ACQUISITION OF SUBPHONETIC VARIATION BY GERMAN L2 LEARNERS OF ENGLISH

Inaugural-Dissertation

zur

Erlangung der Doktorwürde

der Philologischen Fakultät

der Albert-Ludwigs-Universität

Freiburg i. Br.

vorgelegt von

Anne-Julie Maurer

aus Lahr

WS 2013/2014

Erstgutachterin: Prof. Dr. Brigitte Halford

Zweitgutachter: Prof. Dr. Christian Mair

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission der
Philologischen, Philosophischen und Wirtschafts-
und Verhaltenswissenschaftlichen Fakultät: Prof. Dr. Bernd Kortmann

Datum der Fachprüfung im Promotionsfach: 16.05.2014

To my friends and family

Acknowledgements

First and foremost, I would like to thank Prof. Dr. Brigitte Halford, who has been a mentor to me for so many years. She is the most supportive and motivating supervisor one could wish for. Without her I would have never become interested in linguistics in the first place, and this dissertation would have never seen the light of day.

I am also very grateful to Prof. Dr. Christian Mair. His support, discussion, and the thoughtful criticism he provided in the *Oberseminar* were greatly appreciated.

To all of my participants – thank you so much! Without you, this dissertation would not have been possible. I am greatly indebted to each and every one of you for dedicating your time to this project. Also, thank you to the four speakers who recorded the stimuli for the perception experiment. You know who you are!

To Philipp Holz – thank you for designing the Soundcase software.

To Dr. Björn Schelter, Dr. Christian Langstrof, Marten Juskan & Jens Cornelis – thank you for all your help regarding equations, statistics, and particularly *R*.

To Dr. Rebecca Davies – thank you for editing and proofreading the early versions of this dissertation with such care. But even more so, thank you for being a truly wonderful, supportive, and loving friend.

To Marlene Greiwe and Stefan Kruppa – thank you for finding the solution to my R plot issues, and for always providing a second home when I needed it.

To my family – mom, dad, Christian, Melanie and Max – for teaching me that following your heart is always the right choice. I owe you so much.

To Andreas – I love you more than words can say.

Contents

Figures

Tables

Abbreviations

ANOVA	Analysis of Variance
AOL	Age of Learning
BC	Catalan Speakers from Barcelona
BICS	Basic Interpersonal Communication Skills
BNC	British National Corpus
C	Consonant
CALP	Cognitive and Academic Language Proficiency
CE	Monolingual Canadian English Speakers
CF	Monolingual Canadian French Speakers
CLIL	Content and Language Integrated Learning
COCA	Corpus of Contemporary American English
CP	Critical Period
ELF	English as a Lingua Franca
ESL	English as a Second Language
EU	European Union
FLA	Foreign Language Acquisition
FLA	Foreign Language Assistant
FMKS	Frühe Mehrsprachigkeit an Kitas und Schulen (*Association for Early Multilingualism in Day Nurseries and Schools*)
GLA	Gradual Learning Algorithm
GLM	Generalized Linear Model
IPA	International Phonetic Alphabet
L1	First Language
L2	Second Language
L2LP	Second Language Linguistic Perception Model
LOR	Length of Residence

LP	Linguistic Perception Model
NLM	Native Language Magnet Model
NZE	New Zealand English
PAM	Perceptual Assimilation Model
PH	Pädagogische Hochschule (*School of Education*)
sd	Standard Deviation
SE	Standard Error
SLA	Second Language Acquisition
SLM	Speech Learning Model
TC	Catalan Speakers from Toronto
TL	Target Language
V	Vowel
VLE	Vowel Length Effect
VLV	Vowel Length Variation
VOT	Voice Onset Time

1. Introduction

Most, if not all, adult second language (L2) learners initially have difficulty perceiving and pronouncing non-native sounds. Although some learners seem to be able to become quite competent speakers, many retain an audible accent. This is particularly the case when the feature in question is subphonemic in the target language, i.e. below the conscious level of even native speakers. The major aim of the present dissertation is to examine and describe how the acquisition of such a subphonemic feature – vowel length variation as a cue to final consonant voicing – proceeds in German learners of English. In English, vowels preceding voiced consonants are lengthened, so that TRAP (cf. Wells (1982)) is longer in *bad* than in *bat*. In contrast, in German vowel length is phonemic, so that *Stadt* 'city' and *Staat* 'state' are only distinguished by vowel length. The topic is complicated further by the fact that many German learners exhibit final devoicing, since German does not have voiced plosives in final position (*Rad* 'bike' and *Rat* 'advice' are pronounced the same). Thus, learners who produce inaccurate vowel length variation paired with devoiced final consonants have great difficulty signaling meaning. For this reason, it is of particular value to investigate how German learners of English acquire vowel length variation both in perception and production.

This thesis is built on an experiment conducted with two groups speaking one of the two most prominently taught varieties of English: American English and British English. Each of the two groups consists of 30 German students of English, 20-30 years old, who started learning English at age 10 or 11, and 10 native speakers of the same age. The German speakers have all completed a stay abroad of at least 4 months in the country of their target accent. Two control groups were recorded as well, each consisting of 3 students with no experience abroad. All speakers completed a speaking experiment consisting of four reading tasks varying in formality (minimal pairs, a word list, sentences and a text) as well as a perception experiment which focused on vowel length variation.

The variables chosen for analysis are the three front vowels KIT, DRESS and TRAP and the back vowel LOT. The vowels are studied in the environment of the three groups of voiced and voiceless plosives, i.e. /b/, /d/, /g/ and /p/, /t/, /k/. The vowels were chosen according to two criteria: Firstly, minimal pairs for all four vowels can be found easily. This is not the case for FOOT and STRUT, which are historically one and the same vowel, and which were excluded from the analysis for this reason. Secondly, LOT was included although it is pronounced differently in American English and British (long /ɑː/ vs. short /ɒ/), because it might give an indication as to how the different speaker groups deal with a phonetically different vowel. Moreover, KIT and DRESS have similar counterparts in German, whereas TRAP and LOT are absent from the German vowel inventory.ɑ

This thesis is designed as a contribution to the growing field of subphonetic studies as well as to that of second language acquisition (SLA). This means that a large part of this dissertation deals with the *whys* of language acquisition. Why do some learners manage to acquire vowel length variation easily while others struggle? Why should learners even try to achieve a native-speaker accent? Studying this is important for a number of reasons. Firstly, it gives hints as to theoretical issues regarding age constraints in SLA. A critical period (CP) has been proposed, after which language learning becomes very difficult or even impossible. Scovel, for example, says that

> [i]t is the nature of the human brain, not its nurture, that crucially affects [L2 pronunciation]. The onset of cerebral dominance, which seems to occur around the age of twelve, inhibits the ability of a person to master the sound patterns in a second language without an impinging foreign accent. (Scovel 1969: 254)

