing languages, and

4. To evaluate existing theoretical models of the voicing contrast in relation to Serbian results.

There has been very little interest in this area in Serbian linguistics. The following section gives a summary of the most important characteristics of the Serbian sound system, relevant for the present study. Because of the lack of literature on this topic, there is no separate literature review, but instead an overview of the most important research is included in this section.

## 2.7 Serbian phoneme inventory

The topic of the present study is Standard Serbian, as spoken in the Republic of Serbia. All participants in the present study are educated speakers of varieties spoken in the north and north-west of the country, which are considered to be the base of Standard Serbian. The speakers had no non-standard features in their speech.

Serbian is a South Slavonic language that is traditionally described as having voiced and voiceless (unaspirated) stops. Three stop pairs have contrast in voicing: /b/-/p/, /d/-/t/, and /g/-/k/. Their realisation is believed to be uniform across dialects. Serbian also has a voicing contrast in fricatives and affricates. The voicing contrast is present in word-initial, word-medial, and word-final position. Table 2.1 shows the phonemic inventory of Serbian consonants.

| | Bilabial | Labio-dental | Dental | Alveolar | Post-alveolar | Palatal | Velar |
|---|---|---|---|---|---|---|---|
| Plosive | p b | | t d | | | | k g |
| Affricate | | | ʦ | | ʧ ʤ[9] | ʨ ʥ | |
| Nasal | m | | | n | | ɲ | |
| Trill | | | | r | | | |
| Fricative | | f | s z | | ʃ ʒ | | x |
| Approximant | | ʋ | | | | j | |
| Lateral approximant | | | | l | | ʎ | |

Table 2.1 Serbian consonant system

Acoustic investigations of the phonetic properties of Serbian have mainly been concerned with the system of accents, and much less with individual sounds or groups

[9] These affricates are alveolo-palatal, as the symbols reflect.

of sounds. The topics that have attracted most attention are articulation and acoustic properties of speech sounds (Krajišnik, 1994; Miletić, 1927-28, 1933; Petrović & Gudurić, 2010), the nature of Serbian affricates (Miletić, 1933; Miller-Ockhuizen & Zec, 2002, 2003; Peco, 1961-1962b; Zec, 2003), assimilatory processes (Kašić, 1980, 1985), investigations of accents (Jokanović-Mihajlov, 1983; Lehiste & Ivić, 1986; Peco & Pravica, 1972; Sokolović, 1997a, 1997b, 1997c, 1997d, 1998), and phonetic research in dialectology (Đurović, 1996; Marković & Sokolović, 2000, 2004). To my knowledge no systematic research has been carried out on acoustic correlates of the voicing contrast in Serbian obstruents. In the text that follows I will only concentrate on the topics of interest for my research.

Textbooks and older studies do not deal with the acoustics of the voicing contrast, except stating that during the closure of voiced stops the vocal folds vibrate, while during the production of voiceless stops they do not (Belić, 1968; Miletić, 1960; Simić & Ostojić, 1989). Experimental studies about the realisation of the voicing contrast are rare, and other studies do not go further than establishing the presence or absence of a voice bar in spectrograms (Petrović & Gudurić, 2010).

Regressive voicing assimilation of obstruents is present within words (across morpheme boundary). Word-finally the voicing contrast is said to be present in Serbian, although a certain degree of phonetic devoicing may occur. The devoicing of voiced consonants in final position has been mentioned in the literature about Serbian, but the extent to which such devoicing is a consistent feature of Serbian has not been fully resolved. There are two opposing views regarding this process in Standard Serbian: one view is that there is no devoicing of final consonants, the other is that there is some degree of devoicing, but there is no agreement about its nature and its scope (Belić, 1960, 1968; Ivković, 1913; Miletić, 1960; Peco, 1961-1962a; Simić & Ostojić, 1989). In the literature about dialects, complete devoicing of voiced final consonants has been observed, but mainly in the dialects that are in contact with the neighbouring languages. Other dialects have either partial devoicing or have no devoicing at all. Detailed discussion about devoicing in non-standard varieties can be found in Peco (1961-1962a). The main shortcoming of these studies is that processes in utterance-final position were not separated from processes in other environments (before a voiced or a voiceless sound), and that there was no clear definition of a phonetically devoiced stop. This topic is further discussed in Chapter 6.

Serbian has a five-vowel system: /i/, /e/, /a/, /o/, and /u/[10], and a syllabic trill /r/. There is a phonemic length contrast so each of the five vowels, as well as the syllabic /r/, can be short and long. The phonemic length distinction appears in stressed syllables or the syllables immediately following the stressed ones.

Quantity, stress, and tone are combined in an accentual system which is traditionally regarded as having four accents: short falling, short rising, long falling, and long rising. Monosyllabic words always have falling accents. Falling accents also occur on the first syllable of a polysyllabic word. Rising accents occur on any syllable except final. Final syllables cannot be accented.

The domain of the accents includes the stressed syllable and the syllable that follows it. In syllables with falling accents fundamental frequency reaches its maximum in the first half of the syllable, and then falls. If there is another syllable after the accented syllable, it continues to fall to a lower value. In syllables with rising accents, f0 rises, sometimes until the end of the syllable, and continues to rise or stays on the same level in the next syllable. The distinction between short falling and short rising accent in disyllabic words is in the f0 level in the second syllable, which is higher for the rising accent. The distinction between long falling and long rising accent is in the f0 contour in the accented syllable, which is falling for the falling accent and level or rising for the rising accent, and in f0 level in the following syllable, which is higher for the rising accent than for the falling accent (Jokanović-Mihajlov, 1983; Lehiste, 1970; Lehiste & Ivić, 1986; Peco & Pravica, 1972; Sokolović, 1997a, 1997b, 1998).

The phonemic length contrast is mainly expressed through duration. Vowels have fairly constant quality in stressed and unstressed syllables (Lehiste, 1970; Lehiste & Ivić, 1986).

The type of accent (falling or rising) does not affect vowel quality. Vowel quantity in stressed syllables generally has no effect on /i, u/, and syllabic /r/, but can have some effect on /e, a, o/, so that short vowels are centralised and lowered compared to long vowels. However, these differences are generally small, and they are speaker-dependent (Lehiste, 1970; Lehiste & Ivić, 1986; Sokolović, 1997d).

[10] I use symbols /a, e, i, o, u/ to denote the five-vowel system in Serbian, as is usually done in literature. They do not represent their exact phonetic characteristics, but I will not address this issue because it is not relevant for the present study.

# Chapter 3 Methods

## 3.1 Linguistic material

Material for the present study consisted of 99 words. They are all real Serbian words (one word is a well-known abbreviation which is read as a CVC word). All six stop consonants in Serbian (/p/, /t/, /k/, /b/, /d/, and /g/) were represented in the material. Five tokens of each stop in word-initial position and eight tokens of each stop in word-final position were included. Another set of words was added for measuring preceding vowel duration. Because the number of minimal pairs suitable for this analysis is limited, in this set two or three minimal pairs of words were used to represent each cognate pair of stops. Monosyllabic words were used to measure vowel duration before word-final stops, and disyllabic words to measure vowel duration before word-medial stops. In addition to this, six words were added to the word list for a pilot study. The full list of words is given in Appendix A. The structure of words that were used for each type of measurement is outlined below.

For measuring acoustic correlates in word-initial position words with the structure SVC(C) were used (S = stop). One word for a combination of each stop followed by each of five Serbian vowels /a, e, i, o, u/ was included in the material (6 stops x 5 vowels = 30 words). All five vowels were included in order to examine the effect of the quality of the following vowel on the realisation of stop voicing, based on the literature review in Chapter 1. In the majority of words the vowel was phonologically short (28 out of 30). There was one word with phonologically long vowel and one with alternate pronunciation (either short or long vowel). Prior to the statistical analysis of results, a pilot study was carried out to check whether phonological vowel length affects realisation of acoustic correlates of voicing contrast in the preceding stops (see Appendix B). Because results from the pilot suggested that this is not the case, results for word-initial stops before short and before long vowels were pooled and analysed together.

For measuring acoustic correlates in word-final position words with the structure (C)CVS were used. In this part of the study each stop was represented with 8 words (6 stops x 8 words = 48 words). Words with phonologically short vowels were used if a suitable word could be found. Out of 48 words, 34 had phonologically short vowels, and

14 had long vowels. Vowel quality was not controlled, but words with vowels of different height before each stop were used, if available.

For measuring preceding vowel duration words with the following structure were used: for vowels before word-final stops (C)(C)CVS, and for vowels before word-medial stops (C)CVSV. For this correlate a set of 17 minimal or near-minimal pairs was chosen, where the difference was in the voicing of the stop. There were eight word pairs with a phonologically short vowel and nine word pairs with a phonologically long vowel, and the two sets of data were analysed separately where appropriate

The number of tokens of target segments provided by the word list described above is given in Table 3.1.

| | |
|---|---|
| Word-initial stops | /b/ - 5, /p/ -5, /d/ - 5, /t/ - 5, /g/ - 5, /k/ - 5 |
| Word-final stops | /b/ - 8, /p/ -8, /d/ - 8, /t/ - 8, /g/ - 8, /k/ - 8 |
| Word-final stops (for preceding vowel duration) | /b - p/ - 3, /d - t/ - 3, /g - k/ - 3 |
| Word-medial stops (for preceding vowel duration) | /b - p/ - 2, /d - t/ - 3, /g - k/ - 3 |
| Pilot study | /b/ - 6, /p/ - 6 |

Table 3.1 The number of tokens provided by the word list for each stop in the present study

In order to shorten the recording session, which was about 50-60 minutes, 19 words were used to measure acoustic correlates of voicing in more than one position. For example, some words with the structure SVS or CCVS were used to measure voicing correlates in both initial and final position, or to measure vowel duration before the final stop and voicing correlates in initial or final position. In addition to this, six words with word-initial stops were used both in the pilot study and in the main study. These words were randomized with other words, and presented in the same way as the rest of the words.

## 3.2 Subjects

Twelve native speakers of Serbian, six male and six female, participated in this study. They were between 23 and 62 years old (the average age was 41). All subjects were speakers of Standard Serbian. They had no noticeable non-standard features in their speech, and they reported no known history of speech or hearing disorders. Six of the subjects are educated to the secondary school level (until the age of 19), and six have a university degree. They will be referred to by their initials: females BCf, DARf, MCf, MRf, MVf, SCf, and males BPm, DRm, IJm, IVm, MPm, RVm.

Ten of the subjects live in Serbia. Eight were recorded in Serbia, and two (IVm and MVf) were recorded while on a short visit to the UK. They are all effectively monolingual. Although they all had a foreign language at school (which would have been one of the following languages: English, French, German or Russian), they are not functionally effective in this other language.

Two subjects, RVm and BPm, who live in the UK, were recorded in the UK. At the time of recording, they had lived in the UK for seven and eight years, respectively. RVm speaks Serbian at home and with his friends, which he reported as about half of his weekly language usage at the time of recording. BPm does not speak Serbian at home and he reported speaking relatively little Serbian, with friends and family members who are Serbian speakers. Both BPm and RVm had English as a foreign language at school. They moved to the UK as adults. Since there could potentially be some influence from English on their production of VOT in Serbian (Sancier & Fowler, 1997; Tobin, 2009a, 2009b), results for these two subjects were analysed both separately and with other results (see Chapter 4).

## 3.3 Recording and analysis

Words for this study were embedded in a longer list of words intended for a follow-up study (the total number of words was 243). One set of words for the follow-up study had the same structure as the words described above, but instead of stops contained fricatives and affricates. The other set contained consonant clusters in word-initial or word-final position. All words were randomized and presented to the subjects for reading in Cyrillic in two conditions: in isolation and in a sentence frame

“Reci____osam puta” /ˈreʦi____osam ˈputa/ (Say____eight times). As outlined in Section 2.6, the first condition, isolation, was used in order to establish acoustic correlates of stop voicing in utterance-initial position and in utterance-final position. The sentence condition was designed to provide intervocalic position for word-initial and word-final stops, and to examine if there is an effect on acoustic correlates of voicing when a word is embedded in a sentence.

Isolated words were presented in 18 blocks: 14 blocks of 14 words, 3 blocks of 13 words, and 1 block of 8 words. Each block started and ended with a couple of fillers to avoid listing effect in reading. In the sentence condition, words were presented in 24 blocks: 21 blocks of 11 sentences, and 1 block of 12 sentences.

Subjects were instructed to read at a habitual, natural rate. If a word was unfamiliar, they were instructed to read it in a way that felt right to them. For minimal pairs of words that differ only in vowel length, it was necessary to give additional instructions to the subjects, in order to elicit the word with the intended vowel length. Below each of such words, an explanation, usually another word with the same meaning (or a preposition plus a word), was given as a clue to what was required of them. Explanations were also given for some less frequent words. This was explained to the subjects prior to the recording. In order to familiarise the subjects with recording, each session started with a trial block that contained five words in isolation, followed by a break, and then five words in the sentence frame. After each trial block sound level and the quality of recording were checked and adjusted before proceeding with the main task. Trial blocks were not used for analysis.

Recordings were made in a quiet room. The utterances were recorded onto a Toshiba laptop (Intel Pentium M Processor 1.6 GHz) via an M-Audio MobilePre USB audio box and a Sony ECM-MS907 electret condenser microphone. The sampling rate was 44.1 kHz.

Reading material was presented to subjects on the laptop screen using Prompt and Record program (ProRec v. 1.0) developed by Mark Huckvale at University College London[11]. ProRec presents timed text prompts on the screen and at the same time records speech onto the hard disc of a computer. For words read in isolation, the program was set to display a new word every 3 seconds, while for the sentence

[11] Available online from http://www.phon.ucl.ac.uk/resource/prorec/.

condition a new sentence was displayed every 5 seconds. These durations were chosen because they were judged as the most suitable to provide a comfortable, habitual rate of speech, especially in the sentence condition. As was discussed in Chapter 1, speaking rate can have an effect on the realisation of some correlates of voicing and this program was used in order to minimise variations in speaking rate. Because of timed prompts no utterance was longer than the pre-set duration, although the minimal length could not have been controlled.

The recording session was about 50-60 minutes long. There was a short break after each block, and the subjects moved to the next block when they were ready. In the sentence condition, one block was omitted by mistake when recording subjects SCf and MCf, so that three word tokens were lost for SCf and six tokens for MCf.

## 3.4 Segmentation and measurements

### 3.4.1 Segmentation criteria

The software package Praat v. 4.5.14. was used for acoustic analysis (Boersma & Weenink, 1992-2012).

In the present study, segment boundaries were determined from visual displays of waveforms and wideband spectrograms of recorded speech tokens. Wideband spectrograms were used for first-hand orientation, and waveform displays for fine details and for the final decisions about the placement of a boundary. Segmentation criteria which were used in the present study are mainly based on Turk et al. (2006). In their approach, segmental durations are defined by oral consonant constrictions (onsets and releases), while duration of a vowel (in a CVC sequence) includes formant transitions, burst, and aspiration of voiceless aspirated stops. For each type of measurement, segmentation criteria are outlined below.

**VOT, voicing onset and stop closure duration**

VOT is defined as time between the release of a stop and the onset of glottal vibration (Lisker & Abramson, 1964). The release of a stop is usually visible on waveform and spectrogram displays as a burst of noise. In the present study, the release was marked at the onset of the burst on the waveform, if it was consistent with the burst on the spectrogram. In case of multiple bursts, the first one was taken as the beginning of the release (following a discussion on Phonet mailing list[12]; for other approaches see Foulkes, Docherty, & Jones, 2010).

The onset of voicing can be determined either from spectrograms or from waveform displays. On spectrograms, several landmarks have been used in the past, such as periodic striations in the F1 region (Peterson & Lehiste, 1960), the first vertical striations corresponding to glottal pulsing (Lisker & Abramson, 1964), or the appearance of energy in higher formants (Klatt, 1975). In recent studies the onset of voicing is often determined from the waveform displays, for example at the start of the first glottal pulse (Kessinger & Blumstein, 1998), at the start of the first complete vibration of the vocal folds (Cho & Ladefoged, 1999), or at the zero crossing before the upward movement of the first full cycle of vibration (Francis, Ciocca, & Yu, 2003). To evaluate different methods, Francis et al. (2003) compared measurements of voicing onsets obtained by electroglottography with five different measurements obtained from the acoustic signal. One measurement was taken from the waveform, at the beginning of the first full cycle of oscillation, and the remaining four from the spectrogram: at the onset of voicing bar, the first formant, the second formant, and the third formant. After aspirated stops, measurements taken from the waveform and from the voicing bar in spectrograms were more accurate and less variable than measurements taken at the landmarks based on higher formants (although after unaspirated stops there was no statistically significant difference). Following these findings, in the present study the voicing onset after a stop was labelled at the beginning of the first full cycle of oscillation on the waveform, at the positive zero crossing (Figure 3.1).

Voicing onsets for prevoiced stops in absolute initial position were determined from waveform displays, and marked at the onset of the first glottal pulse (Figure 3.2).

[12] Available online from http://jiscmail.ac.uk/lists/PHONET.html (8. June 2005)

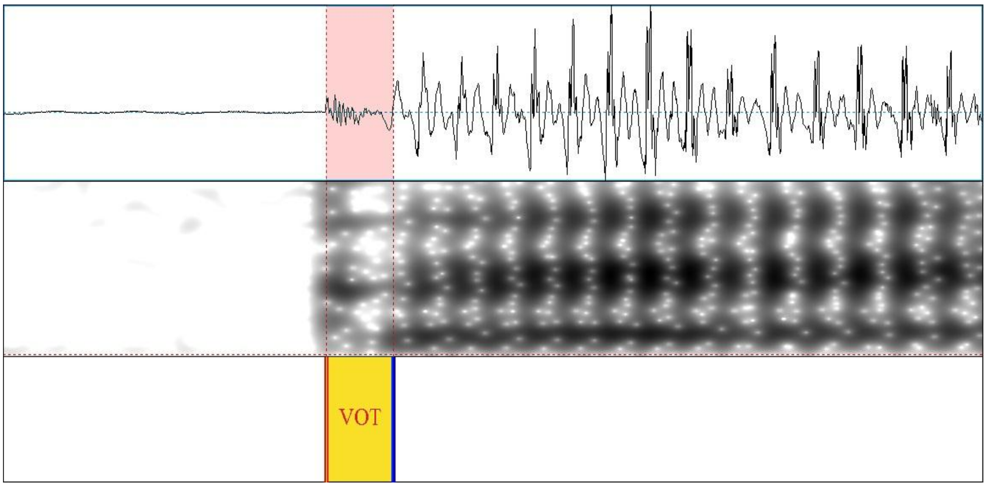

Figure 3.1 Measurement of Voice Onset Time

The figure shows part of the waveform and spectrogram display for one token of *peh*. The area between the two vertical lines represents VOT for the stop [p]. The first line is positioned at the onset of the release for [p] and the second line at the onset of voicing for the following vowel, using the criteria defined in the text.

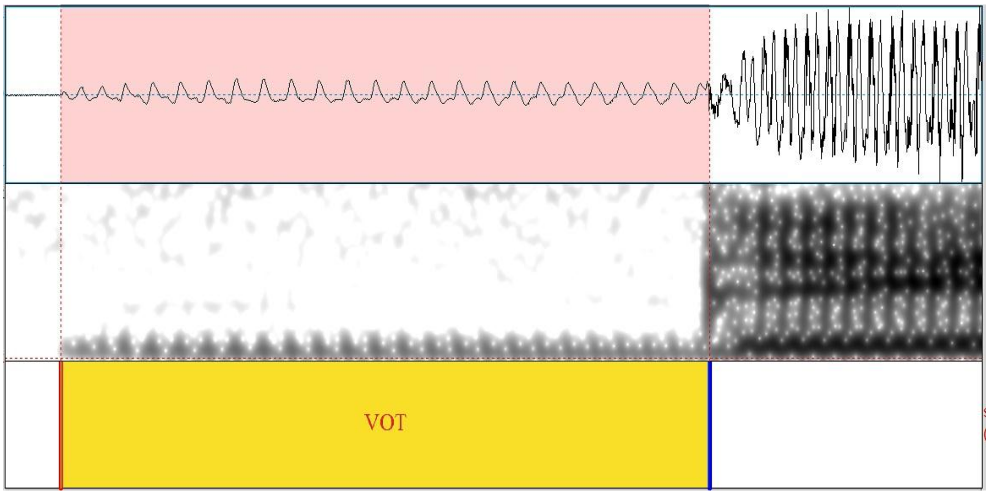

Figure 3.2 Measurement of Voice Onset Time in prevoiced stops

The figure shows part of the waveform and spectrogram display for one token of *ded*. The area between the two vertical lines represents VOT for the first [d]. The first line is positioned at the onset of voicing and the second line at the onset of the release for initial [d], using the criteria defined in the text.

On spectrograms, the onset of stop closure after a vowel is characterised by a decrease in amplitude and a loss of the second formant and the higher formants (Keating, et al., 1983; Turk, et al., 2006). On a waveform display, there is a drop in amplitude and a change in the waveform from complex to more sinusoidal (Docherty, 1992; Mack, 1982). In the present study, spectrograms were used for orientation and the onset of stop closure was marked from waveforms using these criteria (Figure 3.3).

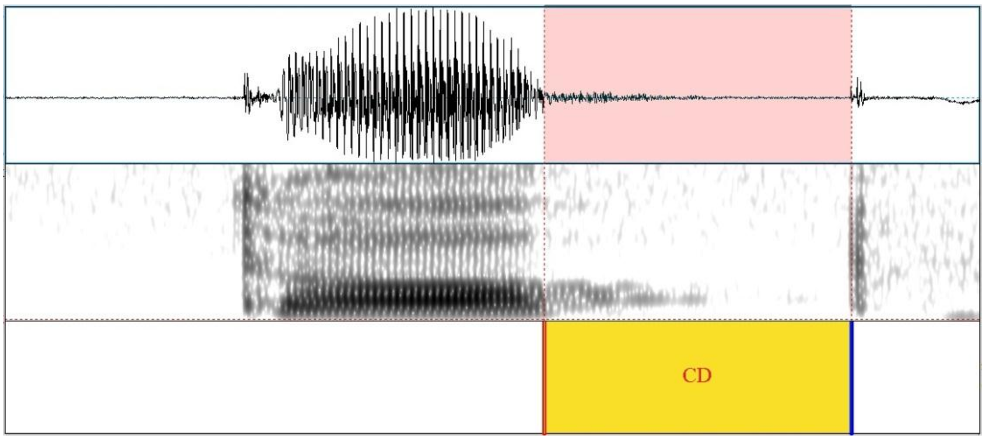

Figure 3.3 Measurement of closure duration

The figure shows waveform and spectrogram display for one token of *top*. The area between the two vertical lines represents closure duration (CD) for the final stop [p]. The first line is positioned at the offset of the preceding vowel and the second line at the onset of the release for [p], using the criteria defined in the text.

**Vowel duration**

In the present study, vowel durations were measured to examine the effect of stop consonant voicing on the preceding vowel. Previous studies on this topic have adopted different approaches to determining vowel duration. In early studies, in stop-vowel-stop sequences the beginning of the vowel was marked at the stop burst and any aspiration, if present, was included in the vowel duration (Peterson & Lehiste, 1960; Zimmerman & Sapon, 1958). In other studies the beginning of the vowel was marked at the onset of voicing, whether on waveforms (Kessinger & Blumstein, 1998; Laeufer, 1992; Mack, 1982), or on spectrograms, for example at the onset of F1 (Cochrane, 1970), at the onset of formant structure (Chen, 1970), or using both presence of voicing and formant structure (Wardrip-Fruin, 1982). In some studies it is not clear which approach was used (Davis & Van Summers, 1989; Edwards, 1981; Sharf, 1962). It seems that a majority of researchers opted for the duration of vowel without preceding aspiration (VOT), so that cross-linguistic comparisons of the effect of the second stop voicing on vowel duration are not affected by VOT differences in the first stop, although this was not always explicitly stated. The same approach was used in the present study and vowel duration did not include VOT of the preceding stop (which is different from Turk et al.'s approach).

In the set of words used for measuring vowel duration, vowels were preceded by a stop, a nasal, the lateral /l/, or the trill /r/, and followed by the stop in question.

The onset of the vowel after a stop was determined as the onset of F2 and higher formants in conjunction with the increase in amplitude and more complex waveform pattern. The offset of the vowel before a stop was marked at the onset of stop closure, as outlined above.

After the lateral /l/ and after nasals there was often a clear spectral discontinuity at constriction release (Figure 3.4 and Figure 3.5), which was used for segmentation. After the trill /r/, whether it was realised as a trill or as a tap, there was a spectral discontinuity and a dip-and-rise in the waveform, which was used to determine the boundary (Figure 3.6).

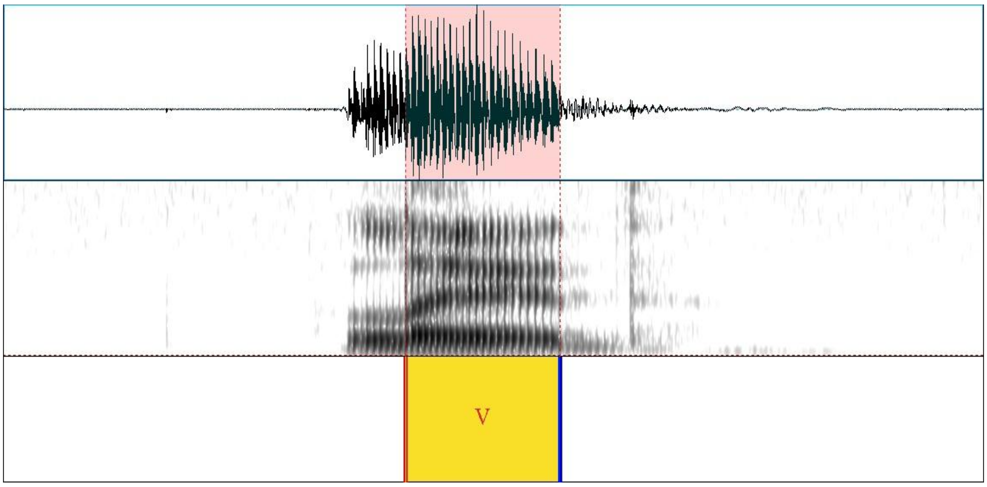

Figure 3.4 Segmentation at the /l/-vowel boundary

The figure shows waveform and spectrogram display for one token of *led*. The area between the two vertical lines represents duration of the vowel [e]. The first line is positioned at the [l]-[e] boundary and the second line at the offset of the vowel [e], using the criteria defined in the text.

In cases where segmentation was not straightforward, the following approaches were adopted. For example, in some tokens with prevoiced stops, there were irregularities in vocal fold vibration either at the beginning of the voicing, or just before the burst. In the latter case the intensity of vocal fold vibration was low or there was a short break in voicing. These irregularities were included in VOT measurement, and the frequency of their occurrence is discussed in Section 4.1. Further, in some utterances with intervocalic /b, d, g/, there was no visible release burst. In these cases closure duration was measured until the onset of the following vowel, which is potentially a small difference of only a few milliseconds (one or two cycles of oscillation), and results were included in analysis. The number of these tokens was too small to be analysed separately.

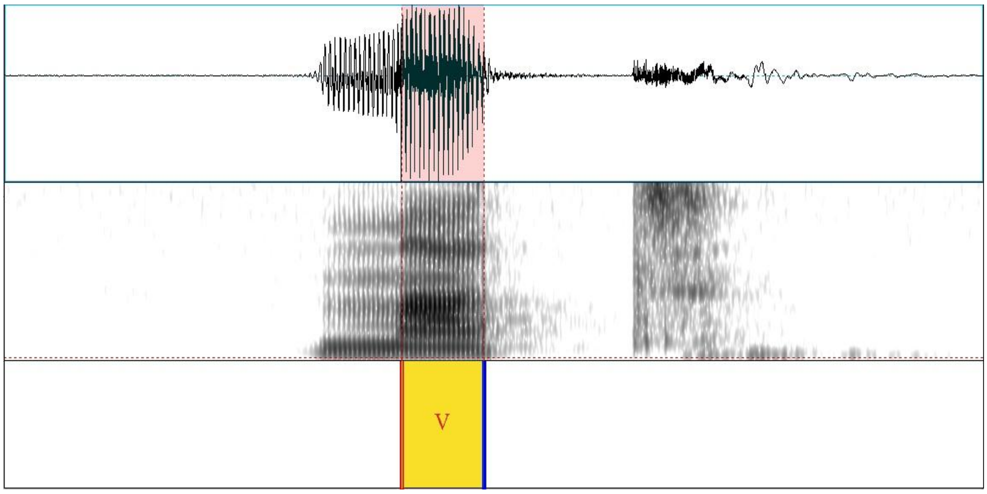

Figure 3.5 Segmentation at the nasal-vowel boundary

The figure shows waveform and spectrogram display for one token of *mat*. The area between the two vertical lines represents duration of the vowel [a]. The first line is positioned at the [m]-[a] boundary and the second line at the offset of the vowel [a], using the criteria defined in the text.

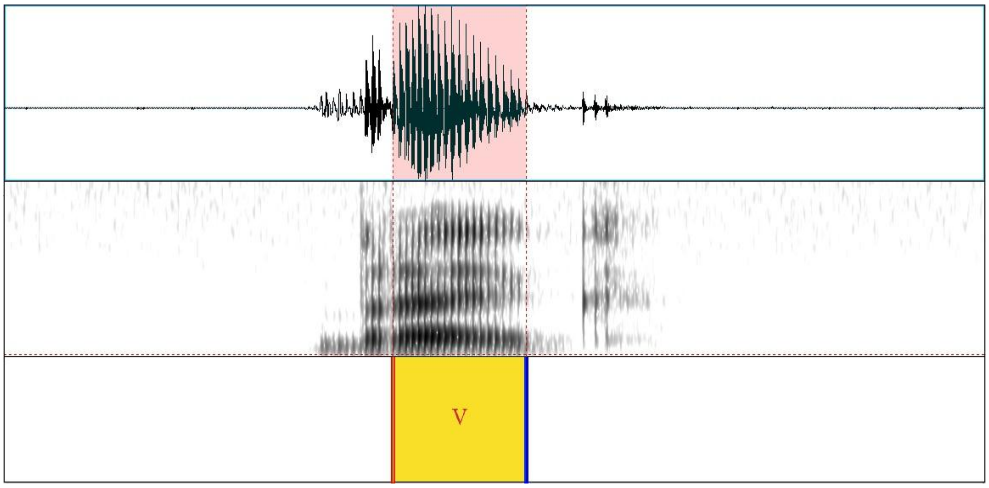

Figure 3.6 Segmentation at the /r/-vowel boundary

The figure shows waveform and spectrogram display for one token of *breg*. The area between the two vertical lines represents duration of the vowel [e]. The first line is positioned at the [ɾ]-[e] boundary and the second line at the offset of the vowel [e], using the criteria defined in the text.

A subset of 190 tokens was re-analysed after a period of time in order to check the consistency with which the above segmentation criteria were applied. A Pearson's correlation coefficient *r* or Spearman's correlation coefficient $r_s$ (for non-normally distributed data) was obtained for each type of measurement. The correlation coefficient between the first and the second measurement was: for VOT $r_s = 0.997$, $p<0.001$ (2-tailed), for closure duration $r_s = 0.981$, $p<0.001$ (2-tailed), for voicing in the closure $r_s = 0.968$, $p<0.001$ (2-tailed), and for vowel duration $r = 0.977$, $p<0.001$ (2-tailed). Strong correlation in all cases suggests that segmentation criteria were applied consistently in the present study.

### 3.4.2 Measurements

To extract measurements from TextGrid files, an existing Praat script by Remijsen (2004) was used, which I slightly modified for this thesis. This script takes durational measurements (in the present study VOT, vowel duration, closure duration etc.) from TextGrid files and displays them in a table. Measurements were rounded to the nearest millisecond.

A total of 2367 utterances were available for analysis for the present study (99 words x 2 conditions x 12 subjects minus 9 that were not recorded), but 256 (or 11%) were discarded, which left 2111 utterances for analysis. There were several reasons for discarding data. First, some tokens were discarded due to background noise or pronunciation problems, such as hesitation, mispronunciation, or exaggeration. Second, in the sentence condition, a number of tokens were discarded because there was a pause before or after the target word, and, as a result, the obstruent in question was not in intervocalic position as intended. Third, in the material intended for measuring voicing-conditioned vowel duration, which consisted of minimal or near-minimal word pairs, if one word from the pair was discarded, the other word had to be discarded from analysis as well. Fourth, some tokens were discarded because certain segments were realised in such a way that accurate segmentation was difficult or impossible. These include tokens with unreleased final stops, and tokens with intervocalic /g/ realised as an approximant, both intended for measuring closure duration. Similarly, accurate segmentation was not possible at some nasal-vowel and lateral-vowel boundaries, and at /r/-vowel boundaries where /r/ was realised as an approximant.

## 3.5 Statistical analysis

Statistical analysis was performed using statistical software SPSS Statistics 17.

For each separate statistical test performed in this study, the distribution of measured values was examined visually in a normal probability plot and tested for normality using a Shapiro-Wilk normality test (Field, 2009), and other parametrical test assumptions were checked, depending on the type of test. If violations of parametric test assumptions were large, a non-parametric test was used, such as Mann-Whitney U-test or Kruskal-Wallis test. In case of small violations, a parametric test was used where possible. There are several reasons for this. First, all the data in this study were ratio type data measured with high precision, and as a consequence all differences are finely grained. When a non-parametric test is used on such data, some of this information is lost when transforming data into ranks (Sheskin, 2000). Second, for some of the analyses in this study, there are no non-parametric alternatives to parametric tests. Third, most parametric tests are robust to small violations of the underlying assumptions (Field, 2009; Stevens, 1996). For example, if a distribution is non-normal but this deviation comes from skewness or kurtosis, a parametric test (such as ANOVA) can be used since both skewness and kurtosis have only a slight effect on level of significance or power (Stevens, 1996). In addition to this, if the test of choice is an ANOVA and there is another violation of assumptions, such as non-homogeneity of data or unequal group sizes, a Post-hoc test which is robust to a particular violation can be chosen (Field, 2009). A corrected level of significance of $\alpha = 0.01$ was used for all tests in this study, due to the high number of tests performed, and also due to the fact that non-normal distribution was frequent in my data, and in some cases the assumption of homogeneity of data was not satisfied.

Another non-parametric procedure, CART analysis (Classification and Regression Tree) was used to test for significant differences induced by variables that were not included in the initial design of the study. CART produces a classification tree by splitting results for the dependent variable into groups on basis of the chosen independent variable(s). In the present study it was used to explore VOT differences between subjects based on their place of birth (Chapter 4). An ANOVA would not have been suitable here due to unequal numbers of tokens in each group. In addition to this, it was used as an exploratory tool to gain more insight into between-subject differences. In this case, they were grouped according to their production results, not based on any pre-defined factor (see Chapters 4 to 7). Results of this analysis are statistically significant at 0.05 level, but the exact *p*-value is not supplied in the SPSS output, and consequently it was not reported in the present study.

To further test the importance of the observed effects, an effect size was calculated for each statistical test: $\omega^2$ for ANOVA, Cohen's *d* for t-test, and effect size estimate *r* for Mann-Whitney U-test. Effect size is a measure that indicates the degree of association between independent and dependent variables. Because it is independent of the size of the sample (unlike *p*-value), it serves as a better basis for comparisons, whether within or between studies. It is usually interpreted as being small, medium, or large, and guidelines are defined for each effect size measure separately.

Effect size $\omega^2$ was chosen over $\eta^2$ because it is an unbiased estimate with more rigorous assumptions. The following guidelines were used for this measure: 0.01 = small, 0.06 = medium, 0.14 = large (Field, 2009, p. 390). For one-way ANOVAs $\omega^2$ was calculated using the following formula[13]:

$$\omega^2 = \frac{\mathrm{SS_b} - \mathrm{df_b} * \mathrm{MS_w}}{\mathrm{SS_t} + \mathrm{MS_w}}$$

where $\mathrm{SS_b}$ = sum of squares between groups, $\mathrm{df_b}$ = degrees of freedom between groups, $\mathrm{MS_w}$ = mean square within groups, $\mathrm{SS_t}$ = total sum of squares, as presented in an ANOVA results table in SPSS. For two-way ANOVAs it was calculated using the formula:

$$\omega^2 = \frac{\mathrm{SS_{factor}} - \mathrm{dffactor} * \mathrm{MS_{error}}}{\mathrm{SS_{tc}} + \mathrm{MS_{error}}}$$

where $\mathrm{SS_{factor}}$ = sum of squares of a factor, $\mathrm{df_{factor}}$ = degrees of freedom of a factor, $\mathrm{MS_{error}}$ = mean square of error, $\mathrm{SS_{tc}}$ = sum of squares total corrected, as presented in an ANOVA results table in SPSS.

Cohen's d was calculated from *t*-values and degrees of freedom: $d = \sqrt{\frac{t^2}{t^2+df}}$ . The following guidelines were used for Cohen's *d*: 0.2 = small, 0.5 = medium, and 0.8 = large (Pallant, 2007, p. 208).

[13] Effect size $\omega^2$ was calculated using a program written by Jalal Al-Tamimi.

The effect size estimate *r* was calculated from Z-value as: $r = \frac{Z}{\sqrt{N}}$ , where N is total number of cases. The guidelines used for *r* were: 0.1 = small, 0.3 = medium, and 0.5 = large effect (Pallant, 2007, p. 223).

Findings are presented in Chapters 4 to 7 for each of the acoustic correlates of the voicing contrast that was investigated: VOT, closure duration, voicing in the closure and preceding vowel duration.

# Chapter 4 Results for Voice Onset Time (VOT)

## 4.1 Realisation and distribution of VOT

In word-initial position in words uttered in isolation phonologically voiced stops were realised with negative VOTs (as prevoiced). This was true for all subjects and for all word tokens without exception. Phonologically voiceless stops were realised with short lag to intermediate VOTs (as unaspirated to slightly aspirated). Measured VOT values range from -311 ms to -44 ms for voiced stops, and from 2 ms to 86 ms for voiceless stops. The distribution of measured VOT values is shown in the histogram in Figure 4.1[14].

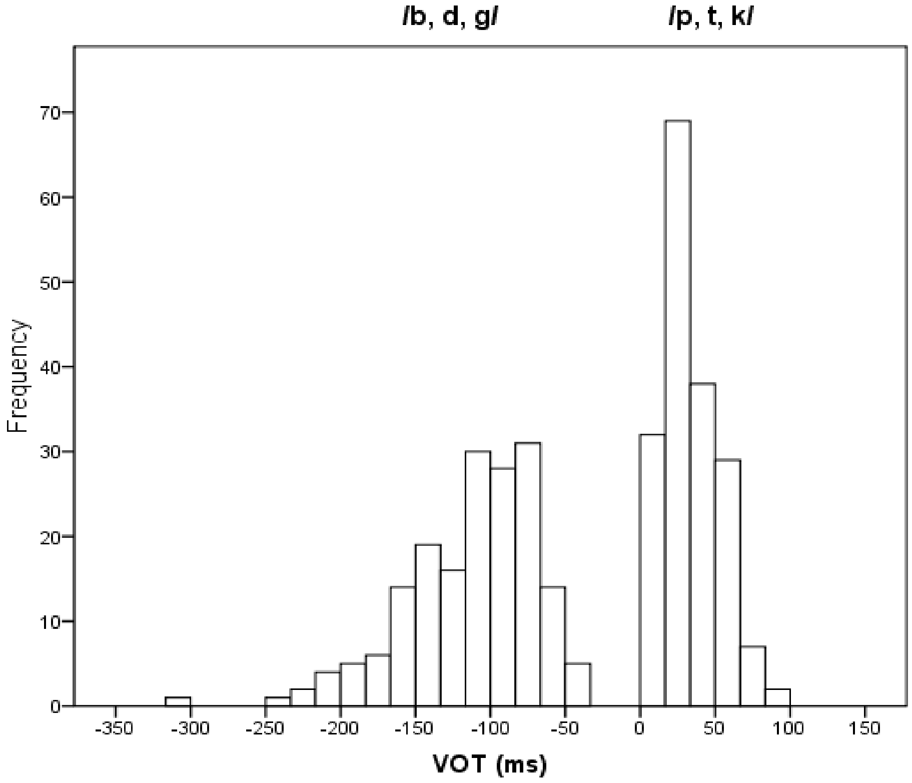

Figure 4.1 Histogram showing the distribution of VOT values in utterance-initial stops

There is no overlap between the two categories. They are clearly separated in the pooled data and in the data for each subject.

[14] As can be seen from Figure 4.1, there is an outlier with extreme VOT value. This outlier was removed before statistical analysis, because of its potential to skew results.

The mean VOT in the data pooled across subjects is -112.15 ms for voiced stops and 33.50 ms for voiceless stops (*SD* = 41.83 and 18.48 respectively)[15]. The width of separation between the two categories expressed as difference between the two means is 145.65 ms in the pooled data, and the distance between medians is 134 ms. The distance between means for individual subjects varies from 94.2 ms (subject SCf) to 183.96 ms (subject MVf). The distance between the measured VOT values on each side closest to the other category is 46 ms in the pooled data, and for individual subjects it varies from 49 ms (subject SCf) to 108 ms (subject IVm).

The difference between the two VOT categories is significant in the pooled data, Mann-Whitney U-test, $p < 0.001$ (2-tailed), $Z = -16.23$, $N = 352$, $r = -0.87$ (large effect), and in the data for each subject (Table C1 in Appendix C).

Some further details about the phonetic realisation of utterance-initial stops deserve to be mentioned. Stops were consistently released in this position. All stops, except one, were produced with a visible release burst. In prevoiced stops voicing was generally maintained without interruption. Exceptions to this include six tokens (out of 176) with a short period of irregular voicing close to the beginning of voicing, and three tokens with a very short break after the beginning. In some tokens the amplitude of voicing is low just before the burst (four tokens, all by BCf), or this period of low-intensity voicing ends in a voicing break just before the release (seven tokens). On the other hand, prevoicing can have high amplitude relative to the signal. In 17 tokens, there is even a vocalic element in the prevoiced part of the stops (subjects MPm and DARf contributed more than other subjects did, with five and six tokens respectively). An illustration of a vocalic type of resonance is given in Figure 4.2 (the figure suggest that in this particular token implosivisation might have been used, which is one of active manoeuvres used to sustain voicing, as is discussed in Section 4.4.2).

In word-initial intervocalic position phonologically voiced stops were mostly realised with fully voiced closures and VOT could not have been measured. Results for voicing in the closure are presented in Chapter 6.

[15] Because results from the pilot study suggested that phonological vowel length does not affect realisation of the voicing contrast in the preceding stops, results for word-initial stops before phonologically short and before phonologically long vowels were pooled together in all analyses.

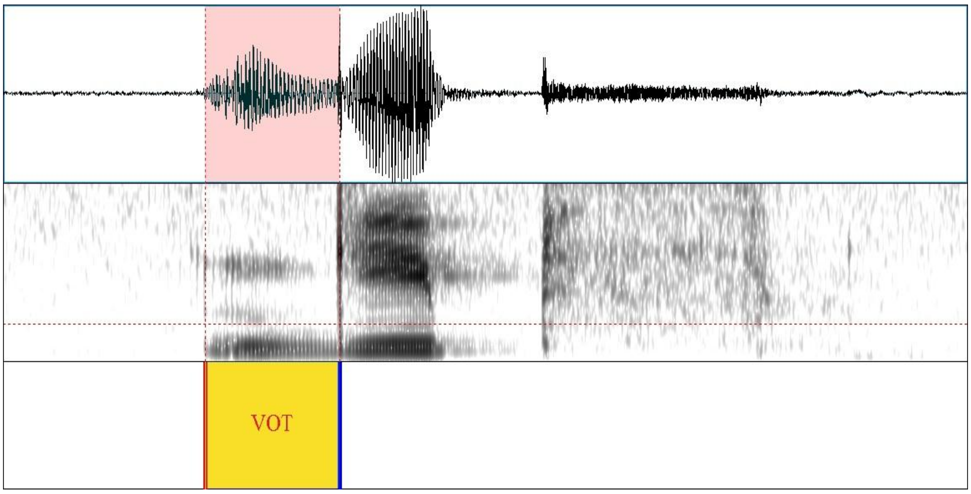

Figure 4.2 Illustration of a vocalic element in [g] in word *gips* uttered by subject DARf

In the sentence condition phonologically voiceless stops were realised in the same way as in the initial position, that is, with short lag to intermediate VOTs, or as unaspirated to slightly aspirated voiceless stops. Measured VOT values range from to 2 ms to 86 ms, and the mean is 32.7 ms. Their distribution is shown in the histogram in Figure 4.3.

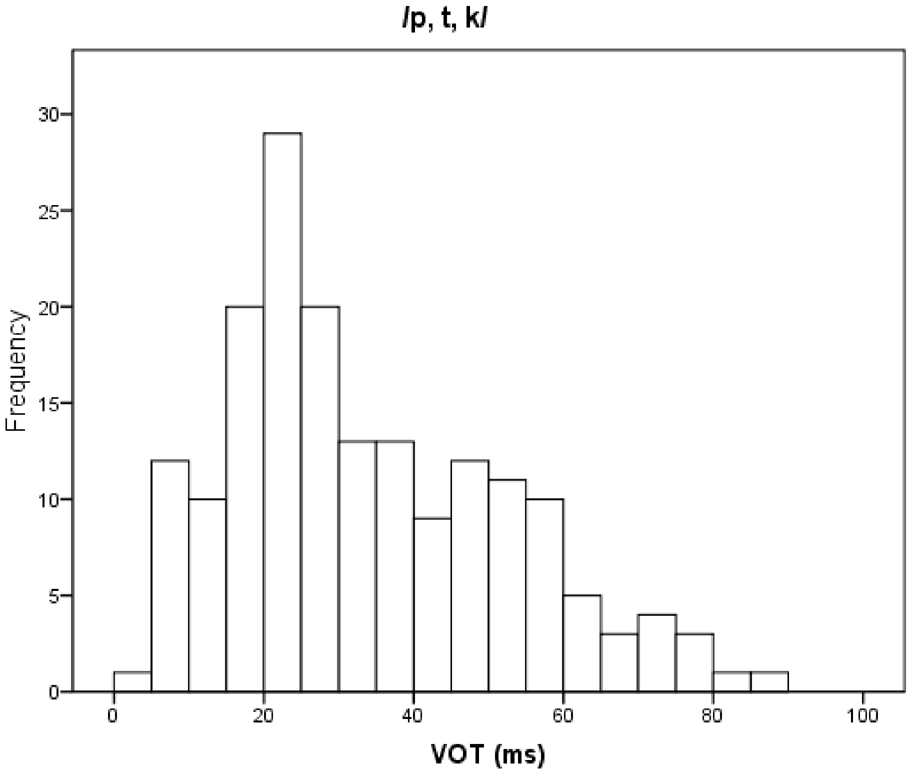

Figure 4.3 Histogram showing the distribution of VOT values for /p, t, k/ in word-initial intervocalic position

The difference between VOT results for voiceless stops in the two conditions, in isolation and in the sentence frame, is not statistically significant (Mann-Whitney U-test, *p* = 0.86, 2-tailed, *Z* = - 0.17, *N* = 349). The difference between their means is very small (0.8 ms). This small difference could be explained by the structure of the sentence frame used in the second condition ("Reci__osam puta" Say__eight times). The target word had a prominent place within the sentence and subjects pronounced it with care, in a similar way as if it was in isolation. It is likely that this is the reason behind very small, almost negligible differences between VOTs in the two conditions.

## 4.2 Linguistic factors affecting VOT

### 4.2.1 Place of articulation and the quality of the following vowel

The distribution of VOT values for /b, d, g/ in utterance-initial position as a function of place of articulation is shown in boxplots in Figure 4.4 (in all boxplot figures in the present study the horizontal bar represents the median). Mean VOT values for each place of articulation are given in Table 4.1 (in this chapter and in Chapters 5 to 7 all figures and tables illustrating the effect of linguistic factors on acoustic correlates of voicing are based on the pooled data; there is further analysis of individual results).

| | | Mean VOT (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | -117.42 | 57 | 44.76 |
| Dental | /d/ | -118.12 | 60 | 40.74 |
| Velar | /g/ | -100.79 | 58 | 36.67 |
| Total | | -112.15 | 175 | 41.38 |

Table 4.1 Mean VOT (ms), number of tokens (N) and standard deviation (SD) for /b, d, g/ in utterance-initial position for each place of articulation

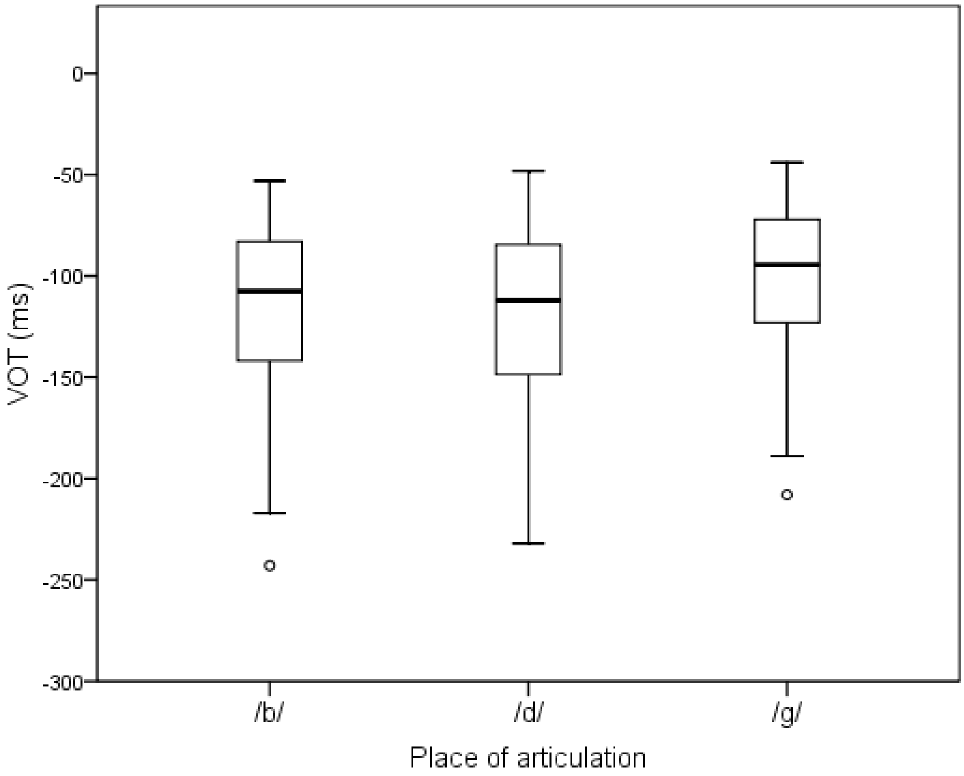

Figure 4.4 Boxplots showing the distribution of VOT values for /b, d, g/ in utterance-initial position as a function of stop place of articulation

The distribution of VOT values before each vowel is shown in boxplots in Figure 4.5, and mean VOT values before each vowel are given in Table 4.2.

A two-way between-groups analysis of variance (ANOVA) was conducted to explore the effect of stop place of articulation and the quality of the following vowel on VOT. There was no significant main effect of place of articulation on VOT, $F(2,160) = 3.25$, $p = 0.041$, $\omega^2 = 0.025$ (small effect)[16]. A Tukey HSD post-hoc test revealed no significant differences in VOT for /b/, /d/, and /g/. The main effect of the following vowel on VOT did not reach statistical significance, $F(4,160) = 0.41$, $p = 0.8$, and there was no significant interaction between the two main factors, $F(8,160) = 1.04$, $p = 0.41$.

[16] The *p*-value is smaller than 0.05, but it is above the adjusted level of significance that was set at 0.01 for this study, as discussed in Chapter 3. Because the effect size is small as well, and results from the post-hoc test were not significant, I regard this as a non-significant effect.

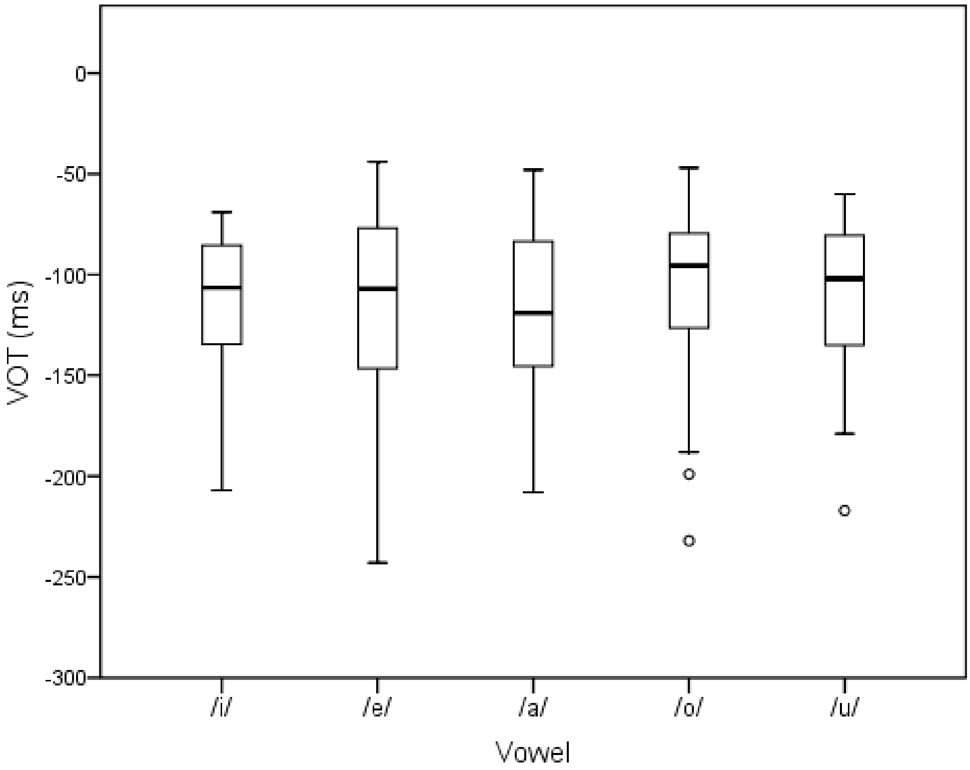

Figure 4.5 Boxplots showing the distribution of VOT values for /b, d, g/ in utterance-initial position before each vowel

| Following V | Mean VOT (ms) | N | SD |
|---|---|---|---|
| /a/ | -119.08 | 36 | 43.44 |
| /e/ | -112.54 | 35 | 47.22 |
| /i/ | -111.97 | 36 | 34.55 |
| /o/ | -108.14 | 36 | 43.48 |
| /u/ | -108.62 | 32 | 38.27 |

Table 4.2 Mean VOT (ms), N and SD for /b, d, g/ in utterance-initial position before each vowel

For /p, t, k/ in utterance-initial position the distribution of VOT values at different places of articulation is shown in boxplots in Figure 4.6, and their means in Table 4.3. There is a tendency for VOT values to increase as the place of articulation moves from front to back, in order /p/ < /t/ < /k/. Results for the velar /k/ are somewhat separated from other results.

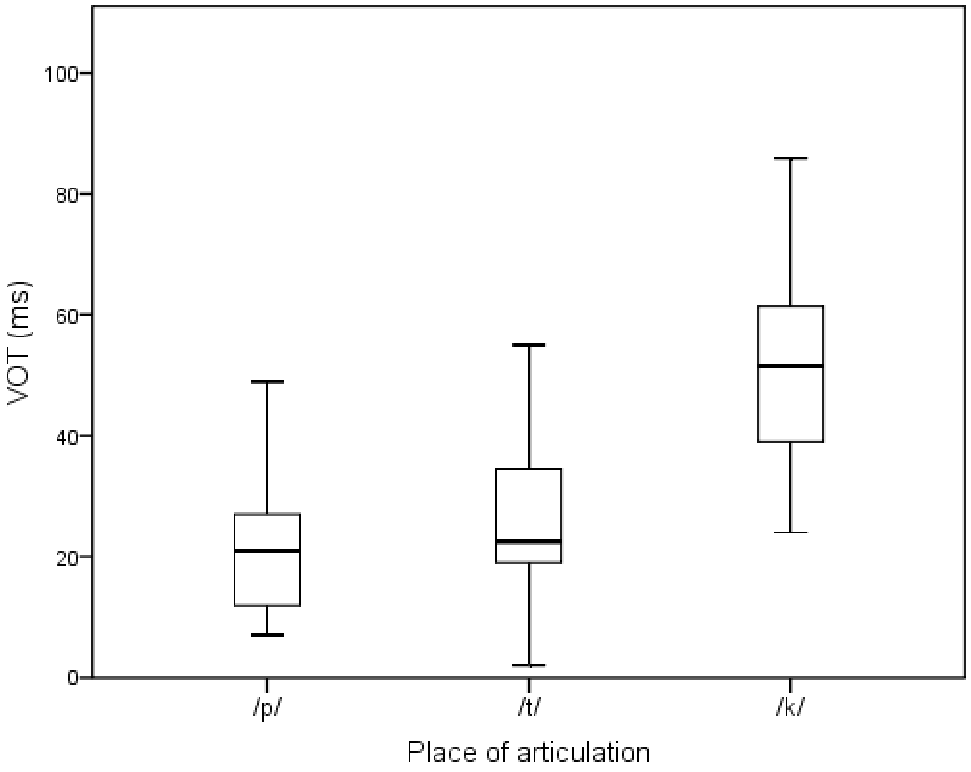

Figure 4.6 Boxplots showing the distribution of VOT values for /p, t, k/ in utterance-initial position as a function of stop place of articulation

| | | Mean VOT (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 21.67 | 57 | 11.01 |
| Dental | /t/ | 26.50 | 60 | 12.08 |
| Velar | /k/ | 51.73 | 60 | 15.32 |
| Total | | 33.50 | 177 | 18.48 |

Table 4.3 Mean VOT (ms), N and SD for /p, t, k/ in utterance-initial position for each place of articulation

The distribution of VOT values before each vowel is shown in boxplots in Figure 4.7, and mean VOT values before each vowel are given in Table 4.4.

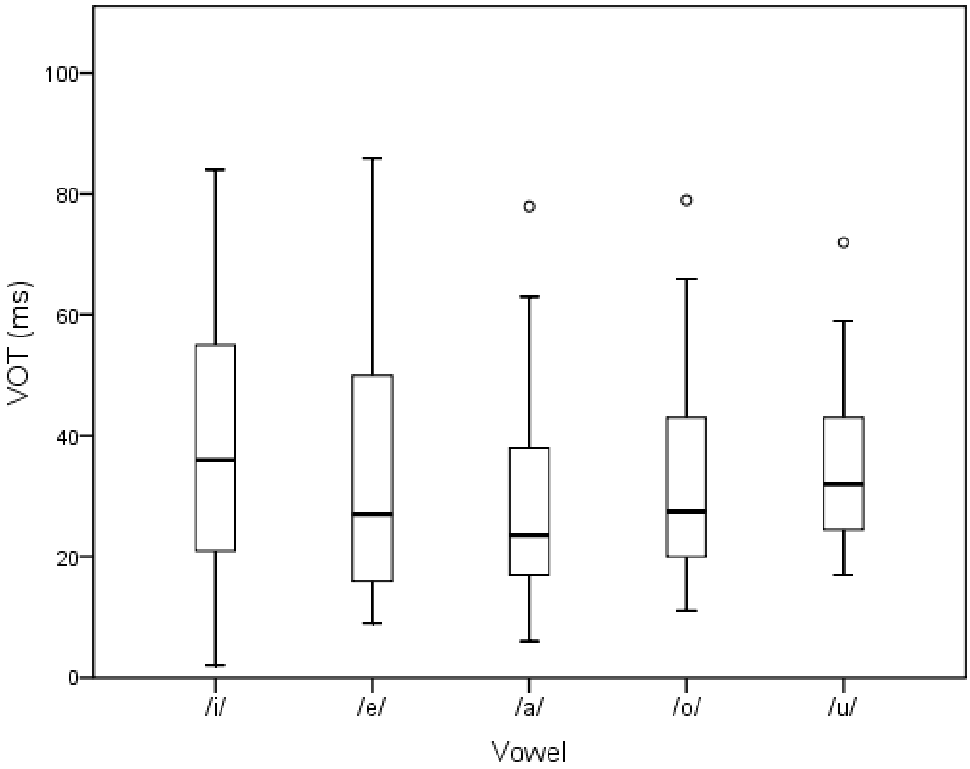

Figure 4.7 Boxplots showing the distribution of VOT values for /p, t, k/ in utterance-initial position before each vowel

| Following V | Mean VOT (ms) | N | SD |
|---|---|---|---|
| /a/ | 28.69 | 36 | 17.73 |
| /e/ | 33.69 | 36 | 21.13 |
| /i/ | 37.75 | 36 | 21.19 |
| /o/ | 32.62 | 34 | 17.19 |
| /u/ | 34.71 | 35 | 13.78 |

Table 4.4 Mean VOT (ms), N and SD for /p, t, k/ in utterance-initial position before each vowel

A two-way between-groups ANOVA revealed a significant main effect of place of articulation on VOT, $F(2,162) = 103.04$, $p < 0.001$, $\omega^2 = 0.5$ (large effect). According to a Tukey HSD post-hoc test, VOT was significantly longer for /k/ than for /p/, $p < 0.001$, and for /k/ than for /t/, $p < 0.001$. The main effect of the following vowel on VOT did not reach statistical significance, $F(4, 162) = 2.68$, $p = 0.034$, but there was a statistically significant interaction between the two factors $F(8, 162) = 3.14$, $p = 0.002$,

$\omega^2$ = 0.042 (small effect). Mean VOT values as a function of stop place of articulation and the following vowel are shown in Figure 4.8.