Whether this is truly the case has been hotly discussed for decades, with no satisfactory solution having been proposed as yet (cf. Lenneberg (1967), Krashen et. *al* (1979), Walsh & Diller (1981), Scovel (1988), Long (1990), Patkowski (1990), Hurford (1991); cf. Long (2005) for an overview). Researchers disagree as to the age at which the CP ends. Scovel (1969) proposes an age of 12 years; Patkowski (1990) argues for an age of 15 years. Long (1990) posits the CP even earlier, namely at six years.[1] Other studies have shown, however, that an early onset of language learning (*Age of learning*, AOL) before age six does not guarantee accent-free speech, either (cf. Flege *et al.* (1997), Thompson (1991)). Yet others, such as Flege *et al.* (1995), Bongaerts (1999) and Moyer (1999), report on a few adult learners who manage to attain a native-like accent. This might mean that a CP, should it exist, ends even earlier, or that it can be influenced or even deactivated by other factors (see below). Due to these findings, the critical period hypothesis is largely seen as outdated and too simplistic. However, there is still the general consensus that pronunciation is the first ability to be affected by age. This would mean that all of the speakers in the present study should have a foreign accent to some degree, since they started learning English at age 10 or 11.[2] Still, it is unclear whether the CP relates only to phonemes or to subphonemic contrasts –

---

[1] Note that different linguistic levels have been researched, the general consensus being that phonology is affected earlier than morphosyntax and lexicon.

[2] Except five bilingual speakers, cf. Appendix A.1.

such as vowel length variation – in addition. Moreover, older learners have been shown to be superior to younger learners in an institutional setting (cf. Perales *et al.* (2004), Muñoz (2004)). This might indicate that the learners in the present study might be quite competent and quick in the acquisition of vowel length variation after all.

Secondly, studying accent is important for teaching. Knowing which factors lead to or inhibit success in an L2 is imperative for instructional planning. For instance, should research show that exposure to the L2 is crucial, this might be difficult for speakers who learn an L2 at school rather than in an immersion setting, which is predominantly the case regarding English in Europe. However, a first step might be to ensure that learners are mainly taught by native speakers or – in the absence of native speakers – that media such as recordings or videos are used to expose learners to native speakers of the L2 early on. Again, research cannot give a clear answer as to the factor of exposure yet (often measured as LOR – *length of residence* in the L2 setting). Positive effects have been reported by Purcell & Suter (1980), Flege & Fletcher (1992), Flege *et al.* (1995), whereas Flege (1988), Thompson (1991) and Moyer (1999) did not find such a facilitating effect. However, Flege (1988) claims that there is a rapid initial effect of LOR, which has disappeared in adult learners. It thus seems that AOL is a stronger factor than LOR. Studies which tested the influence of formal instruction showed a rather disheartening picture for language teachers: Thompson (1991) and Flege *et al.* (1995) did not find any effect of *number of years of instruction*, and even though Flege & Fletcher identified it as a significant factor, it only accounted for 5.2% of the variance in a foreign accent (1992: 356). This might also be due to the fact that teachers at school do not usually include pronunciation training in their sessions. In Germany, much attention is focused on teaching students the correct pronunciation of the interdental fricatives, which German lacks. Other sounds are not usually given much attention, even if they are missing from the German inventory (e.g. TRAP). Instead, the focus is on grammar rules and communicative fluency. Efficiency and success in communication rely strongly on correct pronunciation, however: “Learners with good English pronunciation are likely to be understood even if they make errors in other areas, whereas learners with bad pronunciation will not be understood, even if their grammar is perfect” (Pourhosein Gilakjani 2012: 96). Bent *et al.* find that vowel errors impact intelligibility more severely than do consonant errors (2007: 331). Moyer (1999: 92 ff.) reports that English learners of German who had received segmental as well as suprasegmental phonetic training attained ratings closer to a native level. She proposes this to be one of the three main factors for an L2 accent. Several other researchers have also suggested that pronunciation instruction should be part of language teaching (e.g. Pennington & Richards (1986), Boyle (1987), Chun (1988), Pennington (1989), Perlmutter (1989), Macdonald *et al.* (1994), Derwing *et al.* (1997; 1998), Couper (2003)). Therefore, it might be time to revise the focus of classroom teaching and put more emphasis on pronunciation training. Piske's phrases this as follows:

> L2 pronunciation still receives relatively little attention in most foreign language classrooms. […Studies] indicate that this is rather unfortunate, because just like exposure to high-quality input, intensive training in the perception and the production of L2 sounds could also help foreign language students develop a more accurate L2 pronunciation. (Piske 2007: 308)

This dissertation strongly supports Piske's view. Some ideas on how to incorporate vowel length variation in pronunciation training can be found in chapter 5.

Thirdly, many learners themselves place a high priority on attaining a native-like accent. For instance, the participants in the present study were asked to rate the importance of a native-like accent on a scale from 1 to 10, where 1 equals *not at all important* and 10 equals *highly important*. The average rating was 8.5 ($sd = 1.8$). This importance obviously goes hand-in-hand with motivational factors, which have been established to be a significant factor by e.g. Suter (1976), Purcell & Suter (1980), Elliott (1995), Flege *et al.* (1995) and Bongaerts *et al.* (1997). In addition, non-native speakers may experience undesirable consequences as a result of their accent, as Flege states:

> [Foreign accents] may make non-natives difficult to understand, especially in non-ideal listening conditions […]. They may cause listeners to misjudge a non-native speaker's affective state […], or provoke negative personal evaluations, either as the result of the extra effort a listener must expend in order to understand, or by evoking negative group stereotypes. (Flege 1995: 233 f.)

One area of discussion is the explanation for foreign accents. Three major hypotheses have been suggested. Researchers such as Sapon (1952), Penfield (1965) and Lenneberg (1967) have argued that the maturation of the brain and resulting loss of plasticity make it impossible for late learners to acquire a native-like accent due to "a diminished ability to add or modify sensorimotor programs for producing sounds in an L2" (Flege 1995: 234). Other researchers claim that a foreign accent results from insufficient and/or inadequate phonetic input. Kuhl (2004) proposes that accents develop due to previous experience with native language (L1) sounds (more on this in section 2.3.2.1). Psychosocial reasons such as lack of motivation have also been suggested. Piske *et al.* summarize that

> a wide variety of variables influence degree of foreign accent. These variables include age of L2 learning, length of residence in an L2-speaking country, gender, formal instruction, motivation, language learning aptitude and amount of native language (L1) use. Age of L2 learning appears to be the most important predictor of degree of foreign accent. However, the relative importance of the other variables is uncertain. This is because many variables relating to subject characteristics tend to be confounded.
> (Piske *et al.* 2001: 191)

It is obvious that some of these variables are much easier to determine than others. For instance, motivation is quite difficult to assess. Nevertheless, it has been reported to be a major factor in the successful acquisition of an L2 (cf. Suter (1976), Purcell & Suter (1980); Bongaerts *et al.* (1995)). Furthermore, comparisons across studies remain difficult due to the different nature of elicitation and rating techniques employed (cf. Piske *et al.* 2001: 193 f.; Moyer 2007: 113 f.). This means that "observation that L2 foreign accent studies sometimes yield divergent results should not lead one to conclude that degree of L2 foreign accent cannot be scaled reliably and validly" (Piske *et al.* 2001: 193). Later studies which focused on factors which are often confounded with age showed that these factors might actually be much stronger than previously thought. For instance, Moyer reports that the influence of age of immersion is marginal when other variables are factored in (cf. 1999: 94). In her study, age of immersion alone only accounts for 1% of the variation in accent, while professional motivation accounts for 32%, segmental and suprasegmental input for an additional 12% (Moyer 1999: 94 f.). She concludes that there is[3]

> a balance of socio-psychological and exposure-type variables for predicting accent, including contact with native speakers, length of residence and age of onset, as well as intention to reside in the TL-speaking environment permanently or long-term, comfort with assimilation to the TL culture, desire to improve accent, and sense of overall fluency. (Moyer 2007: 112)

DeKeyser stresses the importance of aptitude in the ultimate attainment of the L2, showing that late learners with high verbal aptitude performed to a similar high standard as child learners (cf. 2000: 514). He states that since the implicit learning mechanisms of children are no longer available to adults, adults who want to achieve a near-native level of language competence must possess "above-average analytical abilities" (DeKeyser 2000: 515).

In the end, the ultimate question seems to be one of *nature* vs. *nurture*. Does nature prevent native-like language learning at some point in a person's lifetime, for example through maturational constraints (lack of neural plasticity) or firm establishment of the L1 which influences the acquisition of an L2? Or is it nurture which hinders acquisition, e.g.

---

[3] TL = Target Language.

through lower quality input, less motivation and altered affect?[4] Bialystok & Hakuta state that "social factors conspire to ease the effort for young children by providing a nurturing environment, simplified input, educational opportunities, cooperative peers, and other supporting aspects of a social context that facilitate the acquisition of any language" (1999: 178). In any case, we must bear in mind that AOL might go hand-in-hand with other factors, which might be the true reason why some early learners are more successful in learning the pronunciation of their L2:

> Early onset does not simply predict greater exposure or contact, but a more advantageous balance of exposure types – experiential *quality*, so to speak – positively affecting both affective and cognitive strategies for improving pronunciation abilities. […U]nderstanding the age factor requires an appreciation for how developing L2 experience feeds into learner goals, decisions and behaviors […], which directly and/or indirectly lead to attainment outcomes. (Moyer 2007: 199 f., original emphasis)

As such, learners who learn their L2 early, perhaps even in an immersion setting, might be more likely to use the L2 in various domains, adopt the culture and language of the L2 as dominant, and restrict their L1 to the home and family life. This has been shown to have a strong influence on their accent (cf. Flege *et al.* (1997), Singleton (2000), Yeni-Komshian *et al.* (2000)). Regarding the German participants in this study, it is noteworthy that all of them except five bilingual speakers acquired English in a classroom setting at age 10 or 11. This means that they can be said to have the same AOL, meaning that this cannot be a major factor influencing their performance.[5] Instead, other factors such as the effect of a stay abroad and the reason for it, as well as phonological proficiency (measured as absence of devoicing and separate categories for TRAP and DRESS) will be evaluated as factors which influence other aspects of accent – in this case the acquisition of the vowel length variation preceding voiced and voiceless plosives.

Many factors which may influence the strength of a foreign accent have been mentioned. The following list is a brief overview of the factors which were found to be highly influential in the acquisition of accent-free pronunciation by L2 learners:

- Learner orientation to the target language (Dörnyei & Skehan (2003))
- Motivation (Thompson (1991), Flege *et al.* (1999), Moyer (1999; 2004; 2007))
- Concern for pronunciation accuracy (Elliott (1995), Purcel & Suter (1980))

---

[4] DeKeyser argues that input cannot be seen as a factor in phonology, because "it is precisely in the linguistic domain where input varies the least – phonology – that the age effects are most readily apparent, and it is at the stage where the comprehensibility of input should be the least problematic – in the later stages of acquisition – that adults clearly perform worse than children" (2000: 519).

[5] Unless perhaps one considers the fact that individuals vary in their degree of biological and neurological maturity, but then it would be impossible to compare any two individual speakers.

- Attitudes towards the culture of the L2 (Major (1993), Stokes (2001))
- Accent self-rating (Moyer (2004; 2007))
- Oral mimicry aptitude (Purcell & Suter (1980), Thompson (1991))
- Segmental and suprasegmental phonological feedback (Moyer (1999; 2004))

The participants in the present study were asked to fill out a questionnaire on their language background before the experiment, which included some of the factors mentioned above. For instance, they were asked how important they found it to attain a native-like accent and how good they judged their accent to be. Since all participants are university students of English, their attitudes towards the culture of the L2 can be said to be widely positive. Motivation, however, might differ. See section 3.4.1 for a closer description of the questionnaire data, and Appendix B.1 for a copy of the questionnaire.

Chapter 2 will outline the theoretical background on which the present dissertation is based. First, the system of vowel length and vowel length variation (section 2.1.1) will be illustrated, i.e. the way in which vowel length and vowel length variation is coded in American English, British English and German. Section 2.2 will introduce factors which influence vowel length variation and will outline the way in which these were or were not controlled for in the present study. Next, section 2.3 will give an overview of how perception and production develop in an L2, and it will present four different speech perception models. Section 2.4 will examine cues relevant to vowel length and vowel length variation. This section will also discuss how children acquire vowel length variation as a function of postvocalic consonant voicing. Finally, section 2.5 outlines the research questions which emerge from the theoretical background.

Afterwards, chapter 3 will introduce the Methodology of the dissertation at hand. The participants will be presented (3.1), and the production (3.2) and perception (3.3) experiments will be explained. Section 3.4 gives information on the data which was gathered for the analysis: Details on the questionnaire can be found in section 3.4.1; the three sections that follow illustrate how vowel length and vowel length variation (3.4.2), the merger of TRAP and DRESS (3.4.3) as well as devoicing (3.4.4) and word frequency (3.4.5) were measured.

Chapter 4 presents the results and analysis. First, section 4.1 will give an overview of the statistical methods used in this dissertation, namely boxplots, (generalized) linear models, the Wilcoxon rank-sum test and the Kruskal-Wallis analysis of variance, the chi² test, and other statistical notations which are reported on. After this general introduction, the chapter offers an analysis of the production and perception experiments in sections 4.2 and 4.3, respectively. Both sections are sub-divided into results from the British group and the American group. Section 4.4 answers the question of whether production and perception are linked. The role phonemic success plays in the acquisition of subphonemic variation is examined in section 4.5. Lastly, section 4.6 is designed as a summary of the findings.

The discussion in chapter 5 will return to the research questions outlined in section 2.5 before considering where the information gained through the perception and production experiments leads us. It will be argued that vowel length variation is a highly useful feature for German learners of English and that this needs to be reflected in pronunciation teaching. In this respect, this dissertation will take an applied perspective and discuss when and how vowel length variation should be taught. Ideas for suitable pronunciation material are also included. As a last point, the political dimension exemplified by the European Commission's "mother tongue + 2" Action Plan will also be considered. A final conclusion will be drawn, and possibilities for further research will be indicated.