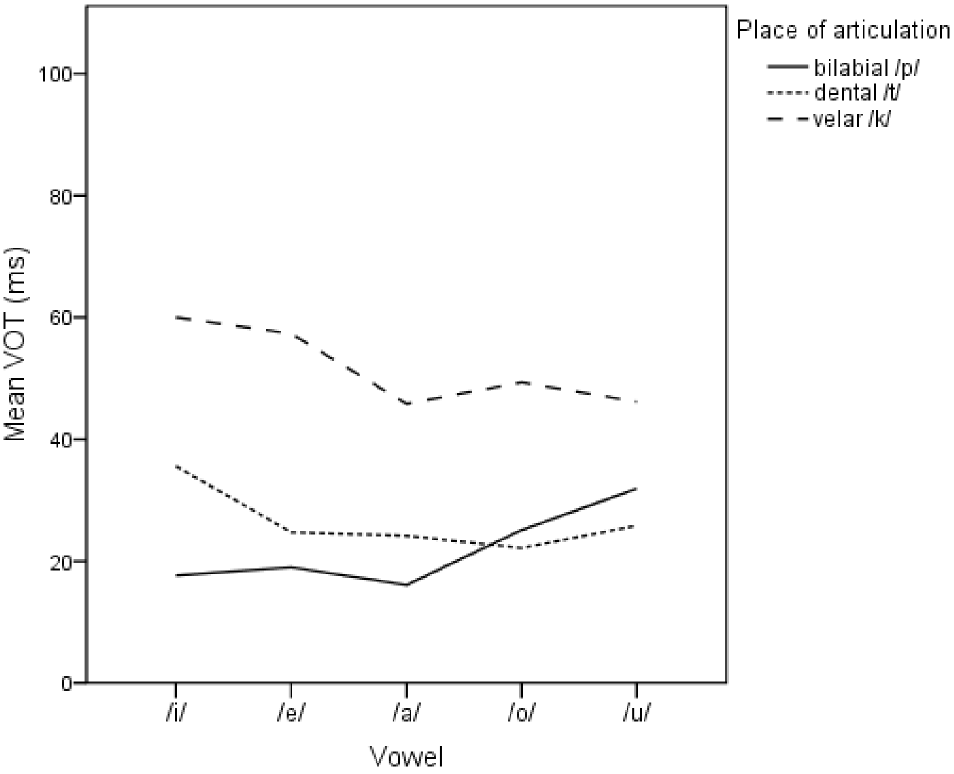

Figure 4.8 Mean VOT values as a function of stop place of articulation and the following vowel for /p, t, k/ in utterance-initial position

Because of the interaction, an additional one-way ANOVA was performed for each stop to separate the effect of place of articulation and the following vowel. The following vowel has a statistically significant effect on stop VOT only for /p/, with $F(4,52) = 5.14$, $p = 0.001$, $\omega^2 = 0.225$ (large effect). A Tukey HSD post-hoc test revealed that VOT for /p/ is significantly higher before /u/ than /i/, /e/ and /a/ ($p = 0.002$, $p = 0.019$, and $p = 0.008$, respectively); the corresponding mean VOT values are 31.91 ms before /u/, 17.67 ms before /i/, 19 ms before /e/ and 16.08 ms before /a/.

The distribution of VOT values for /p, t, k/ in word-initial intervocalic position at each place of articulation is shown in boxplots in Figure 4.9, and mean VOT values in Table 4.5. VOT increases in order bilabial < dental < velar, but values for all three stops overlap, as in utterance-initial position.

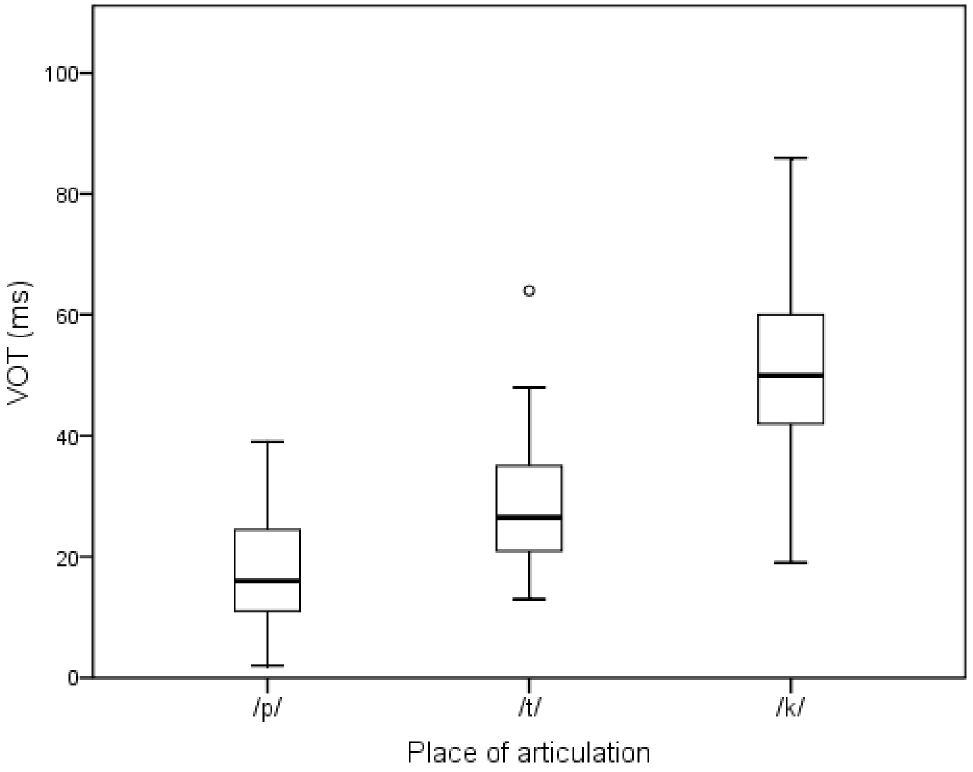

Figure 4.9 Boxplots showing the distribution of VOT values for /p, t, k/ in word-initial intervocalic position as a function of stop place of articulation

| | | Mean VOT (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 18.20 | 56 | 9.42 |
| Dental | /t/ | 28.86 | 58 | 10.02 |
| Velar | /k/ | 50.53 | 58 | 12.84 |
| Total | | 32.70 | 172 | 17.28 |

Table 4.5 Mean VOT (ms), N and SD for /p, t, k/ in word-initial intervocalic position for each place of articulation

The distribution of VOT values before each vowel is shown in boxplots in Figure 4.10, and mean VOTs before each vowel are given in Table 4.6.

A two-way between-groups ANOVA revealed a significant main effect of place of articulation on VOT, $F(2,157) = 151.38$, $p < 0.001$, $\omega^2 = 0.596$ (large effect). All three pairwise comparisons were significant, in order /p/ < /t/ < /k/, according to a Tukey HSD post-hoc test ($p < 0.001$ for each comparison).

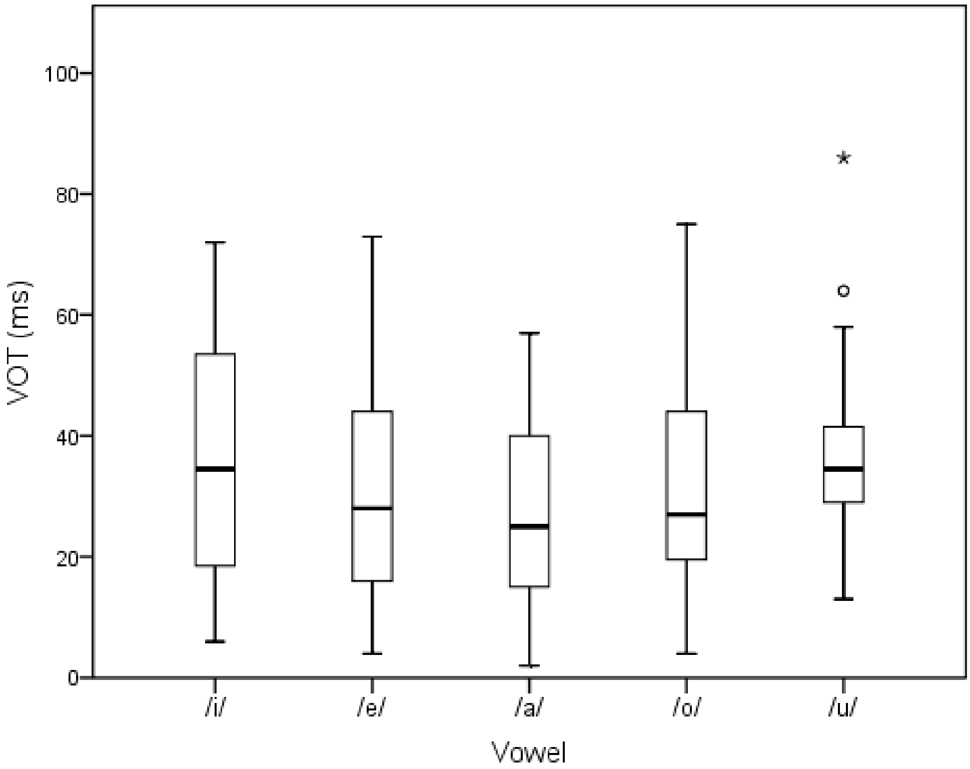

Figure 4.10 Boxplots showing the distribution of VOT values for /p, t, k/ in word-initial intervocalic position before each vowel

| Following V | Mean VOT (ms) | N | SD |
|---|---|---|---|
| /a/ | 27.51 | 35 | 15.23 |
| /e/ | 31.44 | 34 | 18.24 |
| /i/ | 35.28 | 36 | 19.44 |
| /o/ | 31.89 | 35 | 17.06 |
| /u/ | 37.69 | 32 | 15.07 |

Table 4.6 Mean VOT (ms), N and SD for /p, t, k/ in word-initial intervocalic position before each vowel

There was a significant main effect of the following vowel on VOT, $F(4, 157) = 5.33$, $p = 0.001$, $\omega^2 = 0.034$ (small effect). A Tukey HSD post-hoc test revealed that VOT was significantly longer before the vowel /u/ than before /a/ ($p = 0.001$, mean VOT = 37.69 ms and 27.51 ms respectively), and before /i/ than before /a/ ($p = 0.012$, mean VOT = 35.28 ms and 27.51 ms respectively).

The interaction between the two factors fell just short of reaching significance $F(8, 157) = 2.45$, $p = 0.016$, $\omega^2 = 0.023$ (small effect). Relationship between mean VOT values and stop place of articulation and the following vowel is shown in Figure 4.11.

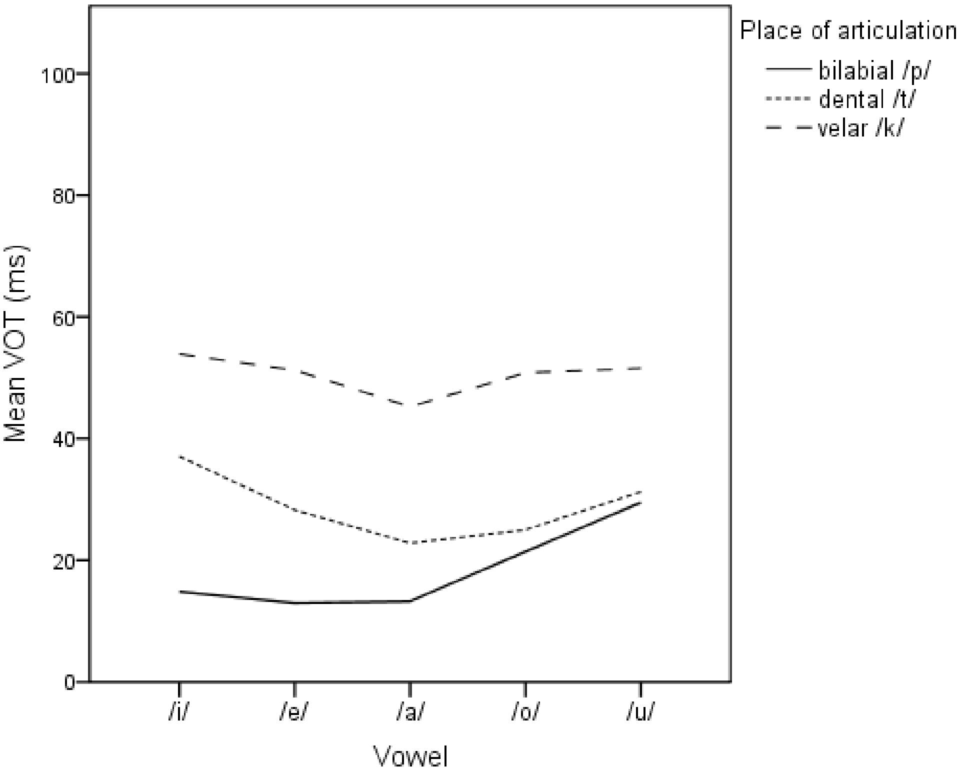

Figure 4.11 Mean VOT values as a function of stop place of articulation and the following vowel for /p, t, k/ in word-initial intervocalic position

Because *p*-value for the interaction was just above the significance level, the interaction was further examined for each stop separately using one-way ANOVAs. The following vowel has a statistically significant effect on VOT for /p/: $F(4,55) = 9.64$, $p < 0.001$, $\omega^2 = 0.382$, large effect; and for /t/: $F(4,57) = 4.68$, $p = 0.003$, $\omega^2 = 0.203$, large effect. A Games-Howell post-hoc test revealed that VOT for /p/ is statistically higher before /u/ than /i/, /e/, and /a/, $p = 0.001$, $p < 0.001$, and $p = 0.001$, respectively (with the mean VOT values of 29.5 ms, 14.83 ms, 13.00 ms and 13.27 ms, respectively). A Tukey HSD post-hoc test showed that VOT for /t/ is significantly higher before /i/ than before /a/, and before /o/, with $p = 0.002$, and $p = 0.014$, respectively (mean VOT values are 37.08 ms for /i/, 22.83 ms for /a/, and 25 ms for /o/).

There is hardly any difference between VOT results for /p, t, k/ in the two conditions. Mean VOTs are shorter in the sentence frame for /p/ and /k/, but not for /t/. None of pairwise comparisons was statistically significant.

### 4.2.2 Summary of findings

Place of articulation has limited effect on prevoicing duration. Although the velar /g/ has the shortest mean prevoicing, there is a lot of overlap in the distribution of VOT values measured at the three places of articulation, and these differences are not significant. The quality of the following vowel has no effect on the duration of prevoicing.

On the other hand, both place of stop articulation and the quality of the following vowel, as well as their interaction, affect VOT in phonologically voiceless stops. VOT of voiceless stops increases in order bilabial < dental < velar. The difference between each pair is significant in the sentence frame, while in isolation /k/ has significantly longer VOT than /p/ and /t/.

The effect of the following vowel on lag VOT is twofold: in the pooled data VOT is higher before high vowels /i/ and /u/ than before low vowel /a/, but this effect interacts with the effect of place of articulation in a way that is specific for each stop. Results for /p/ are the most consistent. In both conditions VOT for /p/ is significantly higher before the back vowel /u/ than before /i, e, a/. The same trend is present for /o/ vs. /i, e, a/, although it is not significant. For /t/ there is a tendency for /i/ and /u/ to be associated with higher VOT values. Because of different direction of this influence, VOT values for /p/ and /t/ before vowels /o/ and /u/ are similar, or they overlap. The effect of the following vowel on /k/ is small, and present as a tendency for VOT before /a/ to be lower than before other vowels.

## 4.3 Speaker factors affecting VOT

The following variables were explored as possible factors affecting the VOT values for each subject: speaker identity, age, gender, place of birth and place of living.

### 4.3.1 Individual differences between subjects

Figure 4.12 shows the distribution of VOT values for /b, d, g/ in utterance-initial position for each of the twelve subjects (in ascending order from the shortest mean prevoicing to the longest), and gives an illustration of between-subject differences. Individual mean VOTs vary from -72.07 ms (for subject SCf) to -162.29 ms (for DARf).

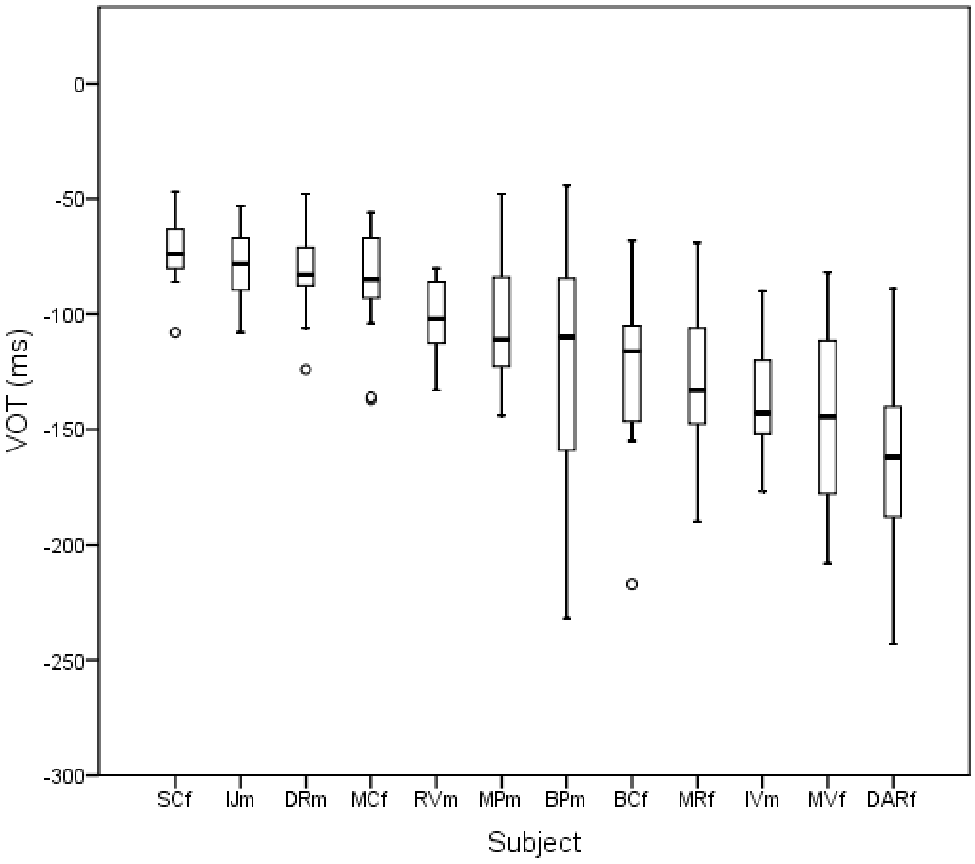

Figure 4.12 Boxplots showing the distribution of VOT values for /b, d, g/ in utterance-initial position for each subject

A CART analysis was performed to examine individual differences in VOT production[17]. There are two groups with significantly different results: Group 1, with shorter prevoicing: mean VOT = -87.89, *SD* = 23.38, *N* = 87 (subjects SCf, IJm, DRm, MCf, RVm, MPm), and Group 2, with longer prevoicing: mean VOT = -137.26, *SD* = 41.05, *N* = 86 (subjects BPm, BCf, MRf, IVm, MVf, DARf). VOTs of the subjects from

[17] For CART analysis outliers were included, since non-parametric tests are less sensitive to them.

the first group were more compact and generally clustered around -100 ms (mainly between -70 ms and -125 ms). The subjects from the second group produced longer prevoicing with a wider range. The majority of the data was in the region from -100 ms to -200 ms.

Figure 4.13 shows the distribution of VOT values for /p, t, k/ in utterance-initial position for each of the twelve subjects (in ascending order from the shortest mean VOT to the longest mean VOT). Individual differences between subjects are reflected in their means, which range from 20.33 ms (for MCf) to 48.73 ms (for RVm).

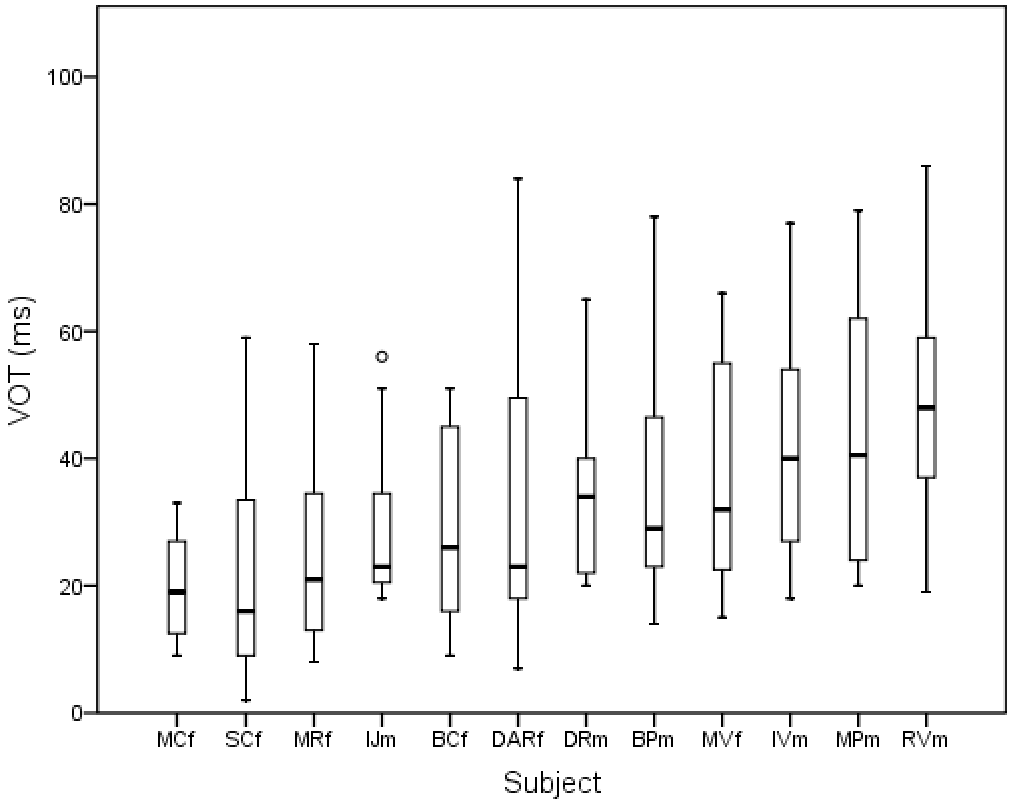

Figure 4.13 Boxplots showing the distribution of VOT values for /p, t, k/ in utterance-initial position for each subject

A CART analysis was performed to examine these individual differences in VOT values for /p, t, k/. There are three groups of subjects, whose results are significantly different: Group 1, with the shortest mean VOT values, mainly between 20 ms and 30 ms, mean VOT = 25.07 ms, *SD* = 14.1, *N* = 73 (subjects MCf, SCf, MRf, IJm, BCf); Group 2, with means between 30 ms and 40 ms, mean VOT = 35.43 ms, *SD*

= 18.26, *N* = 60 (subjects DARf, DRm, BPm, MVf); and Group 3, with the longest mean VOTs, roughly between 40 and 50 ms, mean VOT = 44.84 ms, *SD* = 18.7, *N* = 44 (subjects IVm, MPm, RVm).

Despite between-subject differences, the effect of place of articulation on VOT follows the same pattern for majority of the subjects. Individual means for /p/ and /t/ are below 35 ms for most subjects, which is considered to be within the range for unaspirated voiceless stops (the only exception is subject RVm who pronounced /t/ with longer VOT). Mean VOTs for the velar /k/, however, are higher, and above 40 ms for most speakers. More than half of the subjects have mean VOT for /k/ that is close to 50 ms, or higher (up to nearly 70 ms), which is usually not expected in voiceless stops in a voicing language. These relationships are illustrated in Figure 4.14.

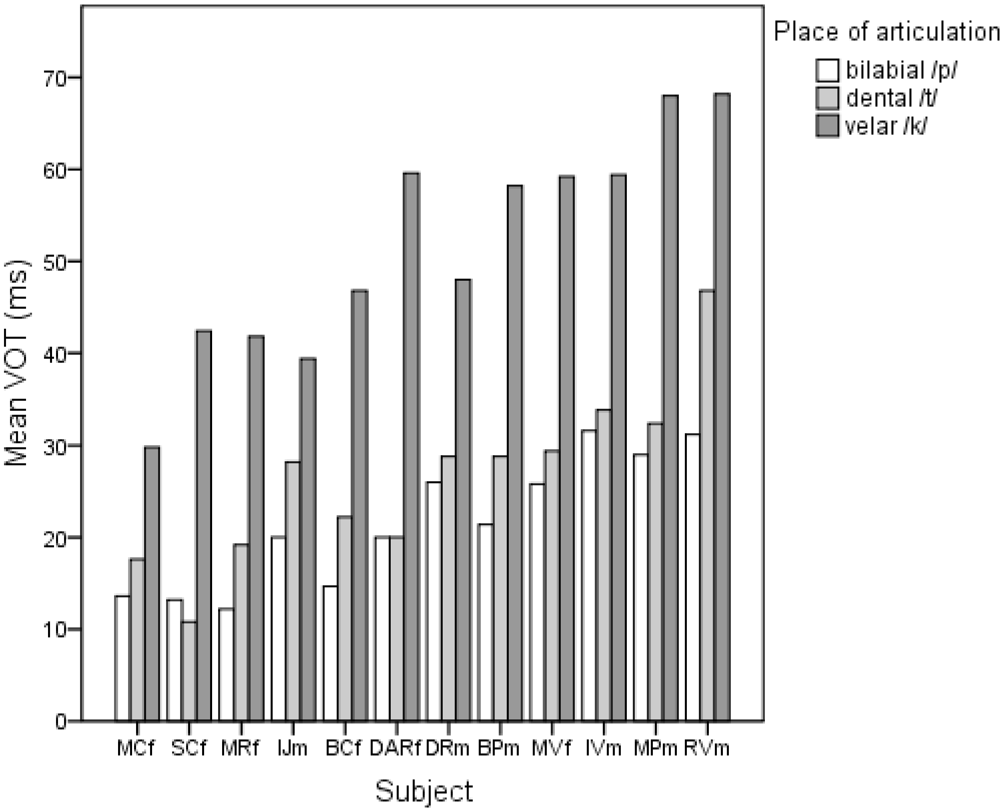

Figure 4.14 Mean VOT for /p, t, k/ in utterance-initial position as a function of stop place of articulation and subject

To sum up, results for Serbian voiceless stops in initial position in isolated words suggest that there is a split between the bilabial and dental vs. velar place of articulation, with bilabial and dental stops being produced mainly as unaspirated, and the velar stop as slightly aspirated. This is combined with the effect of individual subjects' results, where there is a group of subjects with generally longer VOT, which is especially true for the velar /k/. For this reason VOT results for Serbian voiceless stops straddle short lag and long lag (unaspirated and aspirated) category.

Observed differences in VOT production could be caused by differences in individual speaking rates. Because the isolated word material is less suitable for measuring speaking rate, correlation between speaking rate and VOT was examined on the VOT data produced in the sentence frame (see below).

Individual differences in VOT production of /p, t, k/ in word-initial intervocalic position are shown in Figure 4.15 (in ascending order from the shortest mean VOT to the longest mean VOT).

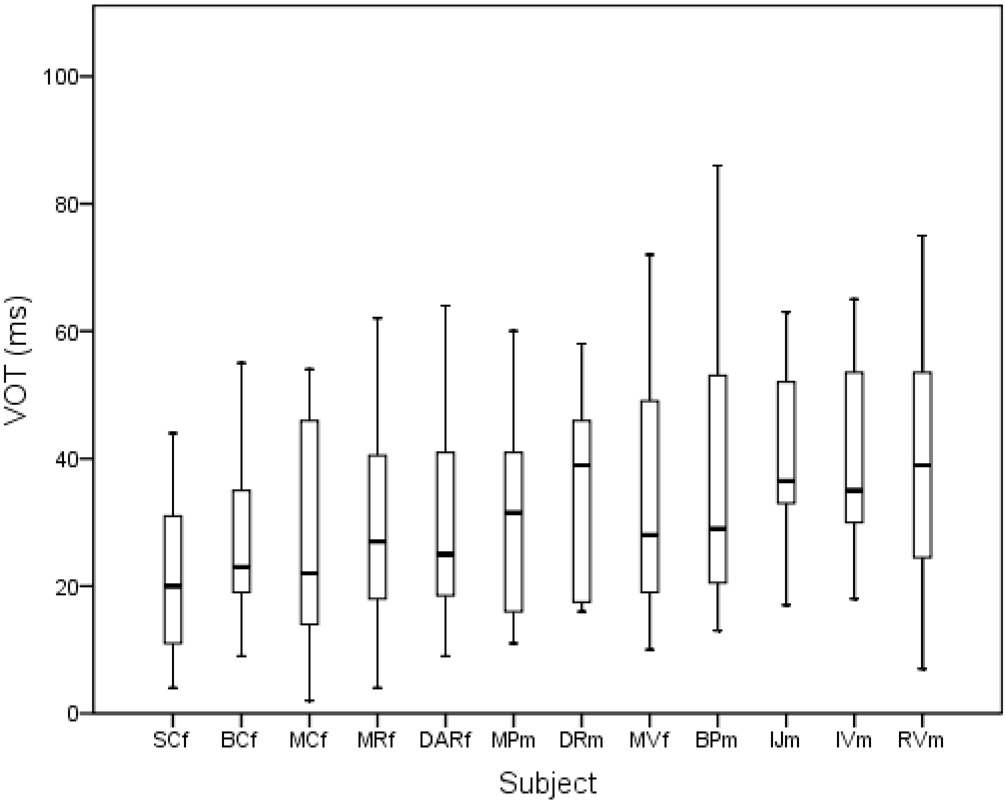

Figure 4.15 Boxplots showing the distribution of VOT values for /p, t, k/ in word-initial intervocalic position for each subject

A CART analysis divided subjects into two groups according to their individual results: Group 1, with shorter means, up to about 30 ms, mean VOT = 27.61 ms, *SD* = 15.29, *N* = 84 (subjects SCf, BCf, MCf, MRf, DARf and MPm), and Group 2: with longer mean VOT values, roughly between 35 ms and 40 ms, mean VOT = 37.56 ms, *SD* = 17.76, *N* = 88 (subjects DRm, MVf, BPm, IJm, IVm, and RVm).

These results for distributions and groupings of individual results are not the same as those obtained for /p, t, k/ in isolated words (Figure 4.13), although there are large similarities. It is noteworthy that in both conditions the subjects with lower mean VOTs tend to be females, and subjects with higher mean VOTs tend to be males (with the exception of MVf). This suggests that in this sample individual differences in VOT production of /p, t, k/ might interact with gender differences, but it was not possible to further examine this interaction using CART analysis (for analysis of the effect of gender on VOT see Section 4.3.2)

Figure 4.16 shows the effect of stop place of articulation on VOT in the sentence frame, and illustrates within-subject variation between the two conditions. The VOT values for /k/ straddle unaspirated and aspirated category in the sentence condition as well, but the order and the magnitude of difference between /p/, /t/, and /k/ for some subjects is different. In the sentence condition VOT increases as the place of articulation moves from front to back (/p/</t/</k/), and, unlike in the first condition, this is true for each subject.

There is a possibility that between-subject differences in VOT production are caused by their different speaking rates. As was discussed in Section 1.1.5, speaking rate has an effect on VOT, so that at faster speaking rates lag VOT values generally decrease. What is more, research by Theodore et al. (2009) suggested that speaking rate affects VOT in a speaker-specific way, so that the same change in speaking rate results in different degree of VOT change, depending on the speaker. Applied to the present study, this means that statistically significant differences between subjects could result from different speaking rates of subjects, and not genuine individual differences in VOT production. By factoring out speaking rate differences, a better assessment of individual differences can be achieved.

In order to test if speaking rate has any effect on VOT in the present study, speaking rate was calculated for each sentence (containing words with /p, t, k/ in initial position) as number of syllables per second. The relationship between VOT and speech rate was investigated using Spearman's Rank Order correlation. The correlation between VOT and speaking rate is very weak, and is not statistically significant (*r* = -0.067, *N* = 172, *p* = 0.19). This result suggests that speaking rate was not a factor influencing VOT production, so further analysis was conducted without speaking rate as a factor.

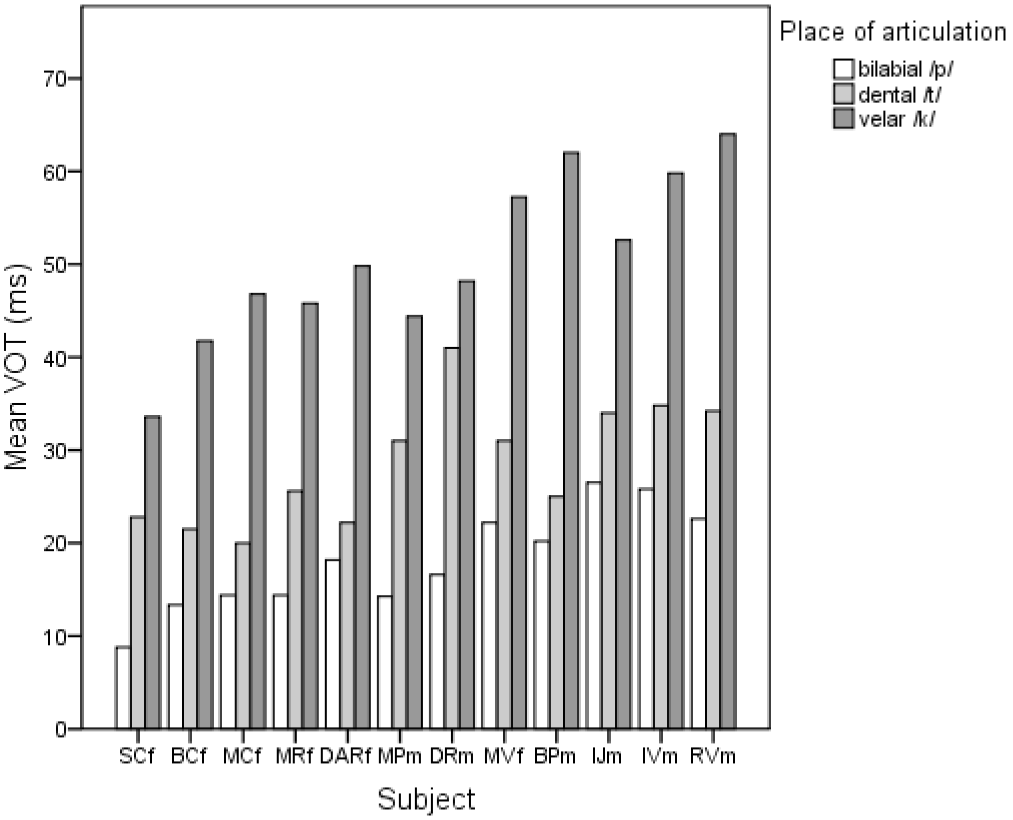

Figure 4.16 Mean VOT for /p, t, k/ in word-initial intervocalic position as a function of stop place of articulation and subject

These individual differences were further explored by separating the effects of several factors: age, gender, place of birth, and place of living.

### 4.3.2 Gender and age

A two-way between-subjects ANOVA was conducted to explore the impact of gender and age on VOT in /b, d, g/ in utterance-initial position. For this analysis the subjects were divided into two equal groups according to their age, with equal numbers of males and females in each group (and the same division was used for all subsequent analyses where there were two age groups): Group 1, with age ≤ 35 years, mean age 30.17 years (subjects MPm, DARf, SCf, DRm, MCf, and RVm), and Group 2, with age > 35 years, mean age 51.83 years (subjects BCf, IJm, BPm, MVf, MRf, and IVm).

The main effect of gender did not reach statistical significance (at the adjusted level of 0.01): $F(1,171) = 5.83$, $p = 0.017$, $\omega^2 = 0.025$ (small effect). Female subjects produced voiced stops with longer prevoicing than male subjects did (mean VOT for females = -119.48 ms, $SD = 45.02$, $N = 85$; mean VOT for males = -105.22 ms, $SD = 36.54$, $N = 90$).

There was a significant main effect of age on VOT, $F(1,171) = 14.053$, $p < 0.001$, $\omega^2 = 0.068$ (medium effect). Older subjects produced voiced stops with significantly longer prevoicing than younger subjects (mean VOT for older subjects = -123.30 ms, $SD = 41.49$, $N = 87$; mean VOT for younger subjects = -101.13 ms, $SD = 38.41$, $N = 88$). There was no statistically significant interaction between the two main factors, $F(1,171) = 0.55$, $p = 0.46$.

However, when VOT distributions are plotted as a function of age of each subject (Figure 4.17), it can be seen that, although it is true that subjects above 35 years of age as a group have longer prevoicing, the actual boundary between the two groups is between 45 and 52 years. Subjects aged 52 years and older (BPm, MVf, MRf, IVm) as a group have longer prevoicing than subjects who are 45 years or younger, and their production is more variable than in younger subjects. Prevoicing produced by younger subjects is shorter and values more compact, with the exception of subject DARf, who has longer mean prevoicing and wider range than other subjects in the same group, and is similar to older subjects. In fact, this subject has the longest mean prevoicing of all subjects, which is probably caused by her individual speaking style. Mean standard deviations reflect this tendency. The group of four subjects over 52 years of age has higher mean standard deviation than all younger subjects without DARf (39.63 vs. 28.54), while DARf has standard deviation similar to the older group (36.09).

A two-way between-subjects ANOVA was conducted to explore the impact of gender and age on VOT in /p, t, k/ in utterance-initial position. The subjects were divided into two equal groups according to their age, as before.

There was a significant main effect of gender on VOT, $F(1,173) = 17.29$, $p < 0.001$, $\omega^2 = 0.019$ (small effect). Male subjects produced voiceless stops with significantly longer VOT than female subjects (mean VOT for males = 38.96 ms, $SD = 17.89$, $N = 89$; mean VOT for females = 27.98 ms, $SD = 17.49$, $N = 88$).

The main effect of age on VOT was not significant, *F*(1,173) = 0.04, *p* = 0.84 and there was no significant interaction between the two main factors, *F*(1,173) = 5.48, *p* = 0.02.

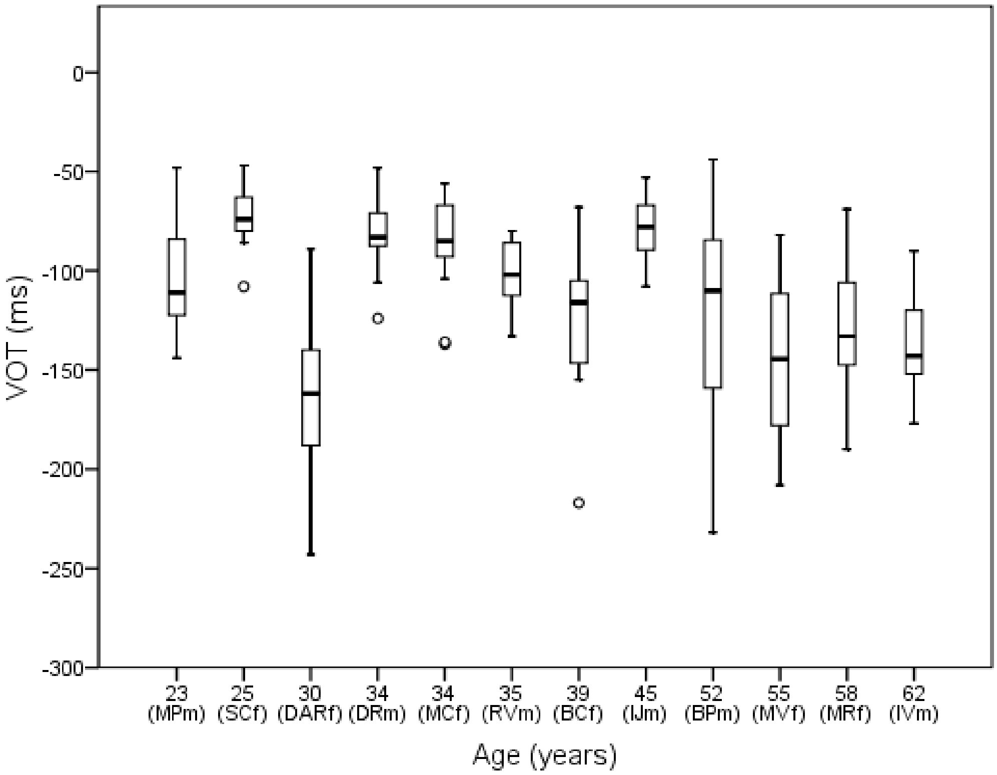

Figure 4.17 Boxplots showing the distribution of VOT values for /b, d, g/ in utterance-initial position for each subject as a function of their age

A two-way between-subjects ANOVA was conducted to explore the effect of gender and age on VOT in /p, t, k/ produced within the sentence frame. There was a significant main effect of gender on VOT, *F*(1,168) = 10.81, *p* = 0.001, $\omega^2$ = 0.054 (small to medium effect). Male subjects produced voiceless stops with significantly longer VOT than female subjects (mean VOT for males = 36.86 ms, *SD* = 16.97, *N* = 88; mean VOT for females = 28.33 ms, *SD* = 16.61, *N* = 84).

The main effect of age on VOT did not reach statistical significance *F*(1,168) = 1.6, *p* = 0.21, and there was no statistically significant interaction between the two main factors *F*(1,168) = 0.09, *p* = 0.77.

### 4.3.3 Place of birth and place of living of subjects

A CART analysis was used to examine place of birth as a factor that could influence VOT in each condition.

For /b, d, g/ in utterance-initial position, CART analysis divided subjects into two groups: Group 1, with subjects born in Čačak, who produce stops with shorter prevoicing (subjects DRm, IJm, MCf, SCf), mean VOT = -79.71 ms, *SD* = 19.45, *N* = 59; Group 2, with subjects born in Belgrade, Valjevo and Užice, who have longer prevoicing (subjects BCf, BPm, DARf, IVm, MPm, MRf, MVf, RVm), mean VOT = -128.65 ms, *SD* = 39.82, *N* = 116.

According to their VOT results for /p, t, k/ in utterance-initial position, the subjects were grouped as follows: Group 1, Čačak and Valjevo, with subjects DARf, DRm, IJm, MCf, MRf, SCf, who produced shorter VOTs in /p, t, k/ (mean VOT = 27.26 ms, *SD* = 15.79, *N* = 90); Group 2, Užice and Belgrade, with subjects BCf, BPm, IVm, MPm, MVf, RVm, who produced longer VOTs (mean VOT = 39.95 ms, *SD* = 18.91, *N* = 87).

For /p, t, k/ in word-initial intervocalic position, the resulting two groups were: Group 1: Čačak and Valjevo (mean VOT = 30.23 ms, *SD* = 15.95, *N* = 88), and Group 2: Užice and Belgrade (mean VOT = 35.29 ms, *SD* = 18.31, *N* = 84). This result replicates the result obtained for voiceless stops in isolation.

To explore the possibility that place of living affects VOT, the two subjects who live in the UK were compared to the subjects who live in Serbia, in order to establish if their daily use of English has any consequences for their VOT production in Serbian. Since both subjects who live in the UK are male, and males in this study produced shorter prevoicing and longer positive VOTs, they were compared only to the other four male subjects, to avoid confounding of this effect with the effect of gender.

Male subjects living in the UK produced longer prevoicing, with the mean VOT of -113.83 ms, *SD* = 43.03, *N* = 30, while male subjects living in Serbia produced the mean VOT of -100.92 ms, *SD* = 32.37, *N* = 60. These differences did not reach statistical significance: Mann-Whitney U-test, *p* = 0.19 (2-tailed), *Z* = -1.31. The difference in means is most likely caused by three outliers in the UK data (two outliers of -207 ms and one of -232 ms). Without the outliers the mean VOT for the UK subjects is -102.56 ms, which is almost the same as for the remaining male subjects.

For /p, t, k / in utterance-initial position, male subjects living in the UK produced longer VOT values (mean VOT = 42.43 ms, *SD* = 19.59, *N* = 30) than male subjects living in Serbia (mean VOT = 37.19 ms, *SD* = 16.88, *N* = 59). These differences did not reach statistical significance: Mann-Whitney U-test, *p* = 0.2 (2-tailed), *Z* = -1.28.

For /p, t, k/ in the sentence condition, the two male subjects who live in the UK produced VOT values that were not significantly different from those produced by the male subjects who live in Serbia (Mann-Whitney U-test, *p* = 0.97, 2-tailed, *Z* = - 0.04). Differences between them are very small: mean VOT = 38.00 ms, *SD* = 20.93, *N* = 30, for the UK group, and mean VOT = 36.28 ms, *SD* = 14.63, *N* = 58 for the Serbian group.

### 4.3.4 Summary of findings

Results presented in this chapter suggest that there is a lot of between-subject variation in VOT production, with individual means for phonologically voiced stops ranging from about -162 ms to -72 ms, and for phonologically voiceless stops from about 20 ms to 49 ms, both in isolated words. What is more, the magnitude of these differences is such that for prevoiced stops there are two groups of subjects whose results are significantly different, while for voiceless stops there are three such groups in isolation and two in the sentence frame. Several factors have been found to contribute to this variability: gender, age, and place of birth.

Gender as a factor is relevant for voiceless stops only, where male subjects produced significantly longer VOTs than female subjects, by about 11 ms in isolation, and by 9 ms in the sentence frame.

Age, on the other hand, affects only prevoicing duration, but not the duration of positive VOTs. The four oldest subjects (52-62 years) produced voiced stops with longer prevoicing than younger subjects, and they were also more variable in their production, having larger ranges and standard deviations then younger subjects (except DARf).

In addition to this, there are also statistically significant differences in VOT production related to the place of birth of subjects. Subjects from Čačak produce shorter prevoicing and shorter positive VOTs, while subjects from Belgrade and Užice produce longer prevoicing and longer positive VOTs. Subjects from Valjevo are in between, with longer prevoicing and shorter values of positive VOT.

Two male subjects who live in the UK do not differ significantly in VOT production from the remaining four male subjects, who live in Serbia. It is interesting that they have, in fact, slightly longer prevoicing duration than other subjects (by 13 ms), which is contrary to what could be expected from the literature. They also have 2-5 ms longer positive VOTs than other male subjects (5 ms in isolated words, and 2 ms in the sentence frame), although this is far from significant. It is tempting to attribute this difference to the influence of English, since English voiceless stops are aspirated. However, this could also be a consequence of their individual results. According to their place of birth they belong to the Užice + Belgrade group with higher VOT values, and they are both male, and males as a group produced longer VOTs. Therefore, although it is possible that this result represents a genuine effect of English on VOT production in Serbian, it could also be caused by several other factors.

## 4.4 Discussion

### 4.4.1 Effect of phonological voicing category on VOT in Serbian

Voice Onset Times for phonologically voiced and voiceless stops in Serbian are well separated, because all /b, d, g/ tokens were realised as prevoiced, and there is no overlap between the two categories. This is in line with findings for some other voicing languages, such as Hungarian (Gósy & Ringen, 2009), Canadian and European French (Caramazza & Yeni-Komshian, 1974; Ryalls, Provost, & Arsenault, 1995), Japanese (Shimizu, 1989), Polish (Keating, et al., 1981), Russian (Ringen & Kulikov, fc), Castilian and Latin American Spanish (Rosner, et al., 2000; Williams, 1977), and Swedish (Helgason & Ringen, 2008), where there were no positive VOT values in phonologically voiced stops, or very few of them.

The duration of prevoicing found in Serbian is among the longest reported in the literature. Overall mean prevoicing of 112 ms found in Serbian is similar to that found in Canadian French by Ryalls, Cliché et al. (1997), some Latin American Spanish dialects (Lisker & Abramson, 1964; Williams, 1977) and in Dutch (van Alphen & Smits, 2004, Experiment 1 only). In the majority of voicing languages, prevoicing is shorter than in Serbian (see Section 1.1.2, Table 1.2 and references there), and the same is true for prevoiced tokens of /b, d, g/ reported for English by Lisker and Abramson (1964) and Smith (1978).

Serbian voiceless stops /p/, /t/, and /k/ have higher VOT values than would be expected for typical unaspirated (short lag) stops in voicing languages. In fact, the number of VOT measurements of zero or just above zero is relatively small in this study, and while mean VOTs for /p/ and /t/ are within the unaspirated range, there is a tail of values for some tokens that go up to 60 ms, which is in the aspirated (long lag) range. This is even more the case with VOT values for /k/, which straddle the unaspirated and the aspirated category, with the range of up to 80 ms, which is reflected in the overall mean VOT of 52 ms. These higher VOT values are akin to intermediate VOT values found in other languages, for example in Hungarian (Gósy, 2001), Japanese (Riney, et al., 2007; Shimizu, 1989), Polish (Keating, et al., 1981), Hebrew (Obler, 1982), and a number of other languages (Cho & Ladefoged, 1999). This is an important finding because it reinforces arguments against the universal and categorical nature of the VOT categories, discussed in Section 1.1.1. Intermediate VOT values are further discussed in Section 8.2.1.

Consistent prevoicing in the realisation of phonologically voiced stops and the degree of separation between the two stop classes found in the present study suggest that the voicing contrast expressed through the measure of VOT is robust in Serbian, and VOT represents a very important acoustic correlate of the voicing contrast.

In contrast to Serbian, a proportion of positive VOT values, instead of prevoicing, was found in Lebanese Arabic (Yeni-Komshian, et al., 1977), Dutch (van Alphen & Smits, 2004), Canadian French (Caramazza & Yeni-Komshian, 1974; Caramazza, et al., 1973), and European Portuguese (Lousada, et al., 2010), and this resulted in some overlap between the two phonetic categories. The number of tokens realised in this way varies. Caramazza and Yeni-Komshian (1974) found that about 40% of /b, d, g/ tokens were realised without prevoicing in Canadian French, which led them to conclude that VOT is not a relevant measure for the voicing distinction (although some other authors found that this was not the case, cf. Jacques, 1987; Ryalls, Cliché, et al., 1997; Ryalls, et al., 1995). In Dutch, some 25% of tokens were realised with positive VOTs and there was a large variation between the speakers: while some speakers produced 100% of their /b, d/ tokens as prevoiced, others prevoiced less than 40% (van Alphen & Smits, 2004). In Portuguese, at least 15% of utterance-initial /b, d, g/ tokens were fully or partially devoiced (Lousada, et al., 2010, p. 266, Figure 3). With regards to the situation found in Canadian French and Dutch, the authors argued that these languages are changing because of the influence of English (Caramazza & Yeni-Komshian, 1974; van Alphen & Smits, 2004), but research on Portuguese suggests that a high degree of devoicing, which occurs not only utterance-initially but in medial contexts as well, might be an important feature of this language (Pape & Jesus, 2011). A certain amount of overlap between the two voicing categories, which was also speaker-specific, was reported for Fenno-Swedish, a minority language spoken in Finland (Ringen & Suomi, 2012). Fenno-Swedish differs from Swedish in two respects. First, although prevoicing is the norm in Swedish, some Fenno-Swedish speakers failed to prevoice consistently (13% of /b, d, g/ tokens in total), and there were between-speaker differences in the number of tokens without prevoicing and in the duration of prevoicing. Second, /p, t, k/ are realised as aspirated in Swedish, but in Fenno-Swedish they are realised as unaspirated, and VOT values are closer to those found in Finnish. Since all Fenno-Swedish speakers are fluent in Finnish, Ringen and Suomi conclude that the overlap between the two VOT categories comes from the influence of Finnish, which is a situation similar to that proposed for Canadian French and Dutch.

Examples of phonetic overlap between the two voicing categories, although not present in Serbian, raise some interesting questions regarding variability in the production of the voicing contrast, and pose a challenge for the existing models of the voicing contrast. Irrespective of whether this variability comes from the influence of another language with a different type of contrast (the situation in Canadian French, Dutch and Fenno-Swedish), or whether it represents a language-specific feature, such as in European Portuguese, both scenarios need to be accounted for in a model of the voicing contrast. Further, in languages where overlap is present, it seems to be a speaker-specific feature. Results for two bilingual Serbian speakers, who do not have any statistically significant influence from English on their VOT production in Serbian, suggest that some speakers do seem to be more susceptible to such influences than others, and raise a question of which factors determine the outcome. Finally, because the overlap between the categories might lead to the loss of salience of the contrast in some cases, it is likely that other acoustic correlates reinforce the contrast, but they are not well researched in voicing languages. These issues and their relevance for the existing theoretical models are further discussed in Chapter 8.

### 4.4.2 Linguistic factors affecting VOT in Serbian

Linguistic factors that were found to affect VOT in the present study are place of stop articulation and the quality of the following vowel, while condition (isolation vs. sentence frame) did not have any effect on VOT in phonologically voiceless stops.

**Effect of place of articulation on VOT**

Place of articulation effect on VOT in Serbian is different for the two stop classes. Place has very little effect on the duration of prevoicing. The velar /g/ has shorter mean prevoicing than /b/ and /d/ (by about 17 ms), but the difference in means between /b/ and /d/ is very small (below 1 ms and within measurement error). These differences did not reach statistical significance and the effect size is small. This result is in line with the majority of findings for other voicing languages and for prevoiced stops in English, where there is a tendency for the velar to have shorter prevoicing than the bilabial and the dental/alveolar (Section 1.1.2, Table 1.2).

However, direction and magnitude of the place effect on VOT and statistical significance of results vary between studies. In studies that reported statistically significant differences, the order of effect was /b/, /d/>/g/ in Castilian Spanish and Swedish (Helgason & Ringen, 2008; Rosner, et al., 2000) and /b/>/d/>/g/ or /d/>/b/>/g/ in Latin American Spanish dialects (Williams, 1977). For English /b, d, g/ realised as prevoiced, Smith (1978) found a significant effect of place on prevoicing in the order /b/>/d/>/g/. A similar discrepancy in the order of this effect is present in studies that did not test for significance: /b/>/d/>/g/ in Hungarian and Puerto Rican Spanish (Lisker & Abramson, 1964), /b/>/g/>/d/ in French, Hebrew and Japanese (Jacques, 1987; Obler, 1982; Shimizu, 1989; Yeni-Komshian, et al., 1977), and /d/>/b/>/g/ or /d/≥/b/>/g/ in Polish, Tamil and English (Keating, et al., 1981; Lisker & Abramson, 1964). As in Serbian, several papers found no significant differences, for example in Dutch (van Alphen & Smits, 2004), French (Ryalls, Cliché, et al., 1997), and Hungarian (Gósy & Ringen, 2009).

In contrast to mixed findings from acoustic studies, explanations that have been offered for this effect tend to concentrate on aerodynamic and articulatory/physiological factors (as discussed in Section 1.1.2), such as place-related differences in supraglottal cavity size (Smith, 1978), and passive expansion of the supraglottal cavity through tissue compliance (Keating, 1984b; Ohala, 1983; Ohala & Riordan, 1979; Westbury & Keating, 1986). These factors are considered to be universal and would be expected to produce the same results in different languages. However, they are unable to explain different order of the place effect on prevoicing across languages or, in some cases, an absence of a place-related effect. On the other hand, these explanations did not consider how active voicing could be related to place-dependent differences in prevoicing duration, although it is generally assumed that in this context voicing is active in voicing languages (Jansen, 2004). Research on active manoeuvres that are used to sustain voicing has suggested that place of articulation can have an effect on the duration of active voicing.

Westbury (1983) carried out a cinefluorographic study on stop production of one speaker of American English, with focus on active cavity enlargement. His speaker frequently employed several manoeuvres to enlarge the oral cavity in voiced stops and to prolong voicing, such as a downward movement of the larynx, more advanced position and a forward movement of the tongue root, a downward movement of the tongue dorsum and tip, and faster downward movement of the upper tongue surface. All manoeuvres, except the first one, were also place-dependent. Westbury pointed out that the most important is the cumulative effect of these actions, which is a function of both place of articulation of a voiced stop and the position in utterance.

Another active mechanism is nasal leakage (prenasalisation), which was found to be related to initiation of voicing in utterance-initial voiced stops in Spanish (Solé & Sprouse, 2011). Two main patterns of nasalisation were used. The first is delayed nasal closure: nasal closure is delayed relative to the oral closure, which results in nasal leak, slowing down the build-up of pressure in the oral cavity. When necessary pressure differential is achieved, voicing starts, and then the velum closes. The second is nasal burst: both nasal and oral cavity close, but the velum opens again, allowing for a brief leakage of air and the initiation of voicing. After this, it closes again. The choice of the pattern used depended on the context, that is, on whether the velum was open at the beginning of the utterance. If it was open, the first pattern was used, but if it was closed, the second pattern was used. There was also between-speaker variability in whether they prefer one pattern, or use both. Further, some speakers used nasal leakage only to initiate voicing, while others used it to both initiate and sustain voicing.

In addition to nasal leakage, Spanish and French speakers were found to use the following mechanisms to initiate or prolong voicing in /b/ and /d/ in absolute initial position: oral leakage (spirantisation), implosivisation (with negative oral pressure before the build-up), or some other active manoeuvre, as well as passive expansion of the oral cavity through wall compliance (Solé, 2011). Majority of speakers used one or a combination of these manoeuvres. Nasal leakage was more frequent than any other mechanism. Oral leakage was used in Spanish, but rarely in French. There were also place-related differences in the use of active gestures. Apicals were less conducive to voicing than labials (presumably due to the smaller area of compliant tissue), which resulted in more cases of nasal leakage than in labials. Finally, there was within- and between-speaker variability in the use of these manoeuvres.

As this research suggests, there is a complex interaction between active and passive manoeuvres that are employed to initiate and to maintain voicing in prevoiced stops. The use of these manoeuvres is not only place-specific, but also language- and speaker-specific, and any attempt to explain place-related differences in prevoicing must take these into account. Future research should be directed at specifying these patterns, as well as differences, in other languages. It would be desirable to have quantitative measures of the degree of cavity expansion due to each active manoeuvre, and due to passive expansion, and their effect on voicing, as well as measures of the effects of nasal and oral leakage.

In phonologically voiceless stops in Serbian, VOT increases as place of articulation moves from front to back, in the order /p/</t/</k/. In isolated words, there is less difference between VOT values for /p/ and /t/ than in the sentence frame. As a consequence, in isolation, only VOT for /k/ is significantly longer than VOT for the other two stops, while in the sentence frame all three pairwise comparisons are significant[18]. The finding that VOT increases for more retracted place of articulation supports findings from the majority of studies in Table 1.1 (Section 1.1.2), although the exact order and magnitude of increase is not consistent across languages.

Serbian results for /p, t, k/ support physiological and aerodynamic explanations for place-related differences in unaspirated or slightly aspirated stops, summarised by Cho and Ladefoged (1999), see Section 1.1.2. The first four explanations are applicable to Serbian (and are likely to be universal): the size of the cavity in front of the

[18] However, when data for the two conditions are pooled together (which is possible because of non-significant differences between the two conditions), the difference between /p/ and /t/ also reaches significance, and all three pairwise comparisons are significant (Tukey post-hoc test, *p* < 0.001 for all three).

constriction, the size of the cavity behind the constriction, velocity of articulators, and the extent of the contact area. Larger cavity in front of velar constriction (i.e. more air causing bigger obstruction) and smaller cavity behind it (i.e. higher pressure build-up during the constriction) can both explain longer VOT in velars, and so can greater contact area in velars (the constriction takes longer to be released). Conversely, faster movement of the lips and the tip of the tongue can help explain shorter VOT in labial and dental stops.

On the other hand, results for aspirated stops in several languages do not follow this pattern. For British English Docherty (1992) found that VOT increased in order /p/</k/</t/, and VOT for /p/ was significantly shorter than that for /t/ and /k/. This finding is in contrast to findings from a number of studies, which for (mainly American) English reported the expected pattern /p/</t/</k/. For German, Jessen (1998) reported a discrepancy between results for utterance-initial position, where the order of VOT increase was /p/</t/</k/, and results for intervocalic position, where it was /p/</k/</t/. Both authors, Docherty for English and Jessen for German, concluded that these findings suggest that VOT production is actively controlled and that both languages must supply a rule that (at least partially) overrides aerodynamic and physiological processes. Since for phonologically voiced stops in German the order of VOT increase was /b/</d/</g/ in all contexts, Jessen proposed that result for /b, d, g/ in German can be explained by passive aerodynamic processes, but that for /p, t, k/ place effect on VOT is actively controlled.

One more factor was proposed by Cho and Ladefoged as a possible explanation for the place effect in both unaspirated and aspirated stops, and that is the tendency to keep duration of the voiceless interval (CD+VOT) uniform across places of articulation. Weismer (1980) argued that this uniform voiceless interval is a consequence of a place-independent devoicing gesture, also called abduction gesture. Results for Serbian do not fully support this explanation. Although in Serbian the voiceless interval is fairly uniform at all three places of articulation, there is no negative correlation between VOT and CD for each place of articulation, which means that these two variables are not inversely related (the relevant statistical analysis is presented in Section 5.2.2). In other words, place-related differences in VOT do not result from the tendency to keep the voiceless interval uniform by balancing out place-related differences in CD.

**Effect of the following vowel on VOT**

The effect of the following vowel on VOT is different for voiced and voiceless stops in Serbian. The quality of the following vowel does not have any influence on duration of prevoicing in /b, d, g/. Other studies reported a mixture of results. Serbian result is in agreement with results for Dutch (van Alphen & Smits, 2004) and for Latin American Spanish dialects (Williams, 1977), where differences were non-significant, and with French data reported by Yeni-Komshian et al. (1977), although they did not test for significance. On the other hand, Smith (1978) found that before high vowels English /b, d, g/ tokens were more often produced as prevoiced, and with longer VOT, than before low vowels. In contrast to Williams, Rosner et al. (2000) found that in Castilian Spanish prevoicing of /b/ and /d/ (not /g/) was longer before /o/ than before /a/. Finally, for Lebanese Arabic, prevoicing was shorter before /i/ than before /a/ and /u/ (Yeni-Komshian, et al., 1977).

This effect has been explained either by differences in supraglottal cavity volume, or by differences in the surface area that can be passively expanded, with high vowels having larger cavity and lager surface area than low vowels (Ohala, 1983; Ohala & Riordan, 1979; Smith, 1978). However, this implies that these influences are automatic and should be expected to apply universally, which does not seem to be the case. Here, as was discussed regarding the effect of place of articulation on duration on prevoicing, active voicing needs to be taken into consideration, as well as its relationship with passive aerodynamic factors, in order to explain the resulting effect, which varies from language to language.

For voiceless stops in Serbian, there is an effect of the quality of the following vowel on VOT, which interacts with place of stop articulation. In the pooled data for all three stops, VOT is higher before high vowels /i/ and /u/ than before the low vowel /a/. Differences are in the range 6-10 ms. This finding is in agreement with results from studies on a variety of languages, both aspirating and voicing, including English, Italian, Hungarian, Portuguese and French (Docherty, 1992; Esposito, 2002; Gósy, 2001; Klatt, 1975; Lousada, et al., 2010; Morris, et al., 2008; Nearey & Rochet, 1994; Smith, 1978), and is consistent with proposed aerodynamic and physiological explanations for this effect. According to one explanation, higher resistance to oral airflow in high vowels makes it more difficult to re-establish transglottal pressure difference and to initiate voicing after the release of a stop, thus lengthening VOT (Chang, et al., 1999; Ohala, 1981). In addition to this, it has been proposed that there is a vertical pull on the focal folds that increases glottal tension and resistance, which takes longer to be overcome and for voicing to start (Docherty, 1992; Morris, et al., 2008).

In Serbian there is also an interaction between the effects of place of articulation and the following vowel on VOT, although the effect size is small. For /p/, VOT is significantly longer before /u/ than before /i, e, a/ in both conditions, with mean differences between vowels of 12-16 ms. For /t/ VOT is significantly longer before /i/ than before /a, o/ in the sentence condition, with differences of 12-15 ms. Interaction with place was also found in American English (Morris, et al., 2008) and French (Nearey & Rochet, 1994), but none of these studies reported the same type of interaction as that found in Serbian. These studies are not easily comparable due to different vowel inventories in the three languages, as well as the number of vowels that was actually investigated (Morris et al. examined only three English vowels, while Nearey and Rochet included nine French vowels). In all three studies, VOTs tend to be longer before high vowels. Serbian and French data agree in that VOT is significantly longer in /pu/ sequences than in /pi/ sequences, which is not the case in English. In all three languages there is no significant difference in VOT between /ti/ and /tu/ sequences. However, in both French and English, the effect of the vowel on VOT is significant for /k/, but this is not the case with Serbian. In trying to account for the finding that VOT is longer in /ki/ than in /ku/ sequence, but for /t/ it is the longest in /tu/ sequence, Morris et al. argue that this pattern results from longer time that is needed for the tongue to move from back position to front position in /ki/ (compared to /ku/) and from front to back position in /tu/ (compared to /ti/). While this explanation could apply to Serbian and French result for /pu/>/pi/ sequences, it cannot explain why the effect for /t/ is not consistent across studies. It is also unclear why there is no effect of vowel quality on VOT for /k/ in Serbian.

In sum, although the finding that VOT tends to be longer before high vowels /i/ and /u/ than before low vowel /a/ in the pooled data for all three Serbian stops generally supports aerodynamic and articulatory explanations for the effect of vowel height on VOT, results for each stop separately require further explanation, because the interaction between the effect of place of articulation and the following vowel seems to be language-specific.

### 4.4.3 Speaker factors affecting VOT in Serbian

An important finding of the present study is that, although in each subject's production VOT values for phonologically voiced and voiceless stops are clearly separated, there is a certain degree of between-subject variation. Out of several speaker factors that have been investigated in this study, age, gender and place of birth of subjects all have an effect on VOT production, while place of living and speaking rate (for positive VOTs) do not.

**Effect of gender on VOT**

Female subjects in this study produced longer prevoicing than male subjects, with difference between means of 14 ms, which just fell short of reaching significance (at the 0.01 level). This finding is consistent with results for Swedish reported by Karlsson et al. (2004), and for Hungarian (Gósy & Ringen, 2009). There was no effect of gender on the frequency of prevoicing, because all tokens were realised as prevoiced in Serbian.

Serbian results do not support anatomical explanation for differences in prevoicing duration and in frequency of prevoicing. According to this explanation, men have larger vocal tracts and larger supraglottal cavity than women, and as a consequence supraglottal pressure increases more slowly during phonation enabling them to sustain voicing for longer. There is ample support for this hypothesis both in the fact that the number of prevoiced tokens was higher for men in English (Smith, 1978) and Dutch (van Alphen & Smits, 2004), and in the significantly longer prevoicing found in men than in women in several languages, such as English (Smith, 1978), Swedish (Helgason & Ringen, 2008), and Dutch (van Alphen & Smits, 2004). However, data from Hungarian (Gósy & Ringen, 2009) and Serbian suggest that anatomical and physiological factors can be overridden. Some authors, such as Helgason and Ringen (2008) and Gósy and Ringen (2009), discuss the possibility that longer prevoicing produced by female speakers comes from their tendency to use clear or more intelligible speech (drawing on research by Bradlow, Torretta, & Pisoni, 1996; Byrd, 1994; Hazan & Markham, 2004).

It is also possible that differences in speaking rate, that is slower speaking rate of female subjects, cause differences in prevoicing duration, although speaking rate is unlikely to have a large effect on monosyllables read in isolation, which were used to measure prevoicing in the present study. What is more, although four out of six female subjects are slower talkers, when the effect of speaking rate on closure duration in the sentence frame was co-varied statistically, female subjects as a group still produced longer closures than male subjects. Most closures were fully voiced in this condition, which means that female subjects produced longer periods of voicing, despite speaking rate differences.

This discrepancy between proposed universal biological factors for male-female differences in production, and the results from production studies, is present in research on phonologically voiceless stops as well.

Speaker gender affects VOT in voiceless stops in Serbian, with male subjects having significantly longer VOTs than female subjects in both conditions (11 ms in isolation and 9 ms in the sentence condition). This result is in agreement with Smith's (1978) results for /b, d, g/ tokens in English realised with short lag VOT, and for results for /p, t, k/ realised with short lag VOT in Hungarian (Gósy & Ringen, 2009), and with long lag VOT in Swedish (Helgason & Ringen, 2008) and Korean (Oh, 2011).

The issue of why gender differences occur in VOT production of voiceless stops is a very interesting one. For English long lag stops, there seems to be an agreement that female subjects produce longer periods of aspiration than male subjects, as was documented in several studies, although differences were not always significant and vary from 5 to 13 ms, depending on the study and the condition (Morris, et al., 2008; Ryalls, et al., 2004; Ryalls, Zipprer, et al., 1997; Sweeting & Baken, 1982).

On the other hand, in other languages, male subjects tend to produce significantly longer VOTs than female subjects, for example in aspirated stops in Korean (Oh, 2011) and Swedish (Helgason & Ringen, 2008). This is also true for short lag stops in Hungarian (Gósy & Ringen, 2009), and for English /b, d, g/ realised with short lag VOT (Smith, 1978; although Sweeting & Baken, 1982, and Morris et al., 2008 reported non-significant differences). Apart from Oh (2011), who found a difference of 13-19 ms, other studies found male-female differences to be smaller than in Serbian and about 2 - 4 ms, depending on condition.

Early accounts of male-female differences in VOT production, because they were based mainly on results for English aspirated stops (where females produced longer VOTs), were focused on finding an aerodynamic or biological explanation as to why it takes longer for female subjects to resume voicing after the release of a stop. The idea that weaker airflow after the release in female subjects is responsible for this delay was not supported by experimental evidence (Karlsson, et al., 2004; Subtelny, et al., 1966). Other proposals, such as lung volume differences, and especially differences in larynx anatomy, physiology, and laryngeal settings, were better supported by experimental results (Hoit, et al., 1993; Koenig, 2000; Stathopoulos & Sapienza, 1997; Whiteside, et al., 2004). However, when recent results from other languages are taken into account, although it is possible that some universal anatomical, physiological, or aerodynamic factors are responsible for (at least some) male-female VOT differences, it is clear that universal factors cannot account for diverse results reported in production studies, and that these universal constraints can be overcome, if needed. Based on similar argumentation, Oh (2011) proposes that observed gender differences are not universal, but that they represent sociophonetic markers of speaker gender, which vary from language to language or even from dialect from dialect. Oh further argues that, while in some instances, such as longer VOTs for aspirated English stops in females, they may have anatomic base, they assume indexing sociophonetic role and need to be learned as such in the process of language acquisition. This, believes Oh, is true for gender differences in both voiceless stops and in prevoiced stops.

### Effect of age on VOT

The effect of age is significant only for prevoiced stops in Serbian, with four oldest subjects (52 - 62 years) having longer prevoicing than the rest of the subjects. They were also more variable in production, as is reflected in wider data ranges and larger standard deviations. These results contradict Ryalls, Cliché et al.'s (1997) results for French, where older subjects produced shorter prevoicing, but they agree in the fact that older subjects were more variable than younger subjects.

There was no effect of age on VOT in voiceless stops in the present study. The same finding was reported by Sweeting and Baken (1982) and Neiman et al. (1983) for both stops classes in intervocalic position in English (but cf. Ryalls et al., 2004 and Ryalls, Cliché et al. 1997, who found that older subjects produced shorter positive VOT values in English and French, respectively). However, Sweeting and Baken (1982) found significantly larger standard deviations in their older group, which was not the case in the present study. They also reported smaller separation between the two voicing categories (short lag and long lag VOT) in older subjects. In contrast to their finding, in the present study the oldest subjects are among the subjects with the largest separation between prevoiced and voiceless stops, because they have longer prevoicing.

It has been hypothesised that some of the changes related to normal ageing could affect VOT production, such as reduced lung volume (Hoit, et al., 1993; Ryalls, Cliché, et al., 1997; Ryalls, et al., 2004), and loss of precision of fine motor coordination needed to control laryngeal-supralaryngeal timing in production (Sweeting & Baken, 1982). In addition to this, it has been proposed that certain phonetic realisations, such as aspirated and prevoiced stops, are more difficult to produce, and that they can become more variable with age, while easier-to-produce short lag stops become less variable with age (Ryalls, Cliché, et al., 1997; Sweeting & Baken, 1982). Finally, the nature of the phonetic realisation of the voicing contrast (prevoiced vs. short lag, or short lag vs. long lag), and the width of separation between the categories, have also been discussed in relation to age-related changes in VOT production. According to this view, in languages such as French the separation between categories is greater, and age-related variability and changes are less likely to lead to the loss of contrast. This also puts less demand on older speakers, and thus the effect of age on VOT production could be bigger in French than in English, for example (Ryalls, Cliché, et al., 1997).

Although limited to prevoiced stops, the effect of age in Serbian supports evidence from Ryalls, Cliché et al. (1997) that older speakers become more variable in stop production. The lack of any effect of age on voiceless stops could tentatively be seen as supporting the idea that voiceless stops, being easier to produce, are not subject to ageing process. On the other hand, smaller lung volume in older speakers does not seem to be a likely explanation in the present study. Furthermore, the width of separation between the two voicing categories does not seem to diminish with age in the present study. It is fairly large for the four oldest subjects (155 ms to 184 ms), which puts them in the group of six subjects with larger separation along the VOT scale.

There is another possible explanation for the limited effect of ageing in the present study – that the effect of ageing is still not clearly present in this age group. The oldest subjects in the present study are younger than subjects in other studies designed specifically to examine the effect of ageing on VOT production, and any effect of ageing might not be easily observable. For example, Sweeting and Baken’s (1982) older subjects were over 75 years old, Neiman at al.'s (1983) subjects were 70-80 years old, while Ryalls, Cliché et al.'s (1997) subjects had the mean age of 67.

This would explain a moderate increase in variability in their production and the lack of any other effect[19]. It cannot, however, explain why they produce longer prevoicing than most other subjects in this study. The explanation for this fact might not be related to age at all, but to the individual features of their production. If VOT results for these four speakers are compared with their results for closure duration in the sentence condition (Chapter 5), it is clear that they produce not only relatively longer prevoicing in /b, d, g/, but also relatively longer /b, d, g/ closures, which are fully voiced (and also longer /p, t, k/ closures). What is more, all four subjects have around 80% of closure voiced in /b, d, g/ tokens in utterance-final position. It is therefore possible that these subjects' phonetic targets include more voicing in the closure than for some of the other subjects. The remaining question is: why do they have different phonetic targets in their production of voiced stops on the whole? It is likely that they share certain aspects of VOT production which are sociolinguistic in nature, specific to their age group, and not necessarily a result of ageing as such. It is therefore possible that variability in production of prevoicing, caused by ageing, interacts with other social factors (as yet unidentified), which may coincide with the age of speakers. This is not unusual – Docherty et al. (2011) found a complex interaction between age and several social factors in VOT production along the Scottish-English border, and argued that any attempt to explain phonetic variation should include both phonetic and social factors. Unfortunately, the design of the present study does not allow for further exploration of these issues, which remain as a topic for further research.

### Effect of place of birth and place of living on VOT

Differences in VOT production related to place of birth of subjects suggest that some regional variation might be present. Subjects from Čačak produced shorter prevoicing and shorter positive VOTs and consequently had the smallest separation

[19] For one subject, BPm, there is also a possibility that his production of prevoicing has become more variable because he lives in an English-speaking country. However, evidence from the literature suggests that this manifests mostly through reduced number of prevoiced tokens (Caramazza & Yeni-Komshian, 1974; Helgason & Ringen, 2008; Heselwood & McChrystal, 1999; Keating, et al., 1983; van Alphen & Smits, 2004), which is not the case with BPm. This is also unlikely because the other subject from the UK, RVm, does not have large standard deviation.

between the categories. Subjects from Belgrade and Užice produced longer prevoicing and longer positive VOTs, which resulted in larger separation between the categories. These results are based on relatively small numbers of subjects, and on unequal numbers in each group (in addition to between-subject differences), and therefore can only be interpreted as a possible topic for further research.

Results for the two speakers who live in the UK, RVm and BPm, deserve further attention. Because they use English on a daily basis, and the amount of Serbian usage is limited to interactions with their families and friends, it is reasonable to expect a certain degree of influence of English on their VOT production. This influence could potentially affect both voiced and voiceless stops in Serbian. Since English /b, d, g/ are mainly produced as voiceless unaspirated, it could be expected that these two subjects exhibit shorter prevoicing in Serbian /b, d, g/ tokens, or the absence of prevoicing in some tokens. Loss of prevoicing has often been attributed to the influence of another language, for example in Canadian French, Dutch, and Fenno-Swedish (Caramazza & Yeni-Komshian, 1974; Ringen & Suomi, 2012; van Alphen & Smits, 2004). Helgason and Ringen (2008) noted the same for Swedish speakers living in the United States in Keating et al.'s (1983) study, who had no prevoiced tokens, as opposed to Swedish speakers from Sweden in Helgason and Ringen's study, who prevoiced consistently.

However, this is not the case with subjects in the present study. All their /b, d, g/ tokens were prevoiced. In this respect, they are similar to a group of Heselwood and McChrystal's (1999) Panjabi speakers from Bradford, who acquired Panjabi in Pakistan, and who had a high percentage of prevoiced tokens (93%). Furthermore, although RVm and BPm produced stops with slightly longer prevoicing then other male subjects, the difference was not significant. Their production of prevoicing in /b, d, g/ does not stand out in any obvious way. In terms of their distributions and their means, they are in the middle of the group of male subjects. They also fit well into their respective age groups.

For /p, t, k/ realised as voiceless unaspirated, previous studies have shown that their VOT can change when speakers are immersed in and use another language on a daily basis, if in that language /p, t, k/ are realised as aspirated. One phenomenon that has received attention is gestural drift, which is defined as "perceptually-guided changes in speech production by a speaker well past the critical period for language acquisition" (Sancier & Fowler, 1997, p. 421). Sancier and Fowler (1997) found that VOT production of their single speaker, a native speaker of Brazilian Portuguese studying in the USA, changed after prolonged stays of several months in either Brazil or the USA. Her VOTs in both languages, although consistently shorter in her Portuguese than in her English, were significantly shorter (by about 5 ms) after several months in Brazil than after several months in the USA.

Tobin (2009b) examined gestural drift in Spanish-English speakers in the USA, and found the same effect in their VOT production in English and in Spanish (but cf. Tobin, 2009a, who reported no change in Spanish). Tobin (2009a) further measured VOT for /p, t, k/ of three Serbian-English speakers, and found that VOT in both languages was shorter after a long stay in Serbia than after a long stay in the USA. The effect was found at all three places of articulation in their English, but only for stops with longer VOTs, that is for /t/ and /k/, in Serbian.

However, there is a difference between these speakers, who divide their time between the two countries, and the two speakers in the present study, who live permanently in the UK and spend very short periods in Serbia (for holidays). A more appropriate comparison would be with speakers in a similar situation.

Major (1992) investigated language attrition in five women born in the USA, who had immigrated to Brazil as adults. They had lived in Brazil from 12 to 35 years and spoke Portuguese with their families, but used English professionally. Their VOTs were shorter when they spoke Portuguese than when they spoke English, but there were also individual differences in the degree of VOT change in their English. Two speakers showed little change in English (below 10 ms), in comparison with an English monolingual control group, and their production of VOT in Portuguese was English-like. Another two speakers had shorter VOT when speaking English, i.e. had more loss (up to about 25 ms), and were closer to the native group in their Portuguese. Finally, the fifth subject had native-like VOT production of Portuguese, very little shortening of VOT in English formal style, but in her conversational English, her VOTs were similar to her VOTs in Portuguese, and about 40 ms shorter than in the control group. Major’s study points out to some very important factors that need to be considered. First, he found large individual differences in the attrition of subject’s first language (L1), as well as in their proficiency in their second language (L2). Second, the proficiency in L2 seems to be correlated with the loss in L1. Third, formal speaking style was less affected than conversational style, and higher level of proficiency in L2 is likely to correlate with greater loss in L1 casual style of speaking.

Results from these studies suggest that some degree of lengthening of VOTs for voiceless stops in Serbian could be expected for the subjects RVm and BPm in the present study.

Looking back at their VOT results, they both produced slightly longer VOTs than other male subjects, in both conditions, but differences were not significant. The effect of age was not significant in both conditions, and neither of the two speakers stands out from the rest of the subjects in this respect. They also do not stand out to any large extent from other subjects with the same place of birth (and both in isolation and the sentence frame it is the group with longer VOTs). RVm does have the highest overall mean VOT of all subjects. His production for each stop individually is similar to that of some of the other subjects living in Serbia, except that he has rather long VOTs for /t/ in isolated words (Figure 4.14), longer than any other subject, although this is not the case in the sentence condition (Figure 4.16). Apart from that, his VOT results are similar to those of, for example, IVm and MPm in isolation condition (Figure 4.14), and of MVf, IJm, and IVm in the sentence frame (Figure 4.16).

To sum up, although both BPm and RVm have VOTs that are at the higher end of the VOT range in this study, it is difficult to attribute this solely to the influence of English, because there are two other factors that may have contributed to this result: they are both male, and belong to the higher VOT group according to their place of birth. All of these factors, combined with some individual specifics of their production, and possible influence from English, could have contributed to their VOT values being slightly higher. However, the effect of English, when separated from other influences, seems to be very small, and does not reach statistical significance. This finding does not support findings by Major (1992), Sancier and Fowler (1997), and Tobin (2009a, 2009b). This could partly be due to fact that BPm and RVm speak a voicing language at home, which could counterbalance the effect of English. In addition to this, at the time of recording they had lived in the UK for a shorter period of time than Major’s subjects.

It is also likely that other factors, as discussed by Major (1992), could be involved. First, the reading task they performed for the present study was controlled, and any differences, if present, would be smaller in such a sample of speech. Second, the level of phonetic proficiency in English of the two subjects may have had some bearing on the results. Third, as Major suggested in his study (Note 12, p. 205), there are a number of other factors, such as individual differences in languages learning skills, and social factors, such as affect and perceived prestige, as well as the interaction between the individual skill and affect, that could impact the influence of English on VOT production in Serbian. These issues remain outside the scope of the present study, and a topic for further research. As far as the present study is concerned, results from the two subjects who live in the UK are not significantly different from other subjects' results, and they have been discussed together.

# Chapter 5 Results for closure duration

## 5.1 Effect of phonological voicing category on closure duration in word-initial intervocalic stops

The distribution of closure duration (CD) results for stops in word-initial intervocalic position is shown in boxplots in Figure 5.1, and their means in Table 5.1 (words were uttered in the sentence frame, and stops were in intervocalic position, so that CD could be measured). Results are pooled across subjects.

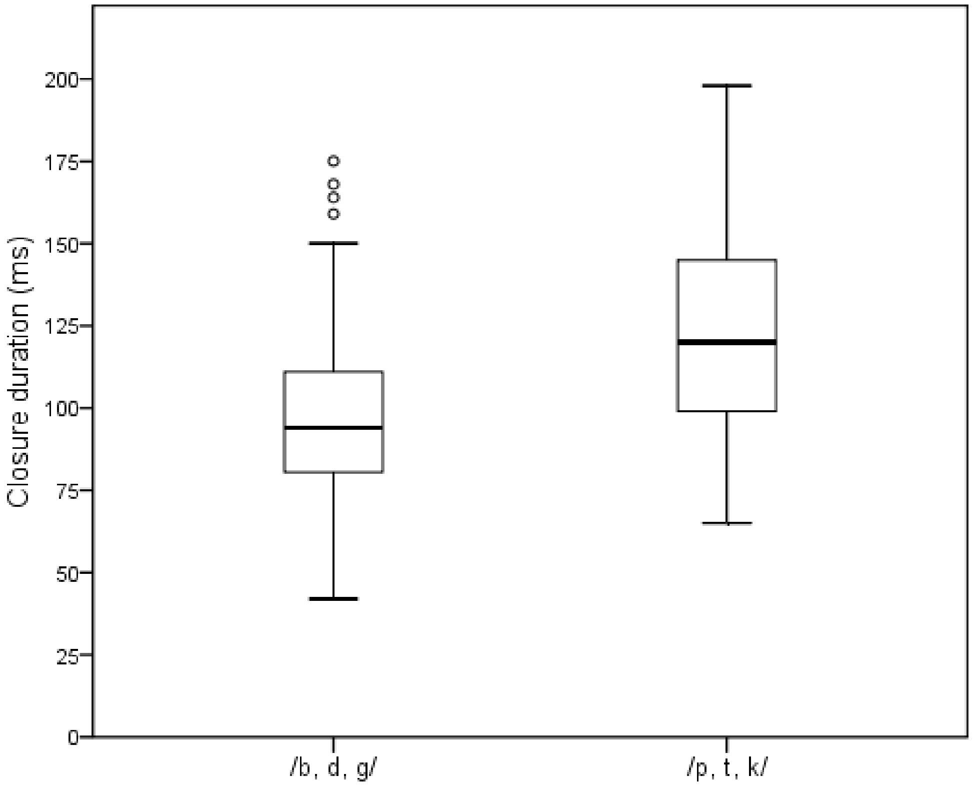

Figure 5.1 Boxplots showing the distribution of CD values for /b, d, g/ and /p, t, k/ in word-initial intervocalic position

Closure durations for stops belonging to the two voicing categories overlap, and range from 42 ms to 175 ms for /b, d, g/, and from 65 ms to 198 ms for /p, t, k/. Mean CD in the pooled data is 95.86 ms for /b, d, g/, and 122.92 ms for /p, t, k/. According to a Mann-Whitney U-test, this difference is statistically significant in the pooled data: $p < 0.001$ (2-tailed), $Z = -8.27$, $N = 340$, $r = -0.45$, medium effect. Statistical analysis of individual results revealed that for seven subjects difference in CD of phonologically voiced and voiceless stops is significant at the adjusted significance level of p < 0.01

(subjects MCf, SCf, DARf, MRf, IVm, BPm, DRm), while for the remaining subjects it is with $0.01 < p < 0.05$ (Table C2 in the Appendix C). The effect size is large for all subjects.

| | Mean CD (ms) | N | SD |
|---|---|---|---|
| /b, d, g/ | 95.86 | 168 | 22.86 |
| /p, t, k/ | 122.92 | 172 | 29.6 |

Table 5.1 Mean CD (ms), N and SD for /b, d, g/ and /p, t, k/ in word-initial intervocalic position

The difference between means for individual subjects varies from 14.6 ms for DRm to 48.94 ms for DARf. However, some of this variation could be due to differences in individual speaking rate. A ratio of CD for /b, d, g/ divided by CD for /p, t, k/ (or expressed as a percentage) is an alternative measure that could eliminate speaking rate differences (assuming that CDs for both classes are equally affected by speaking rate). When differences are expressed as ratios, the same two subjects have the smallest and the largest difference in CD: DRm has a ratio of 0.86, or a 14% difference, while DARf has a ratio of 0.7, or a 30% difference.

## 5.2 Linguistic factors affecting closure duration in word-initial intervocalic stops

### 5.2.1 Place of articulation and the quality of the following vowel

CD results for /b, d, g/ in word-initial intervocalic position, for each place of articulation, are shown in boxplots in Figure 5.2 and mean CD values are given in Table 5.2. There is a tendency for CD values to decrease the further back the place of articulation in the order /b/>/d/>/g/, although there is a lot of overlap in their distributions.

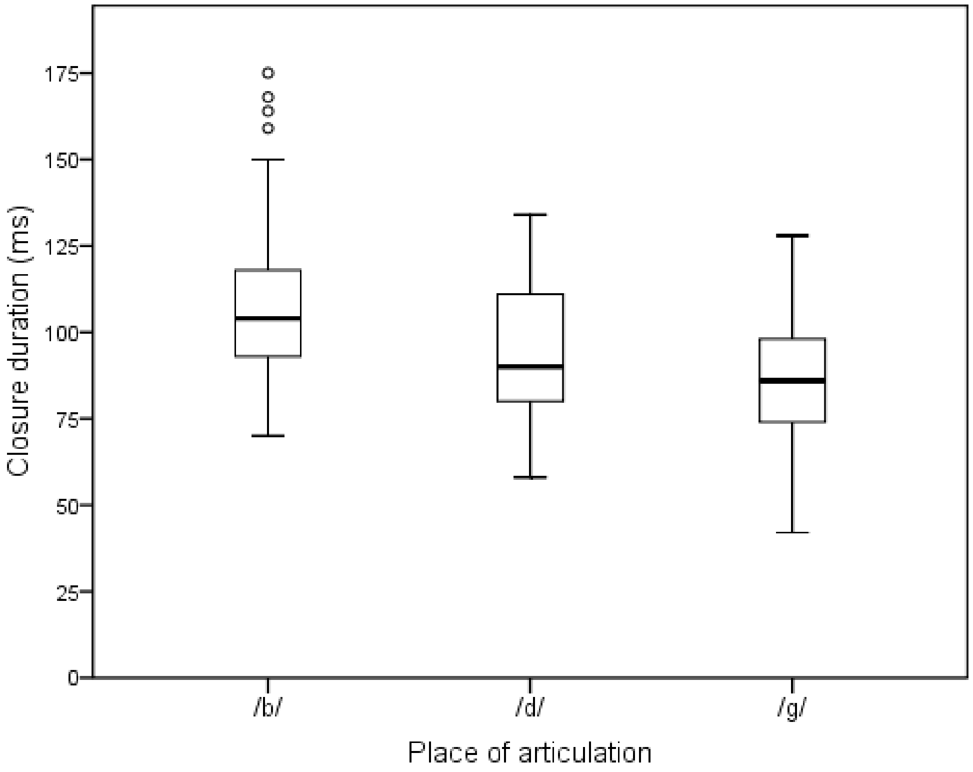

Figure 5.2 Boxplots showing the distribution of CD values for /b, d, g/ in word-initial intervocalic position as a function of stop place of articulation

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 108.39 | 54 | 24.49 |
| Dental | /d/ | 93.49 | 57 | 19.75 |
| Velar | /g/ | 86.37 | 57 | 18.75 |
| Total | | 95.86 | 168 | 22.86 |

Table 5.2 Mean CD (ms), N and SD for /b, d, g/ in word-initial intervocalic position for each place of articulation

A two-way between-groups ANOVA was carried out to explore the effect of stop place of articulation and the quality of the following vowel on CD. There was a significant main effect of place of articulation, $F(2,153) = 15.58$, $p < 0.001$, $\omega^2 = 0.148$ (large effect). A Tukey HSD post-hoc test revealed CD was significantly longer for /b/ than for /d/ and /g/ ($p = 0.001$, and $p < 0.001$ respectively). The main effect of the following vowel did not reach statistical significance, $F(4,153) = 2.11$, $p = 0.82$, and there was no statistically significant interaction between the two main factors, $F(8,153) = 0.38$, $p = 0.93$. Mean CD values before each vowel are given in Table 5.3.

| Following V | Mean CD (ms) | N | SD |
|---|---|---|---|
| /a/ | 87.52 | 33 | 18.62 |
| /e/ | 101.52 | 33 | 26.93 |
| /i/ | 95.85 | 34 | 23.93 |
| /o/ | 94.66 | 35 | 20.94 |
| /u/ | 99.85 | 33 | 21.83 |

Table 5.3 Mean CD (ms), N and SD for /b, d, g/ in word-initial intervocalic position before each vowel

Pooled results for CD of /p, t, k/ in word-initial intervocalic position are presented in Figure 5.3 and in Table 5.4, for each place of articulation. As in phonologically voiced stops, CD decreases in the order bilabial > dental > velar, but the distributions for the three stops overlap.

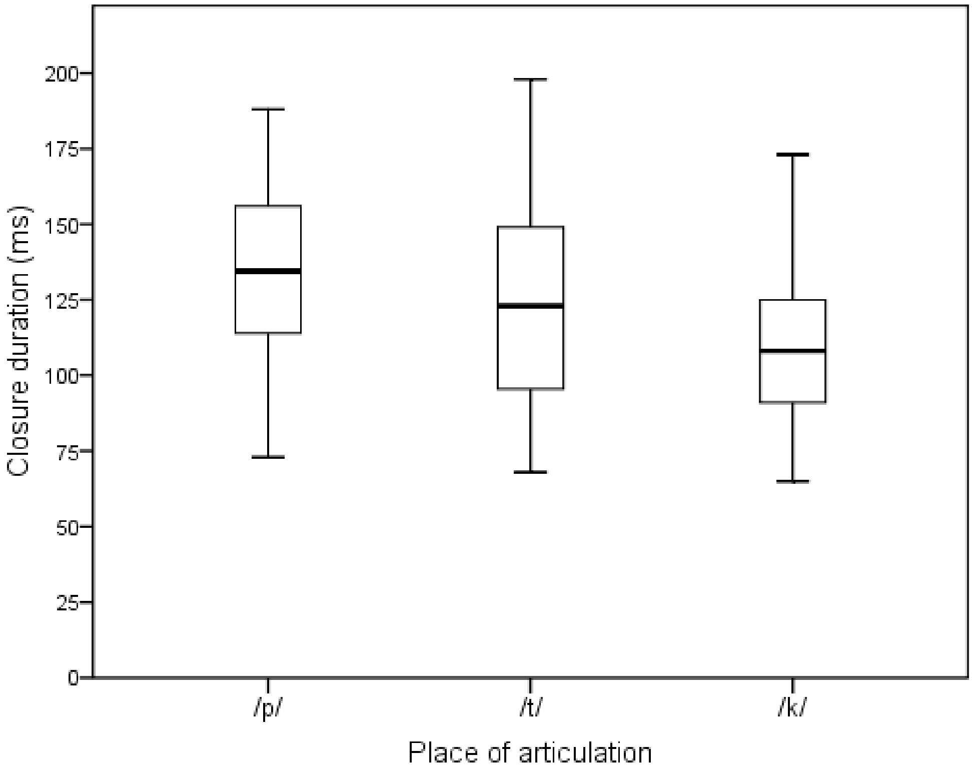

Figure 5.3 Boxplots showing the distribution of CD values for /p, t, k/ in word-initial intervocalic position as a function of stop place of articulation

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 133.69 | 58 | 26.74 |
| Dental | /t/ | 124.82 | 56 | 31.75 |
| Velar | /k/ | 110.31 | 58 | 25.63 |
| Total | | 122.92 | 172 | 29.56 |

Table 5.4 Mean CD (ms), N and SD for /p, t, k/ in word-initial intervocalic position for each place of articulation

A two-way between-groups ANOVA revealed that there was a significant main effect of place of articulation on CD, *F*(2,157) = 10.15, $p < 0.001$, $\omega^2 = 0.099$ (medium effect). According to a Tukey HSD post-hoc test, closures were significantly longer for /p/ than for /k/, and for /t/ than for /k/ ($p < 0.001$, and $p = 0.02$ respectively). The main effect of the following vowel did not reach statistical significance, *F*(4,157) = 0.77, $p = 0.55$, and there was no statistically significant interaction between the two main factors, *F*(8,157) = 0.58, $p = 0.79$. Mean CD values before each vowel are given in Table 5.5.

| Following V | Mean CD (ms) | N | SD |
|---|---|---|---|
| /a/ | 119.76 | 34 | 29.28 |
| /e/ | 126.74 | 34 | 28.89 |
| /i/ | 127.06 | 34 | 28.88 |
| /o/ | 118.11 | 35 | 32.18 |
| /u/ | 123.06 | 35 | 29.2 |

Table 5.5 Mean CD (ms), N and SD for /p, t, k/ in word-initial intervocalic position before each vowel

A summary of results for the effect of place of articulation on CD for both stop classes is presented in boxplots in Figure 5.4.

Phonologically voiced stops at all three places of articulation were realised with significantly shorter closures than their voiceless cognates - for the pair /b/-/p/: Mann-Whitney U-test, $p < 0.001$ (2-tailed), *Z*= -4.74, $r = -0.45$, medium effect; for the pair /d/-/t/: t-test, p < 0.001 (2-tailed), *t*(111) = -6.29, Cohen's *d* = -1.19, large effect; for the pair /g/-/k/: t-test, p < 0.001 (2-tailed), *t*(113) = -5.72, Cohen's *d* = -1.08, large effect.

CDs for each stop class decrease from bilabial to velar place of articulation, although not all pairwise differences are statistically significant, as was shown in the previous analysis (there is a non-significant difference between /d/ and/g/, and between /p/ and /t/).

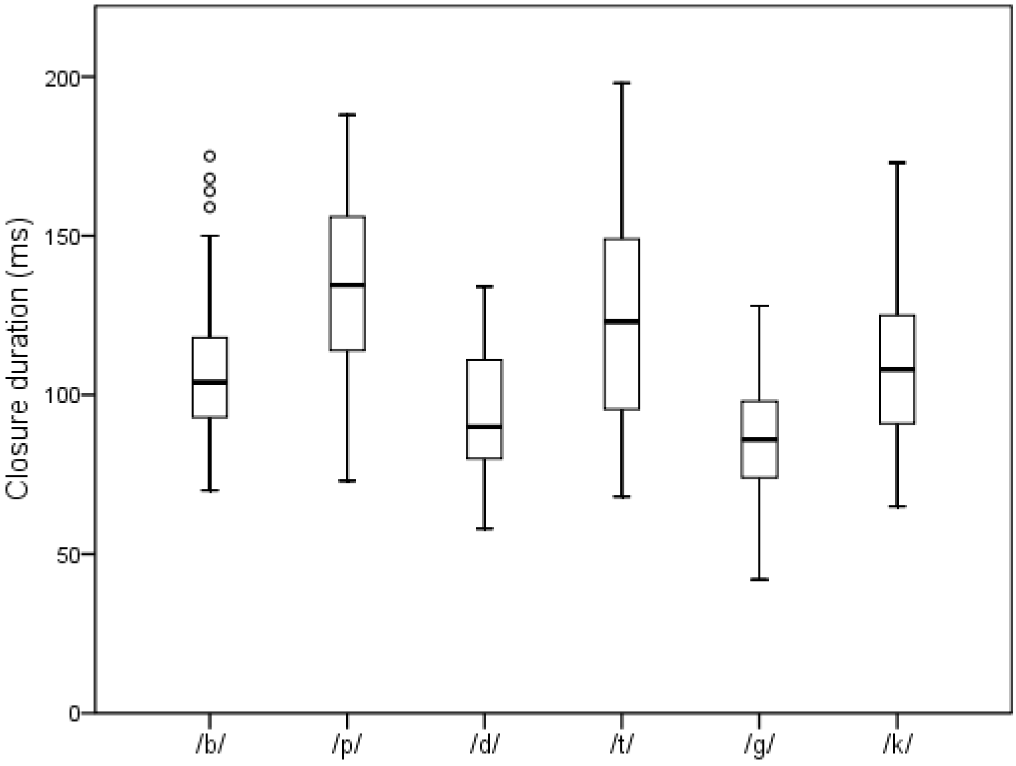

Figure 5.4 Boxplots showing the distribution of CD values for each stop in word-initial intervocalic position

### 5.2.2 The voiceless interval

The effect of place of articulation on CD in voiceless stops is in the opposite direction from the same effect on VOT, as was found in a number of studies on other languages. While CD tends to decrease from labial to velar place of articulation, VOT tends to increase, which suggests that these variations might not be independent. Weismer (1980) proposed that there exists a devoicing gesture with constant duration (abduction gesture), which results in constant duration of the voiceless interval, defined as CD + VOT. He found that in American English this voiceless interval indeed seems to be fairly uniform across places of articulation. A similar finding was reported for French /p, t, k/ by Abdelli-Beruh (2009). Docherty (1992), on the other hand, found less consistency in the duration of abduction gesture in British English, both across speakers and across places of articulation. He suggested that a better measure of the relationship between CD and VOT would be obtained by using a correlation analysis. A negative correlation between CD and VOT, as well as lack of positive correlation between VOT and duration of abduction gesture, would indicate that there exists a uniform abduction gesture. He found a small negative correlation between CD and VOT, which was significant for one speaker, and a larger positive correlation between VOT and duration of abduction gesture, which was significant for four out of five speakers, and in the pooled data. These results do not support the hypothesis that there is an invariant abduction gesture in voiceless stops. Abdelli-Beruh, on the other hand, found no significant correlations between VOT and CD, and between VOT and the voiceless interval, which suggests that variations in VOT are not caused by variations in CD, although the duration of the voiceless interval is relatively uniform in French.

To test this hypothesis on Serbian data, CD and VOT duration for word-initial voiceless stops were examined in the sentence condition (this is the only condition where both variables were measured on the same set of words). Recall that in this condition CDs for /p/ and /t/ are significantly longer than for /k/, while VOT increases in the opposite direction, /p/</t/</k/, and all three pairwise comparisons are significant. When duration of the voiceless interval is calculated by adding CD and VOT for each stop token, overall means are similar: the mean voiceless interval is 153 ms for /p/, 154 ms for /t/ and 161 ms for /k/ (with *SD* of 28.19, 30.56, and 29.49, respectively). These differences failed to reach significance, according to a one-way ANOVA, $p = 0.3$, $F(2,164) = 1.2$. This finding is in agreement with Abdelli-Beruh's result for French and Weismer's result for American English.

However, a correlation analysis revealed that there is a negative correlation between VOT and CD for /t/ (Pearson correlation, $r = -0.3$, $p = 0.027$), but not for /p/ ($r = 0.018$, $p = 0.9$) or /k/ ($r = 0.05$, $p = 0.7$). VOT and duration of the voiceless interval are positively correlated, but this is significant for /p/ (Pearson correlation, $r = 0.35$, $p = 0.009$) and /k/ ($r = 0.48$, $p < 0.001$), and not for /t/ ($r = 0.018$, $p = 0.9$). Correlation analysis was not performed for each subject separately because of the relatively small number of tokens for each stop.

Despite the fairly uniform duration of the voiceless interval in Serbian, there is no inverse relationship between VOT and CD at the three places of articulation, which suggests that place-related differences in VOT are not a consequence of place-related differences in CD. This finding is consistent with results for French and British English, and suggests that Weismer’s explanation cannot account for these results.

## 5.3 Speaker factors affecting closure duration in word-initial intervocalic stops

The following speaker variables were explored as possible factors affecting CD: speaker identity, age, and gender. Because there is no indication in the literature that place of birth or place of living can have an effect on CD, these two variables were not investigated.

### 5.3.1 Individual differences between subjects

Figure 5.5 shows the distribution of CD values for /b, d, g/ for each of the twelve subjects (in ascending order from the shortest mean CD to the longest). Their CDs span a wide range of values of over 130 ms, and the individual means range from 64.8 ms for MPm to 127.7 ms for MVf.

In order to examine individual differences between subjects, a CART analysis was performed. There are two groups of subjects with significantly different CD values: Group 1, subjects MPm, SCf, IJm, DRm, RVm, with the mean CD of 81.25 ms (*SD* = 16.1, *N* = 75), and Group 2, subjects IVm, BCf, BPm, MRf, MCf, DARf, MVf, with the mean CD of 107.65 ms (*SD* = 20.63, *N* = 93).

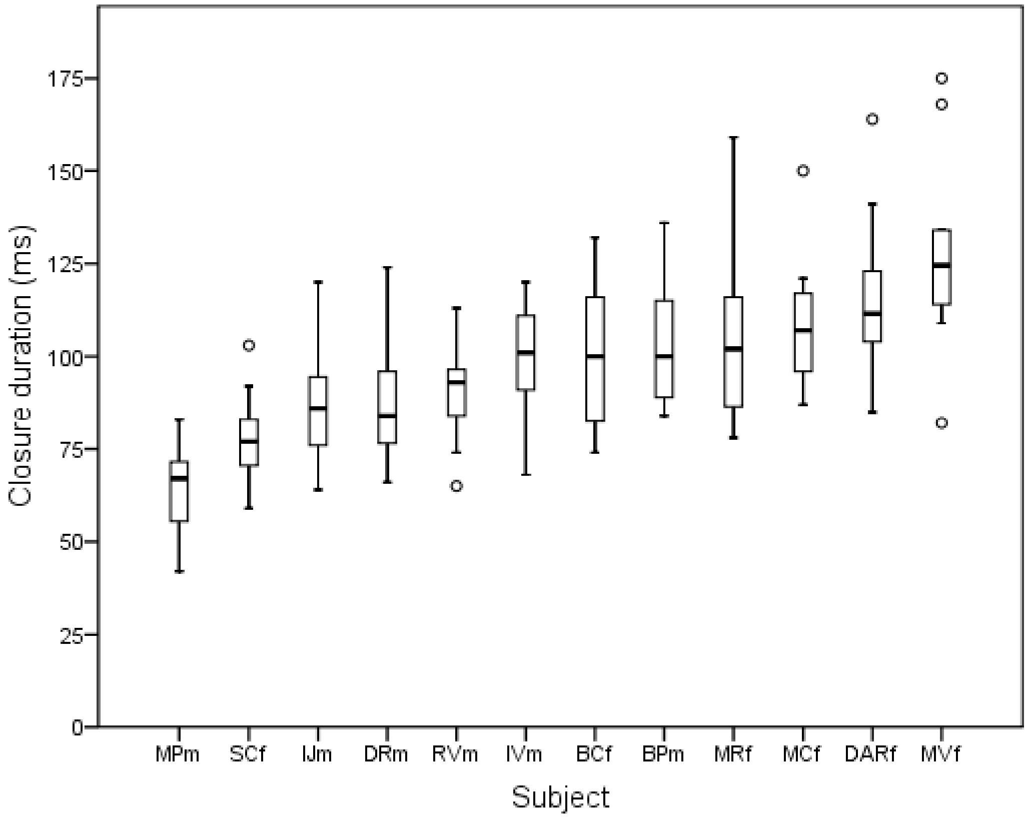

Figure 5.5 Boxplots showing the distribution of CD values for /b, d, g/ in word-initial intervocalic position for each subject

For /p, t, k/, Figure 5.6 shows the distribution of CD values for each subject (in the same order). As was the case with /b, d, g/, measured CDs have a wide range (over 130 ms), as do their individual means, which range from 81.58 ms for MPm to 164.87 ms for DARf.

A CART analysis revealed two groups of subjects whose CDs were significantly different. In the first group, with shorter closures, are subjects MPm, SCf, IJm, DRm, RVm, BCf, BPm (mean CD = 105.11 ms, *SD* = 20.66, *N* = 99), and in the second group, with longer closures, are subjects IVm, MCf, MRf, MVf and DARf (mean CD = 147.07 ms, *SD* = 21.72, *N* = 73).

A comparison of figures 5.5 and 5.6, and CART results, reveal that subjects who produce shorter CDs  for /b, d, g/ tend to have shorter CDs for /p, t, k/ as well, for example subjects MPm, SCf, IJm, DRm, and RVm. On the other hand, subjects MCf, MRf, MVf, and DARf tend to produce relatively longer CDs for both classes. This consistency could represent individual differences in CD production, or could be caused by differences in speaking rate.

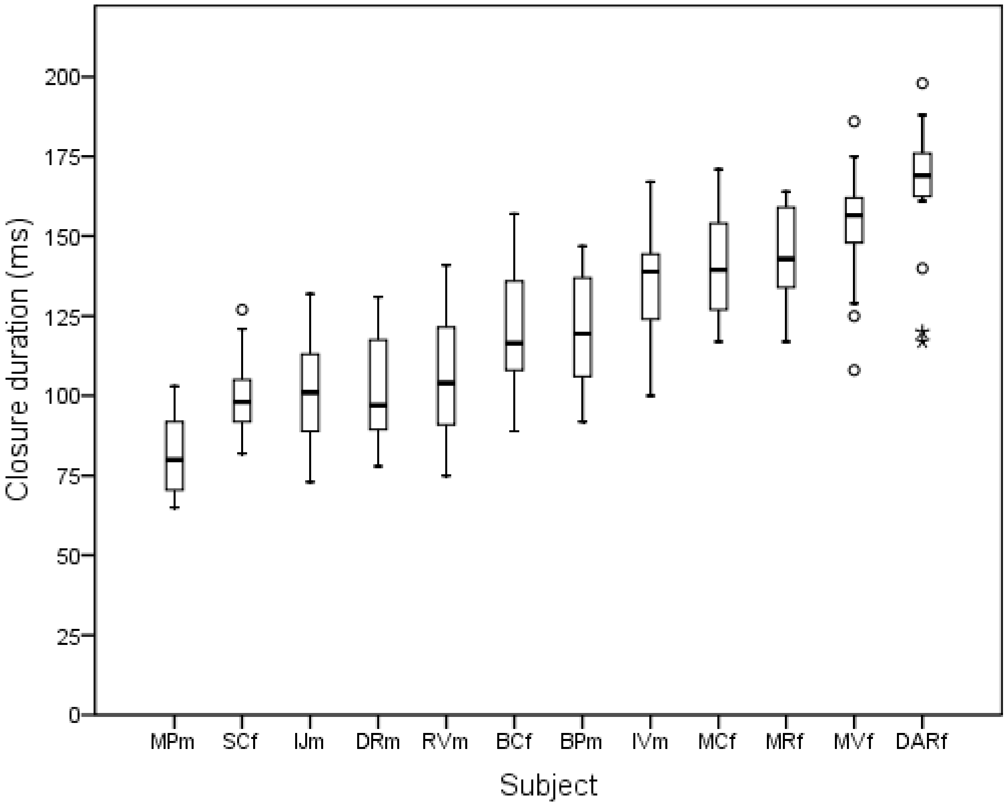

Figure 5.6 Boxplots showing the distribution of CD values for /p, t, k/ in word-initial intervocalic position for each subject

To evaluate the effect of speaking rate, speaking rate in syllables per second was calculated for each sentence. Speaking rate and CD in the pooled data for all stops are correlated, so that as speaking rate increases, CD decreases (Figure 5.7). This was confirmed by Spearman’s correlation analysis, $r_s$ = -0.61, $p$ < 0.001, $N$ = 340.

Individual mean speaking rate varies from 3.64 syll/s (MVf) to 5.67 syll/s (SCf). Although these speaking rates are not out of the ordinary in any way, it is still possible that they can account for at least some of the variability in CD production. Slower speakers tend to produce longer closures than faster speakers do (for both phonologically voiced and voiceless stops), and vice versa, as illustrated in Figure 5.8.

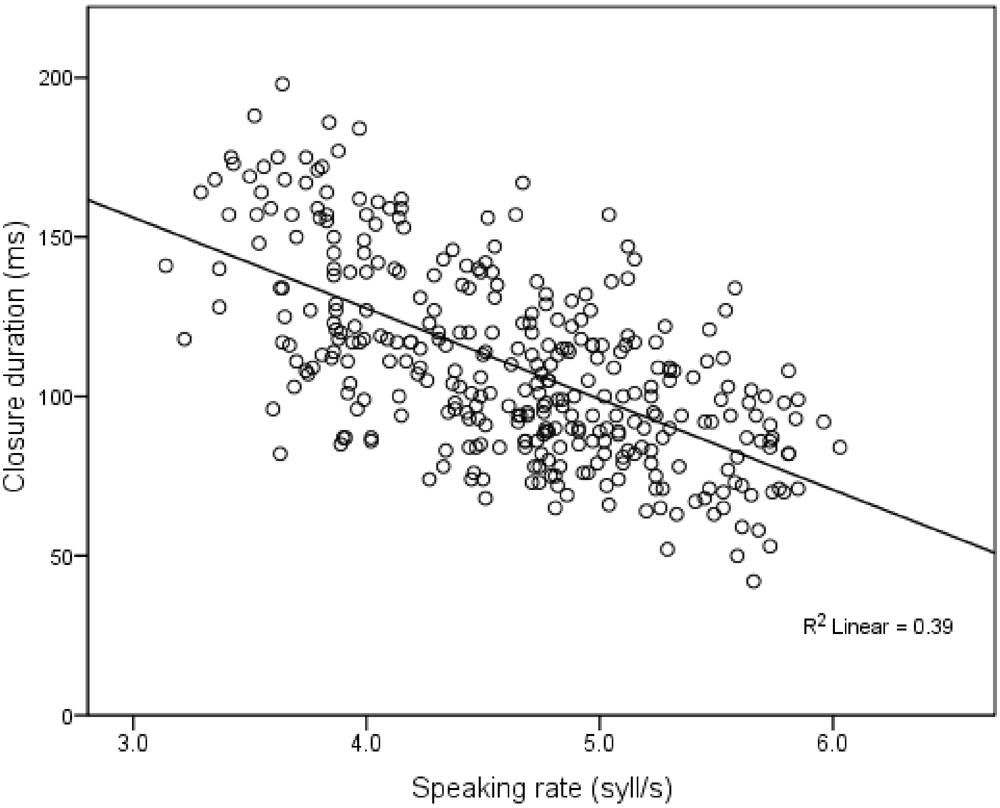

Figure 5.7 Relationship between CD and speaking rate in word-initial intervocalic position (in the pooled data for both stop classes)

Because of the design of the present study it is not possible to statistically determine what proportion of individual CD differences is attributable to between-speaker differences in speaking rate. It is, however, possible to look at individual differences expressed as a duration ratio or percentage. In Figure 5.9, a ratio of mean CDs for phonologically voiced and voiceless stops for each subject is plotted against each subject’s mean speaking rate. If speaking rate was the only factor causing individual differences in CD production, then all subjects would be expected to have approximately the same ratio, which is not the case. When the effect of speaking rate is neutralised by using CD ratio as a measure, there is still some individual variation, and some speakers produce larger differences in CD than others. Two further factors that can contribute to these differences, gender and age of speakers, are investigated in the next section.

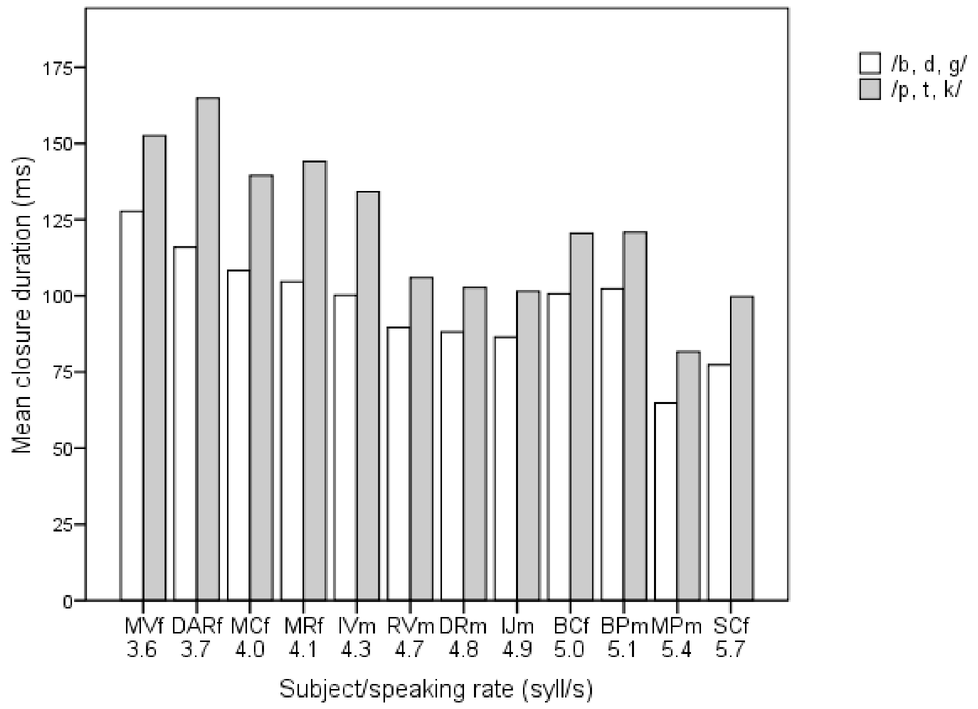

Figure 5.8 Mean CD (ms) for /b, d, g/ and /p, t, k/ in word-initial intervocalic position as a function of individual speaking rate

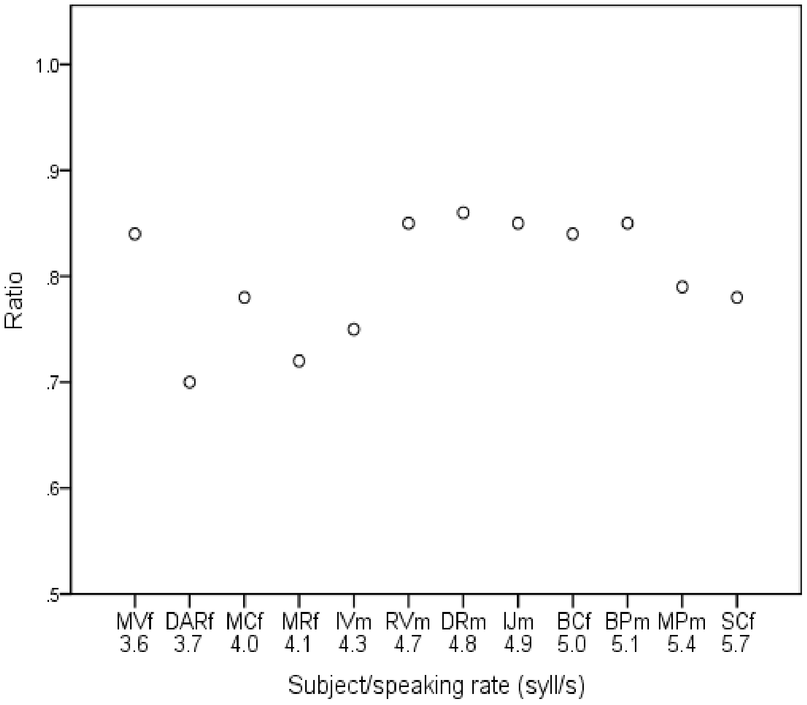

Figure 5.9 Ratio of mean CDs for /b, d, g/ and /p, t, k/ in word-initial intervocalic position as a function of individual speaking rate

### 5.3.2 Gender and age

To investigate the possible influence of gender and age on CD results for /b, d, g/, a two-way analysis of covariance (ANCOVA) was performed, with gender and age as independent variables (subjects were grouped as before), and speaking rate as a covariate. All ANCOVA assumptions were satisfied.

After adjusting for speaking rate, the main effect of age was significant, $F(1,162) = 13.67$, $p < 0.001$, $\omega^2 = 0.041$ (small effect). The effect of gender was just above the significance level, $F(1,162) = 6.64$, $p = 0.011$, but the effect size was large, $\omega^2 = 0.213$. There were no statistically significant interactions.

Table 5.6 presents mean CD and adjusted mean CD as a function of gender, and Table 5.7 presents mean CD and adjusted mean CD as a function of age. Adjusted mean represents the mean CD value after the effect of speaking rate was statistically removed.

| | Mean CD (ms) | N | SD | Adjusted mean CD (ms) |
|---|---|---|---|---|
| Female subjects | 104.22 | 79 | 24.39 | 99.86 |
| Male subjects | 88.45 | 89 | 18.61 | 92.69 |
| Difference | 15.79 | | | 7.17 |

Table 5.6 Mean CD (ms), N, SD, and adjusted mean CD (ms) for /b, d, g/ in word-initial intervocalic position as a function of speaker gender

Female subjects as a group produced longer closures than males. Four out of six female subjects are slower talkers, which may have contributed to this result. When the effect of speaking rate was co-varied, the difference was reduced by about half to about 7 ms (just above the significance level of 0.01, but with a large effect size).

| | Mean CD (ms) | N | SD | Adjusted mean CD (ms) |
|---|---|---|---|---|
| Older subjects | 102.17 | 81 | 21.55 | 101.13 |
| Younger subjects | 89.99 | 87 | 22.59 | 91.41 |
| Difference | 12.18 | | | 9.72 |

Table 5.7 Mean CD (ms), N, SD, and adjusted mean CD (ms) for /b, d, g/ in word-initial intervocalic position as a function of speaker age

Subjects older than 35 years produced voiced stops with significantly longer closures than younger subjects, by about 10 ms. After the effect of speaking rate was co-varied, the difference between means for the two groups was reduced only by 2.5 ms, which suggests that these two groups did not differ in speaking rates to a great extent. When CD results for individual subjects are plotted as a function of their gender and age, a tendency for CD to increase with age is observable in the male group, but not in the female group (Figure 5.10, females are on the left, males on the right, ordered from youngest to the oldest within each group).

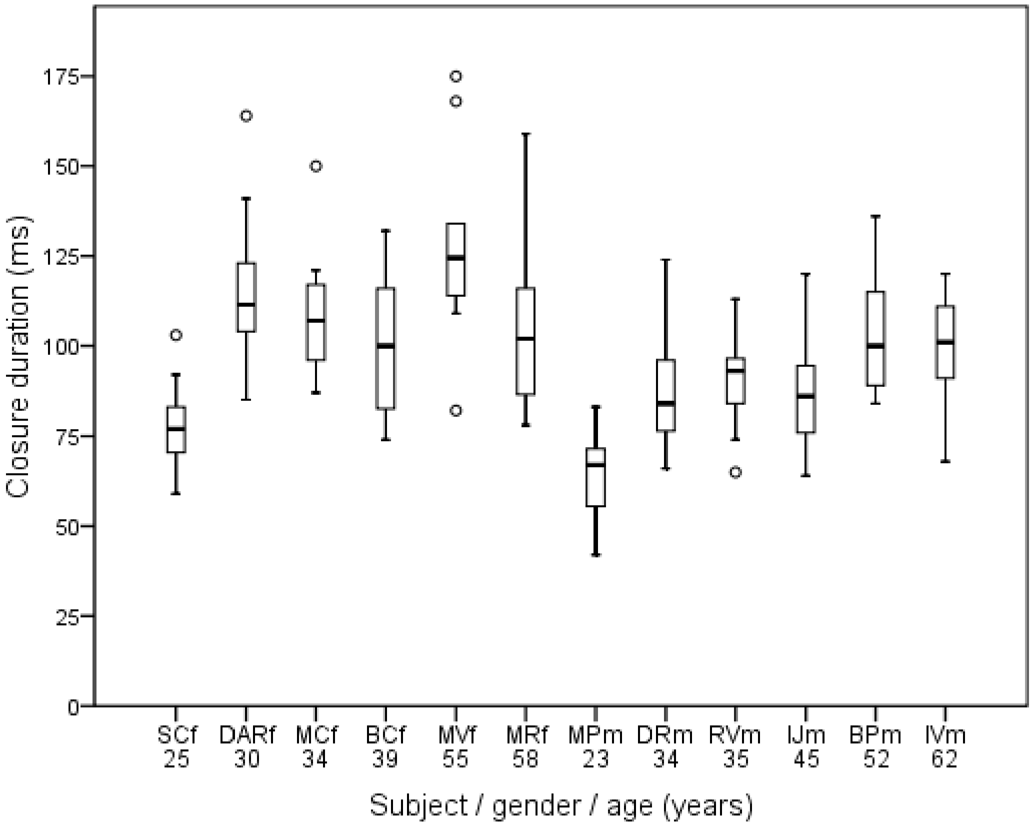

Figure 5.10 Boxplots showing the distribution of CD values for /b, d, g/ in word-initial intervocalic position as a function of speaker gender and age

A two-way ANCOVA was performed to examine the effect of gender and age on CD of /p, t, k/, with speaking rate as a covariate. All ANCOVA assumptions were satisfied.

After adjusting for speaking rate, the main effects of gender and age were statistically significant, for gender $F(1,167) = 30.1$, $p < 0.001$, $\omega^2 = 0.065$ (medium effect), for age $F(1,167) = 9.58$, $p = 0.002$, $\omega^2 = 0.019$ (small effect). There was also a statistically significant interaction between gender and age, $F(1,167) = 8.7$, $p = 0.004$, $\omega^2 = 0.017$ (small effect), Figure 5.11.

Overall, female subjects as a group produced longer closures than the male subjects, and after adjusting for speaking rate the mean difference was reduced from about 28 ms to 16 ms (Table 5.8). There was also less between-subject variation in the male group (Figure 5.12).