2. Theoretical Background

2.1. Vowel Length and Vowel Length Variation

2.1.1 Vowel Length and Vowel Length Variation in English

English uses double cues to signal differences between adjacent vowels: quantity and quality. Vowels are qualitatively different from each other in vowel space, i.e. they differ in F1 and F2. As such, FLEECE is higher and more front than KIT. Moreover, vowels also show duration differences: FLEECE is longer than KIT. The question, therefore, is whether both quantity and quality are distinctive in English, or whether quality alone cues the identity of vowels, as has been proposed by Pinker (1994: 168). Hillenbrand *et al.* (2000) phrase it in the following way: "Duration has long been a key feature in the description and analysis of vowels. The chief phonological question concerns whether duration should be considered a contrastive or redundant feature" (2000: 3013). The answer seems to be that duration is a distinctive feature in English, since a growing number of studies have shown solid durational differences between vowels, both in standard and dialectal variants of English (cf. Peterson & Lehiste (1960), Crystal & House (1982; 1988a,b,c), Labov & Baranowski (2006), Jacewicz *et al.* (2007), Tauberer & Evanini (2009)). Maclagan & Hay (2007) present a particularly convincing argument in their examination of DRESS and FLEECE in New Zealand English (NZE). The two vowels are merged in F1/F2 space, but are clearly distinct in duration, and they are perceived as distinct vowels by listeners.

This duality of cues has led researchers to reconsider the systemic distinction between vowels. For instance, Labov (1994) uses the terms *tense* and *lax* to capture the difference between adjacent vowels such as FLEECE and KIT. According to him, *tense*, as a term referring to muscular tension required to articulate a sound, "is based on more indirect acoustic evidence. The feature [tense] will appear as an abstract assembly of several phonetic features" (cf. Labov 1994: 174). Such features are, for instance, that tense vowels are located at the peripheries of vowel spaces, that they are longer than lax vowels, have greater amplitude, etc. (cf. Labov 1994: 175). Labov goes on to show that tense and lax vowels also behave differently in the history of sound change: Tense vowels tend to rise, whereas lax vowels tend to fall. However, Langstrof (2006a) has shown that DRESS and TRAP cannot fully be considered lax vowels, since they rise in varieties such as Australian English and NZE, which is a clear counterexample to the principles proposed by Labov (1994). Peterson & Lehiste add that the difference between tense and lax vowels "may be described as a difference in the articulatory rate of change associated with the movement from target position to the following consonant" (1961: 274). According to them, tense vowels maintain the target position longer and then rapidly move to the consonant, whereas lax vowels involve "a short target position and a slow relaxation of the hold" (Peterson & Lehiste 1961: 274). Thus, there seems to be some validity to the tense/lax distinction after all, even if it is not unproblematic. Van der Feest & Swingley conclude that "[i]n English, coda voicing may affect vowel duration more strongly than the tense/lax distinction does" (2011: 57).

The picture is further complicated by the fact that English also includes subphonemic vowel length variation. The fact that English vowels are longer before certain segments – their so-called *extrinsic* duration – was commented on as early as 1903. In his book *Englische Lautdauer*, Meyer remarks that vowels are shorter before tense final consonants than lax ones, and that the higher the vowel, the shorter its length (1903: 107 f.). Similar studies were undertaken by Jones (1950), House & Fairbanks (1953), Wells (1962; 1982), and many others. A number of phonetic studies have measured the duration of American English vowels before voiced and voiceless plosives (cf. Peterson & Lehiste (1960), Port (1981), Crystal & House (1982; 1988a,b,c) and Luce & Luce (1985)). Table 2.1-1 below lists the average duration ratios between the four vowels examined in the present study – KIT, DRESS, TRAP, LOT – before voiced and voiceless plosives. The shortest vowel, KIT preceding voiceless plosives, serves as a baseline and is set at 1. This means that the table has to be read as e.g. "TRAP before voiced plosives is 2.56 times the duration of KIT before voiceless plosives."

Table 2.1-1. Duration ratios between KIT, DRESS, TRAP and THOUGHT preceding voiced and voiceless plosives in American English. LOT was not included, which indicates that the speakers in the studies might have merged LOT with THOUGHT. THOUGHT is therefore given in this table. Numbers averaged from the studies of Peterson & Lehiste (1960), Port (1981), Crystal & House (1982; 1988a,b) and Luce & Luce (1985). Source: Langstrof 2006b: 109.

		Vowel preceding voiceless plosives	**Vowels preceding voiced plosives**
Short vowels	KIT	1.0	1.4
	DRESS	1.1	1.54
Long vowels	TRAP	1.83	2.56
	THOUGHT	1.72	2.41

The table exemplifies the following generalizations: TRAP and THOUGHT contrast with KIT and DRESS in that they have much higher average durations.[6] In addition, duration and openness of the vowel correlate: lower vowels are intrinsically longer (cf. Meyer 1903: 107 f., Peterson & Lehiste 1960: 703 f.).[7]

De Lacy also finds that duration varies according to the voicing of the consonant that follows: for instance, the length of KIT before a voiceless consonant is approximately 73% of the length before a voiced consonant (cf. de Lacy 1998: 5). Peterson & Lehiste put the figure at 66% (1960: 702), van Santen, 67% (1992: 538), Chen, 61% (1970: 138) and Mack, 54% (1982: 175). However, the last three researchers' percentages are based on studies including all vowels; Peterson & Lehiste (1960) as well as de Lacy (1998) only include short vowels such as KIT. Peterson & Lehiste also state that the variation in vowel length is greater in long vowels than in short vowels: Short vowels preceding a voiceless consonant are 71% of the length of short vowels preceding voiced consonants, while this difference only amounts to 66% in long vowels (1960: 702).

Another finding which must be mentioned here, because it has direct consequences for the present study, is that there seems to be evidence that syllable structure influences vowel length. Port (1981) and Klatt (1973) publish data showing that vowels in monosyllabic words are longer than vowels in polysyllabic words (counterevidence can be found in Umeda (1975)). This effect also relates to the vowel length variation between vowels preceding voiced or voiceless plosives. For instance, Klatt found that in monosyllables, vowels preceding voiced plosives are 34% longer than when they precede voiceless plosives, whereas this difference is reduced to only 12% in polysyllabic words (cf. Klatt 1973: 1103). These findings result in the fact that the present study only includes monosyllabic words as tokens in the production and perception experiment in order to control for this effect.[8]

Interestingly, vowel length variation has also been shown to apply for coda clusters such as *nasal* + *C* or *liquid* + *C*[9] (cf. Peterson & Lehiste (1960), Chen (1970), Mack (1982), Crystal & House (1988a), van Santen (1992), de Lacy (1998)). In these clusters, it is "the voicing of the C [… which] affects vowel duration, the voicing of the sonorant has no effect" (de Lacy 1998: 1). Examining KIT before a number of such consonant clusters, de Lacy finds that vowels are always significantly shorter before clusters ending in a voiceless consonant

---

[6] Peterson & Lehiste list TRAP under long vowels (1961: 274), as do Crystal & House (1988b: 265).

[7] This effect is discussed in more detail in section 2.2.3.

[8] There are two exceptions which are listed and explained in footnote 31.