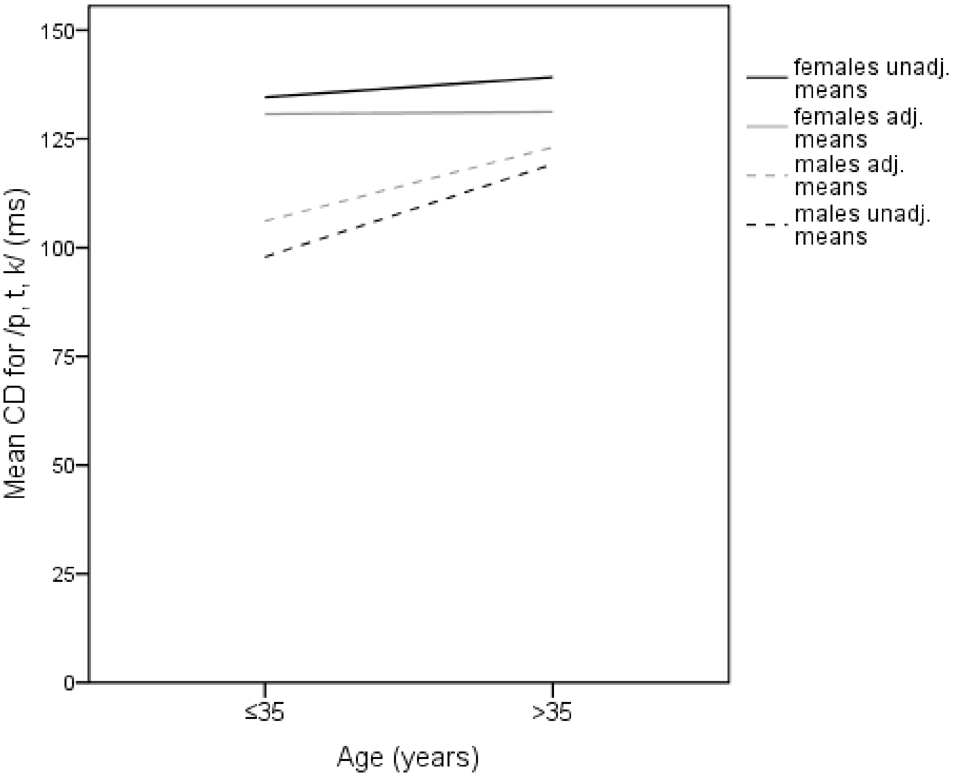

Figure 5.11 Mean CD (unadjusted and adjusted) as a function of gender and age of subjects for /p, t, k/ in word-initial intervocalic position

| | Mean CD (ms) | N | SD | Adjusted mean CD (ms) |
|---|---|---|---|---|
| Female subjects | 136.85 | 87 | 28.25 | 130.93 |
| Male subjects | 108.66 | 85 | 23.64 | 114.63 |
| Difference | 28.19 | | | 16.3 |

Table 5.8 Mean CD (ms), N, SD, and adjusted mean CD (ms) for /p, t, k/ in word-initial intervocalic position as a function of speaker gender

Subjects older than 35 years produced voiceless stops with significantly longer closures than younger subjects. By removing the effect of speaking rate, the difference in means between the two groups was reduced by 4 ms to about 9 ms (Table 5.9). The tendency for younger speakers to produce shorter closures is also evident in the male group (Figure 5.12, females are on the left, males on the right, ordered from youngest to the oldest within each group). In fact, both for /b, d, g/ and /p, t, k/, the CD values seem to increase with age for male subjects, but this is not the case with female subjects, where there is more between-subject variation.

| | Mean CD (ms) | N | SD | Adjusted mean CD (ms) |
|---|---|---|---|---|
| Older subjects | 129.19 | 86 | 24.59 | 127.14 |
| Younger subjects | 116.65 | 86 | 32.83 | 118.41 |
| Difference | 12.54 | | | 8.73 |

Table 5.9 Mean CD (ms), N, SD, and adjusted mean CD (ms) for /p, t, k/ in word-initial intervocalic position as a function of speaker age

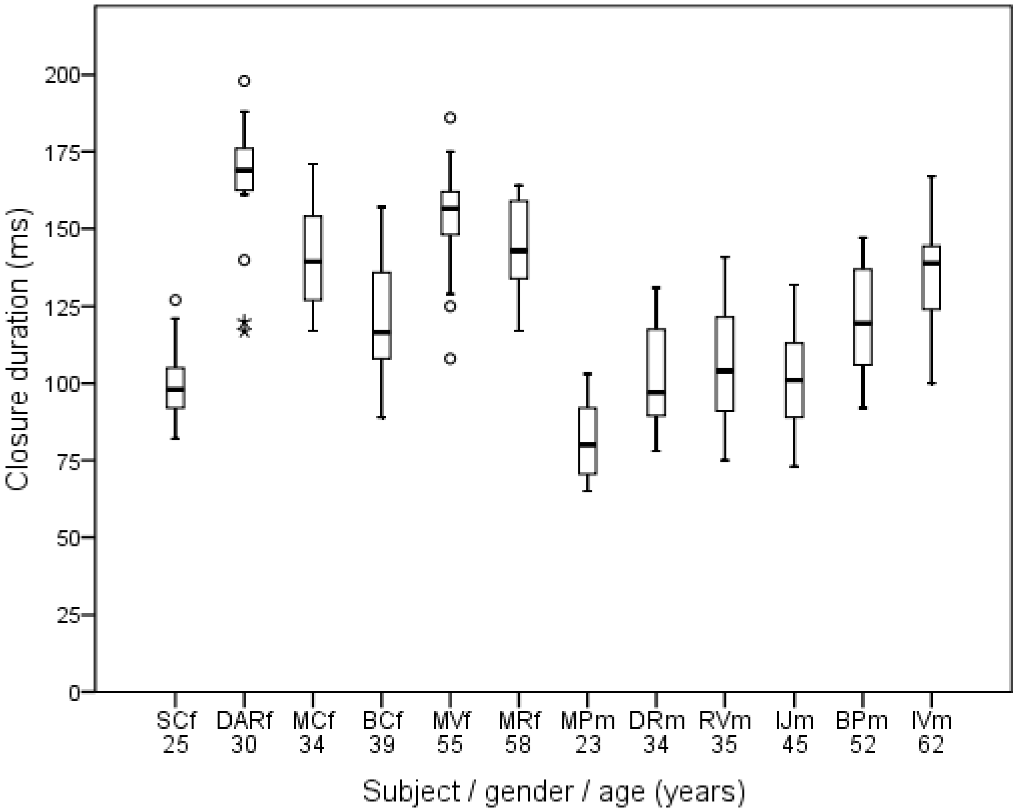

Figure 5.12 Boxplots showing the distribution of CD values for /p, t, k/ in word-initial intervocalic position as a function of speaker gender and age

In order to examine the interaction between gender and age, two further ANCOVAs were performed, for females and males separately. The effect of age was significant for the males, *F* (1, 82) = 17.48, $p < 0.001$, but not for the females, *F* (1, 84) = 0.009, $p = 0.93$. The difference in adjusted CD means between older and younger male subjects was 17.5 ms, but for the female subjects this difference was only 0.5 ms.

## 5.4 Summary of findings for closure duration in word-initial intervocalic stops

Mean CDs for /p, t, k/ and /b, d, g/ in word-initial intervocalic position (in the sentence frame) differ by about 27 ms or 22%, a difference which is statistically significant, with a medium effect size. Results for individual speakers are significant at either 0.01 or 0.05 level, and effect size is large for each speaker. These results suggest that CD is relevant as a correlate of the voicing distinction in Serbian, despite overlap in measured CD values for the two stop classes.

There is a significant effect of place on articulation on CD word-initially, with CD becoming progressively shorter for more retracted place of articulation for both stop classes, although not all pairwise comparisons are significant (/b/ has significantly longer closures than /d/ and /g/, while /k/ has significantly shorter closures than /p/ and /t/). Differences between means for each of cognate pairs are similar and about 25-30 ms, or 20-25% for all three places of articulation.

The quality of the following vowel has no effect on CD.

Duration of the voiceless interval for /p, t, k/ has a fairly uniform duration. However, there is no inverse relationship between VOT and CD, which suggests that place-related VOT differences are not caused by place-related differences in CD.

Despite large between-subject differences in CD for each stop class, subjects tend to produce relatively longer or shorter closures for both classes, while still maintaining the contrast. Three factors were found to induce this between-speaker variability: speaking rate, gender, and age of speakers.

For both /b, d, g/ and /p, t, k/ older speakers as a group produced closures that were about 9-10 ms (or 10% and 7%, respectively) longer than those produced by younger speakers (after adjusting for speaking rate). These differences were statistically significant. However, age interacts with gender so that age effect on CD is only present among the male subjects. Female subjects produce longer closures than male subjects do. Gender differences are somewhat exacerbated by differences in speaking rate, because females tend to be slower talkers in this study. After adjusting for speaking rate, closures produced by female speakers were 7 and 16 ms longer, for /b, d, g/ and /p, t, k/ respectively, which is a difference of 7% and 12%. This is a significant effect. There is more variability in results between the members of the female group. Some of it could be due to differences in speaking rate, but unlike in the male group, correlation with age is less visible.

## 5.5 Effect of phonological voicing category on closure duration in utterance-final stops

The distribution of results for CD in utterance-final stops (in word-final position in isolated words) is shown in boxplots in Figure 5.13 and their means in Table 5.10.

Stops were consistently released in this condition. Only three tokens with final /b/ did not have a visible release burst.

CD values for the two categories overlap in this condition. Measured values range from 47 ms to 170 ms for /b, d, g/, and from 85 ms to 261 ms for /p, t, k/, with means of 104.72 ms and 162.66 ms, respectively. This difference is statistically significant in the pooled data, according to a Mann-Whitney U-test, $p < 0.001$ (2-tailed), $Z = -17.35$, $r = -0.7$ (large effect). Difference between means for individual subjects varies from 26.38 ms for RVm to 88.17 ms for DARf. When differences are expressed as ratios, in order to eliminate possible effect of speaking rate, the same two subjects have the smallest and the largest difference in CD: RVm has a ratio of 0.78, or a 22% difference, while DARf has a ratio of 0.57, or a 43% difference.

Statistical analysis of individual results revealed a significant difference in CD of phonologically voiced and voiceless stops for each subject, with $p < 0.001$ (Table C3 in the Appendix C). The effect size is large for all subjects.

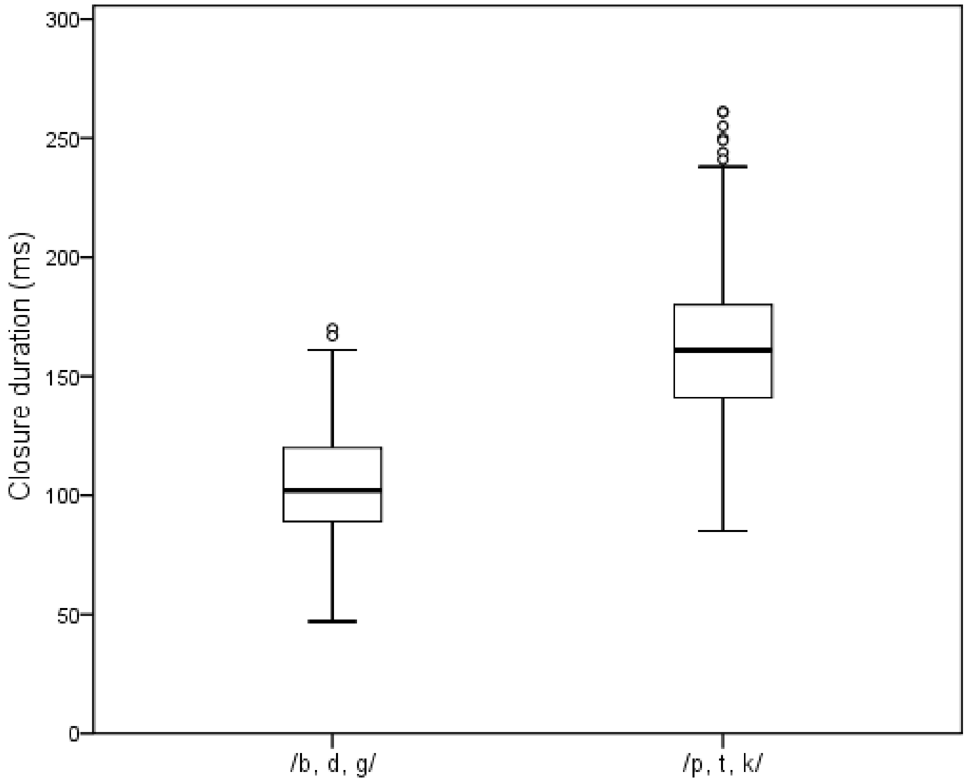

Figure 5.13 Boxplots showing the distribution of CD values for /b, d, g/ and /p, t, k/ in utterance-final position

| | Mean CD (ms) | N | SD |
|---|---|---|---|
| /b, d, g/ | 104.72 | 282 | 22.91 |
| /p, t, k/ | 162.66 | 287 | 35.06 |

Table 5.10 Mean CD (ms), N and SD for /b, d, g/ and /p, t, k/ in utterance-final position

## 5.6 Effect of place of articulation on closure duration in utterance-final stops

In this part of the study the quality of the vowel preceding final stops was not controlled and the vowels were represented with different numbers of tokens. For this reason, the effect of the preceding vowel on CD was not investigated.

CD results for /b, d, g/ in utterance-final position are presented in Table 5.11 and Figure 5.14. There was no significant effect of place of articulation on CD, according to a one-way ANOVA, *p* = 0.33, *F*(2,279) = 1.27.

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 106.77 | 92 | 21.45 |
| Dental | /d/ | 105.49 | 96 | 23.82 |
| Velar | /g/ | 101.91 | 94 | 23.3 |
| Total | | 104.72 | 282 | 22.91 |

Table 5.11 Mean CD (ms), N and SD for /b, d, g/ in utterance-final position for each place of articulation

CD results for /p, t, k/ in utterance-final position in isolation are presented in Figure 5.14 and Table 5.12. A one-way ANOVA revealed no statistically significant effect of stop place of articulation on CD, *p* = 0.43, *F*(2,284) = 0.85.

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 163.66 | 95 | 27.84 |
| Dental | /t/ | 165.37 | 96 | 34.73 |
| Velar | /k/ | 158.99 | 96 | 41.3 |
| Total | | 162.66 | 287 | 35.06 |

Table 5.12 Mean CD (ms), N and SD for /p, t, k/ in utterance-final position for each place of articulation

Phonologically voiced stops at all three places of articulation have shorter closures than their voiceless cognates. For all three pairs a Mann-Whitney U-test revealed significant differences: *p* < 0.001 (2-tailed) for all three pairwise comparisons, and *Z* = -10.68, *r* = -0.78 (large effect) for the pair /b/-/p/, *Z* = -10.09, *r* = -0.73 (large effect) for the pair /d/-/t/, *Z*= -9.4, *r* = -0.68 (large effect) for the pair /g/-/k/.  Lack of any interaction between the voicing distinction and place of stop articulation is illustrated in Figure 5.14.

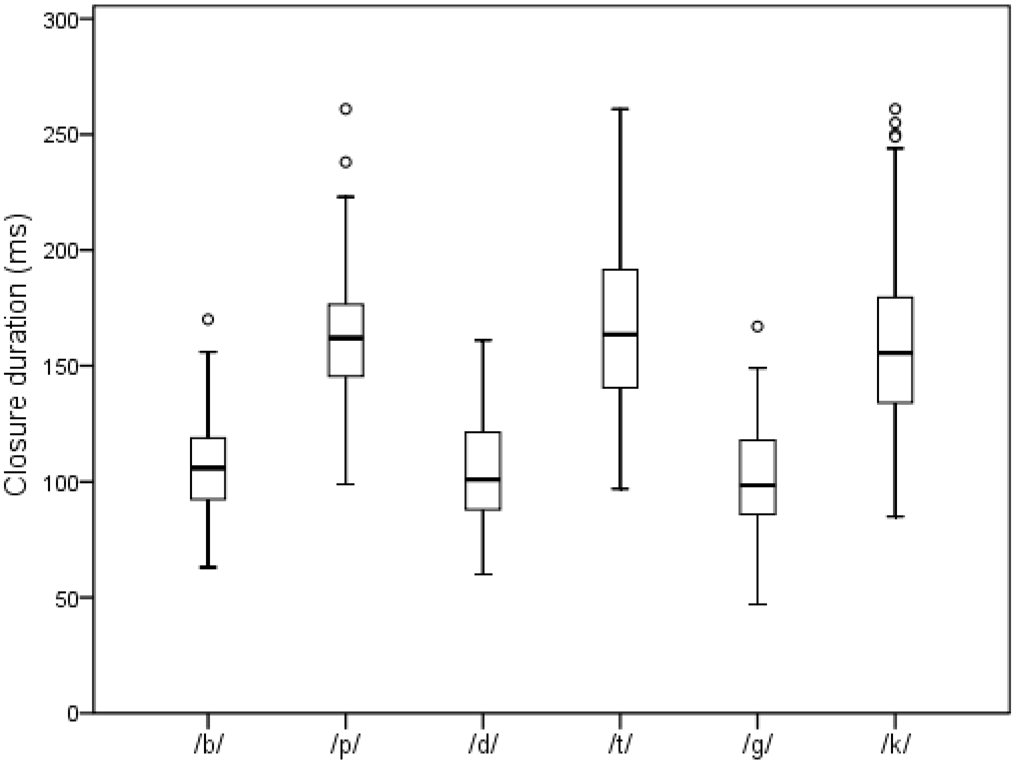

Figure 5.14 Boxplots showing the distribution of CD values for each stop in utterance-final position

## 5.7 Speaker factors affecting closure duration in utterance-final stops

The speaker variables that were explored as possible factors affecting CD are speaker identity, age, and gender.

### 5.7.1 Individual differences between subjects

Figure 5.15 shows the distribution of CD values for /b, d, g/ for each of the twelve subjects (in ascending order from the shortest mean CD to the longest). It also illustrates the extent of between-subject variability in CD production in this context. The mean values vary from around 80 ms (DRm) to around 140 ms (MRf).

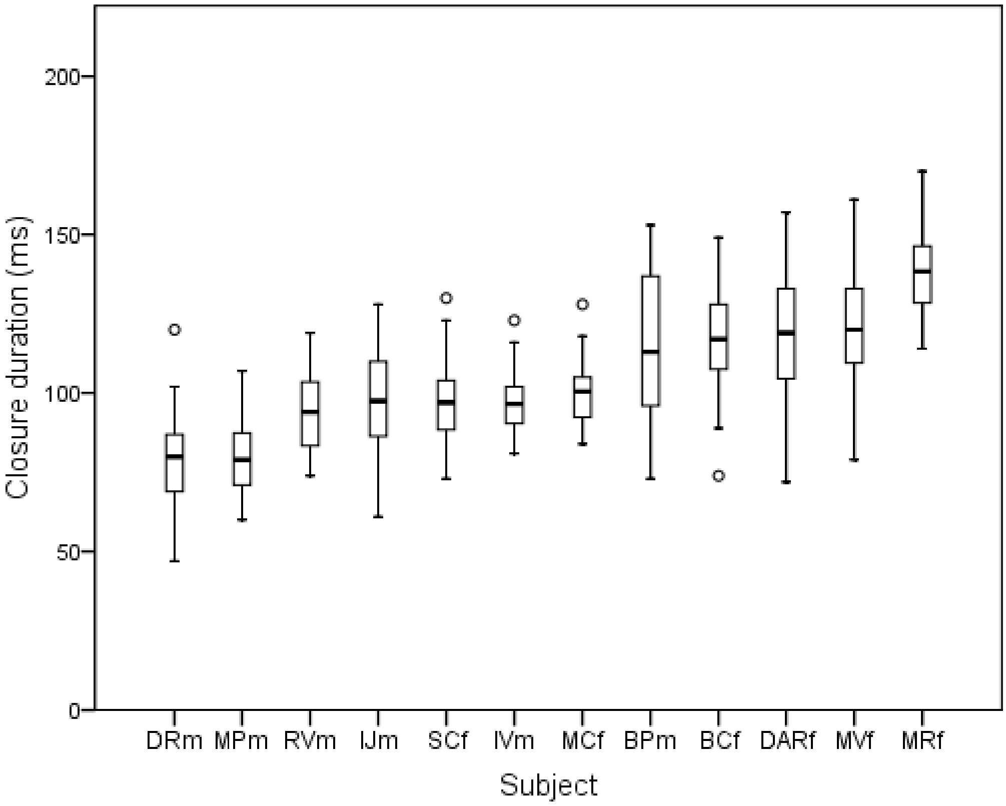

Figure 5.15 Boxplots showing the distribution of CD values for /b, d, g/ in utterance-final position for each subject

A CART analysis performed to test for individual differences showed that there are two significantly different groups of subjects: Group 1, subjects DRm, MPm, RVm, IJm, SCf, IVm, and MCf, with the mean CD of 92.45 ms (*N* = 165, *SD* = 15.6); Group 2, subjects BPm, BCf, DARf, MVf, MRf with the mean CD of 122.02 ms (*N* = 117, *SD* = 20.28).

Figure 5.16 shows the distribution of CD values for /p, t, k/ for each of the twelve subjects (in ascending order from the shortest mean CD to the longest). Mean CD values for individual subjects vary from 121 ms for RVm to 206 ms for DARf.

A CART analysis was performed to examine individual differences in CD production. There are three significantly different groups in this respect: Group 1: subjects RVm, MPm, DRm, with mean *CD* = 129.42 ms (*N* = 71, *SD* = 24.09); Group 2: subjects BPm, IVm, MCf, BCf, SCf, IJm, with mean CD = 160.78 (*N* = 144, *SD* = 21.15); Group 3: subjects MVf, MRf, DARf, with mean CD = 199.21 (*N* = 72, *SD* = 31.79).

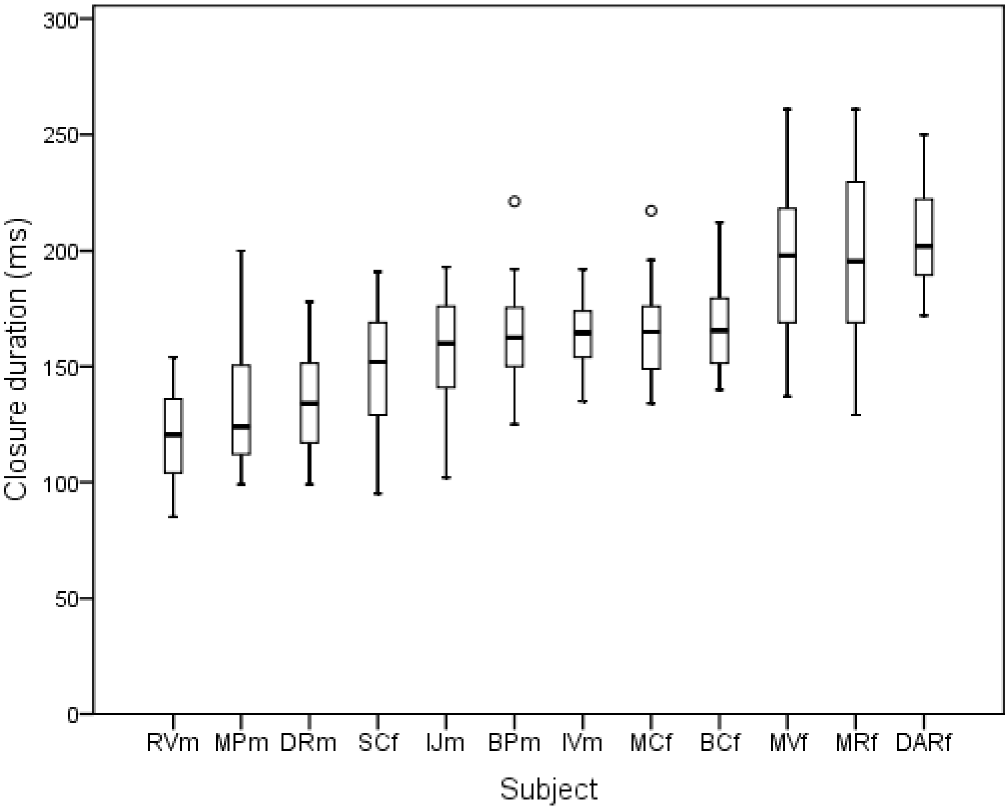

Figure 5.16 Boxplots showing the distribution of CD values for /p, t, k/ in utterance-final position for each subject

A comparison of CART results for both /b, d, g/ and /p, t, k/ suggests that some subjects have consistently longer (MVf, MRf, DARf) or shorter closures (DRm, MPm, RVm) for both stop classes, relative to the others, which could be caused by other factors, such as gender and age, or by differences in speaking rate. Speaking rate in syllables/second was not measured in this condition because it is of questionable validity for monosyllabic words uttered in isolation. Consequently, it was not possible to perform an ANCOVA to determine if there is an interaction between speaking rate and the other two subject factors, gender and age of speakers. Instead, an ANOVA was performed, with gender and age as independent variables.

However, when mean CDs and ratios of mean durations for the two stop classes are plotted as a function of individual speaking rate (Figure 5.17 and Figure 5.18), it is clear that some individual differences remain, irrespective of speaking rate (for illustration only, subjects are ordered according to their mean speaking rate measured in sentences with stops in word-initial, not in word-final position).

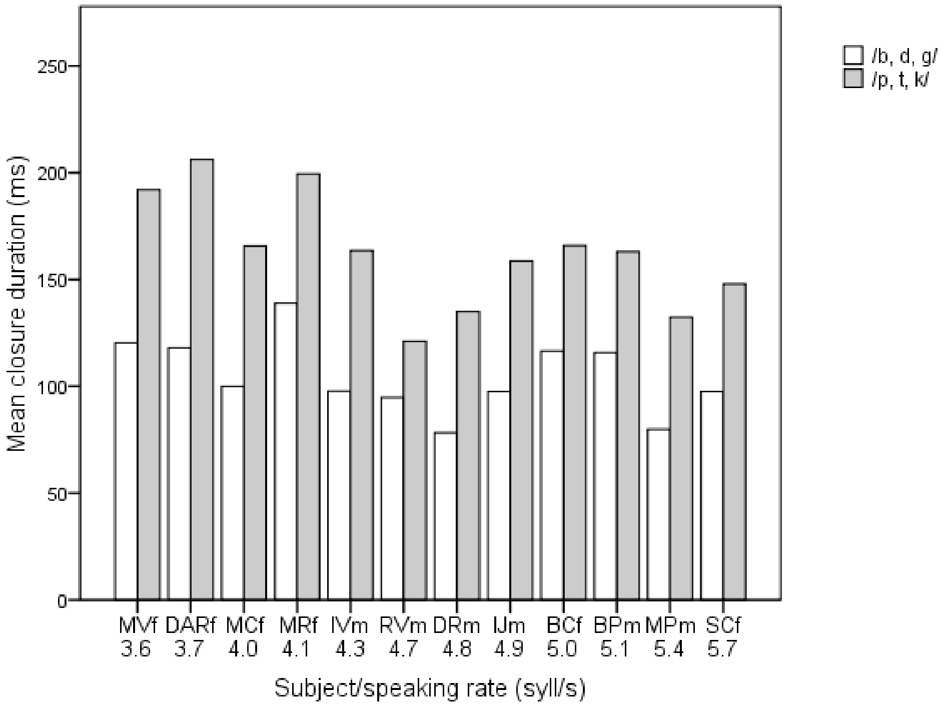

Figure 5.17 Mean CD (ms) for /b, d, g/ and /p, t, k/ in utterance-final position as a function of individual speaking rate

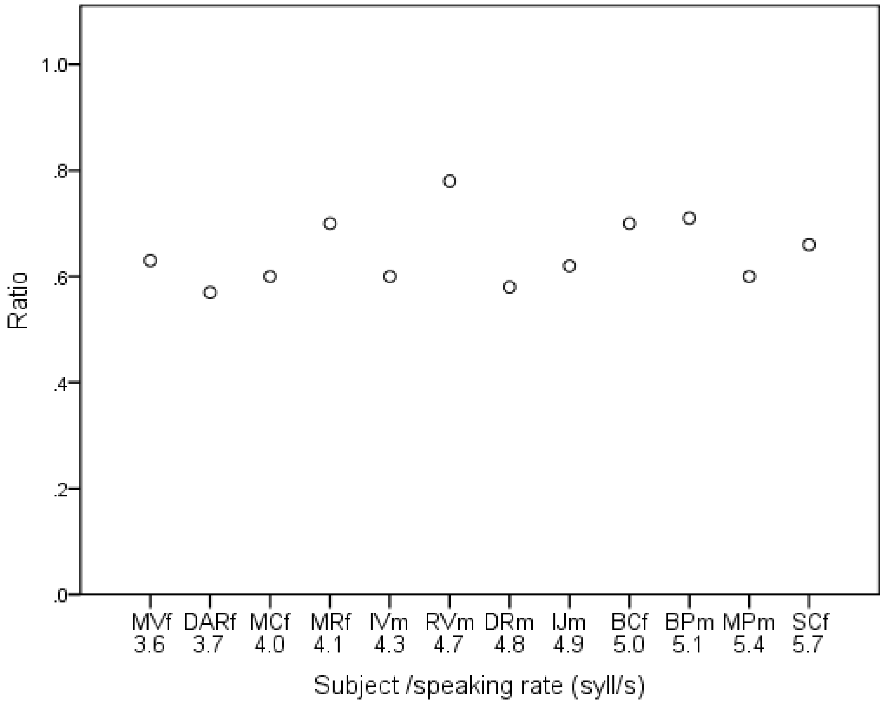

Figure 5.18 Ratio of mean CDs for /b, d, g/ and /p, t, k/ in utterance-final position as a function of individual speaking rate

### 5.7.2 Gender and age

A two-way ANOVA was carried out to explore the effect of gender and age on CD in utterance-final /b, d, g/. The subjects were divided into two equal groups according to their age, with equal numbers of males and females in each group, as before.

The main effect of gender was statistically significant $F(1,278) = 100.52$, $p < 0.001$, $\omega^2 = 0.215$ (large effect). Female subjects produced final voiced stops with longer CD than male subjects did, by 21 ms (or 18%) on average (mean CD for females = 115.17 ms, *SD* = 21.07, *N* = 143, mean CD for males = 93.96 ms, *SD* = 19.54, *N* = 139).

There was a significant main effect of age on CD, $F(1,278) = 83.89$, $p < 0.001$, $\omega^2 = 0.179$ (large effect). Subjects older than 35 years produced voiced stops with significantly longer CD than younger subjects, by 19 ms (or 17%) on average (mean CD for older subjects = 114.38 ms, *SD* = 22.07, *N* = 141; mean CD for younger subjects = 95.05 ms, *SD* = 19.45, *N* = 141).

There was no statistically significant interaction between the two main factors, $F(1,278) = 0.1$, $p = 0.75$.

Boxplots in Figure 5.19 show that there were no large individual variations in either group (females are on the left, males on the right, ordered by age within each group). Figure 5.19 also confirms that within each group older speakers produced longer closures than younger speakers did (although DARf and IVm somewhat stand out from their respective groups).

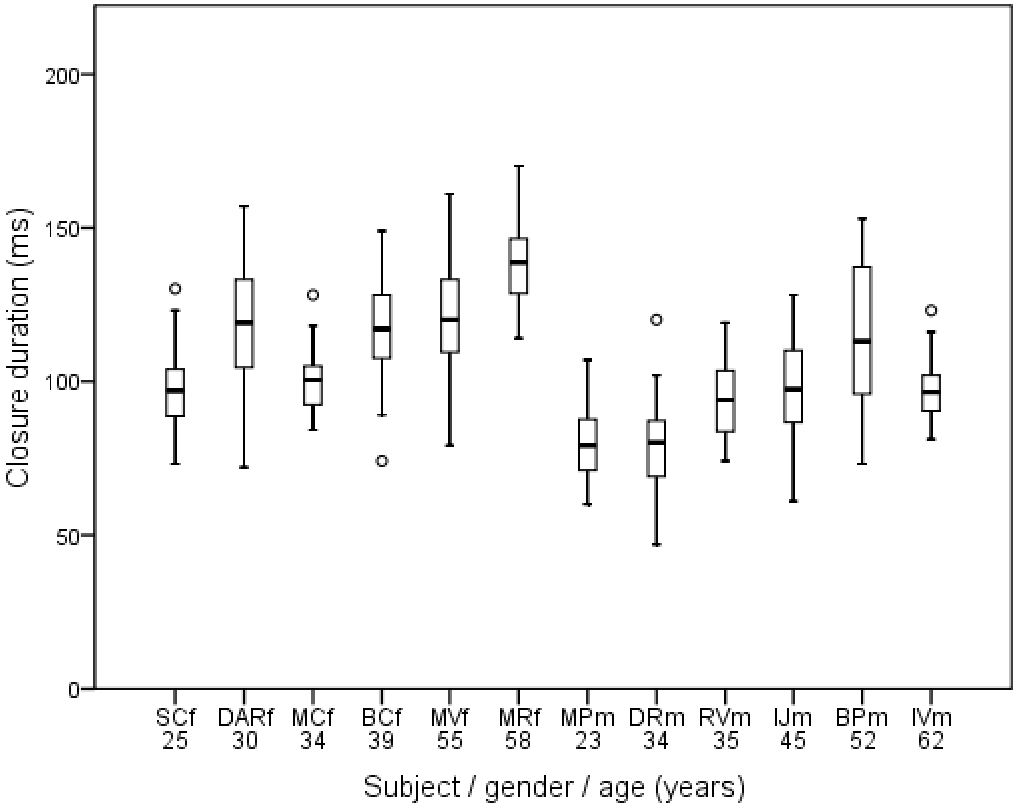

Figure 5.19 Boxplots showing the distribution of CD values for /b, d, g/ in utterance-final position as a function of speaker gender and age

A two-way between-subjects ANOVA was carried out to examine the effect of gender and age on CD in /p, t, k/.

The main effect of gender was statistically significant *F*(1,283) = 103.44, $p < 0.001$, $\omega^2 = 0.233$ (large effect). Female subjects produced voiceless stops with longer CD than male subjects did, with difference of about 34 ms, or 19% (mean CD for females = 179.53 ms, *SD* = 33.87, *N* = 144, mean CD for males = 145.68 ms, *SD* = 27.23, *N* = 143).

There was a significant main effect of age on CD, *F*(1,283) = 44.97, $p < 0.001$, $\omega^2 = 0.1$ (medium effect). Subjects older than 35 years produced voiceless stops with significantly longer CD than younger subjects, with difference of about 22 ms, or 13% (mean CD for older subjects = 173.74 ms, *SD* = 30.06, *N* = 144; mean CD for younger subjects = 151.5 ms, *SD* = 36.27, *N* = 144).

There was a statistically significant interaction between the two main factors, *F*(1,283) = 8.78, $p = 0.003$, $\omega^2 = 0.018$, small effect (Figure 5.20).

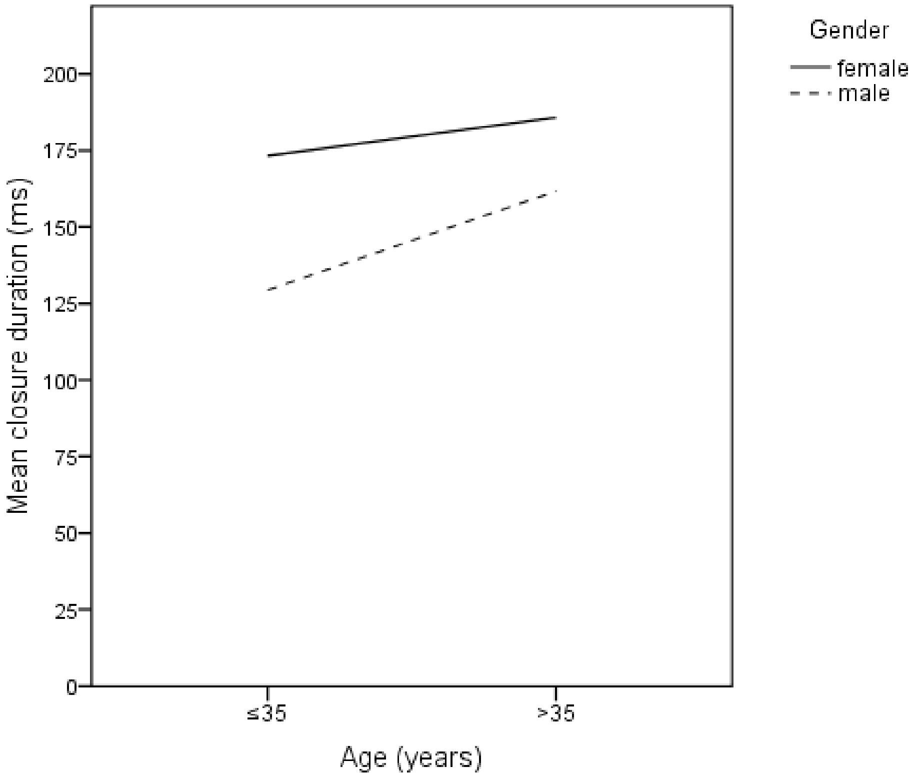

Figure 5.20 Mean CD as a function of gender and age of subjects for /p, t, k/ in utterance-final position

Because of the interaction, the effect of age on each gender was examined separately. For female speakers the effect of age was not statistically significant, Mann-Whitney U-test, *Z* = -1.882, *p* = 0.06 (2-tailed), *N* = 144, although older female subjects produced longer closures (mean CD = 185.78 ms, *SD* = 33.76, *N* = 72) than younger female subjects (mean CD = 173.28 ms, *SD* = 33.03, *N* = 72).

On the other hand, older male subjects produced significantly longer closures than younger male subjects, t-test, *t* (141) = -8.78, *p* < 0.001 (2-tailed), Cohen's *d* = -1.48 (large effect), with mean CD for older males = 161.71 ms, *SD* = 19.64, *N* = 72, and mean CD for younger males = 129.42 ms, *SD* = 24.09, *N* = 71), Figure 5.22.

Boxplots in Figure 5.21 (with females on the left, and males on the right, ordered by age within each group) illustrate that female subjects as a group produced longer closures, although there is more individual variation within the female group. There is also a tendency within each group for younger speakers to produce shorter closures than older speakers do, with the exception of DARf (as was the case with phonologically voiced stops).

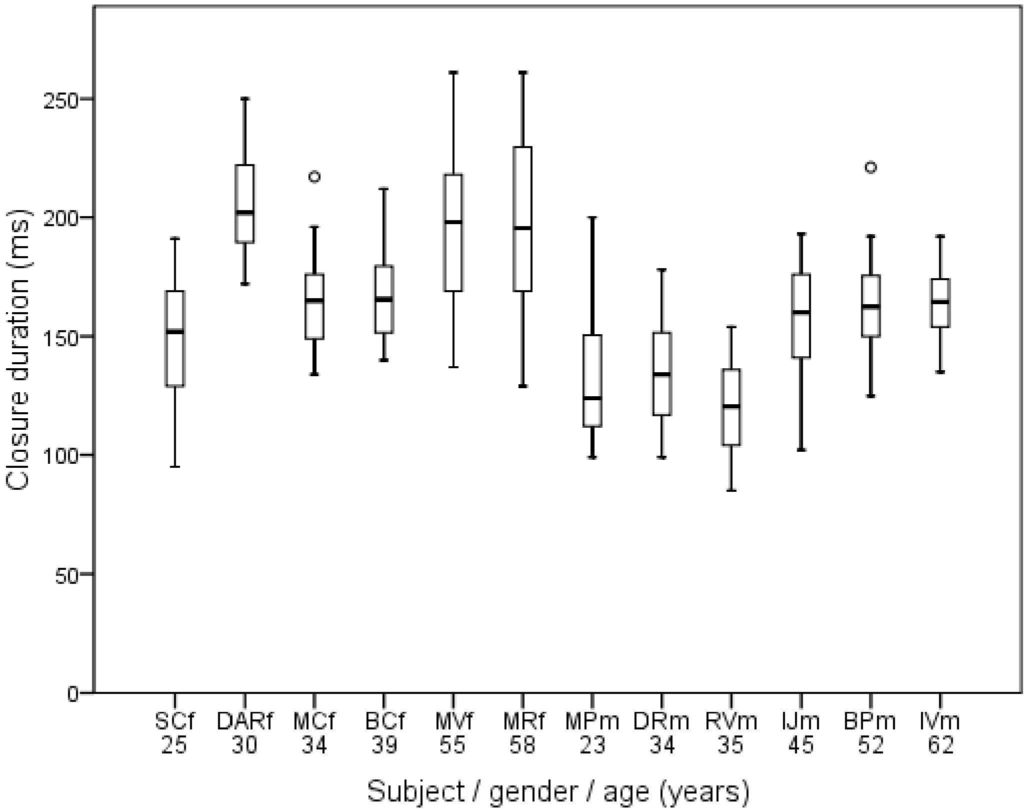

Figure 5.21 Boxplots showing the distribution of CD values for /p, t, k/ in utterance-final position as a function of speaker gender and age

## 5.8 Summary of findings for closure duration in utterance-final stops

Closure duration is a reliable correlate of the voicing distinction in Serbian word-final stops in isolated words. Although CDs for /b, d, g/ and /p, t, k/ overlap, there is a statistically significant difference in CD for each individual subject, and in the pooled data, where it is almost 60 ms or about 37%. The effect size is large in all cases.

Of the linguistic factors that can induce variability in CD, place of articulation does not have any effect utterance-finally in Serbian.

On the other hand, between-subject differences in CD are present for each stop class. Because in this condition words were uttered in isolation, speaking rate was not measured, and it was not possible to adjust for the effect of the speaking rate on CD using an ANCOVA. An ANOVA was run instead, and both gender and age were found to have an effect on CD production.

Female subjects produced longer closures than male subjects. Differences in mean CD between the female and the male group of speakers are about 21 ms for /b, d, g/ and about 34 ms for /p, t, k/ in the pooled data. They represent 18% and 19% difference, respectively, and are statistically significant. The male group is more homogeneous in CD production than the female group, for both stop classes.

The older subjects as a group produced longer closures than the younger subjects. For /b, d, g/ difference in means is about 19 ms, and for /p, t, k/ about 22 ms, or 17% and 13%. Both are statistically significant. However, there is a discrepancy between male and female subjects: the effect of age on CD is more consistent for the male subjects. For the female subjects, individual results for subject DARf, who has longer closures, stand out from other younger females. This might be caused by her speaking rate, since she is one of the slowest speakers, but it could also be an individual feature of her production.

The same pattern of the effect of gender and age on CD is present in word-initial and word-final stops: for /b, d, g/ there is no interaction between gender and age, but for /p, t, k/ they interact in both conditions, so that age differences are significant for the males, but not for the females, where there is more individual variation. This suggests that, since speaking rate was included as a factor in statistical analysis for word-initial stops, but not for word-final stops, observed gender and age differences are not caused only by speaking rate differences, but represent genuine effects, which only interact with speaking rate to a small extent (as shown in Section 5.3).

## 5.9 Effect of phonological voicing category on closure duration in word-final intervocalic stops

The same word tokens were recorded in isolation and in the sentence condition, but fewer tokens were measured in the sentence condition, because some subjects uttered a short break after the target word. Consequently, the stop under investigation was not in intervocalic position as intended, and planned measurements could not have been taken. Out of the twelve subjects, subjects BCf, SCf, and MPm were able to produce most or all of the target words in intervocalic position and their data make up the most of the data in this section (if some tokens were discarded, it was for other reasons). Of the remaining subjects, only IJm had more than five tokens for each stop class. The analysis that follows is based on data from these four subjects: BCf, SCf, MPm and IJm. Subjects who did not produce at least five tokens for each stop class in intervocalic position were not included because it would invalidate statistical analysis.

In this sample, only three stops did not have visible release burst, one /b/ token, and two /g/ tokens. All remaining final stops were released.

The distribution of CDs for final voiced and voiceless stops in the sentence frame for the four subjects is shown in boxplots in Figure 5.22 and their means in Table 5.13. Results are pooled across all four subjects. Results for CD of the two stop categories overlap, as was the case with results for final stops in isolated words. Mean CD for /b, d, g/ is 65.59 ms, and for /p, t, k/ it is 98.38. This difference is statistically significant, according to a t-test, $p$ < 0.001 (2-tailed), $t$(138) = -11.19, Cohen's $d$ = -1.91 (large effect). Differences between means for individual subjects are between 24.09 ms (for MPm) and 40.03 ms (for SCf). Ratios of mean CD for /b, d, g/ and mean CD for /p, t, k/ vary from 0.63 (or 37%) for SCf to 0.73 (or 27%) for IJm.

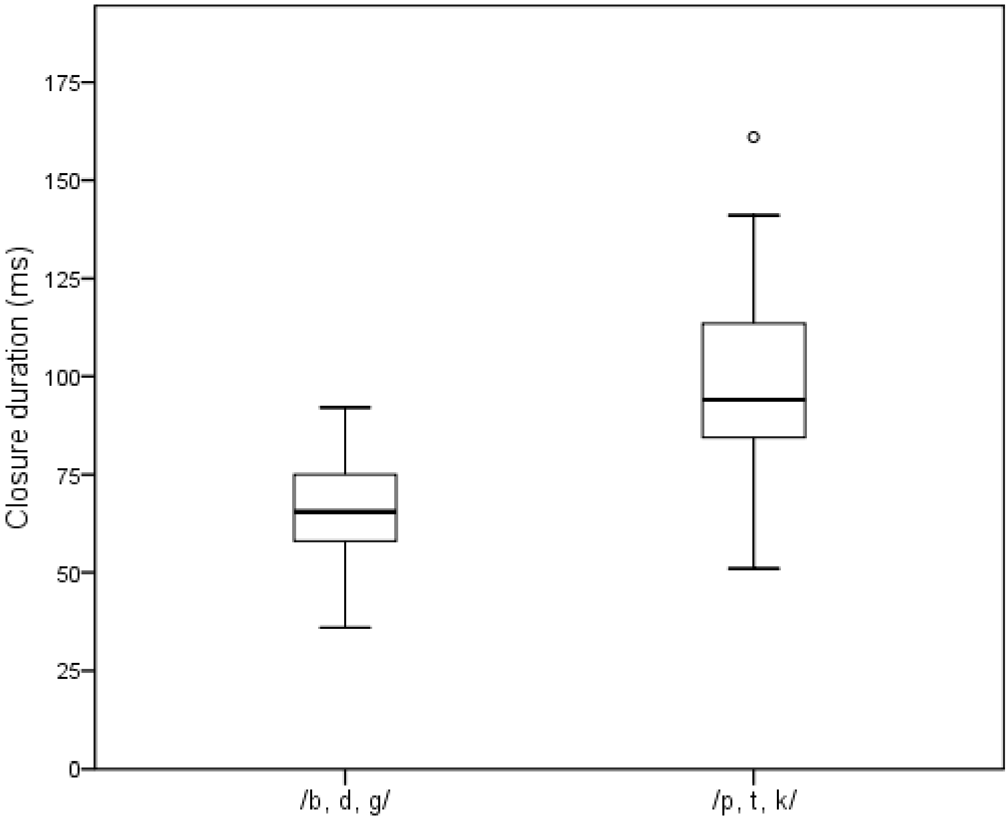

Figure 5.22 Boxplots showing the distribution of CD values for /b, d, g/ and /p, t, k/ in word-final intervocalic position

Statistical analysis of individual results revealed a significant difference in CD of the two stop classes for each subject (Table C4 in the Appendix C). The effect size is large for all subjects.

| | Mean CD (ms) | N | SD |
|---|---|---|---|
| /b, d, g/ | 65.59 | 68 | 12.64 |
| /p, t, k/ | 98.38 | 94 | 21.18 |

Table 5.13 Mean CD (ms), N and SD for /b, d, g/ and /p, t, k/ in word-final intervocalic position

## 5.10 Effect of place of articulation on closure duration in word-final intervocalic stops

CD results for /b, d, g/ in word-final intervocalic position are presented in Table 5.14 and Figure 5.23.

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 71.3 | 23 | 9.89 |
| Dental | /d/ | 63.08 | 24 | 14.14 |
| Velar | /g/ | 62.19 | 21 | 11.86 |
| Total | | 65.59 | 68 | 12.64 |

Table 5.14 Mean CD (ms), N and SD for /b, d, g/ in word-final intervocalic position for each place of articulation

As can be seen from Figure 5.23, there is a slight tendency for CD to increase at more forward places of articulation. Values for the three places of articulation overlap. An ANOVA revealed that these differences were not statistically significant at the adjusted level of 0.01, $p$ = 0.025, $F$(2,65) = 3.89.

CD results for /p, t, k/ in word-final intervocalic position are presented in Table 5.15 and Figure 5.23, for all three places of articulation. Place-related differences in CD were not statistically significant, according to a one-way ANOVA, $p$ = 0.28, $F$(2,69) = 2.49.

| | | Mean CD (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 105.57 | 21 | 23.53 |
| Dental | /t/ | 98.67 | 27 | 17.78 |
| Velar | /k/ | 91.75 | 24 | 21.29 |
| Total | | 98.38 | 72 | 21.18 |

Table 5.15 Mean CD (ms), N and SD for /p, t, k/ in word-final intervocalic position for each place of articulation

In Figure 5.23 these results are presented for each stop pair. For all three pairs a t-test revealed that differences in CD are significant, for the pair /b/-/p/: *p* < 0.001 (2-tailed), *t(42)*= -6.19, Cohen's *d* = -1.91, large effect; for the pair /d/-/t/: *p* < 0.001 (2-tailed), *t(49)*= -7.84, Cohen's *d* = -2.24, large effect; for the pair /g/-/k/: *p* < 0.001 (2-tailed), *t(43)*= -5.84, Cohen's *d* = -1.78, large effect.

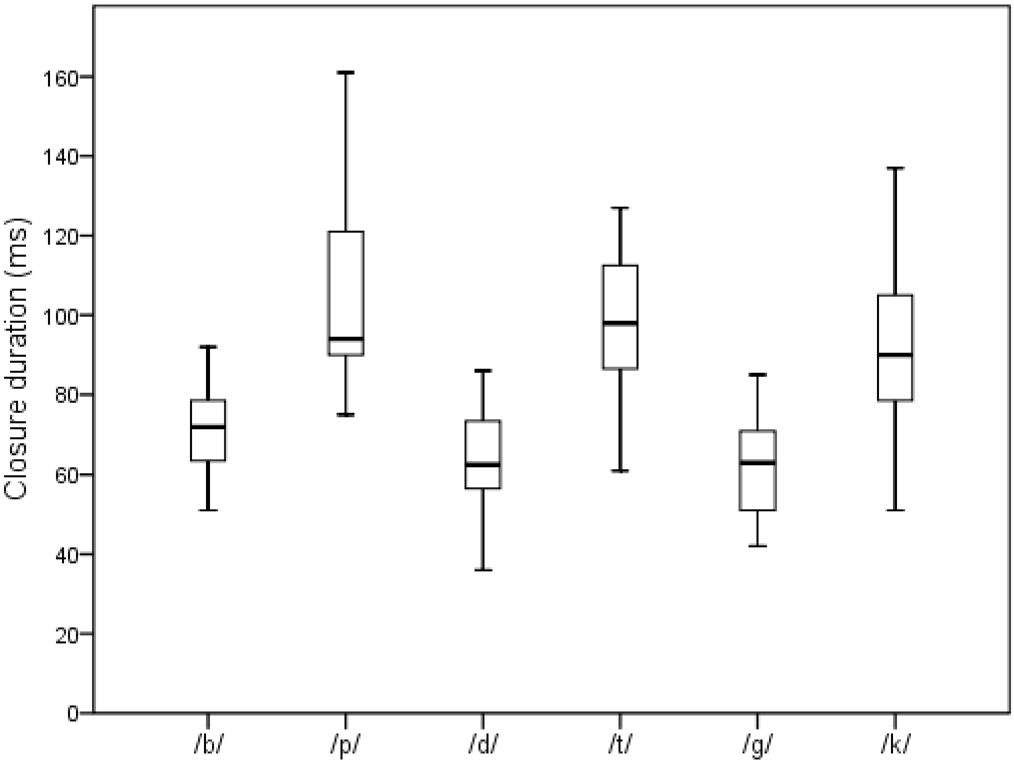

Figure 5.23 Boxplots showing the distribution of CD values for each stop in word-final intervocalic position

## 5.11 Speaker factors affecting closure duration in word-final intervocalic stops

### 5.11.1 Individual differences between subjects

Results for /b, d, g/ in word-final intervocalic position for each subject individually are shown in Figure 5.24 (in ascending order from the shortest mean CD to the longest mean CD). There are no large differences between subjects, except MPm, who tends to have somewhat shorter CDs (with mean CD of 54 ms, versus the other three subjects whose mean CD is around 70 ms). Subject IJm is represented with a smaller number of tokens than others, and consequently has a narrower distribution of CD values. No CART analysis was performed because of the small number of subjects.

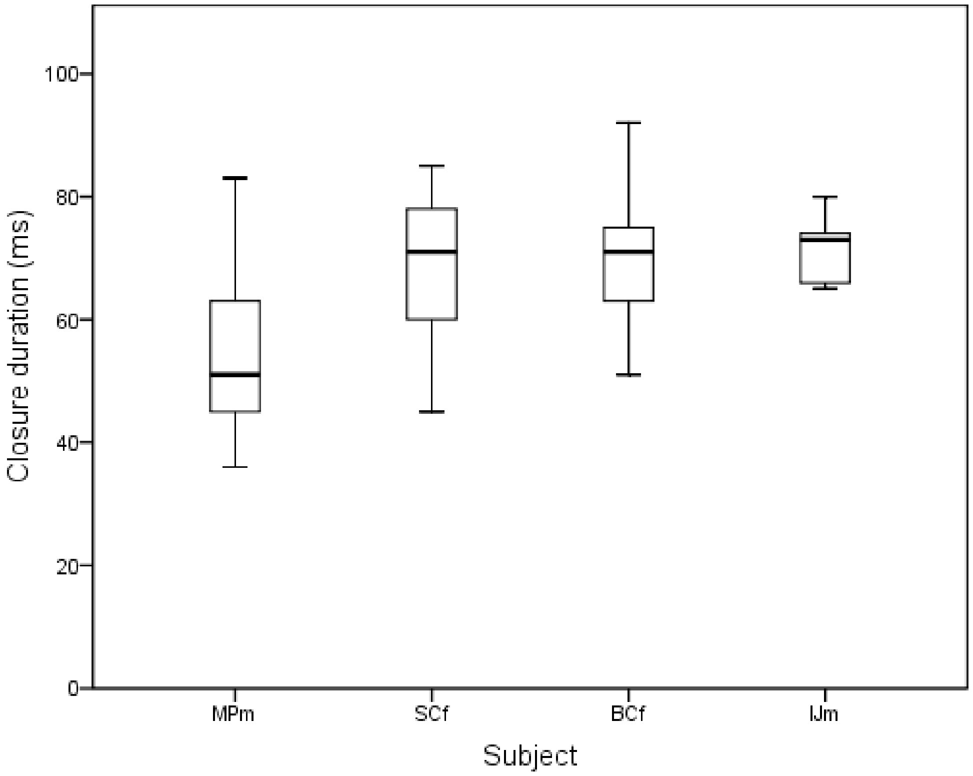

Figure 5.24 Boxplots showing the distribution of CD values for /b, d, g/ in word-final intervocalic position for each subject

For /p, t, k/, individual CD results are shown in Figure 5.25 (in the same order). The two female subjects have longer CDs (with means around 110 ms) than the male subjects (with means below 100 ms), but it is difficult to generalise because subject IJm is represented with fewer tokens than the other three subjects. No CART analysis was performed on these results either, because of the small number of subjects.

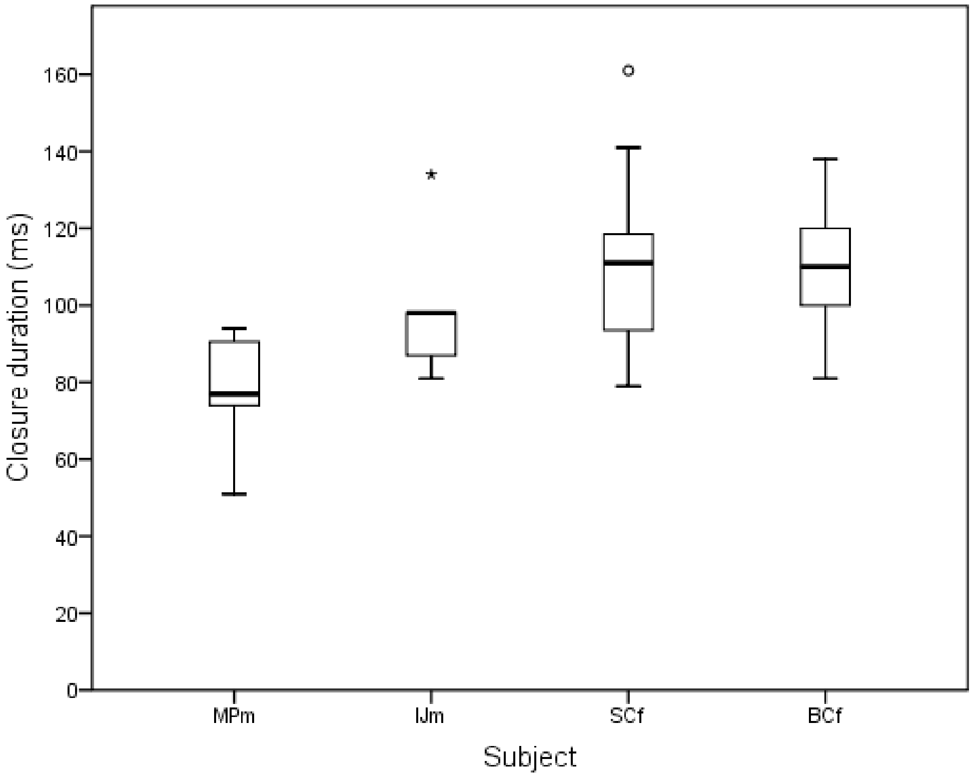

Figure 5.25 Boxplots showing the distribution of CD values for /p, t, k/ in word-final intervocalic position for each subject

### 5.11.2 Gender and age

In this section it was not possible to repeat the kind of analysis that was applied on CD in initial stops in the sentence frame. There are several reasons for this. First, the age span of the four subjects is small, and it is therefore unlikely that age could have an effect (for VOT the effect of age was present in the four oldest subjects). Second, all four subjects are among the faster speakers, so speaking rate, although measured for this sample of speech (in syllables/s), was not expected to have an effect. Third, it is not possible to perform an ANCOVA (with speaking rate as a covariate) because of unequal numbers or tokens at different levels of independent variables for gender and age. A preliminary investigation found that assumption of homogeneity of regression slopes for an ANCOVA was violated, so this test could not have been run.

However, it is possible to draw some tentative conclusions about possible effects of gender on CD production in this condition.

Female subjects produced longer closures for /b, d, g/ than did male subjects, with means of 69.33 ms (*N* = 45, *SD* = 10.39) and 58.26 ms (*N* = 23, *SD* = 13.63), respectively, which is a statistically significant difference: t-test, $t(66) = 3.73$, $p < 0.001$, Cohen's *d* = 0.92 (large effect).

Female subjects also produced longer closures for /p, t, k/, with means of 109.31 (*N* = 42, *SD* = 17.69) for the females and 83.07 ms (*N* = 30, *SD* = 15.51) for the males. These results differ significantly, according to a Mann-Whitney U-test, $Z = -5.53$, $p < 0.001$, $r = -0.65$ (large effect).

Figure 5.26 shows mean CD for phonologically voiced and voiceless stops for each subject as a function of their speaking rate from the slowest (IJm) to the fastest (SCf). There is no monotonic decrease in CD as speaking rate increases. The figure also shows that females have a larger separation between the means for the two categories. This can further be illustrated by their ratios in Figure 5.27. The two female subjects, SCf and BCf, have ratios of 0.63 and 0.64, while the two male subjects, MPm and IJm, have ratios of 0.69 and 0.73. Although these ratios do not differ to a large extent (and there is also no baseline from the previous research to suggest what should be considered as a large difference), the two female speakers not only produced longer closures but also have larger separation between the two categories. This is not likely to be an effect of speaking rate, nor of age differences (the order of age of subjects is MPm < SCf < BCf < IJm).

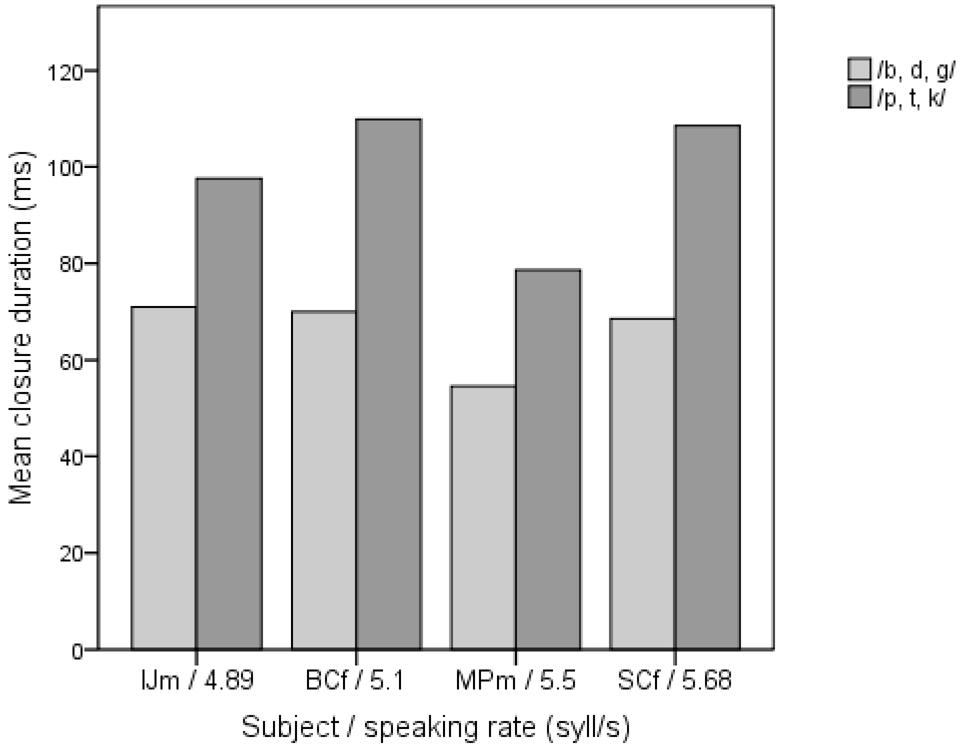

Figure 5.26 Mean CD (ms) for /b, d, g/ and /p, t, k/ in word-final intervocalic position as a function of individual speaking rate

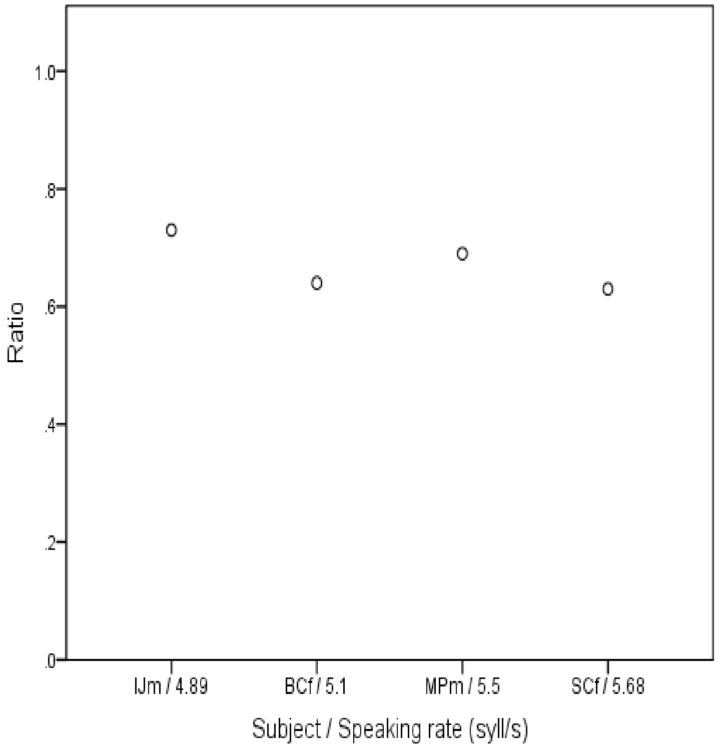

Figure 5.27 Ratio of mean CDs for phonologically /b, d, g/ and /p, t, k/ in word-final intervocalic position as a function of individual speaking rate

## 5.12 Summary of findings for closure duration in word-final intervocalic stops

In word-final intervocalic position, CD is also a correlate that is able to separate phonologically voiced stops and their voiceless cognates in Serbian. CD values for the two categories overlap, but the difference between their means of about 33 ms (or 33%) is significant, with large effect size. This is the case with results for each individual subject as well, where differences vary from 24 to 60 ms (or 27 to 37%).

A place of articulation effect, with longer closures at more forward places of articulation, is present in this sample as a tendency, in contrast to words in isolation, where it was absent. This could be due to different samples used in the two conditions.

The two female subjects produced longer closures than the two male subjects, and also achieved a larger distance between the two stop classes. There does not appear to be an effect of speaking rate on gender-related differences in CD, although this was not tested statistically. The effect of age on production of CD was not tested because of the relatively small age span across the sample.

The above findings, although limited to data from four subjects, generally support findings for word-final stops in isolated words for all twelve subjects.

Closures are shorter in the sentence frame than in isolated words, as could be expected because of word-final lengthening in isolation. Mean differences in CD between /p, t, k/ and /b, d, g/ are reduced in the sentence condition, compared to isolation, but while for the two female subjects this reduction was small (from around 50 ms to around 40 ms), for male subjects mean difference was more than halved (from roughly 50-60 ms to about 25 ms), and achieved by proportionally much larger reduction in the duration of /p, t, k/ closures than in the /b, d, g/ closures. As a consequence, separation of mean CD values expressed as ratio or percentage increased for female subjects in the sentence condition (by 3%) but decreased for male subjects (by nearly 10%). This suggests that female subjects BCf and SCf used more distinctive speech, as was the case with VOT production.

The mean ratio for word-initial stops for these four subjects is larger than that for word-final stops (both in the sentence condition). In other words, there is less durational difference in CD for initial stops than in final stops (18% vs. 33%). This result supports the finding that CD as a correlate of voicing is more important word-finally than word-initially, as reported by Stathopoulos and Weismer (1983) for English.

On the other hand, although both Abdelli-Beruh (2004) and Lousada et al. (2010) found that initial closures were longer than final ones in French and Portuguese, CD differences were similar in both positions.

## 5.13 Effect of phonological voicing category on closure duration in word-medial intervocalic stops

Measurement of CD in word-medial intervocalic position was not included in the design of the present study. However, it was measured in a subset of words, all minimal or near-minimal pairs uttered in isolation, which were used to measure the effect of stop voicing on preceding vowel duration, with structure (C)CVSV. Stops in this sample were syllable-initial in the second, unstressed, syllable of a disyllabic word. This position is often referred to as post-stressed intervocalic position and is frequently used for measuring CD differences in word-medial position.

The mean CD for this set of minimal pairs was 127 ms for /p, t, k/, and 73 ms for /b, d, g/ (with *SD* = 20.89, *N* = 78, and *SD* = 17.21, *N* = 78, respectively). This is a statistically significant difference, according to a t-test, *t* (77) = 24.16, *p* < 0.001 (2-tailed), Cohen's *d* = 2.82, large effect. Difference between the two means is 54 ms or 42% (or the ratio of 0.58). This finding replicates the results for other word positions, and is, in terms of percentage of CD, the largest difference found in the present study.

## 5.14 Discussion

**Closure duration as a correlate of voicing in Serbian**

In the present study phonologically voiceless stops were realised with significantly longer closures than phonologically voiced stops in all contexts that were investigated: in word-initial intervocalic position the difference between means was 27 ms or 22% (of the duration of the longer closure), in utterance-final position 60 ms or 37%, in word-final intervocalic position 33 ms or 33%, and in word-medial intervocalic position 54 ms or 42%. Differences in CD are consistently present not only across contexts, but also across subjects, and the effect size is large in all cases, except in the pooled data in word-initial position, where it is medium. Based on these results, it can be concluded that CD is a very relevant correlate of voicing in Serbian in all word positions.

These findings are in agreement with results from previous studies on voicing-related differences in CD reviewed in Chapter 1, but their consistency across contexts and the size of the effect challenge the assumption that CD as a correlate of voicing is more strongly associated with the fortis/lenis or tense/lax dimension (that is, with aspirating languages), and predominantly with word-final and word-medial position. In fact, both of these assumptions can be questioned based on evidence from Serbian and some other voicing languages.

Results for word-initial position in Serbian are in line with findings from other voicing languages, such as French (Abdelli-Beruh, 2004, 2009; Jacques, 1987), Portuguese (Lousada, et al., 2010), and Arabic (Flege & Port, 1981). Mean CD difference of 27 ms found in Serbian is comparable to that found in French by Abdelli-Beruh (2004), who found a 22 ms difference (or, calculated as a ratio, 0.75, i.e. 25% difference). Jacques (1987) and Abdelli-Beruh (2009) reported somewhat smaller mean differences for French (5-8 ms and 9-16 ms respectively), similar to those reported by Flege and Port (1981) for Arabic (9-10 ms). Portuguese results, on the other hand, are much higher, with the mean difference of about 55 ms (Lousada, et al., 2010).

However, results from several studies on English suggest that in word-initial position in sentence condition CD is not a reliable correlate of the voicing distinction. Docherty (1992) found that /p/ and /k/ closures were 4 ms longer than /b/ and /g/ closures, but for /t/ they were 3 ms shorter than /d/ closures. Stathopoulos and Weismer (1983) reported that in stressed position mean voiceless closures were shorter than the corresponding voiced closures, and no difference or the opposite result in unstressed position. In continuous speech, Crystal and House (1988a) and Umeda (1977) found a similar (small) range of differences and inconsistency in the direction of the effect across conditions and places of articulation.

Serbian results for word-final position are in agreement with findings from several languages, including English (Chen, 1970; Luce & Charles-Luce, 1985; Smith, et al., 2009; Stathopoulos & Weismer, 1983; Umeda, 1977; Wolf, 1978), French (Abdelli-Beruh, 2004), and Portuguese (Lousada, et al., 2010). Serbian results for CD in isolated words are in the same range as Chen’s (1970) results for English in the same condition - 58 ms difference in Serbian and 52 ms in English. In both cases this is about 37% difference, although closures were longer in Serbian than in English. In the sentence condition, difference of 33 ms in Serbian is equal to that found in English by Smith et al. (2009), while Stathopoulos and Weismer (1983) reported differences smaller than those in Serbian (and mostly no higher than 20 ms), as well as closures that were somewhat shorter than in Serbian. The same is true for the overall result reported by Luce and Charles-Luce (1985). CD difference in French is smaller (21 ms), and that in Portuguese larger (47 ms) in the sentence condition than it is in Serbian; the same relationship holds for mean CDs in these three languages.

Serbian results for word-medial position (in isolated words) are comparable to those reported by Lisker (1957) for the pair /p/-/b/ in post-stressed intervocalic position in English. The difference in means is slightly higher in Serbian (54 ms, or 42%) than in English (45 ms or 37%). Sharf (1962), however, found shorter closures for /p/ than /b/ in English, but the opposite relationship for the other two cognate stop pairs. Shorter closures and a smaller mean difference of 13 ms were measured by Stathopoulos and Weismer (1983) in medial post-stressed position in a sentence condition, which could explain this discrepancy.

Several other studies found significant differences in CD in word-medial post-stressed position in isolated words. Slis and Cohen (1969a) reported a mean difference of 28 ms for Dutch, while in Polish Keating (1985) found a 38 ms difference in mean CD for /t/ and /d/. In German, Jessen (1998) found that intervocalic voiceless stops in German have longer closures, by 26 ms. Word-medially in the sentence frame Lousada et al. (2010) found that /p, t, k/ closures were 46 ms longer (42%) in Portuguese, which is in the same range as the Serbian results (in isolated words), but they did not perform any statistical analysis. In German, Fuchs (2005) found significantly longer closure for /t/ than for /d/ in post-stressed position (but the opposite in stressed position).

Very few studies investigated the voicing effect on CD in all three positions within the word. Lousada et al. (2010) found larger difference in percentage of CD word-medially than word-initially and word-finally, as is the case in the present study. Stathopoulos and Weismer (1983), on the other hand, found bigger effect word-finally than word-medially, but word-initially it was inconsistent across conditions.

The fact that Serbian results reported here are consistent with patterns seen across languages, points to a somewhat different role of CD as a correlate of the voicing contrast than previously thought. CD is a relevant correlate in voicing languages, with differences in CD comparable to those in aspirating languages (despite the lack of uniformity in methodology used between studies), and a correlate that is present in all word positions, unlike in English, where word-initially it does not appear to be very relevant. These findings pose a challenge for theoretical models proposed by Kingston and Diehl and Jessen because these two models include CD as one of the key correlates but underestimate its role in voicing languages (this issue is further discussed in Chapter 8).

The reason behind voicing-related difference in CD is not often discussed in the literature. According to one view, CD difference could be explained in terms of physical and physiological constraints during the production of the voicing contrast. Namely, because of the oral pressure build-up during the stop closure, the trans-glottal pressure difference necessary to maintain vocal fold vibration cannot be sustained for a long time, which makes voiced closures shorter than voiceless ones (Fuchs, 2005; Ohala, 2011; Ohala & Riordan, 1979; Pickett, Bunnell, & Revoile, 1995).

This explanation was criticized by Kluender et al. (1988), who argue, following findings by and Ohala and Riordan (1979) and Westbury (1983), that passive vocal tract enlargement can only account for voicing of 50 to 100 ms, but that speakers use active enlargement in voiced stop production. This argument reinforces the view expressed in the present study that active maneuvers for voicing must be taken into consideration when discussing prevoiced stops. Kluender et al. suggest that the voicing effect on CD differences is not an automatic result of physical constraints in production. Although they do not explicitly offer any other explanation for it, Kluender et al. argue that language communities choose to exploit certain durational differences in order to enhance phonological contrasts, and that “ the closure-duration correlate has in part a perceptual rationale” (1988, p. 166).

Another shortcoming of the above explanation is that it does not take into account whether closure of phonologically voiced stops is actually fully voiced or not, and if it is not, to what extent is the voicing present during the closure. While this might not be an issue word-medially between vowels, in word-final position, and especially in utterance-final position, the extent of closure voicing might not be correlated with the measured CD. Unfortunately, studies that have reported CDs rarely commented on the extent of closure voicing in phonologically voiced stops. More data for a large number of languages would help to gain a better understanding of this issue and whether both physiological and speaker-controlled factors determine voicing-related CDs.

### Effect of place of articulation on CD

As mentioned in the literature review in Chapter 1, a place of articulation effect on CD has been reported in a number of studies, but findings are diverse and often inconsistent. Differences in experimental conditions, number of subjects and general methodology could, at least in part, explain discrepancies in results regarding the presence or absence of the place effect on CD and its relationship to the voicing contrast, as well as its magnitude and direction. The majority of studies suggest that labials have the longest closures.

The place of articulation effect found in Serbian word-initially, where the order of CD is bilabial > dental > velar, was not found in English, except by Umeda (1977). Flege and Port (1981) observed this tendency in Arabic, and it is also present in Portuguese (Lousada, et al., 2010), but neither study provided statistical analysis of this effect.

In Serbian, a place effect is absent in word-final stops in isolated words, but present as a non-significant tendency in the sentence condition, where CD decreases in order bilabial > dental > velar (although only data for four subjects were available). In word-initial stops, however, this effect was statistically significant, with the order /b/ > /d/, /g/ and /p/, /t/ > /k/. Several other studies found different patterns for /p, t, k/ and /b, d, g/. Esposito (2002), for example, reported the following significant differences in Italian: /p/ > /t/ > /k/ and /b/ >/d/ = /g/, the same as Byrd (1993) for English.

There are hardly any explanations for place-related differences in CD. The above-mentioned articulatory explanation for voicing-related differences could also account for the finding of some studies that CD of voiced stops decreases with more posterior place of articulation. In stops produced with larger cavity volume and with bigger surface area available for passive vocal tract expansion, trans-glottal pressure drop is slower and as a result voicing can be maintained for longer, as was discussed in relation to maintenance of prevoicing and its duration in Chapter 4 (Keating, 1984b; Ohala, 1983; Ohala & Riordan, 1979; Westbury, 1983). However, this explanation assumes that all voiced closures are fully voiced, which is not necessarily the case. In addition to this, this does not explain why the same relationship would be found in voiceless stops, unless it is maintained to parallel that in voiced stops. Another problem is that not all studies reported the same order for voiced and voiceless stops, including the present study. Few studies actually found CD to decrease in order labial > alveolar/dental > velar, which leads to the conclusion, expressed by Abdelli-Beruh (2009), following Docherty (1992), that place effect on CD “could well be attributed to language-specific processes” (p. 68). This view can be seen as supporting Kluender et al.’s (1988) argument that CD production is partly under speaker control, although it does not necessary support their idea of auditory enhancement as such.

### Effect of gender and age on CD

As for the other effects on CD, the present study found that female subjects produce longer closures for both stop classes, which is largely independent from differences in speaking rate between women and men. This is in agreement with Zue and Laferriere (1979), who reported that women in their study produced longer segments. Their explanation, which is very likely to hold for the present study, is that in careful speech women generally prefer correct forms of pronunciation.

In sum, CD is a relevant correlate of the voicing distinction in Serbian in initial, final and medial stops. Voiceless stops have longer closures than voiced stops, and despite overlap this difference is statistically significant in each word position. Furthermore, the present study found that stops are consistently released in Serbian, even in word-final position, unlike in English, for example. This result is in agreement with results for French, where final stops are also frequently released (Laeufer, 1992). Laeufer argued that word-finally in French the voicing contrast is expressed mostly through the properties of the stop, such as CD, presence or absence of voicing in the closure, consistent releases and frequent vocalic releases, as opposed to English, where preceding vowel duration is very important in signaling the voicing value of the stop, and stops are often unreleased and partially devoiced. A pattern similar to that found in French seems to be present in Serbian as well. The next chapter presents results for voicing in the closure in Serbian, and how it relates to CD.

# Chapter 6 Results for voicing in the closure

## 6.1 Voicing in the closure in word-initial intervocalic stops

### 6.1.1 Effect of the phonological voicing category on voicing in the closure

In word-initial intervocalic position (in the sentence condition) majority of phonologically voiced stops were realised with fully voiced closures, where voicing continued unbroken from the previous vowel into the stop closure and then into the following vowel. There were only eight tokens out of 168 (or 4.8%) with a voiceless interval during the closure. This interval occurred at the end of the closure when voicing subsided for a short period just before the burst. The mean duration of this voiceless interval (for eight tokens) is 22.5 ms. These eight tokens were produced by the following subjects: MVf (1 token), MCf (1), DARf (4), IVm (1) and IJm (1), and except IJm they all belong to the group of subjects who produced longer closures (in fact, subjects MCf, DARf and MVf produced the longest closures of all subjects). This is not unexpected, because the longer the closure, the more difficult it is to sustain voicing. Out of these eight tokens, there was one /b/ token, one /d/ token and six /g/ tokens, a distribution which is consistent with findings that it is more difficult to sustain voicing in velars than in stops produced at other places of articulation (Keating, 1984b; Ohala, 1983; Ohala & Riordan, 1979).

Phonologically voiceless stops were realised as voiceless, either with completely silent closures (39 tokens out of 172 or 22.67%), or with some voicing carried over from the previous vowel, usually for a few cycles (133 tokens). This carry-over voicing has lower amplitude than voicing in the closure of voiced stops. The range of values for carry-over voicing is 5 to 36 ms, and its mean duration is 17.15 ms.

Durations of voicing in the closure for both stop classes are presented in boxplots in Figure 6.1 (in the pooled data). There is no overlap between the two categories. Results for mean durations of voicing in the closure in phonologically voiced stops replicate results for mean CD measured in the same condition, reduced proportionally by the shorter voicing of the eight tokens with incomplete voicing. Mean duration of voicing in the closure for voiced stops is 95.13 ms (*N* = 168, *SD* = 23.01), and for voiceless stops 13.26 ms (*N* = 172, *SD* = 9.42). Difference between their means is 81.87 ms. For individual speakers differences between means vary from 62.3 ms (MPm) to 108.59 ms (MVf). A summary of results for each subject is given in Table C5 in Appendix C.

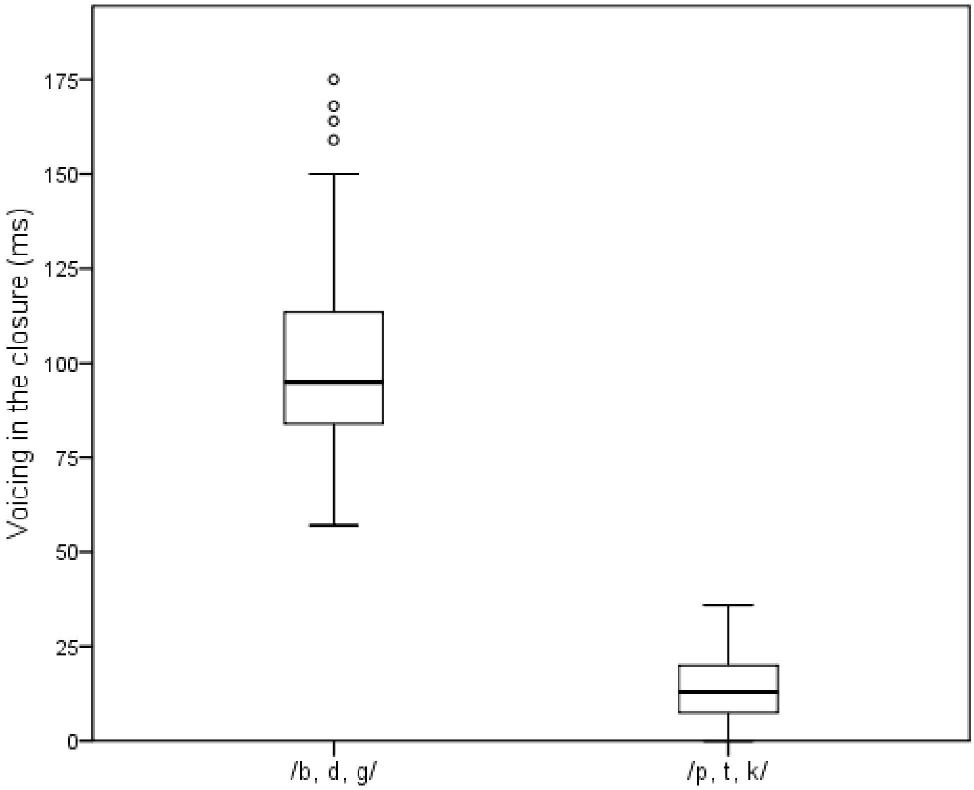

Figure 6.1 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for word-initial intervocalic stops

When duration of voicing in the closure is calculated as a percentage of CD for each token, voiced stops have 99.26 % of their closures voiced, while for voiceless stops it is 10.28% (Figure 6.2).

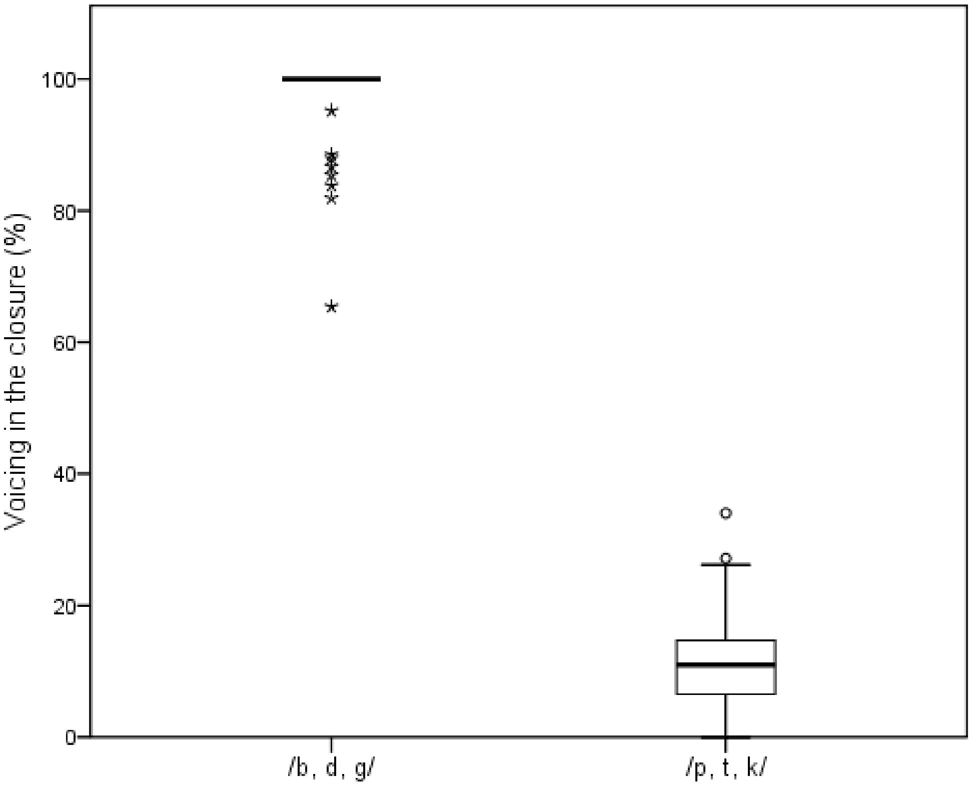

Figure 6.2 Boxplots showing the distribution of values for duration of voicing in the closure (%) for word-initial intervocalic stops

### 6.1.2 Effect of place of articulation on voicing in the closure

Boxplots showing the distribution of voicing duration at each place of articulation for initial /b, d, g/ are given in Figure 6.3 and their means in Table 6.1. Mean durations of voicing reflect mean CDs measured in the same condition, and the effect of place on duration of voicing in the closure is significant (one-way ANOVA, $p < 0.001$, $F(2, 165) = 17.66$, $\omega^2 = 0.166$, large effect), as was the case with CD in Section 5.2.1. A Tukey post-hoc test revealed that duration of voicing is significantly longer in /b/ than in /d/ ($p = 0.001$) and in /b/ than in /g/ ($p < 0.001$).

Mean percentages of voicing in the closure at each place of articulation are similar, and close to 100%, because most of the stops are fully voiced: 99.79% for /b/, 99.78% for /d/, and 98.24% for /g/. There are no individual differences concerning duration of voicing in the closure, since majority of /b, d, g/ closures are fully voiced.

| | | Mean voicing in the closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 108.13 | 54 | 24.42 |
| Dental | /d/ | 93.22 | 57 | 19.96 |
| Velar | /g/ | 84.63 | 57 | 18.38 |
| Total | | 95.13 | 168 | 23.01 |

Table 6.1 Mean duration of voicing in the closure (ms), N and SD for /b, d, g/ in word-initial intervocalic position for each place of articulation

No other factors, either linguistic or speaker-related, were investigated for stops in word-initial position, because the results would repeat findings obtained for CD in this condition.

Results for duration of voicing in the closure for /p, t, k/ in word-initial intervocalic position are presented in Table 6.2 and Figure 6.3. Because place-related differences in voicing duration are very small and because of the skewed nature of the data (many values are zero or close to zero), no statistical analysis was performed to establish the effect of place of articulation on duration of voicing in the closure. Mean percentage of voicing in the closure is 12.24% for /p/, 11.28% for /t/, and 7.35% for /k/. On average, only the first 10.28% of the closure is voiced.

| | | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 16.88 | 58 | 9.98 |
| Dental | /t/ | 14.11 | 56 | 8.73 |
| Velar | /k/ | 8.83 | 58 | 7.7 |
| Total | | 13.26 | 172 | 9.42 |

Table 6.2 Mean duration of voicing in the closure (ms), N and SD for /p, t, k/ in word-initial intervocalic position for each place of articulation

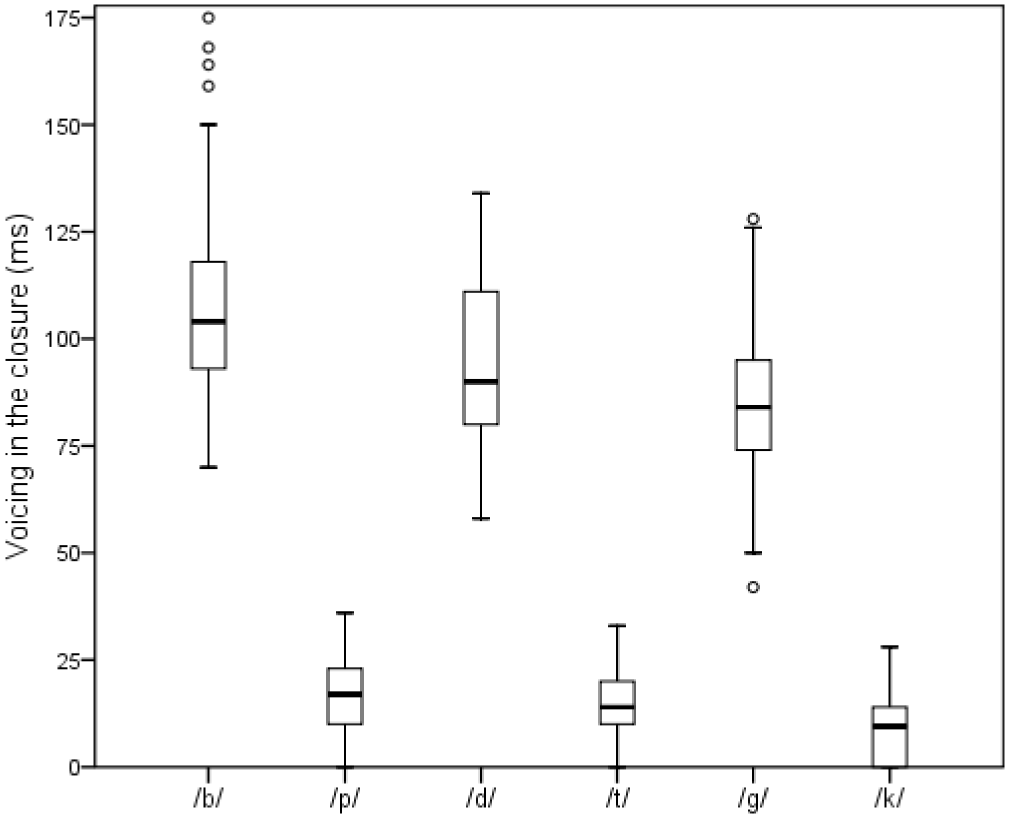

Figure 6.3 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for stops in word-initial intervocalic position as a function of stop place of articulation

### 6.1.3 Summary of findings

Duration of voicing in the closure clearly separates phonologically voiced and voiceless Serbian stops in word-initial intervocalic position. While phonologically voiced stops are realised with closures that are mostly fully voiced, phonologically voiceless stops are realised with a short period of carry-over voicing, which is lower in amplitude and occupies on average 10% of CD. Difference in duration of voicing in the closure is statistically significant in the pooled data, and the effect is large.

This result reinforces the finding for word-initial stops in isolated words, where all voiced stops were prevoiced, and all voiceless stops were produced without voicing during the closure.

There is a place effect on duration of voicing in the closure in phonologically voiced stops. Duration of voicing decreases in order bilabial > dental > velar, and is significantly longer in /b/ than in /d/ and /g/.

Women in the present study produce longer closures and consequently longer duration of voicing in the closure than men, and so do older male subjects compared to younger male subjects.

## 6.2 Voicing in the closure in utterance-final stops

### 6.2.1 Effect of phonological voicing category on voicing in the closure

In utterance-final position (in isolated words) phonologically voiced stops were realised with fully voiced closures in 51 out of 282 tokens (18.09% of tokens). The remaining 231 tokens (81.91%) were realised with partially voiced closures. The range of values of incomplete voicing is 7 to 104 ms. Mean duration of voicing in the pooled data is 64.43 ms.

In the same condition phonologically voiceless stops were realised either with no voicing at all (88 out of 287 tokens or 30.66% of tokens), or with voicing that continued from the preceding vowel into the closure for some time (199 out of 287 tokens or 69.34%). The range of values for carry-over voicing is in 0 to 40 ms range, but mostly below 20 ms. Mean duration of voicing in the pooled data, including instances of zero voicing, is 10.17 ms.

Distributions of results for both stop classes are presented in Figure 6.4 and their means in Table 6.3. Phonologically voiced stops are realised with significantly longer periods of voicing than phonologically voiceless stops, according to a Mann-Whitney U-test: $p < 0.001$ (2-tailed), $Z = -19.5$, $r = -0.82$, large effect. The same is true for each of the twelve subjects, and the effect size is large in all cases. A summary of results for each subject is given in Table C6 in Appendix C. Difference between means for the two categories is 54.26 ms. Difference in means for individual subjects varies from 13.5 ms (MCf) to 99.46 ms (MRf).

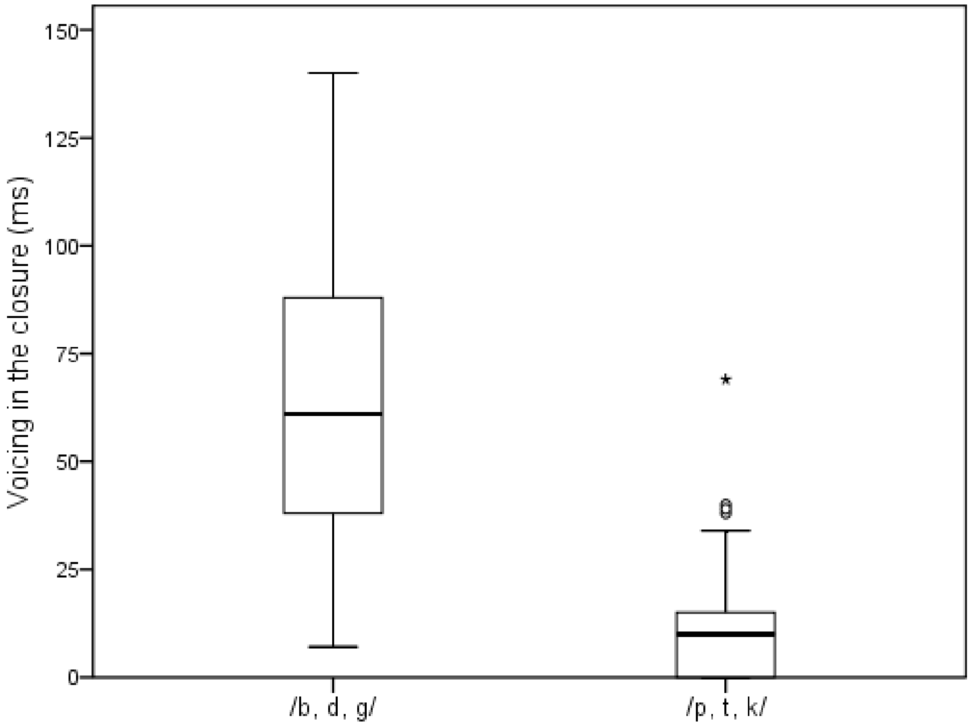

Figure 6.4 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for utterance-final stops

| | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|
| /b, d, g/ | 64.43 | 282 | 31.25 |
| /p, t, k/ | 10.17 | 287 | 9.55 |

Table 6.3 Mean duration of voicing in the closure (ms), N and SD for utterance-final stops

The percentage of voicing in the closure is shown in boxplots in Figure 6.5 and their means in Table 6.4. Overall, about 62% of closure in /b, d, g/ tokens in this condition was voiced, as opposed to only 6.5% of carry-over voicing in /p, t, k/ tokens. This difference is statistically significant in the pooled data (Mann-Whitney U-test, $p$ < 0.001, 2-tailed, $Z$ = -20.3, $r$ = -0.85, large effect), as well as in data for each speaker ($p$ < 0.001, and large effect for all speakers).

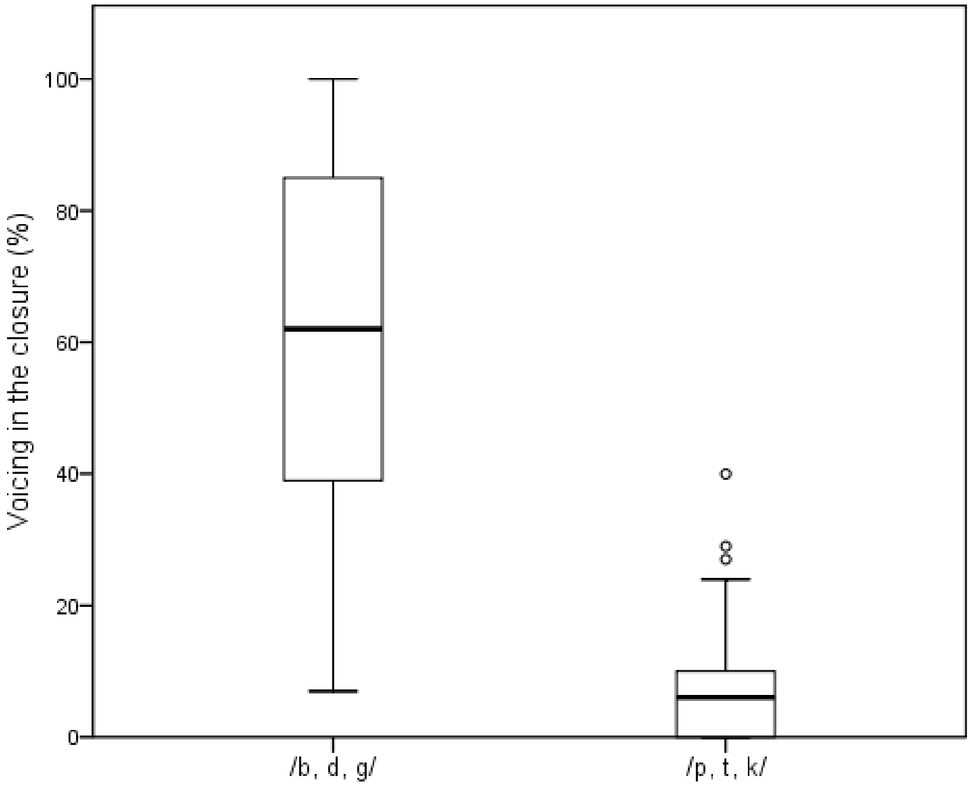

Figure 6.5 Boxplots showing the distribution of values for duration of voicing in the closure (%) for utterance-final stops

| | Mean voicing in the closure (%) | N | SD |
|---|---|---|---|
| /b, d, g/ | 61.84 | 282 | 27.17 |
| /p, t, k/ | 6.45 | 287 | 6.08 |

Table 6.4 Mean duration of voicing in the closure (%), N and SD for utterance-final stops

### 6.2.2 Effect of place of articulation on voicing in the closure

Distributions of results for /b, d, g/ at each place of articulation are presented in Figure 6.6 and their means in Table 6.5. There is very little difference in duration of voicing between the three places of articulation, which was confirmed by a non-significant result of a one-way ANOVA, $p$ = 0.21, $F$(2,279) = 1.56.

| | | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 68.09 | 92 | 31.34 |
| Dental | /d/ | 65.14 | 96 | 31.83 |
| Velar | /g/ | 60.12 | 94 | 30.36 |
| Total | | 64.43 | 282 | 31.25 |

Table 6.5 Mean duration of voicing in the closure (ms), N and SD for /b, d, g/ in utterance-final position for each place of articulation

When voicing in the closure is expressed as a percentage of the corresponding CD, there is still very little difference between the means for the three stops: 64.39% for /b/, 61.2% for /d/ and 59.86% for /g/, which is not significant, Kruskal-Wallis test, *p* = 0.54, $\chi^2(2,282) = 1.25$.

Results for /p, t, k/ in utterance-final position are presented in Table 6.6 and Figure 6.6. Expressed as percentage, /p/, /t/, and /k/ have 7.1%, 6.5%, and 5.7% of closure voiced, respectively.

The effect of the preceding vowel on duration of voicing in the closure was not investigated because neither phonological length nor quality of the vowel preceding final stops was controlled, and as a consequence they are represented with different numbers of tokens.

| | | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 11.43 | 95 | 9.18 |
| Dental | /t/ | 10.25 | 96 | 10.14 |
| Velar | /k/ | 8.85 | 96 | 9.22 |
| Total | | 10.17 | 287 | 9.55 |

Table 6.6 Mean duration of voicing in the closure (ms), N and SD for /p, t, k/ in utterance-final position for each place of articulation

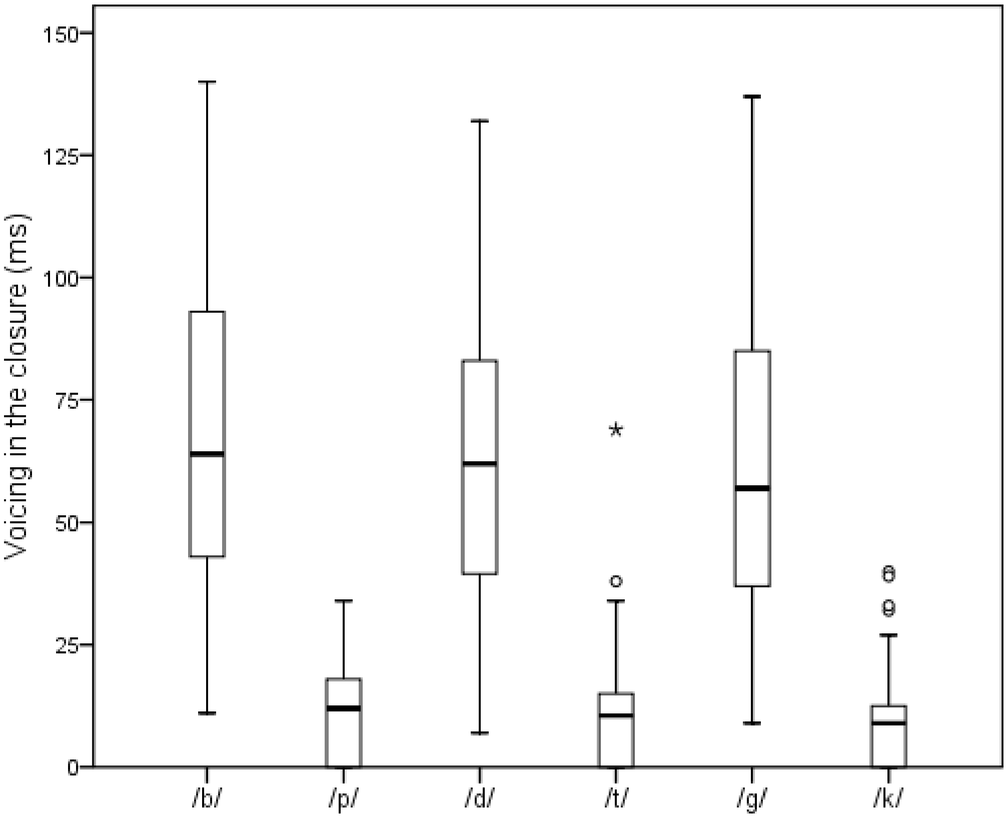

Figure 6.6 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for stops in utterance-final position as a function of stop place of articulation

### 6.2.3 Speaker factors affecting voicing in the closure

**Individual differences between subjects**

In utterance-final position, there are individual differences between subjects in the number of /b, d, g/ tokens produced with fully voiced closures, as well as in the duration of voicing in tokens with broken voicing. As far as tokens with fully voiced closures are concerned, one third of them (33.33%) came from subject RVm, who produced 17 out of 24 words with fully voiced closures, and the rest of tokens came from seven subjects: MVf (2), BCf (1), DARf (5), MRf (6), IVm (6), BPm (4), and MPm (9). Four subjects did not produce any fully voiced closures (MCf, SCf, DRm, IJm).

Figure 6.7 shows mean CD and mean duration of voicing in the closure in /b, d, g/ for each subject, and illustrates differences between subjects, both in terms of mean duration of voicing in the closure in milliseconds, and in terms of the proportion of the closure that is voiced. The subjects are ordered from the subject with the shortest mean CD to the subject with the longest mean CD. It is generally more difficult to sustain voicing in longer closures, but this does not seem to be the reason behind between-speaker variation in this sample of speech. For example, subject MRf, who produced the longest mean CD, had a large proportion of it voiced (80%), and also had the longest mean duration of voicing. What is more, subjects with similar mean CD vary greatly in the duration of the voiced portion, such as subjects RVm, IJm, SCf, IVm and MCf, or subjects BPm, BCf, DARf and MVf. This result suggests that speakers have some control over duration of voicing in the closure that they produce.

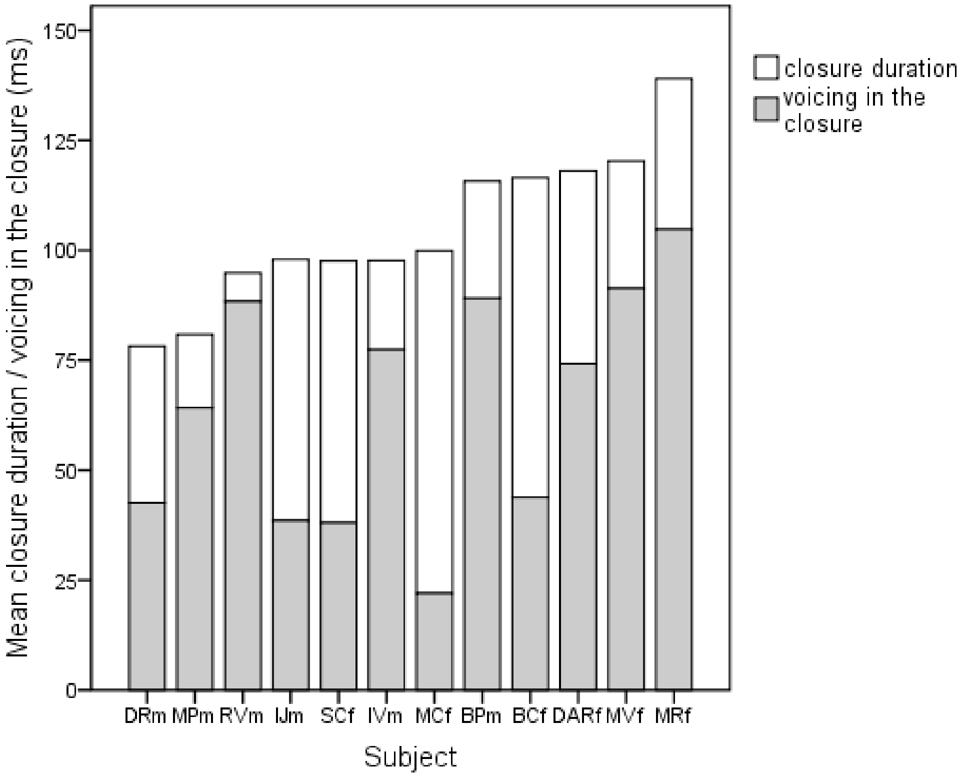

Figure 6.7 Mean CD (ms) and mean voicing in the closure (ms) for /b, d, g/ in utterance-final position for each subject

When expressed as a percentage of CD, duration of voicing in /b, d, g/ tokens varies from subject to subject, as illustrated by Figure 6.8 (subjects are ordered as in Figure 6.7, from the shortest mean CD to the longest). Subject MCf has the lowest mean percentage of voicing (22%). Three subjects have on average around 40% of the closure voiced (BCf, SCf, and IJm), and two subjects have more than half of the closure voiced (DRm 56% and DARf 64%). Five subjects have on average around 80% of closure voicing (MVf, MRf, BPm, IVm, and MPm). Finally, subject RVm has on average 94% of the closures voiced (which includes 17 tokens he produced with fully voiced closures). The figure also illustrates that there is no inverse relationship between percentage of voicing in the closure and CD, which suggests that some other factors are involved.

A CART analysis divided subjects into three groups with significantly different duration of voicing in the closure: Group 1, subjects MCf, SCf, BCf, DRm, and IJm (mean = 36.91 ms, *N* = 118, *SD* = 16.23), Group 2, subjects DARf, IVm, and MPm (mean = 72.07 ms, *N* = 71, *SD* = 19.53), and Group 3, subjects MVf, MRf, RVm, and BPm (mean = 93.51 ms, *N* = 93, *SD* = 22.05).

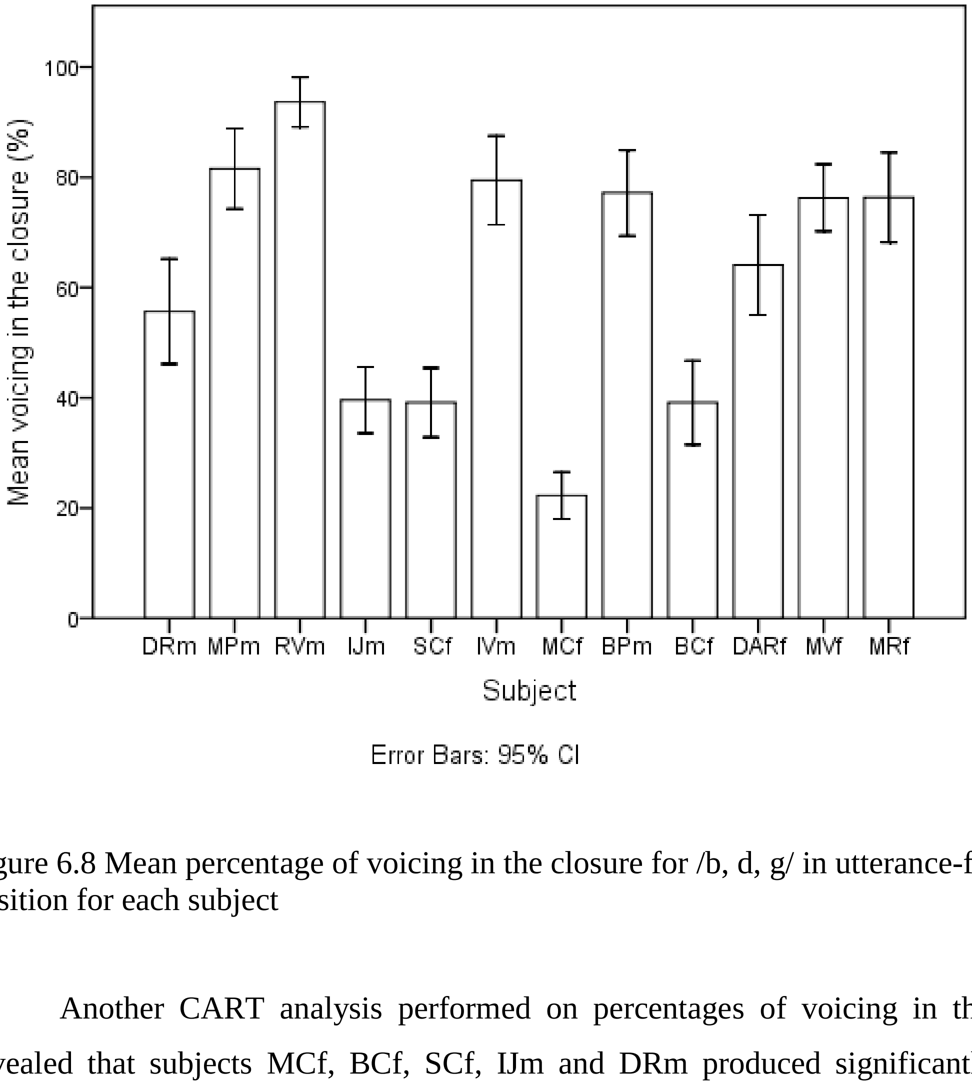

Figure 6.8 Mean percentage of voicing in the closure for /b, d, g/ in utterance-final position for each subject

Another CART analysis performed on percentages of voicing in the closure revealed that subjects MCf, BCf, SCf, IJm and DRm produced significantly shorter percentage of voicing than the rest of the subjects, with mean voicing of 38.9 % (*N* = 118) and 78.4% (*N* = 164), respectively. Except subject BCf, they come from the same town, which suggests that regional differences in the realisation of closure voicing might be present.

Figure 6.9 presents mean CD and mean duration of voicing in the closure for /p, t, k/ for each subject. Irrespective of their mean CD, all subjects produced voicing of similar duration, between 5 ms (MRf) and 18 ms (MVf). This represents less than 10% of CD in all cases.

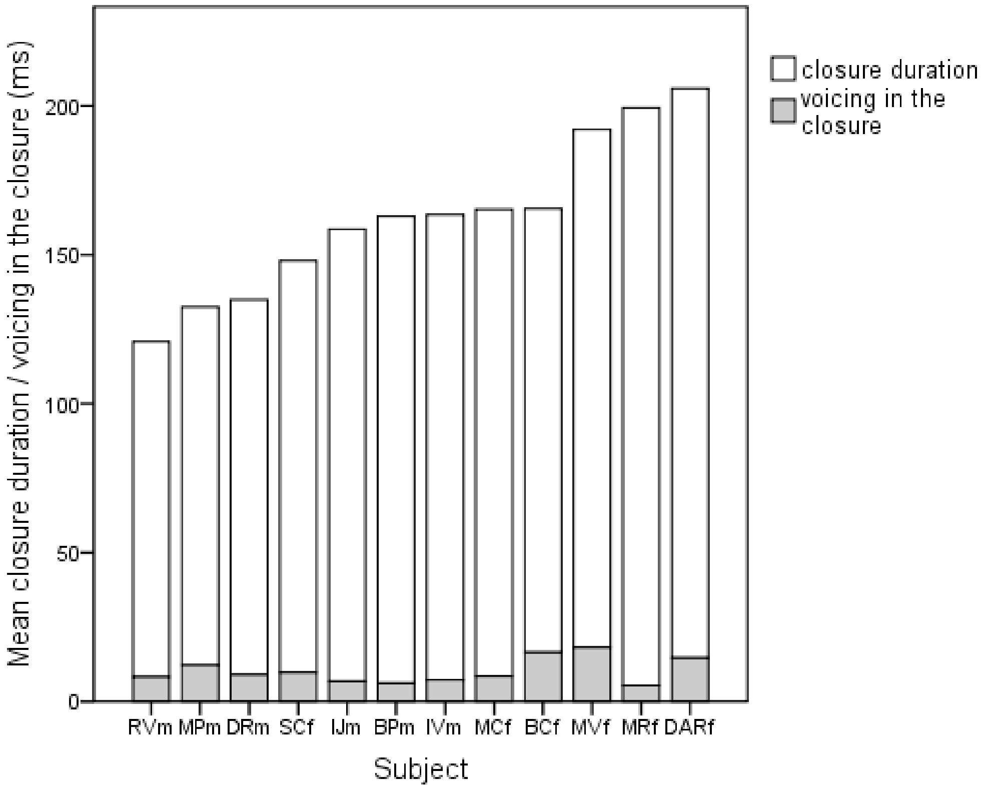

Figure 6.9 Mean CD (ms) and mean voicing in the closure (ms) for /p, t, k/ in utterance-final position for each subject

**Gender and age**

A two-way ANOVA was carried out to examine the effect of age and gender on duration of voicing in the closure in /b, d, g/ (subjects were divided into two groups, as before). The main effect of age was significant: $p < 0.001$, $F(1, 278) = 29.44$, $\omega^2 = 0.085$, medium effect, but the effect of gender was not, $p = 0.2$, $F(1, 278) = 1.69$. There was a significant interaction between gender and age, $p < 0.001$, $F(1, 278) = 23.38$, $\omega^2 = 0.067$, medium effect, Figure 6.10.

Younger speakers as a group produced shorter voicing in the closure than older speakers did: mean for younger speakers = 55.02 ms, *N* = 141, *SD* = 27.58; mean for older speakers = 73.83 ms, *N* = 141, *SD* = 31.95.

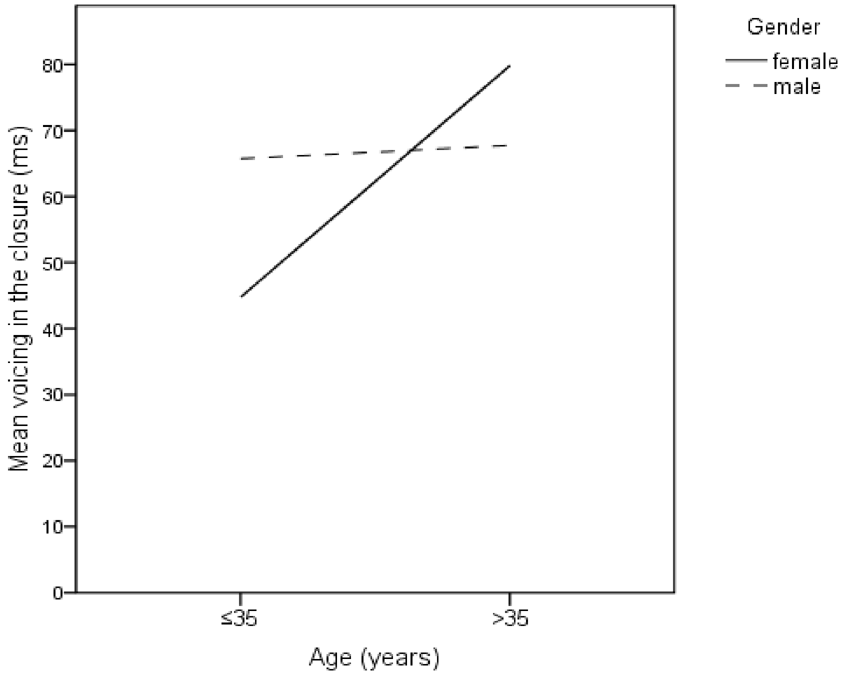

Figure 6.10 Mean duration of voicing in the closure (ms) for /b, d, g/ in utterance-final position as a function of gender and age of subjects

The interaction between age and gender is further illustrated in Figure 6.11, which presents data for each speaker (females on the left, males on the right, ordered from youngest to the oldest within each group). The significant effect of age is likely to be caused by the fact that four out of six older subjects (MVF, MRf, BPm, and IVm) produced longer voicing in the closure than most younger subjects (with the exception of RVm). Apart from this, there is no clear-cut effect of age in either gender group. The difference between younger and older men is very small (mean voicing duration for younger men = 65.72 ms, *N* = 69, *SD* = 23.28; mean for older men = 67.74 ms, *N* = 70, *SD* = 29.85). There is a 35 ms difference between younger and older women (mean voicing duration for younger women = 44.76 ms, *N* = 72, *SD* = 27.54; mean for older women = 79.83 ms, *N* = 71, *SD* = 33.02), but the effect is coming from the fact that two oldest women, MVf and MRf, produced longer voicing in the closure compared to other female subjects.

It is interesting to note from this analysis that females produced voicing in the closure with duration comparable to that produced by males or longer, although it is often hypothesized that it is more difficult for women to sustain voicing. The same has been suggested for older speakers, but this does not hold for the oldest speakers in this study, both male and female, who, in fact, have some of the longest mean voicing durations in this context.

Some of these differences could be caused by differences in speaking rate. A separate analysis was performed on percentage of voicing in the closure, because any effect of speaking rate is minimized by using this measure.

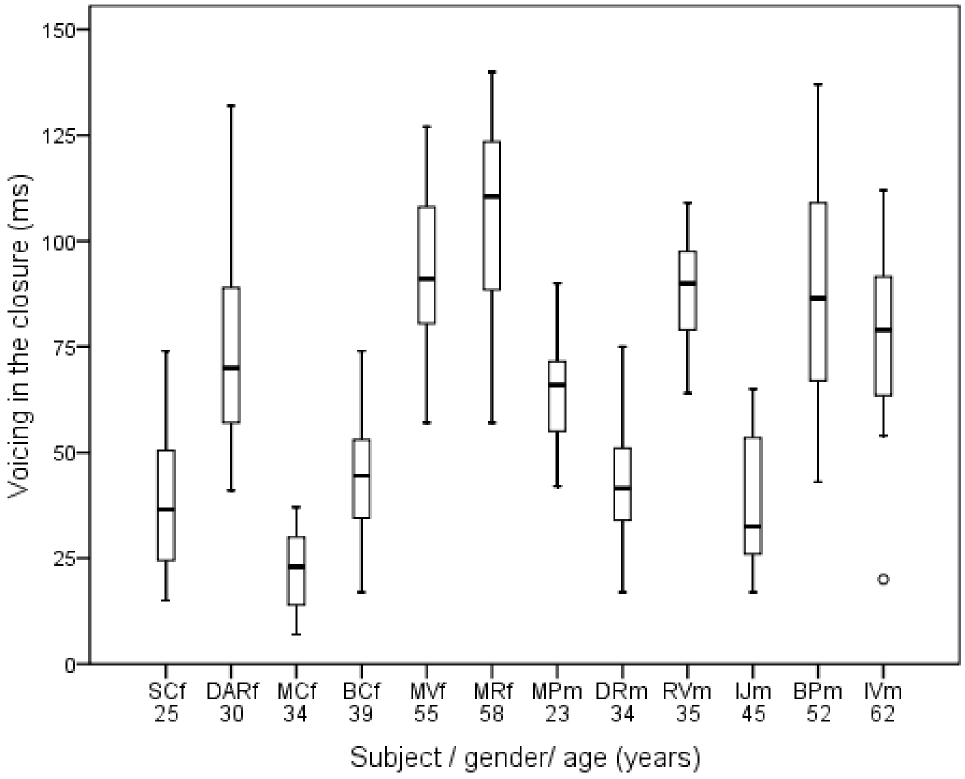

Figure 6.11 Boxplots showing the distribution of values for voicing in the closure for /b, d, g/ in utterance-final position as a function of speaker gender and age

When percentages are analyzed, the difference between male and female speakers is statistically significant, Mann-Whitney U-test, *Z* = -5.7, *p* < 0.001 (2-tailed), *r* = -0.34 (medium effect). Male subjects have longer mean percentage of voicing in the closure than females (the mean for males = 72.12 %, *SD* = 24.7, *N* = 139; the mean for females = 52.7%, *SD* = 35.02, *N* = 143). This finding could be related to the fact that females tend to produce longer closures. On the other hand, there is no significant difference between younger and older speakers, Mann-Whitney U-test, $Z$ = -1.49, $p$ = 0.14 (2-tailed). Figure 6.12 shows the interaction between gender and age: while there is almost no difference in percentage of voicing in the closure between older male and female subjects, there is a large gender-based polarisation within the ≤35 group, where younger male subjects have nearly twice as much of the closure voiced as younger female subjects. This is also illustrated in Figure 6.13, which shows data for percentage of voicing in the closure as a function of age and gender of subjects.

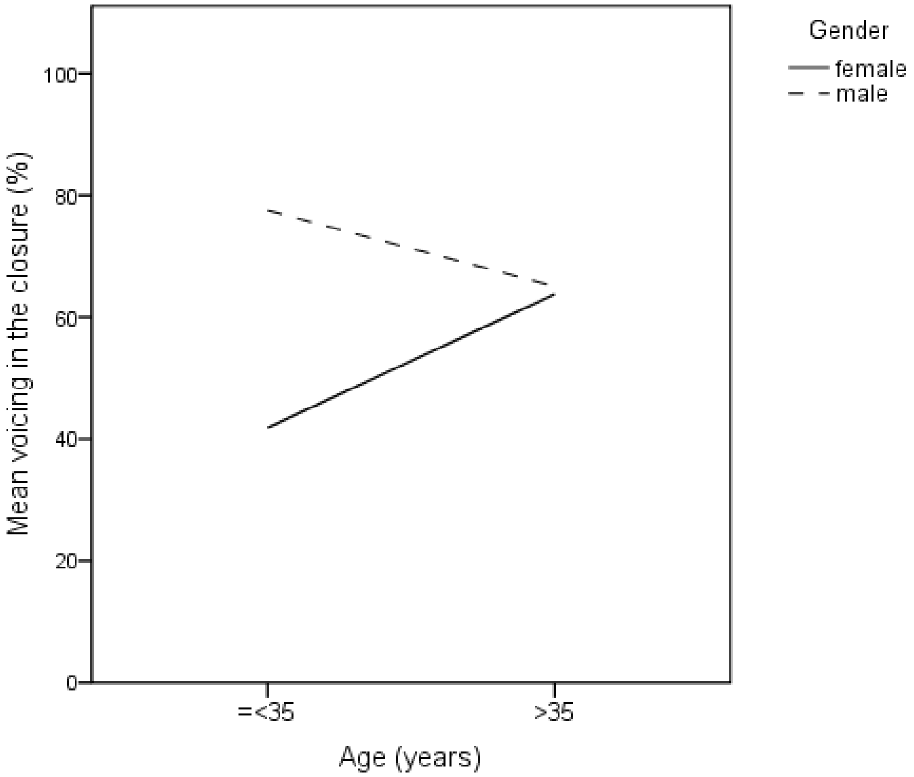

Figure 6.12 Mean duration of voicing in the closure (%) for /b, d, g/ in utterance-final position as a function of gender and age of subjects

In the female group, results for percentage of voicing in the closure replicate those for absolute duration of voicing in the closure, where two oldest subjects, MVf and MRf, and to a smaller extent DARf, stand out from the rest of the subjects (Figure 6.13). In the male group, younger males produced more voicing as a group, compared to older subjects, but they also had shorter closures (see Chapter 5). This difference is not large, and in fact, the effect of age is not obvious in the male group (Figure 6.13).