[9] *C* = 'consonant'.

than a voiced consonant. Moreover, the vowels are shorter before a cluster than before the same single consonant in which the cluster ends. For instance, KIT before /ld/ is 201 ms, whereas it is 213 ms before /d/ alone (cf. de Lacy 1998: 7). He takes this as evidence that the number of consonants in the cluster influences vowel length. However, he admits that this shortening does not fall into the perceptible range (de Lacy 1998: 7; cf. Lehiste 1970: 13, who gives 10-40 ms as a “just noticeable difference”). De Lacy explains that this shortening is caused due to syllables having a target length specified in the grammar, from which the consonant cluster subtracts more than a single consonant. All vowels have minimum inherent lengths, however, which cannot be further shortened (cf. de Lacy 1998: 7 ff.; cf. also Klatt 1973: 1102). In any case, the tokens chosen for the present study do not include coda clusters.

We can conclude that vowel pairs in English are differentiated both by quantity and quality, and that both characteristics are distinctive. Kluender *et al.* posit that “[i]t is not just that more cues are better than fewer (although this may well be true); it is that certain cues can have optimal perceptual effect only in the context of other cues, and vice versa” (1988: 157).

The question remains where this allophonic quantitative difference in duration in vowels preceding voiced or voiceless plosives originates from. Since the 1970s, scholars have been searching for an explanation for this “vowel length effect” (VLE). Two explanations seem plausible: physiological or phonological reasons. Lehiste proposes that the “greater length of low vowels is due to the greater extent of the articulatory movements involved in their production” (1970: 19). Support for the articulatory explanation comes from a study examining vowel length variation in American English in vowels preceding flaps. Fox & Terbeek found that vowels are significantly longer before flapped /d/ than flapped /t/ (cf. 1977: 32). They argue that this is a universal characteristic and that it is “due to the physiological constraints involved in producing a vowel preceding a voiced consonant *vs* voiceless consonant rather than any phonological distinction” (Fox & Terbeek 1977: 33, original italics). Similarly to Lehiste (1970), Belasco (1953) hypothesizes that greater articulatory force is required to produce a voiceless consonant than a voiced one: “The anticipation of a consonant requiring a ‘strong’ force of articulation will tend to shorten the preceding vowel since more of the total energy needed to produce the syllable is concentrated in the consonant” (1953: 1016). However, the effect only holds in syllable-final position; vowels are not lengthened after voiced initial consonants, which calls into question the total energy hypothesis proposed by Belasco (cf. Kluender *et al.* 1988: 154).

Halle & Stevens (1967) and Chomsky & Halle (1968) name "laryngeal adjustments needed to maintain glottal vibration during oral constriction or closure" as the reason for the VLE (Kluender *et al.* 1988: 154). Counterevidence can be found in Lehiste (1970) and Chen (1970), the latter of whom states that if the explanation were true, *vowel – sonorant – obstruent* sequences (e.g. *cart* vs. *card*) should show the sonorant alone to vary in length. "The obvious reason is that in such a sequence the vowel is separated from the [voice] obstruent and is, therefore, shielded from the immediate effect of the laryngeal adjustment which takes place between the sonorant and the obstruent" (Chen 1970: 148). However, Chen found that both the vowel and sonorant in these conditions varied in length.[10] Moreover, vowel length variation has also been shown to be present in whispered speech, where there is no laryngeal adjustment (cf. Scharf (1964)).

Finally, Javkin (1975) proposes the hypothesis that the VLE is due to auditory factors: "Glottal pulsing during the closure interval is perceived by listeners as an extension of the preceding vowel. Therefore, vowels are heard as longer before voiced than before voiceless consonants" (cited in Kluender *et al.* 1988: 155 f.). He reasons that since listeners perceive lengthening before voiced consonants, they reproduce matching lengthening in their speech. However, Kluender *et al.* find this explanation to be lacking, because "it is unclear why detectable errors in apparent length would not be corrected" by speakers, which raises the question "why listeners/speakers would tend to exaggerate perceived vowel lengthening beyond what is afforded by glottal pulsing alone" (Kluender *et al.* 1988: 156).

Other researchers believe that there are phonological reasons behind the VLE. Among them is Lisker (1974), who claims that each phoneme has a duration specified in the grammar of a language. Labov & Baranowski (2006) support the phonological explanation by showing that there are dialects in which vowel length is distinctive. Finally, Tauberer & Evanini (2009) examine vowel shifts, which seem to provide evidence for the phonological explanation as well, since vowels which undergo shifting change their duration accordingly. Even if two vowels overlap in the vowel space, they are not confused with each other by speakers if they differ in length (cf. also Maclagen & Hay (2006) for NZE). In addition, subphonemic vowel length variation such as in English has been found in many other languages as well (German by Maack (1953), Spanish by Zimmerman & Sapon (1958), Norwegian by Fintoft (1961), Danish by Fischer-Jørgensen (1964), Swedish by Elert (1964), Dutch by Slis & Cohen (1969), Russian and Korean by Chen (1970), French by Laefur (1992)). This large number of findings

---

[10] Vowel in *cart* 118 ms, /r/ 138 ms; in *card*: vowel 157 ms, /r/ 169 ms. cf. Chen 1970: 149.

led researchers to believe that the VLE was a phonetic universal. One study which can serve as a particular clear support for the phonological hypothesis is Mack (1982), who showed that in French, the shortening of vowels before voiceless consonants is not nearly as advanced as in English. If physiological reasons alone were behind this variation, no difference should be observable between English and other languages (cf. de Lacy 1998: 14). Chen finds an interesting and valuable midway between the two competing hypothesis, saying that

> [W]e may tentatively conclude that (a) it is presumably a *language-universal* phenomenon that vowel duration varies as a function of the voicing of the following consonant, and (b) the extent, however, to which an adjacent voiced or voiceless consonant affects its preceding vowel durationwise is determined by the *language-specific* phonological structure. (Chen 1970: 139, original italics)

The studies mentioned so far have dealt extensively with American English. Unfortunately, vowel length variation in British English has not been observed steadily, unless it was part of a larger study on a certain phoneme. This means that it is not clear whether vowel length variation is actually the same in American and British English. The general opinion seems to be that subphonemic vowel length variation exists in British English, too, but to a lesser extent.

The following section will outline how vowel length and vowel length variation is encoded in the German phonological system. One similarity to English can serve both as a concluding remark to this section and as an introductory remark to the following one: “[Es] ergiebt sich die interessante tatsache, dass im norddeutschen die dauer des vokals abhängt von der höhe der zungenstellung: am kürzesten sind die höchsten vokale i und u, länger ist e, am längsten o und a” (Meyer 1904: 354).[11]

2.1.2 Vowel Length and Vowel Length Variation in German

German and English share a long history of common ancestry before English developed out of the Low German branch of languages, and German out of the High German branch. Therefore, it stands to reason that both languages should share a number of characteristics, while having developed features unique to the language after separating. German, like English, has a vowel system which is characterized by oppositions both qualitative and quantitative in nature. For instance, the words *Miete* ‘rent’ and *Mitte* ‘middle’ are minimal pairs differing in the length of the first vowel: /ˈmiːtə/ vs. /ˈmɪtə/. Moreover, /iː/ is qualitatively

---

[11] Translation: “The interesting fact emerges that in Northern German, the duration of the vowel depends on the height of the tongue: The shortest vowels are the highest vowels i and u, e is longer, o and a are the longest.”

different from /ɪ/ - it is higher and more front, as in English. In total, German has eight long vowels (/iː, eː, ɛː, yː, øː, aː, uː, oː/)[12] and three diphthongs /aɪ, aʊ, ɔʏ/. There are seven short vowels (/ɪ, ɛ, ʏ, œ, a, ɔ, ʊ/). Vocalized *r* /ɐ/ and schwa /ə/ are an exception in so far as they only appear in unstressed syllables, but can occur both in open and closed syllables. Long vowels also appear in open and closed syllables; short vowels occur only in closed syllables.