The age differences also interact with some other individual factors. There is a possible effect of place of birth. Two out of three female subjects with the lowest duration of voicing and the lowest percentage of voicing are from Čačak (SCf and MCf), as well as the male subject with the lowest duration and percentage of voicing, IJm and DRm. On the other hand, younger subject RVm, and to some extent DARf and MPm, produced voicing duration that is in the range with that of the oldest subjects, which is likely to be an individual feature.

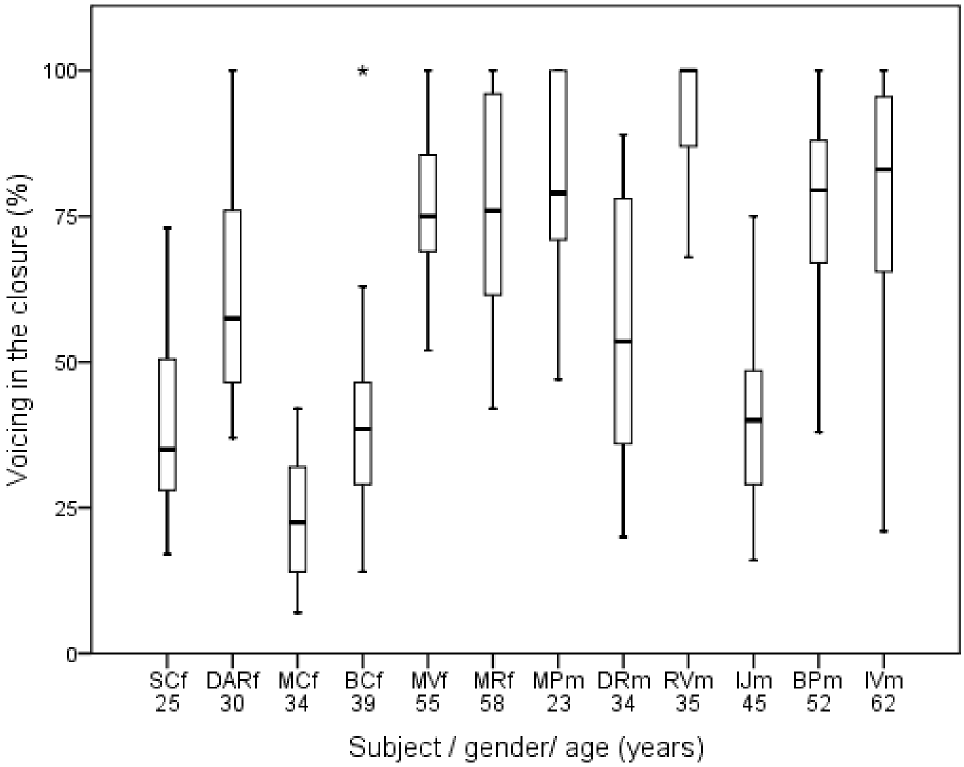

Figure 6.13 Boxplots showing the distribution of values for voicing in the closure (%) for /b, d, g/ in utterance-final position as a function of speaker gender and age

### 6.2.4 Summary of findings

In utterance-final position the contrast between phonologically voiced stops and their voiceless cognates is well maintained in Serbian. An average of 62% of the closure is voiced in /b, d, g/ tokens, while in /p, t, k/ tokens it is about 7%. In addition to this, 18% of voiced stops were fully voiced. Difference in absolute duration of voicing in the closure is significant in the pooled data and in data for each subject, as is difference in percentage of closure that is phonetically voiced, and the effect size is large.

There is no effect of place of stop articulation on duration of voicing in /b, d, g/ in this condition.

Individual differences in percentage of closure that is voiced in /b, d, g/ tokens are large, and vary from 22% to 94%. These differences are only partly attributable to differences in CD. On the other hand, all subjects have less than 10% of closure voiced in voiceless stops.

Gender and age of subjects also have an effect on voicing duration, but they interact with each other, and possibly with some other factors. Older speakers produce voicing of longer absolute duration than younger subjects, by about 19 ms, but this difference mainly comes from the fact that two oldest male and two oldest female subjects (aged above 52 years) have longer duration of voicing in closure than the rest of the subjects.

Men have on average 19% more of the closure voiced than women, because women in this study produced longer closures (Chapter 5). In terms of absolute duration, however, the two groups are similar.

## 6.3 Voicing in the closure in word-final intervocalic stops

### 6.3.1 Effect of the phonological voicing category on voicing in the closure

For this part of the study the same material was used as in Section 5.9 (CD of word-final stops in the sentence frame). A large number of tokens were discarded because there was a pause after the target word. The analysis that follows is based on data from four subjects, BCf, SCf, MPm, and IJm, who produced more than five tokens of /b, d, g/ and /p, t, k/ in intervocalic position.

In the sentence frame, final /b, d, g/ were realised with voicing that continued unbroken in about two thirds of tokens (43 tokens out of 68 that were valid for the analysis). The remaining tokens were realised with voicing that continued into the closure from the previous vowel and then subsided at some point before the burst. The range of incomplete voicing is 31 to 83 ms.

All four subjects produced tokens with incomplete voicing: SCf had 14/21 tokens, BCf had 8/24 tokens, IJm had 2/5 tokens, and MPm 1/18 tokens. Subjects BCf and IJm have on average longer CD then other subjects, so this break in voicing could be caused by the difficulty in maintaining voicing (although, on the whole, differences between subjects were not great).

In the same condition /p, t, k/ were realised either with no voicing at all (29 out of 72 tokens or about 40%), or with voicing that continued into the closure from the preceding vowel for a short period of time. The mean duration of voicing in the pooled data, including instances of zero voicing, is 8.10 ms, which is 8.13% of the mean CD of corresponding word tokens. The range of values for carry-over voicing is 0 to 27ms.

The distribution of voicing duration for both stop classes is shown in boxplots in Figure 6.14 and their means in Table 6.7. There is no overlap between values for the two voicing categories. Their means differ by 50.72 ms. Difference in means for individual subjects varies from 45.33 ms (MPm) to 59.06 ms (IVm). Phonologically voiced stops were realised with significantly longer closure voicing, according to a Mann-Whitney U-test, $p < 0.001$ (2-tailed), $Z = -10.25$, $r = -0.87$, large effect.

Statistical analysis of individual results revealed that for each subject difference in duration of voicing in the closure between phonologically voiced and voiceless stops is significant (Table C7 in the Appendix C). The effect size is large for all subjects.

| | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|
| /b, d, g/ | 58.82 | 68 | 12.68 |
| /p, t, k/ | 8.10 | 72 | 7.96 |

Table 6.7 Mean duration of voicing in the closure (ms), N and SD for word-final intervocalic stops

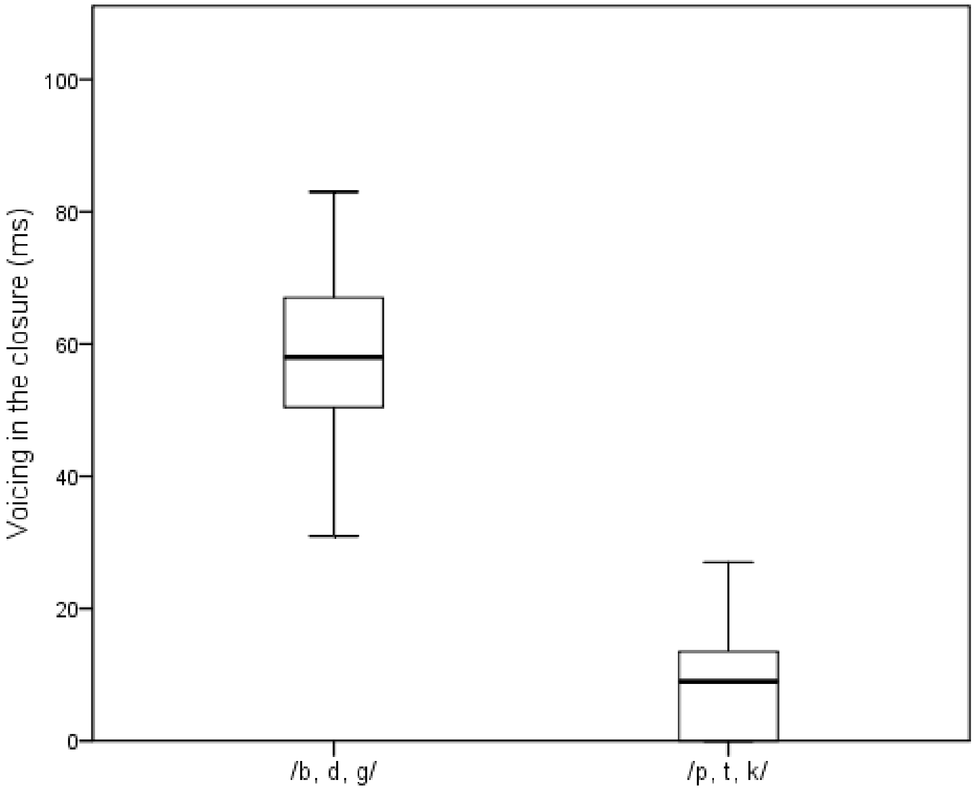

Figure 6.14 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for word-final intervocalic stops

The distribution of results for voicing in the closure expressed as a percentage of corresponding CD is shown in boxplots in Figure 6.15 and the means in Table 6.8. The two categories are well separated in this condition.

| | Mean voicing in closure (%) | N | SD |
|---|---|---|---|
| /b, d, g/ | 90.5 | 68 | 13.8 |
| /p, t, k/ | 8.13 | 72 | 8.04 |

Table 6.8 Mean duration of voicing in the closure (%), N and SD for word-final intervocalic stops

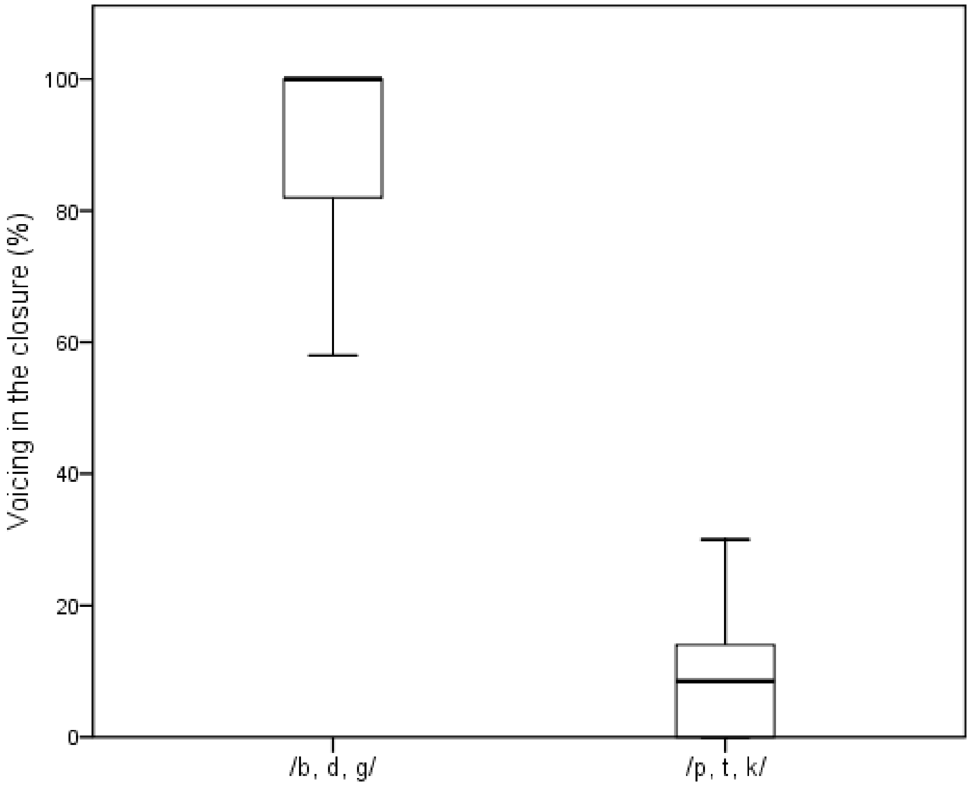

Figure 6.15 Boxplots showing the distribution of values for duration of voicing in the closure (%) for word-final intervocalic stops

Phonologically voiced stops were realised with significantly higher percentage of closure voicing than phonologically voiceless stops, according to a Mann-Whitney U-test, $p$ < 0.001 (2-tailed), $Z$ = -10.41, $r$ = -0.88, large effect. For each subject this is also a significant result ($p$ < 0.001 for SCf, BCf and MPm, and $p$ = 0.003 for IJm, with large effect size for all subjects).

### 6.3.2 Effect of place of articulation on voicing in the closure

Table 6.9 and Figure 6.16 show results for duration of voicing in the closure at the three places of articulation. The effect of place of articulation was significant, according to a one-way ANOVA: $p$ < 0.001, $F(2,65) = 10.21$, $\omega^2 = 0.213$ (large effect). A Tukey post-hoc test revealed that voicing duration was significantly longer for /b/ than for /g/, $p$ = 0.002, and for /d/ than for /g/, $p$ < 0.001. When duration of voicing in the closure is expressed as percentage of CD, mean percentages for /b, /d/ and /g/ are

64%, 63% and 59% respectively. These differences do not reach statistical significance: Kruskal-Wallis test, $p = 0.14$, $\chi^2(2,68) = 3.9$.

| | | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /b/ | 67.3 | 23 | 9.15 |
| Dental | /d/ | 55.7 | 24 | 13.12 |
| Velar | /g/ | 53.1 | 21 | 10.95 |
| Total | | 58.82 | 68 | 12.68 |

Table 6.9 Mean duration of voicing in the closure (ms), N and SD for /b, d, g/ in word-final intervocalic position for each place of articulation

Results for duration of voicing in the closure for /p, t, k/ in word-final intervocalic position are presented in Table 6.10 and Figure 6.16. Percentage of closure that is voiced is 10.73% for /p/, 8.51% for /t/ and 5.42% for /k/, and 8.13% in the pooled data.

| | | Mean voicing in closure (ms) | N | SD |
|---|---|---|---|---|
| Bilabial | /p/ | 10.81 | 21 | 7.46 |
| Dental | /t/ | 8.52 | 27 | 8.99 |
| Velar | /k/ | 5.25 | 24 | 6.35 |
| Total | | 8.10 | 72 | 7.96 |

Table 6.10 Mean duration of voicing in the closure (ms), N and SD for /p, t, k/ in word-final intervocalic position for each place of articulation

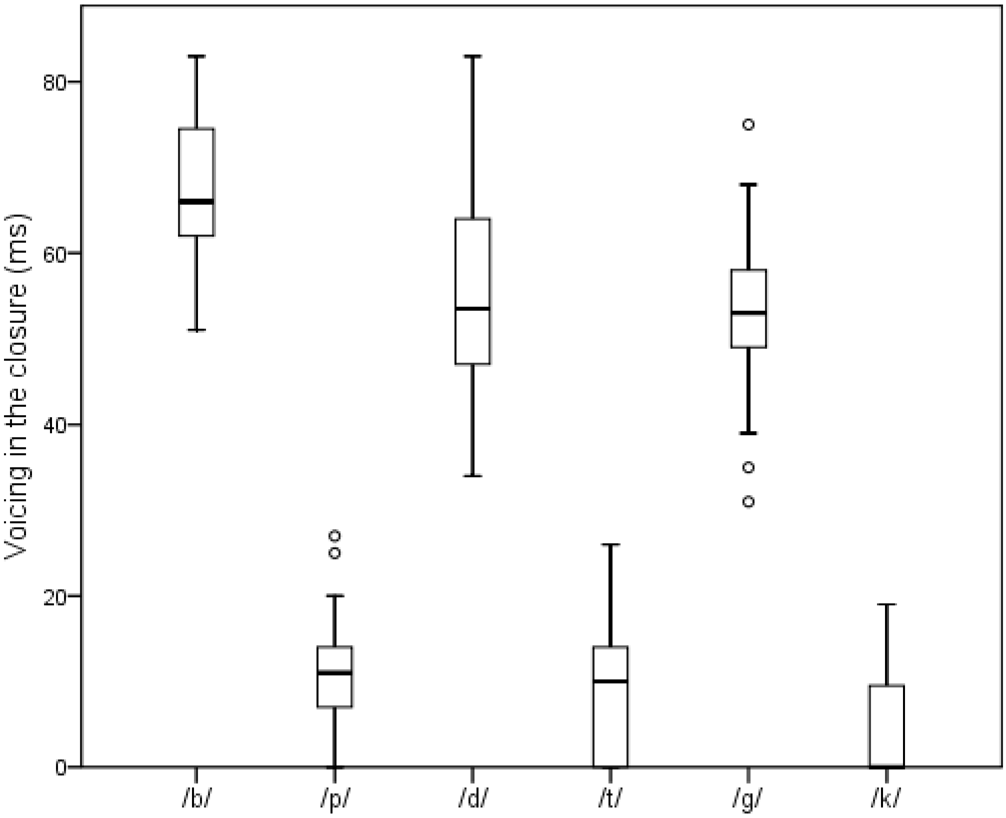

Figure 6.16 Boxplots showing the distribution of values for duration of voicing in the closure (ms) for stops in word-final intervocalic position as a function of stop place of articulation

### 6.3.3 Speaker factors affecting voicing in the closure

**Individual differences between subjects**

Despite the smaller sample of four subjects, some individual differences regarding the number of /b, d, g/ tokens produced with fully voiced closures can be observed. Subjects BCf and MPm produced the majority of their /b, d, g/ tokens with fully voiced closures, 16/24 and 18/19 respectively, while this was not the case with subject SCf, who had 7/21 tokens fully voiced (subject IJm is not represented with enough tokens for such generalisation).

Figure 6.17 shows mean CD and part of the closure that is voiced for each subject, and illustrates between-subject differences. The subjects are ordered from the subject with the shortest mean CD to the subject with the longest mean CD. While subject MPm has somewhat shorter mean CD, absolute duration of voicing in the closure is similar for all four subjects.

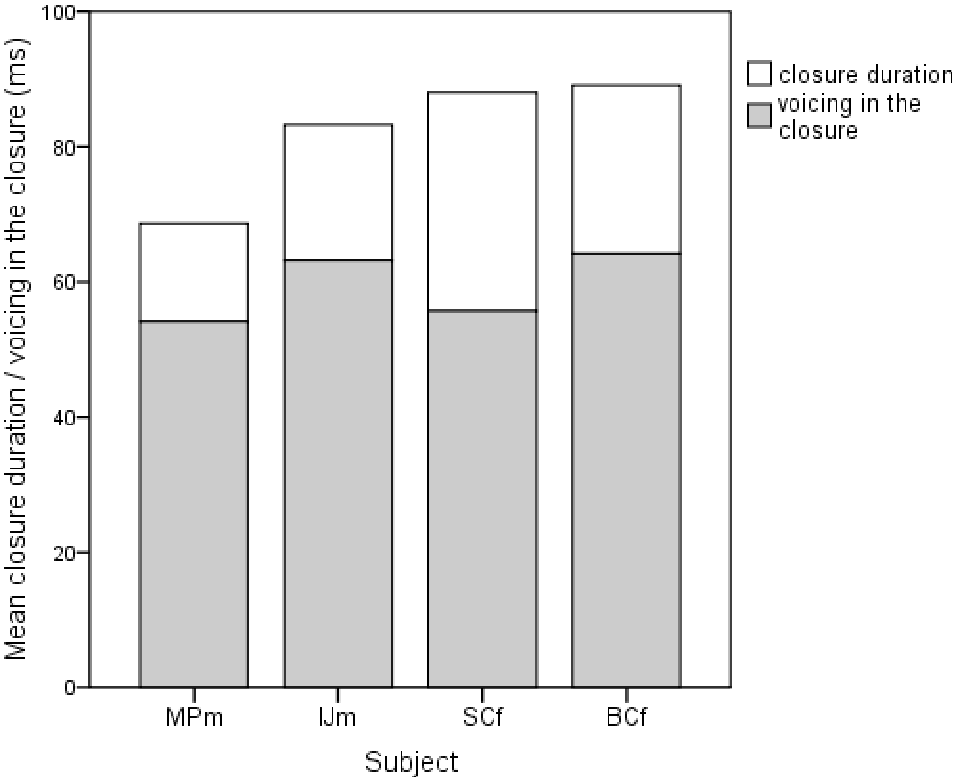

Figure 6.17 Mean CD (ms) and mean voicing in the closure (ms) for /b, d, g/ in word-final intervocalic position for each subject

Results for percentage of closure that is voiced are given in Figure 6.18. There is a contrast between the sentence conditon and the isolation condition in the percentage of closure that is voiced for three out of the four subjects. Subjects SCf, BCf and IJm all have about 40% of their /b, d, g/ closures voiced in isolation, while in the sentence frame that percentage is much higher (82%, 92% and 88%, respectively). Subject MPm, on the other hand, has consistenly high percentage of voicing in the closure: 82% in isolated words, and 99 % in the sentence condition.

CART analysis was not performed because of relatively small number of tokens.

Figure 6.19 presents mean CD and mean duration of voicing in the closure of /p, t, k/ for each subject. On average, between 5% and 11% of the closure is voiced, and the absolute duration of voicing is between 4 and 11 ms.

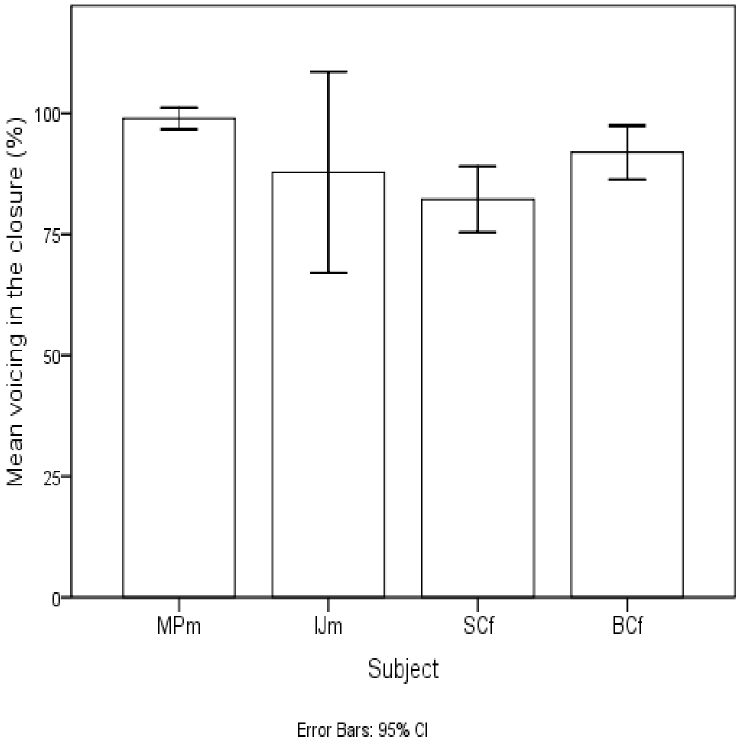

Figure 6.18 Mean voicing in the closure (%) for /b, d, g/ in word-final intervocalic position for each subject

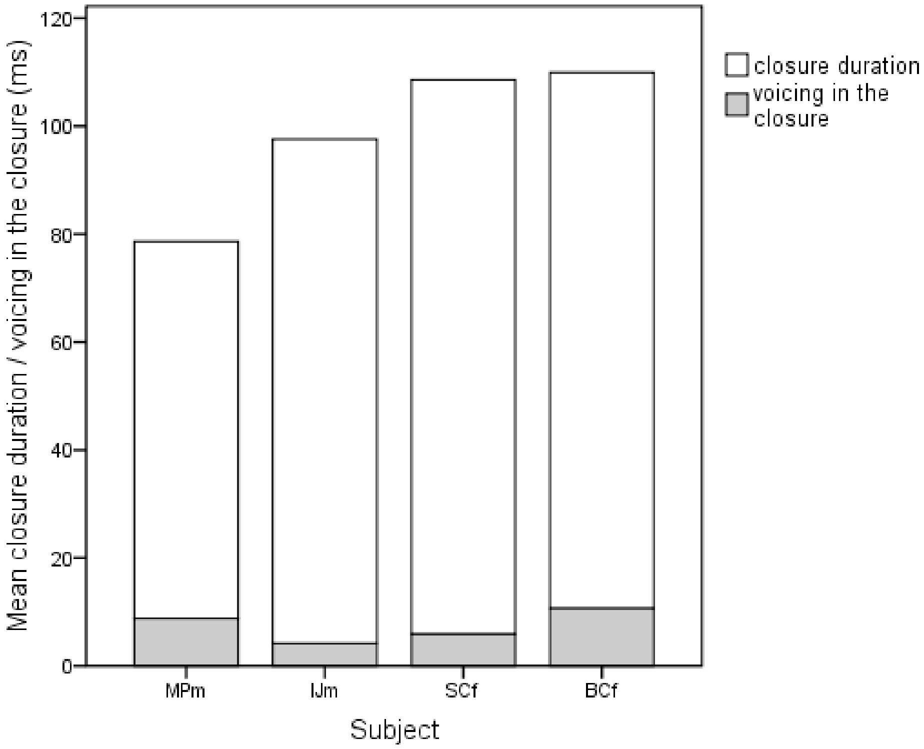

Figure 6.19 Mean CD (ms) and mean voicing in the closure (ms) for /p, t, k/ in word-final intervocalic position for each subject

**Gender and age**

The effect of gender and age on voicing in the closure in /b, d, g/ was not investigated using an ANCOVA (with speaking rate as a covariate) because of smaller number of tokens, similar speaking rate of subjects and because of relatively small age differences between subjects.

The effect of gender on duration of voicing in the closure was not significant, according to a Mann-Whitney U-test, $p$ = 0.15 (2-tailed), $Z$ = -1.43. Mean duration of voicing for female subjects was 60.22 ms ($N$ = 45, $SD$ = 11.93), and for male subjects 56.09 ms ($N$ = 23, $SD$ = 13.9).

However, when voicing in the closure is expressed as percentage of CD, male subjects had longer percentage of closure voiced than female subjects (96.53%, $N$ = 23, $SD$ = 13.9 for males; 87.36%, $N$ = 45, $SD$ = 14.79 for females), and the difference was statistically significant, Mann-Whitney U-test, $p$ = 0.007 (2-tailed), $Z$ = -2.7, $r$ = -0.33, medium effect. This result is due to fact that, although all four subjects had similar absolute durations of voicing in milliseconds, females produce longer closures, and consequently percentage of closures that is voiced is lower.

### 6.3.4 Summary of findings

Results for voicing in the closure in word-final intervocalic position are based on a smaller sample than in the previous two conditions (four subjects), but they generally reinforce previous findings. The voicing contrast is maintained in this condition as well, and voiced stops have significantly longer voicing in the closure than voiceless stops (91% vs. 8% of CD). The same is true for each subject individually. Individual results in this sample are more coherent than in utterance-final position, with all subjects having on average between 82% and 99% of closure voiced. Even subjects who have about 40% of /b, d, g/ closures voiced in utterance-final position, have a high percentage of closure voiced in this condition. In addition to this, male subjects produced a significantly higher percentage of voicing in the closure than female subjects, although the absolute duration of voicing was not significantly different.

There is an effect of place of articulation on voicing duration, with /b/ and /d/ having significantly longer duration of voicing than /g/, but there is no significant difference in the percentage of closure that is voiced.

## 6.4 Discussion

Voicing in the closure is a reliable correlate of the voicing distinction in word-initial intervocalic position in Serbian. Phonologically voiced stops are realised mostly with fully voiced closures (95% of tokens), while phonologically voiceless stops are realised with predominantly silent closures. If there is any voicing present, it is a low-amplitude voicing that is carried over from the preceding vowel, and it occupies on average about 10% of the closure. This finding is in agreement with Abdelli-Beruh’s (2004) findings for French for the same word position, where phonologically voiced stops were mostly realised as completely (or nearly completely) voiced, while in phonologically voiceless stops 25% of closure or less was voiced. Lousada et al. (2010), on the other hand, found not only partially, but also fully devoiced /b, d, g/ tokens word-initially in Portuguese: about 5% of /b/ and /d/ tokens were completely devoiced, and further 5-15% of /b, d, g/ tokens were partially devoiced (2010, p. 266, Figure 3).

This result also replicates the finding for word-initial /b, d, g/ in isolated words in the present study, where all stops were produced as prevoiced. In the sentence condition stops were in intervocalic position, which is conducive to voicing, and the intervening word boundary obviously does not affect production of closure voicing in phonologically voiced stops to any large extent.

Duration of voicing in the closure in word-medial stops was not measured in the present study. This decision was based on the assumption that if voicing is present in phonologically voiced stops word-initially and word-finally in the sentence frame (where voicing continues across word boundary from the preceding vowel into the stop closure, or from the closure into the following vowel), then in word-medial position, where only syllable boundary might intervene, voicing is also highly likely to be present during most of the closure. To check this, two-syllabic words in the sentence frame (used for measuring preceding vowel duration) were visually inspected to assess the amount of voicing present during closure. The stop in question is at the beginning of an unstressed syllable and intervocalic (in post-stressed intervocalic position). In this sample, in 78 out of 94 tokens (or 83%) closures were fully voiced, and in remaining 16 tokens (17%) closures were voiced for the most part, except a short voiceless interval just before the burst. The majority of partially devoiced tokens were /g/ tokens (10 out of 16). They mostly come from three subjects: MCf, SCf, and IJm. Voiceless stops were realised as in the other two intervocalic conditions, with little or no voicing during the closure. This finding confirms that duration of voicing in the closure is a correlate of the voicing distinction word-medially as well. The number of devoiced tokens is higher than in initial intervocalic position, which could be due to different stress patterns of the words investigated. Namely, Keating (1984b) found that if the following vowel is stressed, it prolongs duration of voicing in the stop. Keating argues that stress is associated with greater activity in respiratory muscles, and this increases subglottal pressure during stop production, which results in longer voicing. Since in the present study word-initial stops are before a stressed vowel, and word-medial stops before an unstressed vowel, this could explain differences in the number of tokens with interrupted voicing.

These results are similar to findings from Hungarian, Russian, and Swedish. In word-medial intervocalic position in Hungarian 95.5% of /b, d, g/ tokens were fully voiced, and all tokens in word-initial position were prevoiced (Gósy & Ringen, 2009), as in Serbian. In Russian, over 97% of /b, d, g/ tokens were fully voiced in both positions (Ringen & Kulikov, fc). In Swedish as well, all /b, d, g/ tokens were predominantly voiced in all three word positions (Helgason & Ringen, 2008). In Portuguese, on the other hand, Lousada et al. (2010) found partially devoiced /d/ tokens, and both partially and fully devoiced /g/ tokens word-medially, although the number of partially and fully devoiced stops was lower in initial and medial position than in word-final position. Relatively high incidence of devoicing has emerged as an important characteristic of Portuguese. Pape and Jesus (2011) also found high percentage of devoicing of phonologically voiced stops in medial position in Portuguese. There was no consistent effect of place of articulation and the following vowel on number of devoiced tokens. Pape and Jesus suggested that "it could be the case that the high amount of devoicing is an important feature of this language, and thus overrides the expected higher voicing probabilities for bilabials and dentals" (p.1569).

In word-final intervocalic position in Serbian about 63% of tokens in this condition were fully voiced, and the mean percentage of closure voicing is 91%. All four subjects had on average 82% or more of closure voiced. In both final conditions, as well as in other word-positions, /p, t, k/ were realised with little or no voicing.

In word-final intervocalic position in Portuguese, the number of partially or fully devoiced stops is about 30%, but the percentage of fully devoiced stops is relatively high, with about 15% of /b/ and /d/ tokens and 30% of /g/ being fully devoiced (Lousada, et al. 2010, Figure 3, p. 266). Abdelli-Beruh (2004), on the other hand, found that majority of /b, d, g/ tokens had more than 75% of closure voiced. Results for Serbian are in between these two results, with 15% of tokens partially devoiced, but with no fully devoiced tokens (the lowest percentage of voicing found in Serbian was 58% in 3 out of 68 tokens).

In utterance-final position Serbian voiced stops are realised as either fully voiced (18% of tokens) or as partially devoiced (82% of tokens). Despite some between-subject variability in the mean percentage of the closure that is voiced, for each subject the amount of closure voicing in the cognate stop pairs clearly separates the two voicing categories. In the pooled data, on average 62% of the /b, d, g/ closures is voiced. A slightly higher percentage was reported for Hungarian pre-pausal stops, where about 70-74% of closure was voiced (Gósy & Ringen, 2009).

For English, for all three word positions, Docherty (1992) found that there is significantly more voicing in the closure of phonologically voiced stops than in the closure of phonologically voiceless stops. In phonologically voiceless stops in post-vocalic position there is usually some voicing carried over from the preceding vowel, as in Serbian. However, English is different from the above-mentioned voicing languages because of higher number of phonologically voiced stops that are partially or fully devoiced. In word-initial intervocalic position, 97% of /b, d, g/ tokens had interrupted voicing, with 52% to 67% of closure voiced. In word-final intervocalic position, Docherty found that 46% of /b, d, g/ tokens were without any voicing, and that overall, in most of phonologically voiced stops in final position voicing was interrupted (this includes intervocalic, pre-pausal, and voiceless context).

Results for German show a similar pattern as the English results. In word-initial intervocalic position Beckman et al. (fc) found that mean percentage of voicing was 55% for /b/ and /d/, and 42% for /g/, while in medial intervocalic position about 63% of tokens had more than 90% of the closure voiced. Results from Serbian and other voicing languages, on the one hand, and German and English results, on the other hand, are consistent with the proposal that speakers of voicing languages actively aim to voice closures of phonologically voiced stops, but that in aspirating languages intervocalic voicing is a passive, phonetic process (Beckman, et al., fc; Jansen, 2004).

Presence or absence of voicing in word-final stops, as well as the amount of voicing in the closure, are among the rare topics concerning the voicing contrast in Standard Serbian that have received some attention in the past. Studies were mainly based on impressionistic results, and the main problem was that stops in utterance-final position were not discussed separately from stops in other positions, assimilatory or non-assimilatory. This is probably the reason why there are two opposing views: one, that all phonologically voiced stops are realised as fully voiced, and the other, that they can be realised with some degree of devoicing. It has been suggested that there are three patterns of realisation of the phonologically voiced stops:

1. Stops with fully voiced closures, followed by a voiced release, which can include a vocalic element (Belić, 1968; Ivković, 1913; Miletić, 1960).

2. Stops with partially voiced closures, followed by a voiceless release (Ivković, 1913; Miletić, 1960). These stops are often referred to as partially devoiced. Miletić considered this pattern to be frequent in utterance-final position.

3. Unreleased stops with partially voiced closures, where voicing dies out at some point in the closure (Ivković, 1913). This pattern was considered to be rare.

Results from the present study confirm that word-final intervocalic stops are often released with voicing throughout the closure, and that pre-pausal stops are released, either with fully voiced closures, or with partially voiced closures, as in the second pattern. Instances of a vocalic element after the release were observed in the present study, but were not frequent. The third pattern, where a stop is unreleased, is extremely rare in the present study. This finding is in agreement with Peco (1961-1962a), who noted that the release was always present in his data.

In sum, results from the present study confirm that duration of voicing in the closure is a correlate of the voicing contrast in Serbian stops, in all three word positions. There is a certain amount of devoicing in voiced stops, in particular in utterance-final position, the degree of which is also speaker-dependent, but never at the cost of contrast maintenance. In addition to this, in word-final position the voicing contrast is reinforced by two other correlates that were investigated in the present study, closure duration (Chapter 5) and preceding vowel duration (Chapter 7), which suggests that, despite some devoicing, the contrast is robust in this position.

### Effect of place of articulation

According to the aerodynamic explanation for place-related differences in voicing duration, it is easier to sustain voicing at a more forward place of articulation, because of larger area of compliant cavity walls behind constriction (Keating, 1984b; Ohala & Riordan, 1979; Westbury & Keating, 1986). Following this, it could be expected that stops produced at a more forward place of articulation would be produced with longer periods of vocal fold vibration, and that they would be less likely to devoice.

The numbers of devoiced tokens in Serbian vary with position in the word. In word-initial intervocalic, word-medial, and utterance-final position /g/ is more often devoiced than the other two stops. These results are in agreement with aerodynamic hypothesis, although the number of devoiced stops in the first two conditions is relatively small for generalisations. In word-final intervocalic position the number of devoiced tokens was 5 for /b/, 11 for /d/, and 9 for /g/, which does not support fully the aerodynamic hypothesis.

In Hungarian medial stops /b/ tokens were less likely to devoice than /d/ or /g/ tokens (Gósy & Ringen, 2009), but mixed results were reported for Portuguese (Lousada, et al., 2010). In Portuguese in initial position the percentage of devoicing was in order /b/ = /d/ > /g/ (no /g/ tokens were devoiced), which contradicts the above hypothesis (as noted by Pape & Jesus, 2011 as well). In medial position the order was /b/ < /d/ < /g/, and in final position /b/, /d/ </g/, both of which could be interpreted as supporting the aerodynamic hypothesis.

A place effect is present in Serbian as a tendency for voicing duration to decrease in order bilabial>dental>velar, but in some conditions differences are very small, in the range of few milliseconds. A significant effect of place on voicing duration is only present in word-initial intervocalic stops, with pattern /b/ > /d/, /g/, and in word-final intervocalic position, where voicing duration decreases in order /b/, /d/ > /g/. Although these findings could be interpreted as supporting the aerodynamic hypothesis, absolute values of voicing duration for /b, d, g/ suggest that they could not have been achieved by passive vocal tract expansion only, and that active voicing must have been involved, as was argued for prevoiced stops in utterance-initial position. For example, mean duration of voicing in word-initial intervocalic /b, d, g/ is 85-108 ms, which is much higher than 60 ms that can be achieved through passive cavity expansion, as was proposed by Westbury and Keating (1986). For final pre-pausal position, Westbury and Keating found that only about 30 ms of voicing can be achieved through passive cavity expansion, but in Serbian there is 64 ms of voicing on average in this condition (and even longer in Hungarian, for example, as reported by Gósy & Ringen, 2009). These results confirm that mechanisms for passive cavity enlargement are not sufficient to explain place-related effects in phonologically voiced stops in languages like Serbian, and that active voicing manoeuvres need to be taken into account.

In contrast to Serbian, results for phonologically voiced stops in German support the aerodynamic hypothesis. In word-medial intervocalic stops Jessen (1998) found significantly longer duration of voicing at the more forward places of articulation, which suggests that voicing is passive in German. This is reinforced by Beckman et al.'s (fc) result that in word-initial sentence-medial stops, the velar /g/ is realised with a lower percentage of voicing than /b/ and /d/.

Docherty (1992), on the other hand, found shorter period of voicing in bilabials than in alveolars and velars in word-initial and word-final position (in both stops classes), and no significant place differences in percentage of voicing in the closure, and suggested that duration of voicing in the closure is under active speaker control in English.

## Effect of gender

In Serbian, in word-initial intervocalic position, female subjects produced longer closures and therefore longer periods of voicing in voiced stops, but for all subjects percentage of voicing in closure was almost 100% (female subjects also produced somewhat longer prevoicing in utterance-initial position). In word-final intervocalic position, absolute duration of voicing was approximately the same for male and female subjects, but the percentage of voicing in the closure was significantly longer for male subjects because females produced longer closures. In utterance-final position as well, female subjects had a smaller percentage of the closure voiced than did male subjects, because of the longer closures. In terms of absolute durations of voicing, in both female and male group some of the subjects were able to sustain voicing for a comparable period of time, but there are also individual differences in this respect, combined with age differences, which are difficult to separate. These results do not support the hypothesis that because of universal biological differences female subjects produce shorter voicing in the closure than male subjects.

Male-female differences in duration of voicing in the closure have been investigated by fewer studies and results from other languages are inconsistent. In Swedish, Helgason and Ringen (2008) found that in intervocalic and pre-pausal stops percentage of closure voicing in voiced stops was significantly higher for male subjects than for female subjects (and the same was true for the duration of prevoicing word-initially). The opposite conclusion was reached for Hungarian by Gósy and Ringen (2009), who found that female subjects produced longer voicing in the closure word-medially, i.e. longer closures, which were mostly fully voiced; female subjects also produced longer prevoicing word-initially. These findings for non-initial stops do not necessarily contradict each other, but are difficult to evaluate, because Helgason and Ringen did not report absolute values. From Figure 6 in their paper, it seems that there is a lot of between-subject variation, and that female subjects produced voicing in the closure of duration that is comparable to that of most male subjects. It also seems that, at least in some of the conditions, females produced longer closures, which could, for the same duration of voicing, result in smaller percentage of voicing in the closure. In Russian, there was no gender effect on duration of prevoicing in phonologically voiced stops, but female subjects produced longer voicing in intervocalic position (Ringen & Kulikov, fc).

### Effect of age

The effect of age on duration of voicing in the closure in phonologically voiced stops is present in this study, but it is difficult to summarise, partly because of lack of data in word-final intervocalic position, and partly because of interaction between gender and age in utterance-final position. It can be said with certainty that in utterance-final position four oldest subjects produced longer voicing in the closure than the rest of the subjects. In addition to this, in word-initial intervocalic position, older subjects as a group produced longer closures and hence longer periods of voicing than younger subjects (which is especially true for male subjects), and also produced longer prevoicing in utterance-initial position. Overall, these results suggest that older subjects do tend to produce longer voicing duration. However, it is also important to note that there is generally a lot of between-subject variability in both age groups, such as observed polarisation between younger speakers, where in utterance-final position younger male speakers produced longer voicing and longer percentage of voicing in the closure than younger female speakers. These findings need to be confirmed by further research.

# Chapter 7 Results for preceding vowel duration

Preceding vowel duration was examined in word-final and word-medial position, in the same two conditions as other correlates of the voicing contrast: in words in isolation and in the sentence frame. The number of minimal pairs suitable for this type of investigation is limited in Serbian, so it was not possible to use only words with phonologically short vowels, as was done in the rest of the study. Instead, minimal pairs with both phonologically short and phonologically long vowels were included. For the same reason, it was not possible to have equal numbers of tokens in different conditions and word positions. In the text that follows, mean vowel durations before phonologically voiced and voiceless stops in all environments are presented, as well as their ratios. However, part of the statistical analysis is performed only on ratios. There are two reasons for this. First, by using ratios instead of absolute values in milliseconds, any potential effect of speaking rate on other factors is removed. Second, using ratios also makes it possible to compare results for phonologically short and phonologically long vowels, and to pool them together where necessary.

## 7.1 Word-final position

Results for preceding vowel duration before stops in word-final position (for words with phonologically short and long vowels in both conditions) are presented in Table 7.1. Results are pooled across subjects.

For each word pair absolute difference in duration was calculated by subtracting the duration of the vowel before the phonologically voiceless stop from the duration of the vowel before the phonologically voiced stop, and the ratio was calculated by dividing the first value by the second value for each word pair. Results for duration were rounded to the nearest millisecond.

Phonologically short vowels preceding word-final voiced stops are longer than those preceding word-final voiceless stops, with the mean difference of 27 ms in isolation and 22 ms in the sentence frame (or 20% and 17% respectively). Two paired t-tests confirmed that in both conditions these differences were statistically significant: in isolation $p < 0.001$ (2-tailed), $t(24) = 8.78$, Cohen's $d = 1.45$ (large effect), and in the sentence frame $p < 0.001$ (2-tailed), $t(24) = 7.52$, Cohen's $d = 1.09$ (large effect).

|  | Mean vowel duration (ms), N, SD |  | Mean differ. (ms) | Mean ratio |
|---|---|---|---|---|
|  | before /b, d, g/ | before /p, t, k/ |  |  |
|  | Phonologically short vowel |  |  |  |
| Isolation | 133<br>25; 16.64 | 106<br>25; 20.44 | 27 | 0.80 |
| Sentence | 125<br>25; 21.22 | 103<br>25; 19.02 | 22 | 0.83 |
| Total | 129<br>50; 19.3 | 104<br>50; 19.6 | 25 | 0.81 |
|  | Phonologically long vowel |  |  |  |
| Isolation | 193<br>56; 31.71 | 178<br>56; 26.07 | 15 | 0.93 |
| Sentence | 178<br>53; 37.85 | 165<br>53; 35.86 | 13 | 0.94 |
| Total | 186<br>109; 35.54 | 172<br>109; 31.76 | 14 | 0.93 |

Table 7.1 Mean vowel duration (ms), N, SD, mean difference (ms) and mean ratio before word-final stops

In isolation, phonologically long vowels preceding word-final voiced stops are longer than those preceding word-final voiceless stops, with the mean difference of 15 ms (or 7%), which is statistically significant: paired t-test, $p < 0.001$ (2-tailed), $t(55) = 4.74$, Cohen's $d = 0.52$ (medium effect). In the sentence frame, vowels before voiced stops are longer as well, with the mean difference of 13 ms (or 6%), which is significant: Wilcoxon test, $p = 0.001$ (2-tailed), $Z = -3.3$, $r = 0.32$ (medium effect).

## 7.2 Word-medial position

Results for preceding vowel duration for stops in word-medial position are presented in Table 7.2.

| | Mean vowel duration (ms), N, SD | | Mean differ. (ms) | Mean ratio |
|---|---|---|---|---|
| | before /b, d, g/ | before /p, t, k/ | | |
| | Phonologically short vowel | | | |
| Isolation | 137<br>49; 3.31 | 112<br>49; 21.92 | 25 | 0.82 |
| Sentence | 131<br>45; 22.07 | 110<br>45; 23.18 | 21 | 0.84 |
| Total | 134<br>94; 22.81 | 111<br>94; 22.43 | 23 | 0.83 |
| | Phonologically long vowel | | | |
| Isolation | 208<br>31; 40.74 | 180<br>31; 29.81 | 28 | 0.88 |
| Sentence | 200<br>30; 34.48 | 178<br>30; 35.34 | 22 | 0.89 |
| Total | 204<br>61; 37.69 | 179<br>61; 32.39 | 25 | 0.88 |

Table 7.2 Mean vowel duration (ms), N, SD, mean difference (ms) and mean ratio before word-medial stops

Phonologically short vowels before voiced stops are longer than those before voiceless stops word-medially, both in isolation and in the sentence frame, with mean differences of 25 ms and 21 ms, respectively (or 18% and 16%). In both cases the difference was significant, in isolation: paired t-test, $p < 0.001$ (2-tailed), $t(48) = 10.5$, Cohen's $d = 1.6$, large effect; in the sentence frame: paired t-test, $p < 0.001$ (2-tailed), $t(44) = 9.92$, Cohen's $d = 0.93$, large effect. These results are comparable to results for phonologically short vowels in word-final position.

Phonologically long vowels before voiced stops are also longer than vowels before voiceless stops word-medially, both in isolation and in the sentence frame, with mean differences of 28 ms and 22 ms, respectively (or 12% and 11%). These differences were significant, in isolation: paired t-test, $p < 0.001$ (2-tailed), $t(30) = 7.3$, Cohen's $d = 0.78$ (large effect); in the sentence frame: paired t-test, $p < 0.001$ (2-tailed),

$t(29) = 6.26$, Cohen's $d = 0.63$ (medium effect). These differences are bigger than differences for phonologically long vowels in word-final position.

## 7.3 Linguistic factors affecting preceding vowel duration

Because of relatively small number of tokens, statistical analysis of potential factors that induce variability in the realisation of this correlate of voicing was performed only on the pooled data. A four-way ANOVA was carried out on the pooled data for ratio to examine the effect of the following factors: condition (isolation, sentence frame), position within the word (final, medial), phonological vowel length (short, long), and stop place of articulation (bilabial, dental, velar). The quality of the vowel was not controlled because of the limited number of minimal pairs available, and consequently this factor was not investigated (although it could have an effect on the results for other variables).

The main effect of condition was not significant, $p = 0.36$, $F(1,292) = 0.86$, nor was the main effect of position within the word, $p = 0.613$, $F(1,292) = 0.26$.

The main effect of vowel length was significant: $p < 0.001$, $F(1,292) = 33.22$, $\omega^2 = 0.081$ (medium effect), with phonologically short vowels having smaller ratios than phonologically long vowels, with mean ratios of 0.82 and 0.92 (or 18% and 8% respectively). The main effect of stop place of articulation was also significant: $p = 0.008$, $F(2,292) = 4.91$, $\omega^2 = 0.019$ (small effect). The effect was greater before the velars than before the bilabials and the dentals, with means ratios of 0.84, 0.89, and 0.88 respectively.

There was also a statistically significant interaction between vowel length and place of articulation: $p = 0.001$, $F(2,292) = 6.67$, $\omega^2 = 0.028$ (small effect). While the ratio is similar for short and long vowels preceding bilabial stops, the difference increases as the place of articulation moves further back. The ratio for long vowels increases slightly (and the magnitude of the effect decreases), while the ratio decreases to a larger extent for short vowels. This interaction is shown in Figure 7.1. To investigate the interaction between phonological vowel length and stop place of articulation, further tests were performed on the pooled data for phonologically short and long vowels separately, while all other factors were collapsed together.

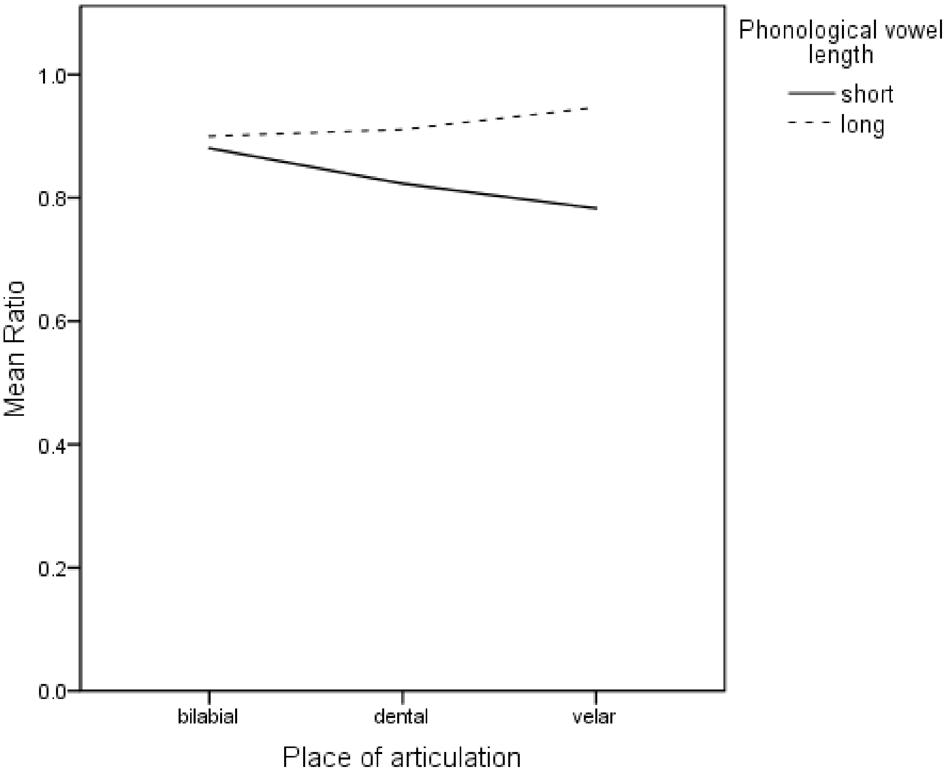

Figure 7.1 Vowel duration ratio as a function of phonological vowel length and stop place of articulation

A one-way ANOVA (ratio by place of articulation) was performed on the data for phonologically short vowels, and the effect of place of articulation on ratio was significant: $p < 0.001$, $F(2,141) = 10.87$, $\omega^2 = 0.121$ (medium effect). A Gabriel post-hoc test revealed that the ratio was significantly higher (i.e. differences smaller) for the bilabials than for the dentals ($p = 0.05$), and for the bilabials than for the velars ($p < 0.001$). Mean ratios at the three places of articulation are 0.88 for the bilabials ($SD = 0.12$, $N = 43$), 0.82 for the dentals ($SD = 0.11$, $N = 37$), and 0.78 for the velars ($SD = 0.1$, $N = 64$).

Another one-way ANOVA was performed on the data for phonologically long vowels, and the effect of place of articulation on ratio was not significant: $p = 0.18$, $F(2,167) = 1.73$. Mean ratios at the three places of articulation are 0.9 for the bilabials ($SD = 0.11$, $N = 51$), 0.91 for the dentals ($SD = 0.11$, $N = 82$) and 0.95 for the velars ($SD = 0.14$, $N = 37$).

## 7.4 Speaker factors affecting preceding vowel duration

**Individual differences between subjects**

The effect of speaker variables on ratio was tested only in the pooled data because of small number of tokens per subject.

Figure 7.2 shows distribution of ratio values for each of the twelve subjects (in ascending order from the lowest mean ratio to the highest mean ratio).

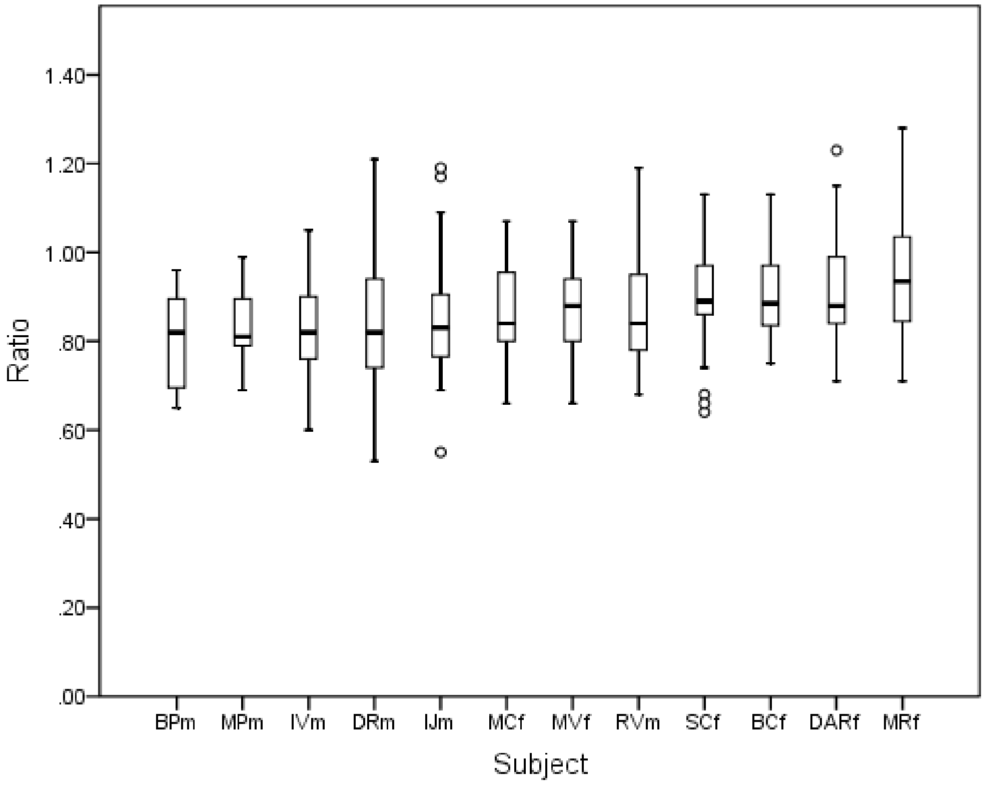

Figure 7.2 Boxplots showing the distribution of ratio values for each subject

A CART analysis revealed that there are three groups of subjects whose results differ significantly: Group 1, with the lowest mean ratio of 0.83, *SD* = 0.11, *N* = 70 (subjects BPm, MPm, IVm); Group 2: with intermediate mean ratio of 0.86, *SD* = 0.12, *N* = 137 (subjects DRm, IJm, MCf, MVf, RVm), and Group 3: with the highest mean ratio of 0.92, *SD* = 0.13, *N* = 107 (subjects SCf, BCf, DARf, MRf).

When results for individual subjects are separated into results for phonologically short and long vowels, the same pattern as that found in the pooled data emerges for each of the twelve subjects. For each subject the effect is larger for phonologically short vowels, and vowel duration difference is significant, with large effect size. Mean individual ratio for phonologically short vowels varies between 0.77 and 0.92 (i.e. vowel duration difference varies between 8% and 23%). Six subjects have more than 20% difference in vowel duration, five subjects have between 10% and 20% difference, and one subject has below 10% difference. For phonologically long vowels mean individual ratio is between 0.84 and 0.99 (vowel duration difference between 1% and 16%). Five subjects have more than 10% difference in vowel duration, and seven subject less than 10% difference. Out of them, only one subject, MRf, has hardly any difference in vowel duration in long vowels (ratio of 0.99), while having ratio of 0.92 (8%) in short vowels. For long vowels, differences in vowel duration are significant only for about half of the subjects, and effect size is small to large. Results for each speaker are given in Table C8 and Table C9 in Appendix C, and a comparison of ratios in Table C10.

**Gender and age**

A two-way ANOVA was carried out to explore the effect of gender and age on duration ratio. The subjects were divided into two equal groups according to their age, as before.

There was a significant main effect of gender, $p < 0.001$, $F(1,310) = 16.46$, $\omega^2 = 0.047$ (small effect). The effect of stop voicing on preceding vowel duration was bigger for male subjects than for female subjects (mean ratio for males = 0.84, $SD = 0.12$, $N = 157$; mean ratio for females = 0.9, $SD = 0.13$, $N = 157$).

Main effect of age did not reach significance, $p = 0.9$, $F(1,310) = 0.02$, and there was no significant interaction between the two main factors, $p = 0.12$, $F(1,310) = 2.44$.

## 7.5 Summary of findings

In all environments investigated the mean duration of vowels before voiced stops was longer than the mean duration of vowels before voiceless stops. Absolute differences for phonologically short vowels are consistent in both word-final and word-medial position, ranging between 21 and 27 ms, with the corresponding ratios between

0.80 and 0.84 (or a difference of 16 - 20%). The total mean ratio for phonologically short vowels is 0.82.

There is more variability in the results for phonologically long vowels. Mean absolute differences in word-final position are smaller than those in word-medial position, with the mean ratio of 0.93 and differences of up to 15 ms (or 6 - 7%) in final position, and the mean ratio of 0.88 and 22 - 28 ms difference (or 11-12%) in medial position. The total mean ratio for phonologically long vowels is 0.92, but this ratio is based on unequal numbers of tokens for word-final and word-medial position, and is skewed towards higher values found in word-final position.

Overall, differences expressed through the ratio indicate that the effect of the following stop voicing is significantly larger on phonologically short than on phonologically long vowels. In all environments, mean differences in vowel duration are slightly higher in isolation than in the sentence frame, and the corresponding ratios are lower, but these differences are small and non-significant. There is a place of articulation effect in phonologically short vowels only, where bilabials have significantly higher ratio (smaller difference) in voicing-conditioned vowel duration than dentals and velars.

For each subject the voicing effect is significantly larger for phonologically short vowels. For phonologically long vowels, however, not all subjects produced significant differences in vowel duration. There is no effect of age on vowel-duration differences, but there is an effect of gender: male subjects produce larger differences than female subjects.

## 7.6 Discussion

A comparison of Serbian results with results for other languages is somewhat difficult because of variability in conditions used in previous studies. A comparison with studies on English that use the same type of words (monosyllables and disyllables) reveals that both absolute differences (in ms) and relative differences (percentages) are smaller in Serbian than in English. For example, in Serbian monosyllables total ratio is 0.81 for phonologically short vowels and 0.93 for phonologically long vowels, but in English it is between 0.53 and 0.67 (Hogan & Rozsypal, 1980; Klatt, 1973; Luce & Charles-Luce, 1985; Mack, 1982). In Serbian disyllables total ratios are 0.83 and 0.88

respectively, while in English it is 0.79 (Klatt, 1973). Studies that reported pooled results for several conditions or word positions in English found the ratio between 0.55 and 0.8 (Chen, 1970; House, 1961; House & Fairbanks, 1953; Laeufer, 1992; Peterson & Lehiste, 1960; Smith, et al., 2009). This difference between Serbian and English is expected, because English stands out from other languages by having larger effect, as was discussed in the literature review in Section 1.5.

There are fewer studies with comparable data from other languages. For French, results are inconsistent. For CVC words in environments comparable to those in the present study, Mack (1982) and Laeufer (1992) reported mean ratios of 0.74 – 0.75. On the other hand, Abdelli-Beruh (2004) found higher ratio of 0.85 for word-final stops, which is in agreement with Chen's (1970) ratio of 0.87 for a set of words with obstruents mainly in word-final position. The Serbian mean ratio of 0.82 for phonologically short vowels falls between Mack's (1982) and Lauefer's (1992) results, on the one hand, and Abdelli-Beruh's (2004) and Chen's (1970) results, on the other hand. However, for phonologically long vowels in Serbian ratios are higher (0.88 – 0.93) and closer to Chen's (1970) result for French.

Some of the factors that have been reported to affect realisation of voicing-conditioned vowel duration in English were also investigated in the present study. First, it has been found that in English the effect was biggest in isolated words and in phrase-final position. In the present study, the effect is slightly larger in isolated words than in the sentence frame in all environments, but this difference is in the range of 5-6 ms and non-significant. In Serbian, there is no significant difference in the magnitude of the effect depending on the position within the word in the pooled data, although the effect is smaller for phonologically long vowels before word-final than before word-medial stops. Second, in English the effect is larger in monosyllabic than in polysyllabic words. Although the number of syllables was not investigated as a separate factor in the present study, words with final stops were all monosyllables and words with medial stops were all disyllables, so the effect of position within the word is confounded with the effect of number of syllables, and non-significant.

Other factors, such as speaking rate and stress could not have been tested in the present study. As for the manner of articulation, it has been found that the effect is bigger in the context of stops than in the context in fricatives, not only in English (Hogan & Rozsypal, 1980; House, 1961; House & Fairbanks, 1953; Peterson & Lehiste, 1960), but also in French, where it exceeds that found in English (Laeufer, 1992). My preliminary research suggested that this is not the case in Serbian (Sokolović-Perović, 2008). In Serbian, in all conditions except for phonologically short vowels word-medially, the effect is smaller before fricatives (10 % or less) and not significant. For phonologically short vowels in word-medial position, however, stops and fricatives have comparable effect (16-18%), and differences are significant in both cases. These findings call for further research beyond the scope of this thesis.

To sum up, out of the factors that have been found to have an effect on the voicing-conditioned vowel duration in English, number of syllables and position within the word/utterance induced variation that did not reach statistical significance in Serbian, nor did differences between the two conditions, isolation and the sentence frame. The factor that induced the most variability was phonological vowel length, and the effect was in the opposite direction to that found in English. In the present study, vowel-duration effect was larger for phonologically short vowels, while in English it was mostly the opposite.

However, there are some additional factors that have surfaced as potentially relevant for voicing-conditioned vowel duration, which have received little attention in the previous research. The effect of stop place of articulation on preceding vowel duration was significant before phonologically short stops in Serbian, with the effect being greater for the dentals and the velars than for the bilabials. It is surprising that this factor was rarely investigated, taking into account that place-related differences in CD have been researched, and considering the proposed temporal adjustment of the vowel duration and CD in the VC sequence. If this hypothesis was true, then place-related differences in CD could be expected to be inversely proportional to differences in vowel duration. Duration of VC sequence in Serbian is discussed below.

Finally, there is a small, although significant effect of gender on voicing-conditioned vowel duration, with male subjects exhibiting larger effect than female subjects. Individual differences are also present, with several subjects having non-significant voicing-related differences in vowel duration for phonologically long vowels.

An important linguistic factor in Serbian is phonological vowel length. The effect is significantly larger for phonologically short vowels, which is in contradiction to results from several studies on English, which found smaller effect in the short/lax vowels than in the long/tense vowels (House, 1961; Klatt, 1973; Luce & Charles-Luce, 1985; Peterson & Lehiste, 1960). For example, House reported a difference of 90 ms for short vowels and 150 ms for long vowels, while Luce and Charles-Luce found that the voicing effect was smaller for the short vowel /ɪ/ than for the long vowel /i/ (42 ms and 66 ms, respectively). In continuous speech, Crystal and House (1982) found almost negligible difference of 5 ms in short vowels, compared to 24 ms in long vowels. In Serbian, on the other hand, the effect of stop voicing on preceding vowel duration is larger in phonologically short vowels than in long vowels in word-final position (25 ms vs. 14 ms, and is also reflected in ratios: 0.81 vs. 0.93). In word-medial position the effect is similar (24 ms and 25 ms), but ratios still reveal bigger effect on short vowels (0.83 for short vs. 0.88 for long vowels).

No hypotheses have been put forward as to why this effect would be smaller in short vowels in English. In a theory of segmental duration in English, Klatt (1973) argued that there is a limit to how much inherently short segments can shorten in certain environments, but that in terms of percentages, the amount of change would be the same as for inherently long segments. He further proposed that if a vowel is shortened by one rule, for example by adding the second syllable to a monosyllabic word, which reduces vowel duration to 66% of its inherent duration, it would then become less compressible when an additional rule is applied, such as changing the final consonant from voiced to voiceless, which, on its own, would also reduce vowel duration to 66%. If both changes occur, the reduction is not cumulative, and vowel is reduced to 54% of its original duration because it cannot be shortened further.

However, none of this reasoning is supported by Serbian data. Changing the voicing value of the final stop from voiced to voiceless does reduce vowel duration, but this effect is bigger for phonologically short vowels than for long vowels. Word-finally, short vowels in voiceless environments shorten more than long vowels, both in terms of absolute values and in terms of ratios. Word-medially, short vowels shorten by about the same amount in milliseconds as long vowels, but more in relative terms. Moreover, adding a second syllable increases vowel duration (although not significantly) instead of reducing it. Total effect of both changes is bigger for short vowels than for long vowels (vowel duration is reduced to 86% of its original value vs. 96% in long vowels).

A proposal along the lines of Klatt’s theory for English cannot resolve the issue of why there is a different effect on vowel duration for phonologically short and long vowels in Serbian. It is possible that it comes from the fact that vowel quality was not controlled in this part of the study, and number of vowel tokens was not balanced across conditions and word positions. An examination of the data that is available suggested that this is not the case.

Another possibility is that it comes from the need to preserve the short-long opposition in vowels. In trying to explain the absence of this effect in Czech, which also has a phonemic vowel length distinction, Keating (1985) proposed that in Czech there is no voicing-conditioned difference in vowel duration because durational differences are reserved for phonemic length contrast. This is an interesting proposal that deserves consideration. It was tested on Southern Serbian, a non-standard variety which, unlike Standard Serbian, does not have phonemic vowel length contrast, but only has short vowels (Sokolović-Perović, 2009). In Southern Serbian, because there is no pressure to maintain phonemic length contrast in vowels, there is potentially more scope for voicing-conditioned variation. If Keating's reasoning is correct, the voicing effect in Southern Serbian could be greater than that in Standard Serbian. However, this is not the case. The overall mean ratio in Southern Serbian is 0.82, and ratios in different environments are similar to those found for phonologically short vowels in Standard Serbian, both in the present study and in Sokolović-Perović (2009), which is based on a different sample. In Southern Serbian the potential for bigger voicing-conditioned vowel duration differences is not utilised.

This could further mean that in Serbian there is no need for vowel duration differences to be exploited to a larger degree in signalling the voicing distinction, since other correlates of voicing are sufficient, such as voicing in the closure and CD. This echoes Esposito's (2002) thoughts about Italian, where she argues that "for Italian, at least, there is no perceptually motivated reason to lengthen the vowel beyond any normal 'physiological' lengthening due to the consonant voicing" (p. 221), and she proposes a production-oriented explanation for this effect in Italian. This might be true for Serbian as well. As the results from the present study have shown, the voicing contrast in Serbian is robust and based not only on the opposition between presence and absence of vocal fold vibration during stop closure, but on differences in CD as well.

However, it is unclear what "normal physiological lengthening" would be and why it occurs in the first place. There are several production-based accounts of the voicing-conditioned vowel duration effect, but they all have certain problems.

Chen (1970) and Kozhevnikov and Chistovich (1966) proposed that because voiceless stops are produced with greater articulatory force, the velocity of movement is greater and the closure is achieved faster, thus shortening the preceding vowel. Apart from the fact that the measurement of articulatory force remains controversial, subsequent studies failed to find supporting evidence for different velocity of the closing gestures of voiced and voiceless stops (see Kluender, et al., 1988 for discussion).

Another account is based on the idea of different vocal cord adjustment rate for the two stop classes (Chen, 1970; Chomsky & Halle, 1968). According to this view, voiced stops require precise laryngeal adjustment needed to sustain active vocal fold vibration, and longer time is needed to achieve this state from the spontaneous vocal fold vibration for the preceding vowel, which increases the duration of the vowel. For voiceless stops, on the other hand, the glottis needs to be wide open and this is achieved through simple abduction of vocal folds, which requires shorter time period. However, fiber-optic and electromyographic studies found no support for the laryngeal adjustment before voiced stops, and acoustic studies found no support for predictions, resulting from this explanation, that vowels before voiced stops would be longer than before nasals (for discussion see Chen, 1970; Kluender, et al., 1988).

The third account is based on the idea of compensatory temporal adjustment, whereby vowel durations are inversely proportional to closure durations in order to keep duration of VC sequence uniform (Port, 1981). An alternative measure is the C/V duration ratio, which is considered to be constant for each stop class across different contexts (Port, 1981). These ideas also received insufficient support from acoustic studies (Chen 1970, Keating 1985, see also Section 1.5 for discussion).

Because of the lack of articulatory data for Serbian, any evaluations of these proposals can only be based on acoustic data. To assess the duration of VC sequence, CD was measured for the same set of minimal pairs used for measuring vowel duration in isolated words. Tables 7.3 and 7.4 show mean vowel duration, mean CD and mean duration of VC sequence for each condition and for each phonological vowel length (for words spoken in isolation).

| Word-final position | Mean preceding vowel duration (ms) | Mean CD (ms) | VD + CD (ms) |
|---|---|---|---|
| | Phonol. short vowel | | |
| /b, d, g/ | 133 | 103 | 236 |
| /p, t, k/ | 106 | 153 | 259 |
| | Phonol. long vowel | | |
| /b, d, g/ | 193 | 101 | 294 |
| /p, t, k/ | 178 | 154 | 332 |

Table 7.3 Mean vowel duration (ms), mean CD (ms), and mean duration of the VC sequence (ms) in word-final position in isolated words

| Word-medial position | Mean preceding vowel duration (ms) | Mean CD (ms) | VD + CD (ms) |
|---|---|---|---|
| | Phonol. short vowel | | |
| /b, d, g/ | 137 | 75 | 172 |
| /p, t, k/ | 112 | 130 | 242 |
| | Phonol. long vowel | | |
| /b, d, g/ | 208 | 70 | 278 |
| /p, t, k/ | 180 | 122 | 302 |

Table 7.4 Mean vowel duration (ms), mean CD (ms), and mean duration of the VC sequence (ms) in word-medial position in isolated words

Results from Tables 7.3 and 7.4 suggest that CDs for the two stop classes are fairly uniform for words with phonologically short and long vowels in each condition, for example mean CD of voiced stops is 101 and 103 ms, and mean CD of voiceless stops is 153 and 154 ms (word-final position, Table 7.3); a similar relationship can be observed for word-medial position (Table 7.4). However, mean vowel durations are different, and the total VC sequence durations are not balanced. Closure and vowel duration do not vary inversely in a systematic way, which means that the VC dyad does not have a uniform duration in Serbian.

It is interesting to point out that CD of voiced and voiceless stops in each of two conditions (word-final and word-medial) are fairly uniform irrespective of the phonological vowel length of the preceding vowel. In addition to this, they are smaller overall in disyllables than in monosyllables, which is not the case with vowel duration, which also points out to the conclusion that they are controlled independently from each other[20]. This finding is similar to that for Polish by Keating (1985), who found significant differences in CD word-medially (mean CD for /t/ = 130 ms, mean CD for /d/ = 92 ms, or about 30%), but no vowel-duration differences. In the present study, durational difference in CD word-medially is larger than in Polish (42%), but is accompanied by vowel-duration differences of about 11-16%.

In addition to this, there is no negative correlation between CD and preceding vowel duration for neither phonologically voiced nor voiceless stops, and (positive) correlation is small in both cases (Spearman's correlation for phonologically voiced stops: $r_s$ = 0.04, $p$ = 0.6, 2-tailed; for phonologically voiceless stops: $r_s$ = 0.21, $p$ = 0.009, 2-tailed).

Finally, the C/V duration ratio was also calculated for each word in this data set. Table 7.5 shows results for the C/V ratio for each condition. It is clear that the C/V duration ratio is very variable, and therefore cannot be considered a relevant parameter for the voicing contrast in Serbian stops.

| | Mean C/V ratio | |
|---|---|---|
| | Word-final | Word-medial |
| | Phonologically short vowel | |
| /b, d, g/ | 0.78 | 0.56 |
| /p, t, k/ | 1.48 | 1.19 |
| | Phonologically long vowel | |
| /b, d, g/ | 0.53 | 0.3 |
| /p, t, k/ | 0.87 | 0.68 |

Table 7.5 Mean C/V ratio in word-medial position isolated words

The C/V duration ratio has a different role within auditory enhancement theory, where it is one of intermediate perceptual properties of the voicing distinction. According to this view, preceding vowel duration is varied by speakers in order to

[20] In addition to this, place of articulation does not have an effect on CD in word-final stops in the present study, but there is an effect of place on vowel duration ratio for phonologically short vowels.

perceptually enhance the closure-duration cue: a longer preceding vowel makes the following consonant closure seem shorter (which suggests a voiced stop), and vice versa. Results from the present study that preceding vowels are longer and closures are shorter for phonologically voiced stops, compared to shorter vowels and longer closures for phonologically voiceless stops, can be interpreted as supporting the auditory enhancement hypothesis. However, these findings do not necessarily suggest that vowels are intentionally lengthened before shorter closures and vice versa, in order to achieve auditory enhancement. The reason for this is that the theory is not explicit about the extent of this effect and whether it is expected to be of the same magnitude in all conditions or not, and also how and to what extent changes in vowel duration should be balanced with changes in CD. As discussed above, vowel-duration differences and closure-duration differences are present in Serbian, but they are of different magnitude and unequal across environments and conditions. Of course, the question remains if this kind of enhancement is needed in Serbian at all, considering the robustness of the voicing contrast, as was argued by Esposito (2002) for Italian, or whether it is a result of production-related constraints.

# Chapter 8 Discussion

In this chapter I discuss my findings in relation to the existing theoretical accounts of the voicing contrast. I examine how well they can represent Serbian results, and which aspects need to be improved. In Section 1, I summarise results from Chapters 4 to 7, outlining which sets of acoustic correlates are relevant in each word position. I further discuss variability induced by both linguistic and speaker factors, pointing out that only some of the variability is coming from universal constraints, and that most of it is specific to Serbian or even to individual speakers. Both these issues are further discussed in relation to the existing theoretical models in Section 2. In this section I examine to what extent predictions from these models outlined in Chapter 2 correspond to Serbian data. Drawing on my findings, I argue that the existing models do not adequately represent the type of voicing contrast found in voicing languages and in Serbian in particular, and I highlight areas that are in need of improvement in each model. Furthermore, in line with previous criticisms of these accounts, I point out that they cannot include various instances of non-universal variability found in Serbian. I argue that, despite some fundamental theoretical differences, the existing models of the voicing contrast have in common that they all fail on those two counts. In Section 3, I propose that an approach that includes elements of exemplar-based models would be suitable to resolve these issues. In Section 4, I highlight a number of important issues for further research, as well as some limitations of the present study.

## 8.1 Acoustic correlates of the voicing contrast in Serbian – summary of results

### 8.1.1 Acoustic correlates that are relevant in Serbian

In this section I present a summary of findings from Chapters 4 to 7. For the purpose of comparison with the theoretical models, acoustic correlates that were found to be significant in Serbian are organised around each word position.

In utterance-initial position, VOT is a reliable correlate of the voicing distinction in Serbian stops. All phonologically voiced stops in this condition are realised as voiced, with negative VOTs. Phonologically voiceless stops are realised as voiceless, with positive (lag) VOTs. The two VOT categories are separated and, unlike in some other languages, there is no overlap between them. Difference in VOT between the two categories is very highly significant, in the pooled data and in data for each subject, and the effect size is large. There is a 146 ms difference between their means in the pooled data, and 94 to 184 ms difference between the means in the data for individual subjects.

In word-initial intervocalic position two acoustic correlates of voicing are relevant in Serbian, closure duration and duration of voicing in the closure. There is some overlap between closure durations for the two stop categories, but differences are significantly different in the pooled data, with medium effect size. Difference between means for the two categories is about 27 ms (or about 22% of longer CD). For seven subjects differences in closure duration are significant at the corrected level of 0.01, and for the remaining five subjects at the 0.05 level. In all cases the effect size is large. In the same environment, phonologically voiced stops are realised with mostly fully voiced closures, while phonologically voiceless stops are realised either with silent closures or with a short period of low-amplitude carry-over voicing. Differences in duration of voicing in the closure are very highly significant both in the pooled data and for each subject, and effect size is large in all cases. Phonologically voiced stops are realised with an average of 99% of the closure voiced, versus 10% in phonologically voiceless stops. This finding replicates the finding for stops in absolute initial position. In both conditions the contrast is between stops with fully voiced closures and stops with mostly voiceless closures.

Word-finally, all three acoustic correlates of voicing that were investigated are relevant in Serbian. In utterance-final position (i.e. word-finally in isolated words), which is considered to be detrimental for maintenance of vocal fold vibration, duration of voicing in the closure is different for the two stop classes. Phonologically voiced stops are realised with longer periods of voicing in the closure than phonologically voiceless stops: about 62% of /b, d, g/ closures is occupied by vocal fold vibration. On the other hand, in /p, t, k/ closures there is no voicing at all, or little voicing that is carried over from the preceding vowel - in total about 7% of closure duration is voiced. These differences are very highly significant in the pooled data and for each speaker, with large effect size in all cases. The second correlate is closure duration. Closure duration of phonologically voiceless stops is longer than that of phonologically voiced stops. This difference is very highly significant in the pooled data and in individual data, and the effect size is large. The separation between the two categories is larger and shows less overlap than in initial position, with difference between the means of 58 ms (or 37%) in the pooled data. Finally, vowels preceding phonologically voiced stops are longer than vowels preceding phonologically voiceless stops. The difference is larger for phonologically short vowels (27 ms in the pooled data or duration ratio of 0.8) than for phonologically long vowels (15 ms or ratio of 0.93). These differences are also very highly significant in the pooled data, with large effect size for phonologically short vowels and medium for phonologically long vowels.

All three correlates are also relevant in word-final intervocalic context. Phonologically voiced stops are realised with longer periods of voicing in the closure than phonologically voiceless stops, with 91% and 8% of the closure voiced, respectively. These differences are very highly significant in the pooled data and for each subject, and the effect size is large. Differences in closure duration are very highly significant in the pooled data and in individual data, with large effect size. Voiceless closures are about 33 ms (33%) longer than voiced closures in the pooled data. Preceding vowel duration is longer in the context of voiced stops than in the context of voiceless stops. Differences are larger for phonologically short vowels than for long vowels, as in isolation, with differences between means of 22 ms (ratio of 0.83) and 13 ms (ratio of 0.94), respectively. They are very highly significant in the pooled data, with large and medium effect size, respectively.

In word-medial intervocalic position, the same three measures function as correlates of the voicing contrast: closure duration, duration of voicing in the closure, and preceding vowel duration. Difference in closure duration between the two stop classes is very highly significant, with large effect size, and there is substantial difference between their means of 54 ms (or 42%). As in word-initial and word-final intervocalic position, phonologically voiced stops are realised with fully voiced closures, or with a short silent interval at the end of the closure, while phonologically voiceless stops are realised with closures that were predominantly voiceless. Preceding vowel duration as a correlate of voicing is relevant in word-medial position as well, both in isolated words and in the sentence condition. Voicing-conditioned vowel differences are 23 ms (ratio of 0.83) for phonologically short vowels, and 25 ms (ratio of 0.88) for phonologically long vowels. They are very highly significant in the pooled data, with large, and medium to large effect size, respectively.

Duration of VC sequence and C/V ratio as correlates of the voicing contrast were also examined in the present study, in word-medial and word-final position, but neither of the two proposals is supported by Serbian data. Duration of the VC dyad is not constant across conditions and phonological vowel length, and there is no negative correlation between closure duration and vowel duration. C/V ratio for each stop class is also variable across conditions.

In sum, the following correlates that are investigated in the present study are relevant for the voicing distinction: VOT/voicing in the closure and closure duration word-initially, and closure duration, duration of closure voicing and preceding vowel duration word-medially and word-finally. They are quite robust in the pooled data, as well as in the data for each speaker, which is confirmed by the statistical significance of results and large effect size in most conditions. Out of these correlates, only preceding vowel duration (in phonologically long vowels) is not used by all speakers to distinguish the two stop classes.

### 8.1.2 Variability in the realisation of the voicing contrast

As mentioned in Chapter 2, in the existing theoretical models of the voicing contrast there is a tendency to associate most of the variability found in the phonetic realisation of the contrast to universal, biological or aerodynamic factors. Language-specific or speaker-specific variation has received little attention and has not been adequately addressed, with the exception of some authors, such as Keating, Cho & Ladefoged, and Kohler, who acknowledged a possibility that a separate set of rules would be needed to account for this variation, but did not develop this further.

In this section, I will summarise the variability found in the present study, and whether it can be explained by universal processes or whether it is language- or speaker-specific. The relevance of these findings for the models of the voicing contrast is further discussed in Section 8.2.

**Linguistic factors**

Results presented in Chapters 4 to 7 reveal that place of stop articulation and the following vowel environment both induce variability in the phonetic realisation of the voicing contrast in Serbian, but that condition (isolation or sentence frame) does not.

Lack of any difference between the two conditions probably comes from the fact that the study is based on controlled speech, which resulted in similar production in both conditions.

The effect of place of articulation on several correlates of voicing in Serbian can only partially be attributed to universal processes. For example, place-related VOT differences in /p, t, k/ support aerodynamic and physiological explanations summarised by Cho & Ladefoged (1999), except the proposal that place related differences in VOT result from the tendency to keep the voiceless interval (VOT+CD) uniform (Weismer, 1980). In Serbian, although duration of the voiceless interval for /p, t, k/ was found to be fairly uniform, VOT and CD are not inversely related.

However, for closure duration and duration of voicing in the closure, the place effect is less straightforward. In /b, d, g/ in utterance-initial and utterance-final position, there is no place effect on either closure duration or on the duration of voicing in the closure/prevoicing. In initial intervocalic and final intervocalic position there is a place effect on both closure duration and on the absolute duration of voicing in the closure (because stops are mostly fully voiced).

Results for the closure duration of phonologically voiceless stops in word-initial intervocalic position and utterance-final position parallel results for phonologically voiced stops – there are significant place-related differences in initial intervocalic position, but there are no differences in final position. However, they differ from results for /b, d, g/ in final intervocalic position, in that for /p, t, k/ there are no place-related differences, but for /b, d, g/ there are.

The effect of the vowel environment was found only in two correlates: VOT and preceding vowel duration.

There is no effect of the following vowel on prevoicing duration or frequency of prevoicing in /b, d, g/, but there is a significant effect on VOT in voiceless stops. In agreement with data from a number of other languages, VOT for /p, t, k/ is higher before high vowels /i/ and /u/ than before the low vowel /a/. This result supports aerodynamic and physiological explanations for this effect discussed in Section 4.4.2. However, the effect of the following vowel interacts with the effect of place of articulation, which is a language-specific effect.

For voicing-conditioned vowel duration phonological vowel length was found to have an effect on the realisation of this voicing correlate in word-final and word-medial position (although it was not possible to control for the vowel quality). The effect is larger for phonologically short vowels than phonologically long vowels. This effect seems to be language-specific, because it is in the opposite direction from the effect found in English (Crystal & House, 1982; House, 1961; Luce & Charles-Luce, 1985; Peterson & Lehiste, 1960), and is not in agreement with Klatt’s (1973) proposal that inherently short and long segments shorten by equal proportions in certain environments (including the position before stops with different voicing).

**Speaker factors**

In the present study individual differences between subjects were found in the production of all correlates of voicing. These differences come partly from differences in gender, age, place of birth, and to a smaller extent from speaking rate, but there is a certain degree of between-subject variation which represents individual features that were not captured by the above factors. In some cases these individual features are such that they dominate other factors, as has been discussed in the results chapters (for example with regard to VOT results by DARf).

On the other hand, despite individual differences, for most correlates in the present study the voicing contrast is robust for each subject (with the exception of preceding vowel duration in phonologically long vowels, as mentioned above).

An effect of gender was found in all acoustic correlates investigated in the present study.

The finding that male subjects produce significantly longer positive VOT, but females tend to produce longer prevoicing argues against proposals that are based on anatomical and physiological differences between men and women (see Literature review in Chapter 1 and Section 4.4.3). Results from the present study indicate that both prevoicing and lag VOT are under speaker control. Further research is needed to establish whether in Serbian longer prevoicing produced by female speakers could be explained by the tendency of female speakers to use clear speech, as was suggested by Helgason and Ringen (2008), or whether gender effect acts as a sociophonetic marker, as was suggested by Oh (2011).

In the present study female subjects produced longer closures for both stop classes in all environments. This difference is partly due to fact that females as a group are slower talkers. When the effect of speaking rate was co-varied statistically in initial intervocalic position, the effect was reduced, but still significant. This finding reinforces findings by Zue and Laferriere (1979) that in careful speech women tend to produce longer segments. In /b, d, g/ produced word-initially in the sentence condition, nearly all closures are fully voiced, and female subjects produce longer closures and consequently longer voicing in the closure.

Finally, there is a small, but significant gender-related difference in the effect on preceding vowel duration, with male subjects having a larger effect than female subjects. This could partly be due to females being slower talkers, but it cannot fully account for the effect. Although females do produce longer vowels than males, they also produce smaller voicing-related differences in vowel duration.

In the present study, the effect of age is present, but not in a systematic way, which is understandable because the oldest subjects in this study are below the age at which physical changes associated with normal ageing can have a considerable effect on the production of the voicing contrast.

Results for the four oldest speakers in the present study support findings that older speakers become more variable in production of prevoicing, but not in production of voiceless stops (Ryalls, Cliché, et al., 1997; Sweeting & Baken, 1982). However, this does not result in the reduction of the separation between the two voicing categories, because at the same time they produce longer prevoicing, thus increasing the separation. On the other hand, results for Serbian do not support the proposal that older speakers produce shorter prevoicing because of smaller lung volumes (Hoit, et al., 1993; Ryalls, Cliché, et al., 1997; Ryalls, et al., 2004). In fact, the oldest four speakers produce prevoicing comparable to that produced by other speakers.

Production of longer prevoicing in the older subjects might be governed by sociolinguistic factors. In the realisation of /b, d, g/, they not only have longer prevoicing, they also produce longer closures in initial intervocalic position, which are fully voiced, as well as longer periods of voicing in word-final position, compared to younger subjects. There is also an effect of age on closure duration, where older subjects tend to produce longer closures for both stop classes. This tendency for older subjects to produce longer prevoicing, longer closures and longer duration of closure voicing suggests that they might have attempted to produce “more clear” or “more correct” speech. Although it has been argued that clear speech characterises female speakers, there is no reason why this could not be true for any other social (or age) group.

In addition to this, there is an indication that certain regional differences in VOT production might be present in this data set. This study was not designed to examine these differences, and the sample of twelve subjects is too small for generalisation, but this might be a topic for further research.

Place of living, on the other hand, does not have any considerable effect on the realisation of the voicing contrast. Results for the two male subjects who live in the UK are similar to results for four male subjects who live in Serbia.

Overall, there seems to be a complex interaction between age, gender, and other sociolinguistic factors, and possibly even some individual features of production, that result in the findings reported in the present study. Acoustic phonetic research has recently brought these issues to attention, with some authors examining interaction between age, gender and race, and some authors investigating a host of social factors, including age, gender, class, ethnicity, country of origin and geographical location (Docherty, et al., 2011; Ryalls, et al., 2004; Ryalls, Zipprer, et al., 1997).

In sum, only a small proportion of observed variation can be attributed to universal, biological or aerodynamic factors. Most of the variability and the interactions between factors are language-specific or controlled by a speaker, which poses a problem for the theoretical models discussed in the next section. These models especially minimise the importance of individual differences in production that are non-distinctive, such as those found in the present study.

## 8.2 Results for Serbian in relation to theoretical models of the voicing contrast

In this section I evaluate whether predictions made by the models about the phonetic realisation of the voicing contrast in Serbian (or a language with the same type of the voicing contrast) can account for my findings. Furthermore, I discuss wider implications of my results for the models.

### 8.2.1 Keating's (1984) model

As was discussed in Section 2.2.1, Keating's model proposes that there are three major phonetic categories, {voiced}, {voiceless unaspirated} and {voiceless aspirated}, which correspond to the traditional VOT categories of voicing lead, short lag and long lag, and which are considered to be fairly uniform across languages.

Keating's model predicts that word-initially in Serbian [+voice] stops are realised as {voiced}, and [-voice] stops as {vl. unasp.}. Keating further argues that in languages of this type there is little allophonic variation and that the {vl. unasp.} category is situated in a narrow VOT area. Results presented in Chapter 4 show that this is not the case in Serbian, where there is a split within the [-voice] category so that /p/ and /t/ are mostly realised as {vl. unasp.}, but /k/ straddles the {vl. unasp.} and {vl. asp.} category. This spreading of the phonetic category for /k/ is consistent for most subjects, with VOT values of up to 80 ms. Even VOT values for /p/ and /t/ are somewhat higher than is usually expected for voiceless unaspirated stops, and for some subjects go up to 60 ms (overall, there is more between-subject variation for /p/ and /t/). Therefore, Serbian exhibits not only phonetic spreading of {vl. unasp.} category as a whole, but also separation within this category in which VOT values for the velar straddle the {vl. unasp.} and {vl. asp.} phonetic category. Keating's model cannot account for such separation.

For a similar result in Polish, where VOT values for /k/ occupy a range between about 20 ms and 100 ms, and even /p/ has VOT values of up to 70 ms, Keating suggested that it resulted from "high vowel contexts or from extra emphasis, or for no apparent reason other than spreading over the phonetic space, as Pol. /k/ does" (1984a, p. 298). However, in Serbian this result does not come from high vowel context, because the effect of the following vowel is not significant for /t/ and /k/ in isolation, and generally the effect of the following vowel on VOT in the pooled data is small. This is supported by results from Cho and Ladefoged (1999), who analysed stops in non-high vowel context in order to avoid possible effect of vowel height on VOT, and found large variation within the voiceless unaspirated category in a number of languages. Extra emphasis as a factor can also be excluded for Serbian, because material that was used and method of data collection were controlled.

Furthermore, although in some languages that contrast {voiced} and {vl. unasp.} stops the {vl. unasp.} category seems to be constrained within a narrow VOT range, a number of studies have reported VOT values for this category that fall between the traditional short and long lag categories. They are often referred to as intermediate VOT values, as mentioned in Section 1.1. These intermediate VOT values occupy position between short lag and long lag VOTs (or, more precisely, short lag values spread towards long lag values). A summary of results from several studies[21] which reported intermediate VOT values for voiceless unaspirated stops in a number of languages is shown in Table 8.1 and Table 8.2[22]. It suggests that this kind of phonetic spreading is not as rare as it may seem. There are many languages, apart from Serbian, that exhibit this spreading, some of them fairly consistently. Table 8.2 presents data for a subset of languages from Cho and Ladefoged (1999) that can be considered as having

[21] Some studies used bilingual speakers, or both monolingual and bilingual speakers. If results for monolingual speakers were available, only they were included in the table. Results for bilingual speakers were included if only bilingual speakers were used (Flege & Port, 1981; Raphael, Tobin, & Most, 1983), or if results for bilingual and monolingual speakers were similar (Caramazza, et al., 1973; Raphael, et al., 1995).

[22] Means of 30 ms or more are included in the tables as representative of intermediate VOT values. This is in agreement with some recent studies, which focus on intermediate VOT values and explicitly use this term (Raphael, et al., 1995; Riney, et al., 2007). Riney et al. stay true to Lisker & Abramson's (1964) definition, and consider values between 25 and 60 ms to be intermediate VOT values. Raphael et al. take the zero onset/short lag category to be between 0 and 30 ms, and long lag category to be 50 ms or more, and all VOT values between 30 and 50 ms are considered to be intermediate VOT values. By using 30 ms as a cut-off point in the present study, all languages that have been discussed by other authors as having intermediate VOT values are included (although they would not always be considered as such according to Keating, who considered the short lag VOT category to be up to about 20-35 ms). Another reason is that if the mean VOT value for a particular category is at least 30 ms, the range of VOT values must include some lower as well as some higher values, which represent phonetic spreading of the {vl. unasp.} category.

intermediate VOT values for the {vl. unasp.} stops (the table includes languages that contrast {voiced} stops with {vl. unasp.} stops).

As can be seen from the tables, intermediate VOT values are not exceptional, but in fact are a frequent phenomenon. On the whole, there is an expected tendency for the velar /k/ to have longer VOT values than the other two stops, and this is true for the most of the data presented here, but in some languages even /p/ and /t/ have values higher than one would expect for short lag category (for example Serbian, Catalan, French, Hebrew, Hungarian, Japanese, Polish).

Although in some of the cases these results could be attributed to effects of bilingualism, it cannot be the only explanation. For example, results for monolingual and bilingual adult speakers of Hebrew are similar to monolingual children's results (10 to 11 years old)[23]. Results from the present study also suggest that even in the cases where subjects live in a country where a language with different VOT categories is spoken, and the subjects are fluent in their second language and use it on a daily basis, VOT categories in their first language are not necessarily affected to a large degree, and many factors determine their VOT production (Section 4.4.3).

[23] In addition to this, Raphael et al. (1995) point out that for some of their bilingual (or L2) speakers, other languages did not always have long lag stops (other languages spoken include Hungarian, Yiddish, Russian and Polish), so the intermediate VOT values in production of Hebrew stops could not be explained by bilingualism.

| Language | | /p/ | /t/ | /k/ |
|---|---|---|---|---|
| Serbian (Present study) | | 22 | 27 | 52 |
| Arabic, S. A. (Flege & Port, 1981) | b | - | 37 (20-65) | 52 (30-85) |
| Catalan (Recasens, 1985) [a] | | 23 | 27 | 47 |
| French (Nearey & Rochet, 1994) $ | | 32 | 35 | 46 |
| French (Yeni-Komshian et al. 1977) $ | m | 20 | 32 | 40 |
| French, Ca (Jacques, 1987) | | 10 | 35 | 33 |
| Dutch (van Alphen & Smits, 2004) | | 19 | 31 | - |
| Greek (Kollia, 1993)[b] | b | 19 | 27 | 49 |
| Hebrew (Obler, 1982) | m | 26 | 34 | 64 |
| Hebrew (Raphael, et al., 1995) | b | 28 | 36 | 56 |
| | m ch | 27 | 25 | 61 |
| Hungarian (Gósy, 2001) | m is | 25 (13-35) | 23 (15-38) | 50 (33-66) |
| | spont | 18 (9-29) | 27 (14-38) | 35 (22-69) |
| Hungarian (Gósy & Ringen, 2009) | m in | 10 (37) | 16 (38) | 37 (77) |
| | m med | 18 (67) | 20 (88) | 43 (73) |
| Japanese (Riney, et al., 2007) | m | 30 | 29 | 57 |
| Japanese (Shimizu, 1989) | | 44 (15-60) | 27 (15-90) | 68(45-100) |
| Japanese (Shimizu 1996 from Gósy 2001) | b | 41 (15-65) | 30 (15-30) | 66(50-100) |
| Polish (Keating et al., 1981) | m | 22 | 28 | 53 |
| Polish (Kopczyński 1977 from Rojczyk 2009) | | 38 | 33 | 49 |
| Portuguese (Lousada et al., 2010) | | 20 | 28 | 51 |
| Spanish, Puerto R (Raphael, et al., 1983)[b] | | 20 | 28 | 39 |

Table 8.1 Mean VOT (ms) for /p, t, k/ with intermediate VOT values

Note. Abbreviations: m = monolingual, b = bilingual, or no label if it was not specified, ch = children; in = word-initial or absolute initial position, med = word-medial position, spont = spontaneous speech. Values are rounded to the nearest ms, and the range is given in brackets, if available. Results marked with $ were calculated from original papers.

[a] Results were reported in Raphael et al. (1995).

[b] Kollia (1993) and Raphael et al. (1983) do not give numerical values, but they are presented in Raphael et al. (1995).

These results cannot be attributed to a high vowel context either. Because of the well-known fact that VOT tends to be higher before high vowels, some studies used non-high vowels (Caramazza, et al., 1973; Cho & Ladefoged, 1999; Flege & Port, 1981; Gósy & Ringen, 2009), while other studies, including the present study, used a balanced data set with both low, mid and high vowels (Gósy, 2001; Riney, et al., 2007), which minimises the possibility that VOT values were affected by this factor. Most importantly, in the present study the effect of high vowel on VOT values for /k/ is non-significant, although /k/ shows most spreading.

| Language | Bilabial | Dental | Alveolar | Velar |
|---|---|---|---|---|
| Banawa | | 22 | | 44 |
| Bowiri | 17 | | 18 | 39 |
| Chicksaw | 13 | | 22 | 36 |
| Defaka | 18 | | 20 | 30 |
| Yapese | 20 | 22 | | 56 |

Table 8.2 Mean VOT (ms) as a function of place of articulation for languages with intermediate VOT values from Cho & Ladefoged (1999)

To sum up, evidence about intermediate VOTs in Serbian and other languages presented in Tables 8.1 and 8.2 does not support Keating’s claim that the {vl. unasp.} category is restricted to a narrow VOT area, and that the three phonetic categories are discrete and well separated acoustically (at least this does not apply to the {vl. unasp.} and {vl. asp.} categories).

When intermediate VOTs are included, the following pattern of VOT distribution emerges. In languages with contrastive aspiration, such as English, [-voice] stops are realised as {vl. asp.} and this category is very stable and positioned in the long lag area of the VOT scale. On the other hand, [+voice] stops can be realised as either {voiced} or {vl. unasp.}, e.g. this category shows phonetic spreading across short lag and voicing lead values. In languages such as Polish and Serbian, [+voice] stops are realised as {voiced}, and this category is very stable in the voicing lead area of the VOT continuum. The [-voice] category, on the other hand, can be realised as {vl. unasp.} and/or with intermediate VOTs, e.g. with possible phonetic spreading into higher VOT values. Not every language uses this possibility of spreading consistently, but there are many languages that use it in at least some conditions and by at least some speakers - examples include data presented in Table 8.1 and Table 8.2, and the bimodal distribution of [+voice] category in English, Turkish and Persian, discussed in Section 1.1. In addition to this, Keating (1984) and Jessen (1998) reported that some speakers of German realise [+voice] stops utterance-initially as {voiced} instead of {vl. unasp}.

This pattern supports Keating's rule of polarisation of adjacent phonetic categories, despite the fact that their discreteness might be controversial. In other words, it could be argued that in aspirating languages the {vl. asp.} category is well separated, or polarised, from the other category, realised through non-aspirated (voiced or voiceless) stops. In voicing languages the {voiced} category is polarised from the category represented by *general-lag* stops (Keating's term)[24]. Each group of languages has one phonetic category firmly placed at one of the two extreme ends of VOT continuum (lead VOT or long lag VOT), and since it is clearly defined acoustically and articulatory, the other category can be variable and can show phonetic spreading without the loss of the contrast (phonetic spreading is optional, or is conditioned by context in some languages). Swedish is a special case – in utterance-initial stops it uses the two categories that are best separated along VOT dimension, that is {voiced} and {vl. asp.} (Beckman, et al., 2011; Helgason & Ringen, 2008), which is an extreme application of the rule of polarisation.

This way of looking at the patterning of phonetic categories, where the contrast is either between voiceless aspirated stops and all other (unaspirated) stops, or between prevoiced and voiceless stops (unaspirated and/or with intermediate VOTs), could be seen as supporting some proposals in phonology that the laryngeal features are privative, not binary. A privative feature [voice] represents a contrast between the presence and absence of a feature, for example between [voice] and [∅]. Beckman et al. (2011) argue that two privative features, [voice] and [spread glottis], can explain the patterning of VOT results in languages with a two-way contrast (and even in some languages with a three-way contrast, such as Thai), as well as the effect of speaking rate on the three phonetic VOT categories, found in a number of languages. Recall from Section 1.1 that previous research has found that VOT values of lead and long lag stops changed with speaking rate in French, Thai, and English, but this was not the case with

[24] A small number of studies reported exceptions to this pattern, as mentioned in Chapter 1 (Caramazza, et al., 1973; Lousada, et al., 2010; Ringen & Suomi, 2012; van Alphen & Smits, 2004).

short lag stops. In Swedish, Beckman et al. found that both VOT categories changed at slower rates. Because only VOT of short lag stops remained unchanged with speaking rate, Beckman et al. argue that this is the unmarked category in these languages, and that prevoiced and long lag categories are marked. In this view, languages such as Serbian, French or Polish have a contrast between [voice] and [∅], English has a contrast between [spread glottis] and [∅], and Swedish has contrast between [voice] and [spread glottis]. If we are to relate this view to Keating's categories, then the contrast between [voice] and [∅] on the phonetic level would be represented as the contrast between Keating's {voiced} category and its absence. In languages that contrast [spread glottis] and [∅] this would represent the contrast between Keating's {vl. asp.} category and its absence. Finally, in Swedish, this would be the contrast between Keating's categories {voiced} and {vl. asp.}. Phonetic data from Table 8.1 and Table 8.2 generally supports this view.

However, the proposal that [voice] is a privative feature has been criticised by phonologists on several grounds (see, for example Kim, 2002; Wetzels & Mascaró, 2001), but one point in particular is relevant for Serbian. Namely, heterosyllabic obstruent clusters in Serbian agree in voicing, so that both obstruents are either voiced or voiceless. This means that the feature value [-voice] is phonologically active in Serbian. In assimilatory processes it exhibits parallel phonological behaviour and has the same role as the feature value [+voice] (and in some other languages, including Romanian, Hungarian, and Yiddish). This suggests that both feature values are necessary and that phonological feature [voice] is binary in Serbian. Referring back to Keating's model, Serbian results give support to Keating's idea that phonetic categories are polarised in their physical realisation, but they do not argue against the binary nature of the feature [±voice] proposed by Keating. A privative feature [voice] (and [spread glottis]) might be more appropriate for some other languages, but this does not seem to be the case with Serbian.

Results from the present study also support the criticisms of Keating's model concerning the lack of phonetic detail on the level of phonetic realisation (Cho & Ladefoged, 1999; Docherty, 1992). As was shown in Chapter 4, there is a lot of variability in VOT in Serbian coming from a number of factors: place of articulation, quality of the following vowel, between-subject differences, gender, age, and place of birth. Only some of them can be explained by universal constraints, such as the effect of place of articulation on VOT in voiceless stops, and the effect of the following vowel on VOT in voiceless stops (to some extent). All other effects are either speaker- or language-specific, and need to be specified separately, but this aspect is not elaborated in Keating's model.

### 8.2.2 Kohler's (1984) model

For utterance-initial position, results from Serbian agree with Kohler's prediction that the voicing contrast would be expressed through presence vs. absence of vocal fold vibration. For utterance-final position, Kohler proposes that articulatory timing is more relevant than laryngeal power and that the voicing contrast is expressed through differences in closure duration and preceding vowel duration, while voicing in the closure is optional (or absent). In Serbian, relevant acoustic correlates of the voicing contrast are not only closure duration and preceding vowel duration, but also voicing in the closure. Phonologically voiced and voiceless stops are realised with systematic and statistically significant differences in voicing in the closure and in closure duration by all subjects (with large effect size). Preceding vowel duration, on the other hand, is somewhat less reliable, because of smaller voicing effect in the pooled data for phonologically long vowels (medium effect size), and because for two subjects overall voicing effect is not significant (as summarised in Section 8.1.1). That is, all three correlates are relevant in Serbian, not only two, as Kohler suggests, but preceding vowel duration is less reliable. Duration of VC sequence, for which Kohler claims to be constant, and even considers it to be a phonological universal, does not have a uniform duration in Serbian, and the two durations are not inversely correlated.

For intervocalic position, Kohler's model predicts that both components, articulatory timing and laryngeal power, are equally important. Thus, closure duration and preceding vowel duration, as well as closure voicing, are expected to be relevant acoustic correlates in this position. In general, this prediction is confirmed by the Serbian results. However, of the three correlates, voicing in the closure and closure duration are likely to be more important than preceding vowel duration, for reasons mentioned above. In word-initial and word-final intervocalic stops closure voicing and closure duration are statistically relevant as acoustic correlates. Preceding vowel duration was not measured in word-initial intervocalic position, but in medial and final position it was also significant, with some exceptions discussed above. This suggests that both articulatory timing and laryngeal power are relevant in Serbian in all word positions, but that both components of articulatory timing are not equally important. This is different from Kohler's proposal.

In addition to this, Kohler's model makes little reference to differences between universal and language-specific sources of variation in the realisation of the acoustic correlates in question, as is the case with Keating's model. As a consequence, variability that was summarised in Section 8.1.2 cannot be included in the model.

### 8.2.3 Jessen's (1998) model

As was discussed in Chapter 2, Jessen proposes that voicing in the closure is the basic correlate for the feature [±voice], and that substitute (non-basic) correlates are closure duration and preceding vowel duration. Substitute correlates are defined as correlates that are contextually more limited than the basic correlate, but can replace it in certain contexts.

In Serbian, voicing in the closure was found to be a relevant correlate in all word positions that were investigated, and this is in agreement with Jessen's proposal. Because voicing in the closure (and prevoicing in utterance-initial position) was a significant correlate in all contexts and for each subject, voicing in the closure is therefore the basic correlate in Serbian. Voicing in the closure also satisfies the strong version of the principle of contextual stability (outlined in Section 2.2.4), and consequently voicing in the closure is also the phonetic invariant (or the common denominator) in Serbian stops.

Further, not only voicing in the closure, but also closure duration is relevant in all contexts that were investigated in the present study, and it was significant for each subject. As a result, closure duration also satisfies the conditions for designation as a basic correlate. In Serbian, the status of closure duration is more important than suggested by Jessen's model, which proposes that it is a substitute correlate. Preceding vowel duration, on the other hand, is not a basic correlate, because the effect was not statistically significant for all subjects.

The perceptual role of these correlates remains to be established for Serbian to confirm their status within Jessens's model. Research by Kingston and Diehl suggests that their perceptual relevance could be universal, but this needs to be confirmed for Serbian.

This analysis reveals that both voicing in the closure and closure duration are basic correlates and phonetic invariants in the realisation of the Serbian voicing contrast. If this is the case, the model is faced with several challenging questions: First of all, if both voicing in the closure and closure duration are candidates for the role of the basic correlates, what is the definition of the common denominator? Jessen's view is that "both the basic and the non-basic correlates are comprised under the common denominator definition of the distinctive feature" (1998, p. 261). While this is true for his feature [±tense], where duration is the common denominator, it does not apply to his feature [±voice], where the common denominator is proposed to be presence vs. absence of voicing.

Second, what is the status of preceding vowel duration in Serbian? Is it a non-basic, i.e. substitute correlate, and is there a need for a non-basic correlate? With two basic correlates that are relevant in all contexts, its role as a replacement for the basic correlate(s) is redundant. In Serbian it seems to have a role of enhancing the contrast in certain contexts. The remaining option within the model is to assume a role of concomitant correlate for preceding vowel duration. According to the definition of concomitant correlates, they occur in the same contexts as the basic correlate because they are a consequence of the basic correlate. However, preceding vowel duration is unlikely to be a concomitant correlate, for two reasons: because it does not occur in all contexts where the two basic correlates occur, and because it is unclear whether it is a consequence of a basic correlate. Although some studies argued that vowels are longer before voiced stops because of the longer time needed for precise laryngeal adjustment in the production of vocal fold vibration, this proposal did not receive support from subsequent studies, as was discussed in Section 7.6. It is also unlikely to be a consequence of differences in closure duration, especially because there is no reciprocal relationship between closure duration and vowel duration in Serbian, as was shown in Section 7.6. Jessen's model would need to be reconstructed to reflect these facts, but it would then lose the symmetry between the feature [±voice] and the feature [±tense].

This is another example how results from aspirating languages were assumed to be relevant for voicing languages, in the same contexts and with the same status, without being supported by a representative body of research. While there is no question that both closure duration and preceding vowel duration have a role as correlates in

Serbian, their importance is different. In fact, results from the present study, obtained following Jessen's methodology, show that closure duration is as important as voicing in the closure, but that preceding vowel duration is not.

Results from the present study further suggest that closure duration might have a more important role than previously thought in voicing languages in general. Closure duration as a correlate was usually associated with aspirated languages and the feature [±tense], so much so that if it was found that a voicing language had significant closure duration differences between the two voicing categories, the possibility that it uses the feature [±tense], not [±voice] was considered, for example for European Portuguese, Spanish and French (Jessen, 1998; Veloso, 1995). In fact, it might be that its relevance in voicing languages has been underestimated. As was discussed in the literature review in Section 1.3.2 and in Section 5.13, in a number of voicing languages phonologically voiceless stops are realised with longer closures than phonologically voiced stops. This is true for all three word positions, and differences that were measured were comparable to those in aspirating languages, and in some cases even larger and more consistent (for example in word-initial intervocalic position closure duration was found to be a reliable correlate of the voicing contrast in Serbian,  French, Portuguese and Arabic, but not in English).

In addition to this, closure duration and preceding vowel duration are usually considered to be relevant in the same contexts and to be inversely related to each other, which is a conclusion based mainly on English data. In Serbian, however, closure duration and closure voicing are relevant in the same contexts. What is more, it is not simply presence or absence of closure voicing that distinguishes the two stop categories in Serbian, but duration of this voicing as well, both in milliseconds and as a percentage of closure duration.

In sum, Jessen's model operates with a comprehensive set of acoustic correlates, most of which have been confirmed to be relevant for both aspirating and voicing languages, and is convincing in the part related to the feature [±tense]. However, it is unable to explain results from the present study. In order to do so, the symmetry between the feature [±tense] and the feature [±voice] needs to be re-examined, as well as some of its crucial elements, such as the role of closure duration and preceding vowel duration, and definition of the common denominator.

### 8.2.4 Auditory enhancement hypothesis by Kingston and Diehl

The auditory enhancement hypothesis by Kingston and Diehl makes a prediction similar to that of Jessen's model, as far as the actual correlates of voicing are concerned (although not how they combine to specify the feature [±voice]): VOT is the relevant correlate in absolute initial position, and closure voicing, closure duration, preceding vowel duration, f0 onset and F1 onset in word-medial and word-final position. This is a very vague prediction for medial and final position, because it includes all possible correlates. However, in terms of how correlates are combined in intermediate perceptual properties (IPPs), the model needs to be revised in order to include results for Serbian.

First, relevance of C/V interaction as a subcorrelate, and the C/V duration ratio as an IPP in Serbian has to be questioned based on the data from the present study. Kluender et al. (1988) argue that the perceptual role of the vowel duration differences is to enhance the closure duration contrast. This conclusion is based on a perceptual experiment with two sets of stimuli, one with a long preceding vowel, and one with a short preceding vowel, where silent closure duration was varied. A change of 90 ms in vowel duration resulted in the /apa/-/aba/ boundary being shifted by about 10 ms. However, vowel duration differences in Serbian are much smaller, and do not exceed 30 ms, while closure duration differences are about 50-55 ms (Chapter 7, Table 7.3 and Table 7.4). Closure duration and preceding vowel duration are inversely related, so that voiced stops are realised with shorter closures and longer preceding vowels and vice versa, but the two measures are not negatively correlated. Based on the production data, the effect of the durational contrast is likely to be small in Serbian. Further perceptual experiments are necessary to establish if there is perceptual enhancement between these two correlates in Serbian and its extent.

Second, although relevance of both closure voicing and closure duration is acknowledged in the model, and they are included as subcorrelates of C/V duration ratio, their relationship is not given sufficient emphasis. Even though Parker et al. (1986) found that presence of voicing in the closure leads to closure duration being perceived as shorter and suggests a voiced percept, there is no separate subcorrelate for this effect. These two correlates are related only indirectly, via the IPP of the C/V duration ratio. This is in contrast to the relationship between closure duration and preceding vowel duration, for which a separate subcorrelate (C/V interaction) is included in the model. Based on Serbian data, it could be argued that this effect was underestimated in the model, and that it should be part of the model.

To sum up, it is likely that the perceptual role of differences in preceding vowel duration is small in Serbian because closure voicing and closure duration are both robust acoustic correlates to the voicing distinction. As a consequence, the existing IPP of C/V duration ratio, with its focus on this effect, and with insufficient attention to the role of closure duration and closure voicing, is unable to represent Serbian data.

### 8.2.5 Summary

There is a striking similarity between the models discussed in this section – they basically operate with the same set of acoustic correlates (with the exception of VOT-based models), although they use various categories/terms, such as IPPs, basic and non-basic correlates, components of articulatory timing and laryngeal power etc. in attempting to organise acoustic correlates into a coherent theory. These models suffer from the same type of problems because they are based on the same way of thinking and essentially represent the same type of model: they assume that on the phonological level there is a small set of invariant and abstract representations, and that on the physical level there are realisations that contain all relevant phonetic details, with a possible set of rules or categories that specify the relationship between the two levels. One of their main problems is their inability to account for variation that is not universal, whether it is language-specific or speaker-specific or even sociolinguistic, such as sub-phonemic variation found in the present study. To overcome this problem, they would benefit from inclusion of an element of an exemplar-based approach, which argues in favour of phonetically rich phonological representations. I discuss exemplar-based models in the next section.

The existing models of the voicing contrast also have in common that they focus on the type of contrast found in aspirating languages such as English and German, and are less able to address patterns of realisation in voicing languages. The choice of acoustic correlates, their relationship, and their relevance reflects this bias, as was discussed above with regard to the position of closure duration in these models, and its relationship with preceding vowel duration and closure voicing.

## 8.3 General discussion

The findings of the present study point to several issues relevant for the area of modelling of the voicing contrast in stops.

First, the literature review in Chapters 1 and 2 and discussion of my results indicate that realisation of the voicing contrast in voicing languages is under-researched and that it is under-represented in the existing theoretical models. This is true not only for individual languages, where comprehensive research is lacking, but also for between-language variability in this group, which was examined in previous chapters (such as, for example, issues of various degrees of devoicing found in Portuguese and some other languages, or differences in the effect of place of articulation on various correlates of voicing, etc.). A growing body of research on voicing languages in recent years has suggested that patterns of realisation of the voicing contrast are more complex than previously thought, which in turn calls for more complex models. Current models cannot cope with these demands without a serious re-examination of some of their basic assumptions, as was shown in the case of Serbian (Section 8.2). The inability of existing models to account for Serbian data further highlights the fact that they are skewed towards the type of contrast found in English and thus unable to account for the patterning found in voicing languages.

Second, results from the present study argue that even variability that was previously considered to be universal, such as the place of articulation effect on duration of prevoicing and voicing in the closure, has to be re-examined in the light of new results. The place of articulation effect is often attributed to passive aerodynamic processes, but in fact in voicing languages these interact with active voicing in complicated ways that have not been sufficiently researched. This is another aspect of the realisation of the voicing contrast in voicing languages that remains unaccounted for by the existing models.

Third, putting more general issues with voicing languages aside, the analysis has shown that current models are unable to account for the patterns of realisation found in Serbian, which presents another challenge for these models. This is true for the choice and hierarchy of acoustic correlates employed in Serbian, and it is particularly true for the several types of non-contrastive variability in Serbian that is non-universal and therefore cannot be explained by existing models.

For instance, this includes language-specific variability, such as the effect of place of articulation on duration of prevoicing and on closure duration, the interaction between the effect of place of articulation and the quality of the following vowel on VOT, difference in voicing-conditioned vowel duration in phonologically short and long vowels, absence of an inverse relationship between VOT and closure duration, and between closure duration and preceding vowel duration. The finding that voiceless stops in Serbian are realised with intermediate VOTs, although present in a number of other languages, could also be regarded a language-specific feature.

This is also the case with the considerable between-subject variability that was found in the realisation of all correlates that were investigated. Although all other potential factors could not have been ruled out, there is a certain degree of variability that is speaker-specific, and which, if not taken into account, can lead to misinterpretation of the group findings.

The same applies to other types of variability discussed under the heading of speaker factors, such as the effect of gender, age, and place of birth. Gender- and age-related differences in production are especially interesting because my results do not fully support explanations that are based on biological differences between men and women and on biological manifestations of normal ageing process, and they suggest that other factors, possibly sociolinguistic in nature, may be relevant as well. At this point it is unclear to what extent this variation is correlated with social categories of gender and age, whether speakers are aware of it, and whether they assign any social meaning to it, but the systematic presence of some of these differences and their statistical significance certainly suggest that they are worth exploring further. In fact, the finding of the present study that biological and social aspects of factors such as gender and age interact in complex ways with each other, and with speakers' expression of their own identity through language, and possibly with some other social factors (such as place of birth in the present study) has already been brought to attention with regard to modelling of sociophonetic variation (Docherty, et al., 2011; Foulkes & Docherty, 2006; Harrington, 2006; Johnson, 2006; Pierrehumbert, 2006).

Moreover, this finding is in agreement with research which, starting from different positions, suggests that social and linguistic information is entwined in lexical representations and which is consistent with an exemplar-based approach to modelling phonological knowledge (Docherty & Foulkes, fc; Foulkes & Docherty, 2006; Pierrehumbert, 2006). Proponents of an exemplar model of phonological knowledge argue that existing models of speech production are unable to account for numerous sources of variability in the speech signal, and for speakers' ability to represent, produce, perceive, and interpret this variability. The basic premise of this approach is that lexical representations consist of detailed traces of previous experiences an individual has had. Each exemplar representation is phonetically rich and contains not only linguistic, but also non-linguistic information, including sociophonetic variability. All these pieces of information are intrinsically connected and therefore variability is an inherent property of stored memories. In other words, such a representation includes, among other things, information about language-specific and speaker-specific variation, and non-universal sub-segmental variation that has been found to be a problem for existing models of the voicing contrast (Hay, Nolan, & Drager, 2006; Pierrehumbert, 2002; Wedel, 2006).

It is noteworthy that, although the present study was not designed to test a single specific theoretical model, it aligns to a degree with an approach that would be taken within an exemplar framework. The methodology used in the present study was such that not only results for the pooled data were discussed, but also between-subject differences and individual results, especially how they contribute to the pooled results and whether the pooled results are representative of the actual production of individual speakers or groups of speakers. It is important to consider individual results in interpreting results in the pooled data, because statistical analysis can obscure individual results and influence the conclusions of an experiment. This is especially true in a situation like that encountered in the present study, where there is a lot of between-subject variation that cannot be fully explained by other factors and is likely to represent speaker-specific features (which can be related to their attitudes, personality, values, identity, or cannot be explained at all). Having said that, it is also important to consider this variability in the face of requirement for contrast maintenance and how they relate to each other. In the present study, despite large between-subject variability in the production of certain acoustic correlates, the voicing contrast was preserved and robust in all speakers. This is an important issue for a description of a contrast in any language.

This view of the importance of individual speakers is shared with the exemplar approach. An exemplar model is a model of "how *individuals* develop and continue to evolve their representation of the meaningful sound patterning to which they are exposed" (Docherty & Foulkes, fc, p. 17). By emphasising the individual, this approach challenges traditional phonetic research which is predominantly based on pooled data of presumably homogeneous groups of speakers. In this view, a certain degree of variability, coming not only from social factors but also from factors such as attitudes, ethnicity, ideology etc., is seen as a way of expressing a speaker's identity (Docherty & Foulkes, fc). This view is supported by my findings in respect of individual variation, especially with regard to VOT results for subject DARf, which suggest that these individual features can override other factors, universal or sociophonetic, and are essentially an expression of the subject's individual speaking style.

One of the challenges for the exemplar model is how it can be related to well-established abstract phonological categories, such as phonemes. Integration of the existing models of phonological knowledge with the exemplar-based model has been proposed in the form of a hybrid model, which combines traditional abstract phonological categories with phonetically rich representations built up from the previous individual experiences. This type of model is still under development, but has a potential of overcoming a number of problems in traditional models (Docherty & Foulkes, fc; Pierrehumbert, 2006).

Although a hybrid model has not been developed in relation to the voicing contrast, current models of the voicing contrast could also be improved using such an approach. The main advantage of a hybrid model of the voicing contrast would be its potential to include any form of variability (and the inability to account for observed variation is one of the main problems of the existing models). According to such a model, individuals obtain knowledge about variability through the process of language acquisition and through language usage. Phonological categories (voicing categories in this case) are established based on generalisations about probability distributions of stored experiences, as is the knowledge of any interaction between these categories and linguistic or non-linguistic factors. This knowledge is continuously updated to reflect each individual's language experience. In such a model it would be possible to account for between-speaker differences in the production of acoustic correlates, such as those found in the present study (each subject's language experience is different, which results in different distributions of stored examples). Further, because of multiple indexing of each stored memory trace, gender- and age-related differences that are not explicable by universal biological factors, and which were also a problem for the existing models of the voicing contrast, would be included in such a model (whether sociophonetic in nature or coming from other factors). Finally, language-specific variation caused by linguistic factors is also encoded in the model. All these types of variation would be explicitly represented in the model through probability distributions and indexing of remembered exemplars.

However, there are many aspects of the hybrid model that are currently undeveloped. Some of the most relevant questions are: how are associations between phonetic patterns in the incoming speech and relevant linguistic and non-linguistic factors formed, what is the role of an individual in the creation of these representations, how is this "bottom-up" process influenced by the existing phonological knowledge, how these representations evolve over the lifespan, and what do hybrid representations look like and how they are constructed (Docherty and Foulkes, fc).

The idea that an individual's phonological knowledge is continuously updated by their on-going language experience is consistent with the research on gestural drift in VOT reported by Tobin (2009a, 2009b) for Serbian-English and Spanish-English speakers and by Sancier and Fowler (1997) for a Portuguese-English speaker. On the other hand, in the present study there was no gestural drift in the production of the two speakers who live in the UK, which poses a challenge to an exemplar-based model. It suggests that some people are more acutely sensitive about what they hear in their environment than others, and raises the question of the necessary conditions for gestural drift to occur. The two Serbian speakers in the present study might be assigning different weight to certain VOT patterns when they occur in English than when they occur in Serbian, but other speakers might not. This means that frequency with which a certain phonetic pattern is present in the surrounding speech is not the only factor that determines how it is stored in the memory, but that it also depends on the weight it is given and how it is indexed by the speaker. This underlines the relevance of questions raised above by Docherty and Foulkes (fc) about the process of association and the creation of stored exemplars, about the role of the individual, and about factors such as the existing phonological knowledge, language background, and attitude.

With respect to the hybrid model of the voicing contrast, it is currently unclear what such a model would look like, and how these phonetically rich representations would be integrated with the traditional abstract phonological categories. It is also unclear whether the concept of acoustic correlates and their relationships would be by-passed in such a model or not, and what the implications would be for the existing models of the voicing contrast in general, and for any particular language. Another question that remains is how to explain the fact that languages seem to adopt consistently the same ways of realisation of the voicing contrast, based on voicing or aspiration. It appears that these patterns are preferred, possibly because they are well adapted in conveying the contrast in noisy environments. The challenge for the hybrid model is how to link extensive variation on the one end and the limited number of patterns on the other. As it stands at the moment, the hybrid model has attractive features, but it raises as many questions as it provides plausible solutions.

## 8.4 Limitations of the present study and directions for further research

Although the voicing contrast is present in Serbian fricatives and affricates as well, due to space limitations this study was limited to stops, and based on a sample of controlled speech. For the same reason only a selection of acoustic correlates that have been found to be relevant in previous research were investigated, and in a limited number of contexts. Future research should be extended to include obstruent classes, correlates, and contexts that have not been researched, as well as speech produced in more naturalistic settings, including spontaneous speech. Some theoretically relevant issues that have arisen from the present study should also be explored further.