Regarding the quality of vowels, long vowels in stressed syllables in German are usually [+tense], while short vowels are [-tense]. An exception is the German front vowel /a/, which does not differ much in quality, but does in quantity (*Stadt* /ʃtat/ 'city' vs. *Staat* /ʃtaːt/ 'state'). Maas (2002: 18) argues that the use of quantitative differences is perceptually much more difficult than the use of qualitative differences, which is why many languages use a combination of both. Languages such as Finnish which rely exclusively on quantitative differences mark the difference in a very robust ratio of 1:3 (cf. Maas 2002: 18). It is quite difficult to ascertain whether vowel quality, vowel quantity, or a combination of both is distinctive in German. This question might be easier to answer if the role of quality and quantity in perception was studied. Weiss (1976; 1977; 1978) and Wängler & Weiss (1975) researched this using stimuli of all 15 German vowels in a /ˈb__tən/ frame. Wängler recorded the tokens himself and changed the vowel quality in six steps from long to short (Weiss 1976: 33 f.). The shortest vowels were shorter than 190 ms; the longest exceeded 400 ms. In total, 287 stimuli were used and played in random order to 20 native speakers of German from different regions. Wängler & Weiss (1975: 196 f.) found a number of striking results:

- Speakers from Northern Germany use qualitative cues more strongly; speakers from Southern Germany rely more on quantitative cues.
- Quantity and quality are inversely proportional: The more clearly a listener can perceive the quality of a vowel, the less important quantity is, and vice versa. Moreover, the closer two vowels are in quality (e.g. /aː/ and /a/), the more important the quantity cue is.
- Quantity seems to be a stronger cue for some vowels (e.g. /œ/ and /øː/), as /œ/ will be perceived as /øː/ if lengthened enough; quality is more important in others (/iː/, /yː/ and /uː/ are never perceived as their lax counterpart, even if they are shortened drastically).
- Low vowels are more often classified by their quality, high vowels more often by their quantity.

---

[12] Many younger speakers have merged /eː/ and /ɛː/ under /eː/.

It thus seems that quality and quantity mutually influence each other, and that both cues are used by speakers in order to distinguish sounds, even though the weighing of the cues may differ in some instances. These two characteristics are not the only ones which are being discussed as distinctive features of vowel length in German, however. Researchers such as Jespersen (1904), Trubetzkoy (1939), Vennemann (1994), Spiekermann (2000), Maas (2002) and others have proposed a third distinctive characteristic, namely syllable cut correlation.[13] This theory states that the length of a vowel is directly connected to accent – long vowels appear primarily in open syllables, while short vowels are used in closed syllables. Moreover, syllable cut correlation considers whether or not the nucleus vowel is linked to a following coda consonant.

Eduard Sievers (1980: 115 f.), following the theory proposed by Jespersen, uses the terms *schwach / energisch geschnittener Akzent* (‘smoothly’ and ‘abruptly cut accent’).[14] In the example of *Miete* and *Mitte* above, /iː/ in Miete can fade out smoothly, while /ɪ/ in Mitte is cut off abruptly by the following /t/. Sievers explains the abruptly cut accent as follows:[15]

> Hier wird der Vocal durch den folgenden Consonanten noch in dem Momente seiner grössten Stärke abgeschnitten. Dies hat zur Folge, dass der Consonant selbst mit starkem Exspirationsdruck gesprochen wird […]. Auf Längen ist er im Deutschen seltner, weil es unbequem ist, den Vocal mit voller Energie längere Zeit auszuhalten. (Sievers 1980: 115 f.)

In contrast, smoothly cut accent is explained as follows:[16]

> Der schwach geschnittene Accent ist den meisten unserer langen Vocale und Diphthonge […] eigen. Hier tritt die Abschneidung des Vocals erst in einem Momente ein, wo dessen Intensität bereits sehr geschwächt ist. […F]ür das Ende des Vocals [wird] der Exspirationsdruck stark herabgesetzt, im nächsten Moment aber für den Consonanten wieder erheblich verstärkt. (Sievers 1980: 116)

In the following, Sievers uses *Schallsilbe* (‘sonority syllable’) and *Drucksilbe* (‘pressure syllable’) in order to describe his concept further. In a sonority syllable, the nucleus of the syllable is characterized by a maximum in sonority, whereas a pressure syllable consists of a nucleus which is articulated by a maximum of energy in the glottal or subglottal area. In smoothly cut syllables, sonority syllable and pressure syllable occur in parallel; abruptly cut

---

[13] Note that English has also been described as a syllable cut language, most notably by Vennemann (2000) and Murray (2000; 2002), who argue that Middle English Open Syllable Lengthening is “derivative of a major prosodic change relating to the phonologization of syllable cut” (Murray 2002: 103). Cf. also Murray 2002: 125.

[14] Note that the 1980 edition is a reprint of the original 1872/1876 edition printed in Weimar/Leipzig.

[15] Translation: “This is when the vowel is cut off during the moment of its energy maximum by the consonant following it. This results in the consonant itself being pronounced with strong expiration pressure. This occurs rarely after German long vowels, because it is inconvenient to hold the vowel with full energy for a longer period of time.”

[16] Translation: “The smoothly cut accent occurs after most of our long vowels and diphthongs. Here, the cut appears only in the moment when the intensity of the vowel is already quite weak. The expiration pressure is lowered towards the end of the vowel, but increases again immediately for the following consonant.”

syllables show a divergence of sonority and energy. This is shown in Figure 2.1-1 below, using the examples of *Rate* ‘rate’ (left) and *Ratte* ‘rat’ (right):

Figure 2.1-1. Smooth (left) and abrupt (right) syllable cut. The gray line indicates the tonal syllable; the black line, the pressure syllable. Source: Maas 2002: 16.

Syllable cut correlation functions as a suprasegmental prosodic phonological opposition. Becker even goes so far as to say that the traditionally used descriptive characteristics “tenseness and duration are phonologically irrelevant allophonic correlates of the prosodic distinction” (Becker 2002: 89, my translation). Instead, the main difference between the vowels is their embedding into the syllable structure and their ensuing difference in length (cf. Becker 2002: 89).