One of the main aims of the present study was to establish the most important acoustic correlates of the voicing contrast in Serbian stops. Future work on Serbian needs to provide a detailed account of acoustic correlates in the remaining two classes of obstruents, and to further evaluate the existing models of the voicing contrast, to the extent that they include other obstruent classes apart from stops. Fricatives are especially interesting, because some authors have proposed the same featural representation for fricatives as for stops (Kohler, 1984), while others have suggested that there is a syncretism between features [±voice] and [±tense] for fricatives that might be universal (Jessen, 1998). In this area, as is the case with the research on stops, voicing languages are under-represented and data from Serbian would be a valuable contribution to this theoretical issue.

Acoustic correlates that have not been included in the present study include properties of the stop burst, f0, and F1 frequency in the vowels preceding or following an obstruent. Further research is needed to establish their role in Serbian and their relevance in the models of phonological representation in general. Among them, properties of the burst have great potential, especially in the light of the findings that in Serbian stops are consistently and strongly released, and that acoustic correlates to voicing found in the stop itself are more important than preceding vowel duration (as was argued by Laeufer, 1992 for French as well).

The present study has also pointed out some areas of research that are likely to be sociophonetic in nature, such as gender and age differences in the realisation of several acoustic correlates in Serbian, and regional differences in VOT production. In addition to this, some of these factors interact with each other and with individual production. More studies, designed specifically for this purpose, would open up a wide area of research. If findings of the present study are supported by larger sets of data, they would have implications for not only study of Serbian, but also for theoretical issues discussed in the present study concerning the connection between universal, sociophonetic and speaker-specific aspects of speech production.

An area that has not been touched upon in the present study is the realisation of the voicing contrast in obstruent clusters across word boundary and related assimilatory processes. Voicing assimilation is present in word-internal clusters, but assimilation across word boundaries has only been researched in a limited number of contexts.

Finally, the perceptual relevance of established acoustic correlates needs to be tested for Serbian, and their role within the model of auditory enhancement re-examined.

## 8.5 Concluding remarks

The aim of the present study was to establish the basic set of acoustic correlates of the voicing contrast in Serbian stops, and to determine which linguistic and speaker factors induce variability in the realisation of these correlates. It further set out to examine language-specific aspects of the realisation of the voicing contrast, how they relate to the way this contrast is realised in other languages, especially voicing languages, and to evaluate the existing models of the voicing contrast in light of these findings.

The experimental results presented in Chapters 4 to 7 identified acoustic correlates of VOT, voicing in the closure, closure duration and preceding vowel duration as relevant in certain word positions in Serbian. The voicing contrast is robust in Serbian in all word positions and for each speaker in the present study. Several linguistic and speaker factors, such as place of stop articulation, quality of the following vowel, and age, gender and place of birth of speakers were found to induce variability in the realisation of these acoustic correlates. The experimental results pointed out to the fact that only some of the observed variability is caused by universal constraints, but that most of it is language- or speaker-specific. Among the most important language-specific features that have arisen from the present study are intermediate VOT values which straddle short lag and long lag VOT category, and lack of any inverse relationship between VOT and CD for phonologically voiceless stops; the effect of place of articulation on duration of closure voicing and on CD, as well as its interaction with the effect of the quality of the following vowel on VOT in the phonologically voiceless stops. Further, some of the variability associated with the age and gender of speakers might be caused by sociolinguistic factors. This is one of the few comprehensive acoustic-phonetic studies of Serbian and hopefully goes some way towards laying a foundation for further research on the voicing contrast in Serbian.

The literature review in Chapters 1 and 2 revealed a discrepancy between the wealth of research on acoustic correlates of voicing in a number of languages and the extent to which this knowledge has been incorporated in the existing theoretical models of the voicing contrast. There seem to be two main problems with these models. The first, and more general problem, is their inability to account for non-contrastive and non-universal variability. The second problem, highly relevant for the present study, is that they are unable to adequately represent the type of realisation of the voicing contrast found in voicing languages. The experimental findings from Chapters 4 to 7 land further support to both criticisms. In Chapter 8, I discussed each of the models in relation to Serbian results and I proposed aspects of these models which need to be improved to include Serbian data. The VOT-based model by Keating has the problem of incorporating intermediate VOT values in the model, while models by Kohler, Jessen, and Kingston and Diehl might be overestimating the role of preceding vowel duration in voicing languages and underestimating the role of CD and its relationship with voicing in the closure. This discussion also highlighted the fact that some of the core assumptions of these models need to be re-assessed in order to achieve this. With respect to the first shortcoming of the existing models, I suggested that an exemplar-based model of phonological knowledge, with phonetically rich lexical representations, has a potential to include various sources of variability. It has some features that for many investigators are very positive: it does not involve any predetermined categories, and allows for linguistic, individual and sociophonetic variation in the realisation of the voicing contrast. On the other hand, it does not explain how this variability relates to abstract linguistic categories, why languages prefer a finite set of overall patterns of realisation, and how these patterns have emerged from exemplar representations. Since the exemplar model is still being developed, these are some of the questions that remain unanswered at the moment, and which call for further research.

# Appendix A

## Lists of the words used for analysis in the present study

| | /b/ | /p/ | /d/ | /t/ | /g/ | /k/ |
|---|---|---|---|---|---|---|
| /a/ | bas | pas | dah | tas | gad | kad |
| /e/ | bek | peh | ded | tek | gest | kelj |
| /i/ | bič | pik | dim | tih | gips | kič |
| /o/ | bob | pop | dok | top | goč | koš |
| /u/ | buć | puč | dud | tuš | gust | kuk |

Table A1 List of words used for analysis with the target stops in word-initial position (each stop before each of the five vowels)

| /b/ | /p/ | /d/ | /t/ | /g/ | /k/ |
|---|---|---|---|---|---|
| slab | čep | gad | sat | prag | bek |
| štab | džip | kad | kmet | trag | tek |
| hleb | hop | ded | let | breg | čik |
| žleb | pop | led | set | mig | pik |
| bob | top | zid | zet | glog | šik |
| rob | cup | plod | hit | smog | dok |
| snob | ćup | dud | sit | zbog | šok |
| zub | SUP | sud | žut | lug | kuk |

Table A2 List of words used for analysis with the target stops in word-final position

| /b/-/p/ | /d/-/t/ | /g/-/k/ |
|---|---|---|
| štab - štap | nad - mat | breg - prek |
| snob - snop | led - let | smog - cmok |
| kub - tup | sprud - prut | lug – luk |

Table A3 List of minimal or near-minimal pairs of words used for analysis with the target stops in word-final position

.

| /b/-/p/ | /d/-/t/ | /g/-/k/ |
|---|---|---|
| snoba – snopa | kada - Kata | nega - neka |
| tuba – tupa | Nada - Nata | nego – neko |
|  | ploda - plota | boga – Boka |

Table A4 List of minimal or near-minimal pairs of words used for analysis with the target stops in word-medial position

# Appendix B

## Pilot study: The effect of phonemic vowel length on acoustic correlates of initial stop voicing

When choosing the word list for the present study, words with phonologically short vowel were used whenever possible, because the duration of syllable nuclei under both short accents is similar and their range is smaller than under long accents (Lehiste, 1970). However, certain combinations of stops and phonologically short vowels are rare in word-initial position in monosyllables, and for this reason one word with phonologically long vowel and one with alternate pronunciation (long or short vowel) were included in the word list for recording. A pilot study was designed to test the hypothesis that in Serbian phonological vowel length has no effect on the realisation of the voicing contrast in the preceding stop, so that all words chosen for the study could be analysed together. This pilot study was carried out before the statistical analysis for the main study.

For this pilot study, six words with an initial stop followed by a phonologically long vowel were added to the word list, randomised with the rest of the words, and presented to the subjects for reading. These six words are: bar, bik, buđ, paž, pir, puž.

Two acoustic correlates of voicing were measured in initial stops in these words: VOT for stops in utterance-initial position and closure duration for stops in the sentence condition, using the same criteria as in the rest of the study (see Chapter 3). Results were compared with the corresponding results for words with short vowels:

bas – bar, bič – bik, buć – buđ
pas – paž, pik – pir, puč – puž.

For each acoustic correlate a t-test (or a Mann-Whitney U-test) was run using statistical software Minitab (v. 13.1). Results are presented in Tables 1 and 2.

|  | Before phonol. short vowel | Before phonol. long vowel | Statistical test and *p*-value |
|---|---|---|---|
| Mean VOT (ms), N, SD for /b/ | -112.8; 28; 43.4 | -112.3; 30; 45.8 | t-test, *p* = 0.97 |
| Mean VOT (ms), N, SD for /p/ | 21.6; 35; 10.8 | 25.8; 32; 14.9 | t-test, *p* = 0.19 |

Table B1 VOT results in utterance-initial position in the pilot study

|  | Before phonol. short vowel | Before phonol. long vowel | Statistical test and *p*-value |
|---|---|---|---|
| Mean CD (ms), N, and SD for /b/ | 106; 26; 25.45 | 108; 29; 18.06 | t-test, *p* = 0.63 |
| Mean CD (ms), N, and SD for /p/ | 133.29; 34; 26.9 | 126.54; 35; 26.94 | Mann-Whitney *p* = 0.32 |

Table B2 Results for CD in word-initial intervocalic position in the pilot study

As can be seen from the above results, measured differences in VOT and closure duration did not reach statistical significance. Results from this pilot study support the hypothesis that phonemic vowel length does not affect voicing correlates in preceding word-initial stops. Consequently, in the main part of this study, results for word-initial stops before phonemically short and long vowels were analysed together.

# Appendix C

## Tables with statistical analysis results for each subject

| | Mean VOT (ms), | Mean VOT (ms), | Statistical test result and |
|---|---|---|---|
| Subject | SD, N for /b, d, g/ | SD, N for /p, t, k/ | effect size[25] |
| MVf | -145.83 | 38.13 | $p < 0.001$ (2-tailed) |
| | 40.87; 12 | 18.48; 15 | $t$ (26) = -14.45, $d$ = -5.67 |
| MCf | -86.79 | 20.33 | $p < 0.001$ (2-tailed) |
| | 24.92; 14 | 8.71; 15 | $t$ (27) = -15.24, $d$ = -5.87 |
| SCf | -72.07 | 22.13 | $p < 0.001$ (2-tailed) |
| | 15.65; 15 | 16.94; 15 | $t$ (28) = -15.82, $d$ = -5.98 |
| BCf | -125.8 | 29.92 | $p < 0.001$ (2-tailed) |
| | 35.21; 15 | 15.74; 13 | $t$ (26) = -15.44, $d$ = -6.06 |
| DARf | -162.29 | 33.2 | $p < 0.001$ (2-tailed) |
| | 41.74; 14 | 22.31; 15 | $t$ (27) = -15.88, $d$ = -6.11 |
| MRf | -130.07 | 24.4 | $p < 0.001$ (2-tailed) |
| | 33.16; 15 | 15.19; 15 | $t$ (28) = -16.4, $d$ = -6.2 |
| IVm | -137.47 | 41.6 | $p < 0.001$ (2-tailed) |
| | 25.55; 15 | 17.43; 15 | $t$ (28) = -22.42, $d$ = -8.47 |
| RVm | -102.20 | 48.73 | $p < 0.001$ (2-tailed) |
| | 16.46; 15 | 17.97; 15 | $t$ (28) = -23.99, $d$ = -9.07 |
| BPm | -125.47 | 36.13 | $p < 0.001$ (2-tailed) |
| | 57.22; 15 | 19.58; 15 | $t$ (28) = -10.35, $d$ = -3.91 |
| DRm | -80.80 | 34.27 | $p < 0.001$ (2-tailed) |
| | 19.64; 15 | 12.96; 15 | $t$ (28) = -18.94, $d$ = -7.16 |
| IJm | -79.67 | 29.20 | $p < 0.001$ (2-tailed) |
| | 15.69; 15 | 11.95; 15 | $Z$ = - 4.67, $r$ = -0.85 |
| MPm | -105.73 | 44.14 | $p < 0.001$ (2-tailed) |
| | 27.5; 15 | 21.25; 14 | $t$ (27) = -16.34, $d$ = -6.29 |

Table C1 VOT results for stops in utterance-initial position for each subject

[25] For *t*-test: *p*-value, *t* and Cohen's *d*; for Mann-Whitney U-test: *p*-value, *Z* and effect size *r*.

|  | Mean CD (ms), | Mean CD (ms), | Statistical test result and |
|---|---|---|---|
| Subject | SD, N for /b, d, g/ | SD, N for /p, t, k/ | effect size[26] |
| MVf | 127.70 | 152.57 | $p = 0.018$ (2-tailed) |
|  | 27.19; 10 | 20.57; 14 | $t$ (22) = -2.56, $d$ = -1.09 |
| MCf | 108.31 | 139.50 | $p < 0.001$ (2-tailed) |
|  | 16.44; 13 | 16.79; 14 | $t$ (25) = -4.87, $d$ = -1.95 |
| SCf | 77.33 | 99.73 | $p < 0.001$ (2-tailed) |
|  | 11.18; 15 | 12.83; 15 | $t$ (28) = -5.1, $d$ = -1.93 |
| BCf | 100.67 | 120.50 | $p = 0.013$ (2-tailed) |
|  | 19.1; 15 | 20.92; 14 | $t$ (27) = -2.67, $d$ = -1.03 |
| DARf | 115.93 | 164.87 | $p < 0.001$ (2-tailed) |
|  | 20.48; 14 | 22.89; 15 | $Z$ = -4.06, $r$ = -0.8 |
| MRf | 104.58 | 144.07 | $p < 0.001$ (2-tailed) |
|  | 22.34; 12 | 16.09; 15 | $t$ (25) = -5.34, $d$ = -2.14 |
| IVm | 100.21 | 134.20 | $p < 0.001$ (2-tailed) |
|  | 14.25; 14 | 19.39; 15 | $t$ (27) = -5.35, $d$ = -2.06 |
| RVm | 89.60 | 106.00 | $p = 0.019$ (2-tailed) |
|  | 12.35; 15 | 21.26; 15 | $t$ (28) = -2.58, $d$ = -0.975 |
| BPm | 102.33 | 120.86 | $p = 0.006$ (2-tailed) |
|  | 15.9; 15 | 17.58; 14 | $t$ (27) = -2.98, $d$ = -1.15 |
| DRm | 88.13 | 102.73 | $p < 0.001$ (2-tailed) |
|  | 15.97; 15 | 17.613; 15 | $t$ (28) = -2.38, $d$ = -0.9 |
| IJm | 86.40 | 101.5 | $p = 0.014$ (2-tailed) |
|  | 15.361; 15 | 15.693; 14 | $t$ (27) = -2.62, $d$ = -1.01 |
| MPm | 64.80 | 81.58 | $p = 0.02$ (2-tailed) |
|  | 11.91; 15 | 12.77; 12 | $t$ (25) = -3.52, $d$ = -1.41 |

Table C2 Results for CD for stops in word-initial intervocalic position for each subject

[26] For *t*-test: *p*-value, *t* and Cohen’s *d*; for Mann-Whitney U-test: *p*-value, *Z* and effect size *r*.

| Subject | Mean CD (ms), SD, N for /b, d, g/ | Mean CD (ms), SD, N for /p, t, k/ | *t*-test result and effect size Cohen’s *d* |
|---|---|---|---|
| MVf | 120.35 | 192.04 | $p < 0.001$ (2-tailed) |
|  | 19.33; 24 | 31.56; 24 | $t$ (46) = -9.44, $d$ = 2.78 |
| MCf | 100 | 165.67 | $p < 0.001$ (2-tailed) |
|  | 9.95; 23 | 19.27; 24 | $t$ (45) = -14.83, $d$ = -4.42 |
| SCf | 97.5 | 148 | $p < 0.001$ (2-tailed) |
|  | 14.2; 24 | 25.4; 24 | $t$ (46) = -8.5, $d$ = -2.51 |
| BCf | 116.54 | 165.88 | $p < 0.001$ (2-tailed) |
|  | 16.93; 24 | 18.7; 24 | $t$ (46) = -9.58, $d$ = -2.83 |
| DARf | 118 | 206.17 | $p < 0.001$ (2-tailed) |
|  | 19.51; 24 | 22.26; 24 | $t$ (46) = -14.59, $d$ = -4.3 |
| MRf | 138.88 | 199.42 | $p < 0.001$ (2-tailed) |
|  | 15.01; 24 | 39.05; 24 | $t$ (46) = -7.09, $d$ = -2.09 |
| IVm | 97.67 | 163.58 | $p < 0.001$ (2-tailed) |
|  | 9.88; 24 | 14.18; 24 | $t$ (46) = -18.69, $d$ = -5.59 |
| RVm | 94.75 | 121.13 | $p < 0.001$ (2-tailed) |
|  | 12.4; 24 | 20.39; 24 | $t$ (46) = -5.41, $d$ = -1.6 |
| BPm | 115.73 | 162.92 | $p < 0.001$ (2-tailed) |
|  | 21.92; 22 | 21.03; 24 | $t$ (44) = -7.45, $d$ = -2.25 |
| DRm | 78.23 | 134.96 | $p < 0.001$ (2-tailed) |
|  | 16.65; 22 | 22.98; 24 | $t$ (44) = -9.51, $d$ = -2.87 |
| IJm | 97.5 | 158.63 | $p < 0.001$ (2-tailed) |
|  | 17.28; 24 | 23.07; 24 | $t$ (46) = -10.39, $d$ = -3.06 |
| MPm | 79.78 | 132.3 | $p < 0.001$ (2-tailed) |
|  | 11.65; 23 | 27.26; 23 | $t$ (44) = -8.5, $d$ = -2.56 |

Table C3 Results for CD for stops in utterance-final position for each subject

| Subject | Mean CD (ms),<br>SD, N for /b, d, g/ | Mean CD (ms),<br>SD, N for /p, t, k/ | Statistical test result and<br>effect size[27] |
|---|---|---|---|
| SCf | 68.57 | 108.6 | *p* < 0.001 (2-tailed) |
| | 10.77; 21 | 20.33; 20 | *t* (39) = -7.82, *d* = -2.5 |
| BCf | 70 | 109.95 | *p* < 0.001 (2-tailed) |
| | 10.23; 21 | 15.38; 22 | *Z*= -5.68, *r* = -0.84 |
| IJm | 71.6 | 97.57 | *p* = 0.004 (2-tailed) |
| | 6.19; 5 | 17.48; 7 | *Z* = -2.86, *r* = - 0.83 |
| MPm | 54.56 | 78.67 | *p* < 0.001 (2-tailed) |
| | 12.82; 18 | 12.1; 23 | *t* (39) = -6.17, *d* = -1.98 |

Table C4 Results for CD for stops in word-final intervocalic position for each subject

[27] For *t*-test: *p*-value, *t* and Cohen’s *d*; for Mann-Whitney U-test: *p*-value, *Z* and effect size *r*.

|  | Mean v. dur. (ms), | Mean v. dur. (ms), | Statistical test result and |
|---|---|---|---|
| Subject | SD, N for /b, d, g/ | SD, N for /p, t, k/ | effect size[28] |
| MVf | 126.3 | 17.71 | $p < 0.001$ (2-tailed) |
|  | 27.87; 10 | 5.44; 14 | $t$ (22) = 12.16, $d$ = 5.19 |
| MCf | 105.46 | 19.36 | $p < 0.001$ (2-tailed) |
|  | 19.59; 13 | 7.73; 14 | $t$ (25) = 14.83, $d$ = 5.93 |
| SCf | 77.33 | 10.33 | $p < 0.001$ (2-tailed) |
|  | 11.18; 15 | 4.44; 15 | $t$ (28) = 21.58, $d$ = 8.16 |
| BCf | 100.67 | 16.43 | $p < 0.001$ (2-tailed) |
|  | 19.1; 15 | 4.75; 14 | $t$ (27) = 16.54, $d$ = 6.37 |
| DARf | 112.29 | 21.2 | $p < 0.001$ (2-tailed) |
|  | 23.73; 14 | 8.16; 15 | $t$ (27) = 13.63, $d$ = 5.25 |
| MRf | 104.58 | 18.33 | $p < 0.001$ (2-tailed) |
|  | 22.34; 12 | 8.64; 15 | $t$ (25) = 12.64, $d$ = 5.06 |
| IVm | 99.43 | 11.8 | $p < 0.001$ (2-tailed) |
|  | 16.31; 14 | 8.54;15 | $Z$ = -4.59, $r$ = -0.85 |
| RVm | 89.6 | 8.0 | $p < 0.001$ (2-tailed) |
|  | 12.35;15 | 8.5; 15 | $Z$ = -4.7, $r$ = -0.86 |
| BPm | 102.33 | 16.93 | $p < 0.001$ (2-tailed) |
|  | 15.9; 15 | 9.39; 14 | $t$ (27) = 17.74, $d$ = 6.83 |
| DRm | 88.13 | 7.13 | $p < 0.001$ (2-tailed) |
|  | 15.97; 15 | 8.84; 15 | $Z$ = -4.71, $r$ = -0.86 |
| IJm | 85.73 | 8.07 | $p < 0.001$ (2-tailed) |
|  | 16.09; 15 | 10.22; 14 | $Z$ = -4.62, $r$ = -0.86 |
| MPm | 64.80 | 2.5 | $p < 0.001$ (2-tailed) |
|  | 11.91; 15 | 6.56; 12 | $Z$ = -4.51, $r$ = -0.87 |

Table C5 Results for duration of voicing in the closure for stops in word-intial intervocalic position for each subject

[28] For *t*-test: *p*-value, *t* and Cohen’s *d*; for Mann-Whitney U-test: *p*-value, *Z* and effect size *r*.

| Subject | Mean v. dur. (ms),<br>SD, N for /b, d, g/ | Mean v. dur. (ms),<br>SD, N for /p, t, k/ | Statistical test result and<br>effect size[29] |
|---|---|---|---|
| MVf | 91.35 | 18.08 | $p$ < 0.001 (2-tailed) |
| | 20.67; 23 | 9.52; 24 | $t$ (45) = 15.5, $d$ = 4.62 |
| MCf | 22.0 | 8.5 | $p$ < 0.001 (2-tailed) |
| | 9.52; 24 | 5.68; 24 | $t$ (46) = 5.97, $d$ = 1.76 |
| SCf | 38.04 | 9.75 | $p$ < 0.001 (2-tailed) |
| | 15.2; 24 | 5.74; 24 | $Z$ = -5.78, $r$ = -0.83 |
| BCf | 48.03 | 16.5 | $p$ < 0.001 (2-tailed) |
| | 14.92; 24 | 13.94; 24 | $Z$ = -5.09, $r$ = -0.73 |
| DARf | 74.25 | 14.58 | $p$ < 0.001 (2-tailed) |
| | 22.9; 24 | 9.84; 24 | $Z$ = -5.94, $r$ = -0.86 |
| MRf | 104.79 | 5.33 | $p$ < 0.001 (2-tailed) |
| | 23.5; 24 | 10.6; 24 | $Z$ = -6.86, $r$ = -0.99 |
| IVm | 77.42 | 7.17 | $p$ < 0.001 (2-tailed) |
| | 19.57; 24 | 8.17; 24 | $Z$ = -5.97, $r$ = -0.86 |
| RVm | 88.38 | 8.29 | $p$ < 0.001 (2-tailed) |
| | 12.89; 24 | 6.7; 24 | $Z$ = -5.96, $r$ = -0.86 |
| BPm | 89.05 | 6.08 | $p$ < 0.001 (2-tailed) |
| | 26.35; 22 | 7.19; 24 | $Z$ = -5.87, $r$ = -0.87 |
| DRm | 42.59 | 9.0 | $p$ < 0.001 (2-tailed) |
| | 16.35; 22 | 9.74; 24 | $Z$ = -5.52, $r$ = -0.81 |
| IJm | 38.54 | 6.75 | $p$ < 0.001 (2-tailed) |
| | 15.31; 24 | 6.94; 24 | $Z$ = -5.89, $r$ = -0.85 |
| MPm | 64.22 | 12.13 | $p$ < 0.001 (2-tailed) |
| | 12.79; 23 | 8.28; 23 | $Z$ = -5.8, $r$ = -0.86 |

Table C6 Results for duration of voicing in the closure for stops in utterance-final position for each subject

[29] For *t*-test: *p*-value, *t* and Cohen’s *d*; for Mann-Whitney U-test: *p*-value, *Z* and effect size *r*.

| Subject | Mean v. dur. (ms), SD, N for /b, d, g/ | Mean v. dur. (ms), SD, N for /p, t, k/ | Mann-Whitney U-test result, and effect size *r* |
|---|---|---|---|
| SCf | 55.76 | 5.9 | $p < 0.001$ (2-tailed) |
| | 10.05; 21 | 6.84; 20 | $Z = -5.52$, $r = -0.86$ |
| BCf | 64.13 | 10.64 | $p < 0.001$ (2-tailed) |
| | 12.26; 24 | 8.53; 22 | $Z = -5.81$, $r = -0.86$ |
| IJm | 63.2 | 4.14 | $p = 0.003$ (2-tailed) |
| | 14.81; 5 | 7.54; 7 | $Z = -2.95$, $r = -0.85$ |
| MPm | 54.11 | 8.78 | $p < 0.001$ (2-tailed) |
| | 13.41; 18 | 7.92; 23 | $Z = -5.46$, $r = -0.85$ |

Table C7 Results for duration of voicing in the closure for stops in word-final intervocalic position for each subject

|  | Mean v. dur. (ms), | Mean v. dur. (ms), | Statistical test result and |
|---|---|---|---|
| Subject | SD, N bef. /b, d, g/ | SD, N bef. /p, t, k/ | effect size[30] |
| MVf | 147 | 121 | $p = 0.005$ (2-tailed) |
|  | 18.53; 10 | 18.9; 10 | $Z = -2.8$, $r = -0.63$ |
| MCf | 147 | 118 | $p < 0.001$ (2-tailed) |
|  | 18.1; 14 | 16.16; 14 | $t$ (13) = 8.49, $d$ = 1.69 |
| SCf | 107 | 90 | $p < 0.001$ (2-tailed) |
|  | 9.85; 14 | 10.84; 14 | $t$ (13) = 4.87, $d$ = 1.64 |
| BCf | 143 | 124 | $p < 0.001$ (2-tailed) |
|  | 20.31; 14 | 13.61; 14 | $t$ (13) = 4.37, $d$ = 1.1 |
| DARf | 160 | 142 | $p = 0.003$ (2-tailed) |
|  | 12.98; 10 | 18.76; 10 | $t$ (9) = 4.06, $d$ = 1.12 |
| MRf | 134 | 123 | $p = 0.009$ (2-tailed) |
|  | 13.43; 12 | 15.08; 12 | $t$ (11) = 3.16, $d$ = 0.77 |
| IVm | 133 | 109 | $p = 0.013$ (2-tailed) |
|  | 15.37; 11 | 8.17; 11 | $Z = -2.5$, $r = -0.53$ |
| RVm | 129 | 102 | $p < 0.001$ (2-tailed) |
|  | 17.75; 14 | 9.11; 14 | $t$ (13) = 6.57, $d$ = 1.91 |
| BPm | 124 | 94 | $p = 0.002$ (2-tailed) |
|  | 14.79; 12 | 10.42; 12 | $Z = -3.06$, $r = -0.62$ |
| DRm | 128 | 97 | $p = 0.001$ (2-tailed) |
|  | 17.39; 10 | 15.24; 10 | $t$ (9) = 4.95, $d$ = 1.9 |
| IJm | 112 | 87 | $p < 0.001$ (2-tailed) |
|  | 17.54; 15 | 16.0; 15 | $t$ (14) = 10.18, $d$ = 1.49 |
| MPm | 136 | 104 | $p < 0.001$ (2-tailed) |
|  | 21.6; 8 | 15.98; 8 | $t$ (7) = 8.19, $d$ = 1.68 |

Table C8 Results for preceding vowel duration for phonologically short vowels for each subject

[30] For paired $t$-test: $p$-value, $t$ and Cohen’s $d$; for Wilcoxon test: $p$-value, $Z$ and effect size $r$.

| Subject | Mean v. dur. (ms), SD, N bef. /b, d, g/ | Mean v. dur. (ms), SD, N bef. /p, t, k/ | Statistical test result and effect size[31] |
|---|---|---|---|
| MVf | 227 | 202 | *p* = 0.005 (2-tailed) |
| | 27.74; 13 | 16.63; 13 | *Z* = -2.8, *r* = -0.55 |
| MCf | 199 | 185 | *p* = 0.008 (2-tailed) |
| | 22.23; 13 | 25.53; 13 | *t* (12) = 3.15, *d* = 0.59 |
| SCf | 164 | 153 | *p* = 0.025 (2-tailed) |
| | 22.12; 12 | 16.63; 12 | *t* (11) = 2.59, *d* = 0.56 |
| BCf | 180 | 166 | *p* = 0.015 (2-tailed) |
| | 30.92; 14 | 26.67; 14 | *t* (13) = 2.79, *d* = 0.49 |
| DARf | 229 | 212 | *p* = 0.061 (2-tailed) |
| | 36.31; 15 | 28.38; 15 | *Z* = -1.87, *r* = -0.34 |
| MRf | 208 | 201 | *p* = 0.4 (2-tailed) |
| | 46.17; 16 | 24.95; 16 | *t* (15) = 0.87, *d* = 0.19 |
| IVm | 213 | 179 | *p* < 0.001 (2-tailed) |
| | 28.46; 16 | 16.62; 16 | *t* (15) = 6.23, *d* = 1.46 |
| RVm | 188 | 175 | *p* = 0.028 (2-tailed) |
| | 25.51; 16 | 28.92; 16 | *t* (15) = 2.43, *d* = 0.48 |
| BPm | 168 | 140 | *p* < 0.001 (2-tailed) |
| | 28.43; 12 | 15.39; 12 | *t* (11) = 4.97, *d* = 1.22 |
| DRm | 175 | 156 | *p* = 0.007 (2-tailed) |
| | 28.43; 16 | 30.43; 16 | *t* (15) = 3.13, *d* = 0.65 |
| IJm | 184 | 167 | *p* = 0.025 (2-tailed) |
| | 30.59; 16 | 21.54; 16 | *t* (15) = 2.49, *d* = 0.64 |
| MPm | 156 | 143 | *p* = 0.002 (2-tailed) |
| | 30.6; 11 | 28.45; 11 | *t* (9) = 4.33, *d* = 0.44 |

Table C9 Results for preceding vowel duration for phonologically long vowels for each subject

[31] For paired *t*-test: *p*-value, *t* and Cohen’s *d*; for Wilcoxon test: *p*-value, *Z* and effect size *r*.

| Subject | Mean Ratio for phonol. short vowels | Mean Ratio for phonol. long vowels |
|---|---|---|
| BPm | 0.77 | 0.84 |
| DRm | 0.77 | 0.90 |
| MPm | 0.77 | 0.88 |
| IJm | 0.78 | 0.93 |
| MCf | 0.80 | 0.93 |
| RVm | 0.80 | 0.94 |
| IVm | 0.82 | 0.85 |
| MVf | 0.83 | 0.90 |
| SCf | 0.84 | 0.94 |
| BCf | 0.89 | 0.93 |
| DARf | 0.89 | 0.94 |
| MRf | 0.92 | 0.99 |

Table C10 Mean vowel duration ratio for each subject

(for phonologically short and long vowels in the pooled data, in ascending order from the shortest mean ratio for phonologically short vowels)

# References

Abdelli-Beruh, N. B. (2004). The Stop Voicing Contrast in French Sentences: Contextual Sensitivity of Vowel Duration, Closure Duration, Voice Onset Time, Stop Release and Closure Voicing. *Phonetica, 61*(4), 201-219.

Abdelli-Beruh, N. B. (2009). Influence of place of articulation on some acoustic correlates of the stop voicing contrast in Parisian French. *Journal of Phonetics, 37*, 66-78.

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. In B. Hala, M. Romportl & P. Janota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences, 1967* (pp. 569-573). Prague: Academia.

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: Fo shift versus voice timing. In V. A. Fromkin (Ed.), *Phonetic linguistics, Essays in honor of Peter Ladefoged* (pp. 25-33). London: Academic Press.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 113*(1), 544-552.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics, 39*, 39-49.

Beckman, J., Jessen, M., & Ringen, C. (fc). Empirical evidence for laryngeal features: German vs. true voice languages.

Belić, A. (1960). *Osnovi istorije srpskohrvatskog jezika, I Fonetika*. Beograd.

Belić, A. (1968). *Savremeni srpskohrvatski književni jezik, Glasovi i akcenat*. Beograd: Naučna knjiga.

Benki, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics, 29*, 1-22.

Bijankhan, M., & Nourbakhsh, M. (2009). Voice onset time in Persian initial and intervocalic stop production. *Journal of the International Phonetic Association, 39*(3), 335-364.

Boersma, P., & Weenink, D. (1992-2012). Praat: doing phonetics by computer [Computer program], Version 4.5.14, retrieved from http://www.praat.org/.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992a). Articulatory phonology: an overview. *Haskins Laboratories Status Report on Speech Research, SR-111/112*, 23-42.

Browman, C. P., & Goldstein, L. (1992b). "Targetless" schwa: an articulatory analysis. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 26-56). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (2010). *Articulatory phonology, Chapters 2-3 (unpublished draft)*. Retrieved on 10. 11. 2010. from http://sail.usc.edu/~lgoldste/ArtPhon/Papers/Week%208-9/AP_Ch2-3.pdf.

Brunner, J. (2005). Supralaryngeal mechanisms of the voicing contrast in velars. *ZAS Papers in Linguistics, 39*.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics, 83*, 97-116.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication, 15*, 39-54.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology, 13*(2), 139-169.

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice Onset Time in two French dialects. *Journal of Phonetics, 2*, 239-245.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America, 54*(2), 421-428.

Castelman, W. A., & Diehl, R. L. (1994). Fundamental frequency effects on [voice] judgments in intervocalic stop consonants (abstract). *Journal of the Acoustical Society of America*, *95*(5), 2977.

Castelman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics, 24*, 383-398.

Chang, S. S., Ohala, J. J., Hansson, G., James, B., Lewis, J., Liaw, L., et al. (1999). Vowel-dependent VOT variation: An experimental study. *Journal of the Acoustical Society of America, 105*(2, Pt. 2), 1400.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22*(3), 129-159.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*, 207-229.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*, 121–157.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology* (2 ed.). Oxford UK & Cambridge USA: Blackwell.

Cochrane, G. R. (1970). Some vowel durations in Australian English. *Phonetica, 22*(4), 240-250.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America, 24*, 597-606.

Crowther, C. S., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92*(2), 711-722.

Crowther, C. S., & Mann, V. (1994). Use of vocalic cues to consonant voicing and native language background: The influence of experimental design. *Perception & Psychophysics, 55*(5), 513-525.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America, 72*(3), 705-716.

Crystal, T. H., & House, A. S. (1988a). The duration of American-English stop consonants: an overview. *Journal of Phonetics, 16*, 285-294.

Crystal, T. H., & House, A. S. (1988b). The duration of American-English vowels: an overview. *Journal of Phonetics, 16*, 263-284.

Davis, S., & Van Summers, W. (1989). Vowel length and closure duration in word-medial VC sequences. *Journal of Phonetics, 17*, 339-353.

Deneš, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27*(4), 761-764.

Derr, M. A., & Massaro, D. W. (1980). The contribution of vowel duration, Fo contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception and Psychophysics, 27*(1), 51-59.

Diehl, R. L., Kluender, K. R., & Walsh, M. A. (1990). Some auditory bases of speech perception and production. In W. A. Ainsworth (Ed.), *Advances in speech, hearing and language processing* (Vol. 1., pp. 243-267). London: JAI Press.

Diehl, R. L., & Molis, M. R. (1995). Effect of fundamental frequency on medial [+voice]/[-voice] judgments. *Phonetica, 52*, 188-195.

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27*(5), 435-443.

Dmitrieva, O., Jongman, A., & Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics, 38*, 483–492.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin, New York: Foris Publications.

Docherty, G. J., & Foulkes, P. (2000). Speaker, speech, and knowledge of sounds. In N. Burton-Roberts, P. Carr & G. Docherty (Eds.), *Phonological knowledge: Conceptual and empirical issues* (pp. 105-129). Oxford: Oxford University Press.

Docherty, G. J., & Foulkes, P. (fc). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*.

Docherty, G. J., Watt, D., Llamas, C., Hall, D., & Nycz, J. (2011). Variation in Voice Onset Time along the Scottish-English border. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 591-594). Hong Kong.

Đurović, R. (1996). *Šest ogleda o srpskim akcentima*. Užice: Učiteljski fakultet Užice, Kulturno-prosvetna zajednica Užice.

Edwards, T. J. (1981). Multiple features analysis of intervocalic English plosives. *Journal of the Acoustical Society of America, 69*(2), 535-547.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. Mackenzie

Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 347-383). Mahwah: Lawrence Erlbaum Associates.

Esposito, A. (2002). On vowel height and consonantal voicing effects: data from Italian. *Phonetica, 59*, 197-231.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.

Fischer-Jørgensen, E. (1954). Acoustic analysis of stop consonants. *Miscellanea Phonetica, II*, 42-59.

Fischer, R. M., & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America, 88*(3), 1250-1259.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics, 10*, 177-192.

Flege, J. E., & Hillenbrand, J. (1987). A differential effect of release bursts on the stop voicing judgments of native French and English listeners. *Journal of Phonetics, 15*, 203-208.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*(2), 125-146.

Foulkes, P., Docherty, G., & Jones, M. (2010). Analysing stops. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: a student's guide* (pp. 58-71). London: Routledge.

Foulkes, P., & Docherty, G. J. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409-438.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*, 113-133.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3-28.

Francis, A. R., Ciocca, V., & Yu, J. M. C. (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America, 113*(2), 1025-1032.

Fuchs, S. (2005). Articulatory correlates of the voicing distinction in alveolar obstruent production in German. *ZAS Papers in Linguistics, 41*.

Fujimura, O. (1971). Remarks on stop consonants - synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson & E. Zwirner (Eds.), *Form and* *substance, phonetic and linguistic papers presented to Eli Fischer-Jorgensen* (pp. 221-232). Copenhagen: Akademisk Forlag.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14,* 339-342.

Gósy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology, 4*(1), 75-85.

Gósy, M., & Ringen, C. (2009). Everything you always wanted to know about VOT in Hungarian (handout). Paper presented at *the 9th International Conference on the Structure of Hungarian (ICSH9)*. Debrecen, 30 August - 1 September 2009. Retrieved on 27. 8. 2010. from http://icsh9.unideb.hu/pph/handout/Ringen_Gosy_handout.pdf.

Gráczi, T. E. (2011). Voicing contrast of intervocalic plosives in Hungarian. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 759-762). Hong Kong.

Gruenenfelder, T., & Pisoni, D. B. (1980). Fundamental frequency as a cue to postvocalic consonantal voicing: some data from speech perception and production. *Perception and Psychophysics, 28*(6), 514-520.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*, 613-617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading Fo cues in the voiced-voiceless distinction. *Journal of Phonetics, 9*, 49-62.

Halle, M., Hughes, G. W., & Radley, J. P. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29*, 107-116.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report No. 101* (pp. 198-213). Cambridge, MA: Research Laboratory of Electronics, MIT.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcast. *Journal of Phonetics, 34,* 439-457.

Hawkins, S. (1992). An introduction to task dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 9-25). Cambridge: Cambridge University Press.

Hawkins, S. (1999). Reevaluating assumptions about speech perception:interactive and integrative theories. In J. M. Pickett (Ed.), *The acoustics of speech* *communication: fundamentals, speech perception theory, and technology*. Boston: Allyn and Bakon.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23,* 351-379.

Hayward, K. (2000). *Experimental phonetics*. Harlow: Longman.

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America, 116*(5), 3108-3118.

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics, 36*, 607-628.

Heselwood, B., & Mahmoodzade, Z. (2007). Vowel onset characteristics as a function of voice and manner contrasts in Persian coronal stops. *Leeds Working Papers in Linguistics and Phonetics*, *12*, 125-142.

Heselwood, B., & McChrystal, L. (1999). The effect of age-group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: An acoustic sociophonetic study. *Leeds Working Papers in Linguistics and Phonetics, 7,* 49-69.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Fledge, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, *76*(1), 18-26.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America, 67*, 1764-1771.

Hoit, J. D., Solomon, N. P., & Hixon, T. J. (1993). Effect of lung volume on voice onset time (VOT). *Journal of Speech and Hearing Research, 36,* 516-521.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. A. Fromkin (Ed.), *Tone: a linguistic survey* (pp. 77-111). New York: Academic Press.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language, 55*(1), 37-58.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America, 33,* 1174-1178.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*(1), 105-113.

Ivković, M. (1913). O zvučnim suglasnicima na kraju reči u srpskom jeziku. *Južnoslovenski filolog, I*(1-2), 66-72.

Jacques, B. (1987). Les indices acoustiques dur trait de voisement dans les occlusives du français parlé à Montréal. *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 4, pp. 40-43). Tallinn.

Jakobson, R., Fant, C. G. M., & Halle, M. (1969). *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, Massachusetts: The MIT Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. 'S-Gravenhage: Mouton & Co.

Jakobson, R., & Waugh, L. R. (1987). *The sound shape of language* (2nd ed.). Berlin, New York, Amsterdam: Mouton de Gruyter.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: a laboratory phonology approach to English, Hungarian and Dutch.* Unpublished PhD thesis, University of Groningen, Retrieved 3. 11. 2005. from http://irs.ub.rug.nl/ppn/264415094.

Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology, 19*, 189-218.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485-499.

Jokanović-Mihajlov, J. (1983). Priroda uzlaznih akcenata u progresivnijim štokavskim govorima. *Srpski dijalektološki zbornik, XXIX*, 295-338.

Kallestinova, E. (2004). Voice and aspiration of stops in Turkish. *Folia Linguistica, Special Issue: Voicing, 38*(1-2), 117-143.

Karlsson, F., Zetterholm, E., & Sullivan, K. P. H. (2004). Development of a gender difference on Voice Onset Time. *Proceedings of the 10th Australian International Conference on Speech, Science & Technology* (pp. 316-321). Macquarie University, Sydney, December 8 to 10, 2004.

Kašić, Z. (1980). Glasovne promene u proklizi u srpskohrvatskom jeziku na osnovu eksperimentalnih istraživanja. *Naš jezik, XXIV*(4-5), 217-246.

Kašić, Z. (1985). Glasovne promene u enklizi. *Naš jezik, XXVI*(4), 228-233.

Keating, P. A. (1980). *A phonetic study of a voicing contrast in Polish.* Unpublished PhD thesis, Brown University.

Keating, P. A. (1984a). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286-319.

Keating, P. A. (1984b). Physiological effects on stop consonant voicing. *UCLA Working Papers in Phonetics, 59*, 29-34.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics, Essays in honor of Peter Ladefoged* (pp. 115-132). London: Academic Press.

Keating, P. A. (1988a). The phonology-phonetics interface. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. I: Linguistic theory: Foundations, pp. 281-302). Cambridge: Cambridge University Press.

Keating, P. A. (1988b). *A survey of phonological features*. Bloomington, Indiana: Indiana University Linguistics Club.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 451-470). Cambridge: Cambridge University Press.

Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics, 11*, 277-290.

Keating, P. A., Mikos, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America, 70*(5), 1261-1271.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*, 143-168.

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics, 26*, 117-128.

Kim, Y. (2002). *Phonological Features: Privative or Equipollent?* Unpublished BA thesis, Department of Linguistics, Harvard University, Harvard.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*(2), 419-454.

Kingston, J., & Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence, Papers in laboratory phonology IV* (pp. 7-27). Cambridge: Cambridge University Press.

Kingston, J., Diehl, R. L., Kirk, C. J., & Castelman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics, 36*, 28-54.

Kingston, J., Diehl, R. L., Kluender, K. R., & Parker, E. M. (1990). Resonance versus source characteristics in perceiving spectral continuity between vowels and consonants (abstract). *Journal of the Acoustical Society of America, 88* (Suppl. 1), S54-S55.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America, 54*(4), 1102-1104.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18,* 686-706.

Klatt, D. H. (1976). Linguistic use of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*(5), 1208-1221.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America, 90*(1), 83-96.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics, 16,* 153-169.

Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language and Hearing Research, 43,* 1211-1228.

Kohler, K. J. (1982). Fo in the production of lenis and fortis plosives. *Phonetica, 39,* 199-218.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica, 41,* 150-174.

Kohler, K. J. (1985). Fo in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America, 78*(1), 21-33.

Kohler, K. J., & van Dommelen, W. A. (1986). Prosodic effects on lenis/fortis perception: preplosive Fo and LPC synthesis. *Phonetica, 43,* 70-75.

Kollia, H. B. (1993). Segmental duration changes due to variation in stress, vowel, place of articulation, and voicing of stop consonants in Greek. *Journal of the Acoustical Society of America, 93*, 2298 (A).

Kopczyński, A. (1977). *Polish and American English consonant phonemes: a contrastive study*. Warszawa: Panstwowe Wydawnictwo Naukowe.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). *Speech: articulation and perception* (English translation, 2 ed.). Washington, DC: Joint Publications Research Service.

Krajišnik, V. (1994). *Kvantititativne i spektralne karakteristike sonanata*. Beograd: Filološki fakultet.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*(4209), 69-72.

Ladefoged, P. (1989). Representing phonetic structure. *UCLA Working Papers in Phonetics, 73*.

Ladefoged, P. (2004). Phonetics and phonology in the last 50 years. *UCLA Working Papers in Phonetics, 103,* 1-11.

Ladefoged, P. (2006). Representing linguistic phonetic structure, draft chapter (in progress before death). Retrieved from http://www.linguistics.ucla.edu/people/ladefoge/phoneticstructure.pdf

Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics, 20,* 411-440.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts, and London, England: The MIT Press.

Lehiste, I., & Ivić, P. (1986). *Word and sentence prosody in Serbocroatian*. Massachusetts Institute of Technology.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419-425.

Liberman, A. M. (1996). *Speech: a special code*. Cambridge, MA: MIT Press.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech, 1*, 153-167.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4*(4), 175-195.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839-862.

Lindau, M., & Ladefoged, P. (1986). Variability of feature specifications. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 464-479). Hillsdale, New Jersey and London: Lawrence Erlbaum Associates.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language, 33*(1), 42-49.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America, 57,* 1547-1551.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word, 20*(2), 385-422.

Lisker, L., & Abramson, A. S. (1965). Stop categorization and voice onset time. In E. Zwirner & W. Bethge (Eds.), *Proceedings of the 5th International Congress of Phonetic Sciences, Munster, 1964* (pp. 389-391). Basel/New York: S. Karger.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on Voice Onset Time in English stops. *Language and Speech, 10*, 1-28.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: some experiments in comparative phonetics. In B. Hala, M. Romportl & P. Janota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences, 1967* (pp. 563-567). Prague: Academia.

Lousada, M., Jesus, L. M. T., & Hall, A. (2010). Temporal acoustic correlates of the voicing contrast in European Portuguese stops. *Journal of the International Phonetic Association, 40*(3), 261-275.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant vowel ratio in speech production. *Journal of the Acoustical Society of America, 78*(6), 1949-1957.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: monolingual and bilingual production. *Journal of the Acoustical Society of America, 71*(1), 173-178.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal, 76*, 190-208.

Malecot, A. (1958). The role of releases in the identification of released final stops. *Language, 34*(1), 370-380.

Marković, J., & Sokolović, M. (2000). Inventar prozodema u govoru Petrovog Sela. *Južnoslovenski Filolog, LVI*(1-2), 635-646.

Marković, J., & Soković, M. (2004). Neke prozodijske karakteristike govora Petrovog Sela. *Proceedings of the Third International Conference: The Life and Work of Academician Pavle Ivić* (pp. 273-281). Subotica - Novi Sad – Belgrade.

Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception and Psychophysics, 23*(4), 331-336.

Miletić, B. (1927-28). Prilog za ispitivanje artikulacije pomoću rendgenovih zrakova. *Južnoslovenski filolog, VII*, 160-200.

Miletić, B. (1933). Izgovor srpskohrvatskih glasova. *Srpski dijalektološki zbornik, V*.

Miletić, B. (1960). *Osnovi fonetike srpskog jezika*. Beograd: Naučna knjiga.

Miller-Ockhuizen, A., & Zec, D. (2002). Durational differences in Serbian palatal affricates. *Proceedings of the First Pan-American/Iberian Meeting on Acoustics*. Cancun, Mexico, 2-7 December 2002.

Miller-Ockhuizen, A., & Zec, D. (2003). Acoustics of contrastive palatal affricates predict phonological patterning. *Proceedings of 15th International Congress of Phonetic Sciences* (pp. 3101-3104). Barcelona.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*, 106-115.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*(6), 505-512.

Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics, 32*, 1-33.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics, 36*, 308-317.

Nagao, K., & de Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *Journal of the Acoustical Society of America, 121*(5), 2882-2898.

Nearey, T. M. (1995). A double-weak view of trading relations: comments on Kingston and Diehl. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence, Papers in laboratory phonology IV* (pp. 28-40). Cambridge: Cambridge University Press.

Nearey, T. M., & Rochet, B. L. (1994). Effect of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1), 1-18.

Neiman, G. S., Klich, R. J., & Shuey, E. M. (1983). Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research, 26,* 111-118.

O' Kane, D. (1978). Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants. *Journal of Phonetics, 6,* 311-318.

Obler, L. (1982). The parsimonious bilingual. In L. Obler & L. Menn (Eds.), *Exceptional language and linguistics* (pp. 339-346). New York: Academic Press.

Oh, E. (2011). Effect of speaker gender on voice onset time in Korean stops. *Journal of Phonetics, 39,* 59-67.

Ohala, J. J. (1981). Articulatory constraints on the cognitive representation of speech. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 111-127). Amsterdam: North-Holland Publishing Company.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189 - 216). New York: Springer-Verlag.

Ohala, J. J. (2011). Accommodation to the aerodynamic voicing constraint and its phonological relevance. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 64-67). Hong Kong.

Ohala, J. J., & Riordan, C. J. (1979). Passive vocal tract enlargement during voiced stops. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers* (pp. 89 - 92). New York: Acoustical Society of America

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*(1), 224-230.

Pallant, J. (2007). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows* (3rd ed.). Maidenhead: Open University Press.

Pape, D., & Jesus, L. M. T. (2011). Devoicing of phonologically voiced obstruents: Is European Portuguese different from other Romance languages? *Proceedings of 17th International Congress of Phonetic Sciences* (pp. 1566-1569). Hong Kong.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics, 39*(2), 129-142.

Peco, A. (1961-1962a). Izgovor zvučnih suglasnika na kraju riječi u srpskohrvatskom jeziku. *Zbornik za filologiju i lingvistiku, IV-V*, 235-244.

Peco, A. (1961-1962b). Priroda afrikata srpskohrvatskog jezika. *Južnoslovenski filolog, XXV*, 161-190.

Peco, A., & Pravica, P. (1972). O prirodi akcenata srpskohrvatskog jezika na osnovu eksperimentalnih istraživanja. *Južnoslovenski filolog, XXIX*(1-2), 195-242.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*(6), 693-703.

Petrović, D., & Gudurić, S. (2010). *Fonologija srpskoga jezika*. Beograd: Institut za srpski jezik SANU, Beogradska knjiga, Matica srpska.

Pickett, J. M., Bunnell, H. T., & Revoile, S. G. (1995). Phonetics of intervocalic consonant perception: retrospect and prospect. *Phonetica, 52*, 1-40.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101-140). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 34*, 516-530.

Pierrehumbert, J. B., Beckman, M. E., & Ladd, D. R. (2000). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr & G. Docherty (Eds.), *Phonological knowledge: Conceptual and empirical issues* (pp. 273-303). Oxford: Oxford University Press.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: the search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics, 57*(3), 291-304.

Pind, J. (1996). Rate-dependent perception of aspiration and pre-aspiration in Icelandic. *The Quarterly Journal of Experimental Psychology, 49A*(3), 745-764.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 51*(5), 1352-1361.

Poch-Olivé, D. (1987). Acoustic correlates for places of articulation in Spanish stop consonants. *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 4, pp. 48-51). Tallinn.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7*, 45-56.

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, *69*(1), 262-274.

Port, R. F., & Dalby, J. (1982). Consonant-vowel ratio as a cue for voicing in English. *Perception and Psychophysics, 32*(2), 141-152.

Port, R. F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America, 66*(3), 654-662.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*(4), 1296-1303.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition. *Phonetica, 38*, 126-147.

Raphael, L. J., Dorman, M. F., Freeman, F., & Tobin, C. (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: spectrographic and perceptual studies. *Journal of Speech and Hearing Research, 18*(3), 389-400.

Raphael, L. J., Dorman, M. F., & Liberman, A. M. (1980). On defining the vowel duration that cues voicing in final position. *Language and Speech, 23*(3), 297-307.

Raphael, L. J., Tobin, Y., Faber, A., Most, T., Kollia, H. B., & Milstein, D. (1995). Intermediate values of Voice Onset Time. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris* (pp. 117-127). New York: AIP Press.

Raphael, L. J., Tobin, Y., & Most, T. (1983). Atypical VOT categories in Hebrew and Spanish. *Journal of the Acoustical Society of America, 74*(S1), S89.

Recasens, D. (1985). *Estudis de fonètica experimental del Català Oriental Central*. Barcelona: Publicacions de l' Abadia de Monserrat.

Remijsen, B. (2004). Bert Remijsens's Praat scripts, Retrieved on 20. 2. 2007. from http://www.ling.ed.ac.uk/~bert/praatscripts.html.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech, 22*(2), 173-189.

Revoile, S., Pickett, J. M., Holden, L. D., & Talkin, D. (1982). Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. *Journal of the Acoustical Society of America, 72*(4), 1145-1154.

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics, 35*, 439-443.

Ringen, C., & Kulikov, V. (fc). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics*.

Ringen, C., & Suomi, K. (2012). The voicing contrast in Fenno-Swedish stops. *Journal of Phonetics, 40,* 419-429.

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

Rojczyk, A. (2009). The voicing contrast in Polish. *Church Slavic Studies, South Korea, 14*(2), 1-12. Retrieved from http://www.dbpia.co.kr/Journal/ArticleDetail/1081933

Rosner, B. S., López-Bascuas, L. L., García-Albea, J. E., & Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics, 28,* 217-224.

Ryalls, J., Cliché, A., Fortier-blanc, J., Coulombe, I., & Prud'hommeaux, A. (1997). Voice-onset time in younger and older French-speaking Canadians. *Clinical Linguistics & Phonetics, 11*(3), 205-212.

Ryalls, J., Provost, H., & Arsenault, N. (1995). Voice Onset Time production in French-speaking aphasics. *Journal of Communication Disorders*, *28*, 205-215.

Ryalls, J., Simon, M., & Thomason, J. (2004). Voice Onset Time production in older Caucasian- and African-Americans. *Journal of Multilingual Communication Disorders, 2*(1), 61-67.

Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on Voice Onset Time. *Journal of Speech, Language and Hearing Research*, *40*, 642-645.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*, 421-436.

Scobbie, J. M. (2005). Flexibility in the face of incompatible English VOT systems. In L. Goldstein & C. T. Best (Eds.), *Papers in laboratory phonology 9: Varieties of phonological competence* (pp. 367-392). Berlin: Mouton de Gruyter.

Sharf, D. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech, 5,* 26-30.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Shimizu, K. (1989). A cross-language study of voicing contrasts of stops. *Studia Phonologica, 23,* 1-12.

Shimizu, K. (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido Publishing.

Shrager, M. (2005). Neutralization of word-final voicing in Russian. *Journal of the Acoustical Society of America, 112*(5), 2419.

Simić, R., & Ostojić, B. (1989). *Osnovi fonologije srpskohrvatskoga književnog jezika* (2nd ed.). Nikšić: Univerzitetska riječ.

Simon, C., & Fourcin, A. J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America, 63*(3), 925-935.

Slis, I. H., & Cohen, A. (1969a). On the complex regulating the voiced-voiceless distinction I. *Language and Speech, 12*(2), 80-102.

Slis, I. H., & Cohen, A. (1969b). On the complex regulating the voiced-voiceless distinction II. *Language and Speech, 12*(3), 137-155.

Slowiaczek, L. M., & Dinnsen, D. A. (1985). On neutralizing status of Polish word-final devoicing. *Journal of Phonetics, 13*, 325-341.

Smith, B. L. (1978). Effects of place of articulation and vowel environment on 'voiced' stop consonant production *Glossa, 12*, 163-175.

Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics, 37*, 257–275.

Snoeren, N. D., Halle, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics, 34*(2), 241-268.

Soković-Perović, M. (2008). *Vowel duration as a function of obstruent voicing in Serbian*. Paper presented at the Annual BASEES Conference, Cambridge, March 2008.

Soković-Perović, M. (2009). Voicing-conditioned vowel duration in Southern Serbian. *Newcastle Working Papers in Linguistics, 15*, 125-137.

Soković, M. (1997a). Fundamental frequency change in two-syllable words with long rising accent. *Facta Universitatis, Series Linguistics and Literature, University of Niš, 1*(4), 283-290.

Soković, M. (1997b). Promena osnovne frekvencije u jednosložnim rečima sa dugosilaznim akcentom. *Zbornik radova XLI konferencije ETRAN-a, II*, 567-569.

Soković, M. (1997c). *Uloga prozodije iskaza u sintezi govora.* Unpublished MPhil thesis, University of Belgrade, Belgrade.

Sokolović, M. (1997d). Uticaj akcenata na formantsku strukturu vokala. *Srpski jezik, II*(1-2), 65-85.

Sokolović, M. (1998). Promena osnovne frekvencije u dvosložnim rečima sa dugosilaznim akcentom. *Naš jezik, XXXII*(3-4), 259 - 270.

Sokolović, M. (2010). Trajanje vokala kao obelezje kontrasta po zvučnosti u srpskom jeziku. *Godišnjak za srpski jezik i književnost, Filozofski fakultet u Nišu XXIII*(10), 423-436.

Solé, M.-J. (2011). Articulatory adjustments in initial voiced stops in Spanish, French and English. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1878-1881). Hong Kong.

Solé, M.-J., & Sprouse, R. L. (2011). Voice-initiating gestures in Spanish: prenasalization. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 72-75). Hong Kong.

Stathopoulos, E. T., & Sapienza, C. M. (1997). Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *Journal of Speech, Language and Hearing Research, 40*, 595-614.

Stathopoulos, E. T., & Weismer, G. (1983). Closure duration of stop consonants. *Journal of Phonetics, 11*, 395-400.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1-38). Hillsdale, New Jersey: Lawrence Erlbaum.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America, 55*(3), 653-659.

Subtelny, J. D., Worth, J., & Sakuda, M. (1966). Intraoral pressure and rate flow during speech. *Journal of Speech and Hearing Research, 9*, 498-518.

Suen, C. Y., & Beddoes, M. P. (1974). The silent interval of stop consonants. *Language and Speech, 17*, 126-134.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1074-1095.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62*(2), 435-448.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America, 82*(3), 847-863.

Summers, W. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America, 84*(2), 485-492.

Sweeting, P. M., & Baken, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research, 25*, 129-134.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America, 125*(6), 3974-3982.

Thomas, E. R. (2000). Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. *Journal of Phonetics, 28*, 1-25.

Tobin, S. (2009a). Gestural drift in Serbian-English speakers. UConn Language & Cognition Brownbag Talk.

Tobin, S. (2009b). Gestural drift in Spanish-English speakers. *Journal of the Acoustical Society of America, 125*(4), 2757.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schlieser (Eds.), *Methods in empirical prosody research* (pp. 1-28). Berlin, New York: De Gruyter.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America, 58*(2), 434-445.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America, 61*(3), 846-858.

van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics, 32*, 455–491.

van den Berg, J. (1958). Myoelastic-Aerodynamic Theory of Voice Production. *Journal of Speech and Hearing Research, 1*(3), 227-244.

Veloso, J. (1995). The role of consonatal duration and tenseness in the perception of voicing distinctions of Portuguese stops. *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 266-269). Stockholm.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, *92*(2), 723-735.

Walsh, T., & Parker, F. (1981). Vowel termination as a cue to voicing in post-vocalic stops. *Journal of Phonetics, 9*, 105-108.

Wang, W. S.-Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research, 2*(1), 66-73.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America, 71*(1), 187-195.

Waters, R. S., & Wilson, W. A. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics, 19*(4), 285-289.

Watson, I. (1990). Acquiring the voicing contrast in French: a comparative study of monolingual and bilingual children. In J. N. Green & W. Ayres-Bennett (Eds.), *Variability and change in French: essays presented to Rebecca Posner on the occasion of her sixtieth birthday* (pp. 37-60). London, New York: Routledge.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*, 247-274.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations. *Journal of Phonetics, 8*, 427-438.

Wells, J. C. (1990). Syllabification and allophony. In S. Ramsaran (Ed.), *Studies in the pronunciation of English, A commemorative volume in honour of A.C. Gimson* (pp. 76-86). London and New York: Routledge. Retrieved on 8. 6. 2011. from http://www.phon.ucl.ac.uk/home/wells/syllabif.htm.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America, 73*(4), 1322-1336.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics, 22*, 145-166.

Wetzels, W. L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language, 77*(2), 207-244.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America, 94*(3), 2151-2159.

Whiteside, S. P., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America, 116*(2), 1179-1183.

Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica, 58*(3), 196-210.

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5*, 169-184.

Wolf, C. G. (1978). Voicing cues in English final stops. *Journal of Phonetics, 6*, 299-309.

Yeni-Komshian, G., Caramazza, A., & Preston, M. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*, 35-48.

Zec, D. (2003). O mestu palatalnih afrikata ć, đ i č, dž u sistemu glasova srpskog jezika. *Južnoslovenski filolog, LIX*, 39-55.

Zimmerman, S. A., & Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America, 30*, 152-153.

Zue, V. W. (1976). *Acoustic characteristics of stop consonants: a controlled study*. Unpublished doctoral thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science. Available from DSpace@MIT http://hdl.handle.net/1721.1/29469.

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America, 66*(4), 1039-1050.