In recent years, researchers have followed the terminology proposed by Jespersen (1904) and spoken of *Vokalischer Silbenanschluss* ‘vocalic syllable contact’ in order to distinguish between syllables which show a smooth or an abrupt cut. The term relates to the way in which the coda consonant is or is not linked to the nucleus of the syllable. Syllables with a smoothly cut (tense) vowel have a consonant in loose contact (cf. Figure 2.1-1, left, and Figure 2.1-2, top), whereas syllables with an abruptly cut (lax) vowel have a consonant in close contact (cf. Figure 2.1-1, right, and Figure 2.1-2, bottom). The following Figure 2.1-2 can serve as further clarification:

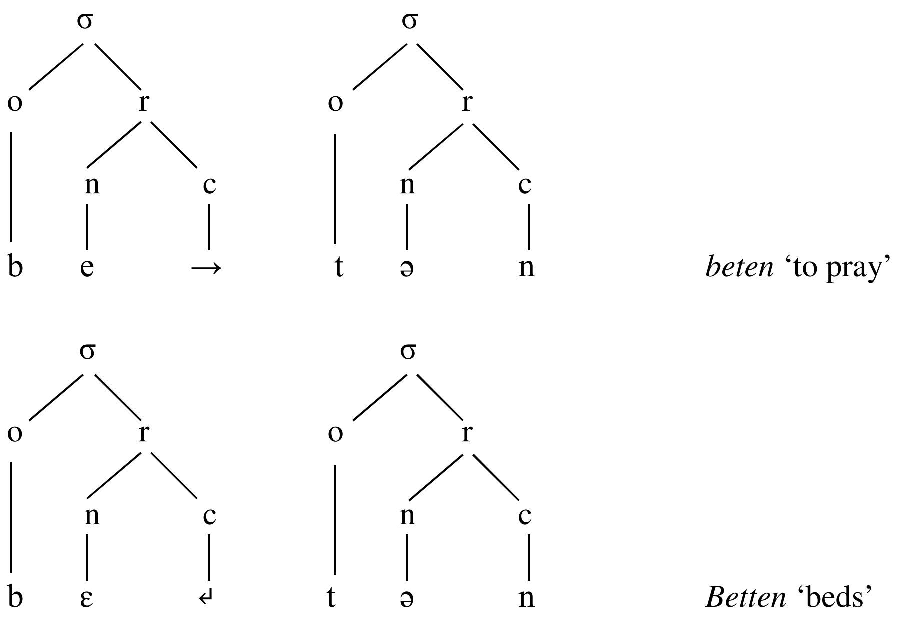


Figure 2.1-2. Smooth syllable cut (top) and abrupt syllable cut (bottom) according to Maas. The character ⟶ indicates a loosely linked vowel, ↵, a closely linked vowel (cf. Maas 2002: 20). *σ* = syllable, *o* = onset, *n* = nucleus, *c* = coda.

Maas uses the special character ⟶ to indicate loose contact of the vowel to the following consonant. Alternatively, the lengthening symbol [ː] may be used in the coda position. This is not necessary, however, since loosely linked vowels are assumed to always be long. In the bottom example, the special character ↵ indicates an empty coda position – the vowel is linked closely to the onset consonant of the following syllable. Similarly to Becker (2002), Maas argues that there is no phonological opposition between long and short vowels, but that phonetic differences such as duration or tenseness are determined by syllable cut (cf. Maas & Tophinke 1993: 140 f.).

Syllable cut correlation played a significant role in linguistics in the early 20th century, but was later abandoned because no acoustic measurements were able to prove a difference between syllable cut correlation and classical vowel quantity. However, since Vennemann (1991) revived this theory, several researchers have managed to show articulatory and acoustic correlates of syllable cut theory, among them Hoole *et al.* (1994: 54 ff.), Kroos (1996: 89 ff.; as cited in Becker 2002: 87) and Jessen (2002: 169 ff.), who showed differences in the spectral balance between smoothly and abruptly cut vowels. Another particularly successful work was published by Spiekermann (2000: 40 f., my translation), who was able to show that syllable cut correlates with

- a continuously high energy level,
- the number of energy maxima of a vowel, and
- the location of said maxima (beginning, middle or end of the vocalic spectrum).

Syllable cut correlation becomes interesting when we look at certain dialects which make a distinction between otherwise homophonous words by means of shortening the vowel. This happens very often with /aː/. As a reminder, this is the only vowel in German which only differs in quantity but not in quality. However, words with other vowels (e.g. *gib* /giːp/ 'give' [2nd ps. sg. IMPERATIVE] or *grob* /gʁoːp/ 'rough', *Dusche* /duːʃə/ 'shower') are also affected. At first sight, it is not quite clear why these words allow shortening, while other structurally similar words (e.g. *lieb* /liːp/ 'dear', *Lob* /loːp/ 'praise' and *Kruscht* /kʁuːʃt/ 'stuff') do not. Some variants have arisen diachronically; others seem to be caused by speaker-internal variation and regional standards. It has also been shown that variation occurs when vowel length is not coded clearly in orthography (cf. Tröster-Mutz 2004: 258 ff.).

This type of dialectal shortening takes place for example in Northern German dialects in words such as *Rad* 'bicycle' and *Rat* 'advice', which, due to final devoicing, are both pronounced /ʁaːt/ in standard German.[17] In Northern Germany, *Rad* can have a short vowel (/ʁat/), whereas *Rat* keeps its long vowel (/ʁaːt/). At first sight, one might argue that the underlying voiced coda in *Rad*, which appears e.g. in the plural form *Räder* /ʁɛːdə/, might be a factor which allows the vowel to be shortened. However, there are a large number of examples which are structurally the same, but do not allow shortening: *Tod* /toːt/ 'death', *Grab* /gʁaːp/ 'grave', *klug* /kluːk/ 'smart' (cf. Tröster-Mutz 2004: 261). In summary, this means that even though syllable cut theory is widely recognized as a feature of German phonology, and acoustic correlates have been found, no fixed rules which explain when syllable cut can occur and which constraints govern it exist as yet. It thus seems that all three features – vowel quality, vowel quantity and syllable cut – must be regarded as distinctive for the moment.

Finally, one observation seems to remain, namely that German does not feature the subphonemic vowel length variation seen in English. However, this view is not uncontested. Braunschweiler (1997), for instance, shows evidence for a lengthening of both long and short

---

[17] Cf. findings of the *Atlas zur deutschen Alltagssprache*, <http://www.philhist.uni-augsburg.de/lehrstuehle/germanistik/sprachwissenschaft/ada/runde_1/f15a-b/>, <http://www.philhist.uni-augsburg.de/lehrstuehle/germanistik/sprachwissenschaft/ada/runde_2/f22a-c/>.

vowels in anticipation of a following voiced consonant.[18] Vowels preceding voiced stops were approximately 20% longer than those preceding voiceless stops (cf. Braunschweiler 1997: 364). Similar findings are reported by Smith *et al.*, who say that "German talkers showed a tendency toward vowel lengthening [of about 10-12%] before voiced stops when speaking German" (2009: 262, 271). Port & O'Dell (1985) not only demonstrated that the acoustic parameters for the underlying voicing contrast were significantly different, even though they overlapped to some degree, but that German listeners were able to distinguish voiced and voiceless word-final stops in about 60% of all instances. Such incomplete neutralization contrasts have also been found in Dutch (Warner *et al.* (2004)) and Polish (Slowiaczek & Dinnsen (1985)). However, this effect seems to be partly due to orthographic differences (cf. Warner *et al.* (2006)) and individual differences between speakers (cf. Piroth & Janker (2004)). This goes to show that the questions of vowel length and neutralization of the word-final voicing contrast in German are notoriously difficult to answer. On the phonological level, however, English and German are described as demonstrating different patterns in the voicing contrast of word-final plosives. Therefore, examining German learners' acquisition of English vowel length variation before voiced and voiceless plosives is a valid choice.

2.2 Factors which Influence Vowel Length and Vowel Length Variation

> Each speech segment occupies a space along the temporal speech stream, and possesses its particular linguistic properties. However, the duration of phonetic segments is not constant; the temporal properties of speech are influenced by many other factors, such as speaking rates, utterance units, syllable context, stress, tone, and position in an utterance. (Jeng & Weismer 2004: 1)

This quote by Jeng & Weismer (2004) illustrates that vowels have an intrinsic, i.e. segment-specific, duration, which is influenced by certain extrinsic, i.e. context-dependent, factors. This section discusses three major extrinsic factors – speech rate (section 2.2.1), word prominence / frequency (2.2.2), and environment / position (2.2.3). The theoretical analysis will be complemented by an explanation of how these factors were or were not controlled for in the present study. First of all, however, a brief account of the inherent length of a vowel – which depends, among others, on the openness of said vowel – will be given. This has been discussed by a number of scholars such as House & Fairbanks (1953), Peterson & Lehiste (1960), Klatt (1973), Lisker (1974), Lehiste (1975), Umeda (1975) and Fourakis (1991), to name but a few. As such, lower vowels have been said to be "intrinsically" longer than higher

---

[18] Braunschweiler (1997) focused on vowels in word-medial position.

vowels (Tauberer & Evanini give +18ms per 100 Hz F1; 2009: 2213).[19,20] Word pairs from the word list of my study show slightly different values, but the same tendency:[21]

Table 2.1-2. Length of KIT, DRESS, TRAP and LOT before voiced and voiceless bilabial plosives. Mean values from the ten American native speakers who participated in the study.

Example word pair	**Phoneme**	**Length in ms preceding /p/**	**Length in ms preceding /b/**
rip – rib	KIT	89	148
ep – ebb	DRESS	91	151
cap – cab	TRAP	110	168
cop – cob	LOT	119	190

Regarding front- and backness of the vowel, Umeda argues that "[a]mong stressed vowels, front vowels show less variability than low or central vowels (back vowels are very infrequent)", and that "KIT is the least sensitive to almost any elongation factor" (1975: 444). This analysis seems hardly innovative, since it only rephrases the effects of openness of the vowel. After all, comparing front vowels to low vowels is comparing apples and oranges, and there are no high central vowels in English. As such, the statement still reads that low vowels show larger variation than high vowels, and thus it makes sense that KIT as one of the highest vowels shows least variation.

Two explanations for inherent vowel length as a function of openness of the vowel have been put forward: physiological and phonological reasons. This dissertation will contribute to this discussion in the following way: Section 4.5 deals with the question of whether phonemic success and the acquisition of vowel length variation are linked. In other words, do learners who have acquired separate categories for the phonemes of TRAP and DRESS also show more native-like vowel length variation in these vowels? This would support the phonological explanation, which states that each vowel has a duration specified in the grammar. This also suggests that when German learners have created a new vowel category (TRAP) and have

---

[19] Note, however, that there are opposing views to the VLE, expressed e.g. by Crystal & House: "[T]he effect was found for long vowels preceding stops, but not for short vowels" and "the mean differences in vowel duration preceding voiced and voiceless consonants in word-final position […] do not signal the voicing characteristic of the consonant" (1988c: 1578).

[20] Some scholars have argued that the ratio of vowel duration to the duration of the vowel + consonant sequence (i.e. the ratio of V/VC) is a stronger correlate of coda voicing than absolute vowel duration alone (cf. Kohler (1979), Barry (1979), Myers (2012)). The V/VC ratio will predictably be greater when C is voiced than when it is voiceless, because the preceding vowel will be lengthened.

[21] The bilabial plosives were selected randomly from the group of plosives to serve as an example here. Likewise, the American native speakers were chosen because their data was available earlier than that of the British native speakers.

managed to separate it from a familiar vowel (DRESS), they should have simultaneously acquired the correct vowel length variation for these vowels.

First, however, three major influential factors on vowel length will be discussed – speech rate (section 2.2.1), word prominence and word frequency (section 2.2.2), and the phonological and syntactical environment of the vowel (2.2.3).

2.2.1 Speech Rate

Speech rate is one of the most difficult features to control with a speaker sample as large as the one on which this dissertation is based. Every person has a distinct speech rate, which may vary from task to task or even within tasks. Different words will be stressed by different speakers, enunciation may also vary. Moreover, within the design of the study, speakers read words both in isolation and in context. All this "renders a statistical study of the measurements excessively difficult" (Heffner 1937: 128).

Shrosbree *et al.* state that "vowel duration is highly sensitive to changes in speech rate, and predictably decreases as speech rate increases" (2011: 1842). Gay names a reduction of vowels of 15-30 ms in fast speech (cf. 1977: 16). This seems quite a small reduction, and indeed Gay states that duration is not reduced as much as one might intuitively expect in fast speech (cf. Gay 1977: 16). He also claims that "the phonetically long vowels […] were not reduced to any greater or lesser per cent during fast speech than the phonetically short vowels" (1977: 16). In contrast, Shrosbree *et al.* find that the duration of tense and lax vowels is reduced to a different extent, which confounds the already overlapping duration distribution between the two sets of vowels (cf. 2011: 1845). One can imagine how confusing this might be for L2 learners, especially because individuals differ in how much duration is reduced in fast speech (cf. Ostry & Munhall 1985: 645). This claim is supported by Moon & Lindblom, who name a shortening of 40% to 60% in rapid speech (cf. 1994: 50).

A second interesting factor is that the vowel is not the only segment to be affected by rapid speech. Gay (1977) reports that although the main reduction happens in the nucleus of the vowel, the initial and final consonant in the CVC syllables he tested accounted for one third of the total reduction (Gay 1977: 16; cf. also Jeng & Weismer (2004)). Like Gay (1977), many researchers argue that vowel and consonant segments in a syllable are compressed to different degrees when speaking rate increases. Whereas Gopal (1990) sees the main reason for this in the speech rate itself, Lindblom & Rapp (1973) as well as Lehiste (1975) argue that the differential compression of vowels and consonants is due to their different syllable position, postvocalic consonant voicing or an increasing number of syllables.[22]

A second group of researchers, among them Garnes (1974), Port (1981) and Crystal & House (1982), however, argue that consonants and vowels are compressed to a similar amount in rapid speech. Interestingly, these researchers name the same factors as the researchers arguing for differential compression rates – an increasing number of syllables, postvocalic consonant voicing and speaking rate itself. This goes to show that the system of compression of the speech signal with increasing speaking rate is fairly complex and far from resolved.

Finally, some researchers more closely investigated the different durational behavior of tense and lax vowels in different speaking rates. In 1973, Lindblom & Rapp (1973) argued that there was a constant ratio of lax and tense vowels, which followed the following formula:

$$D_{\mathrm{lax}} = \frac{V}{\bar{V}} \bullet D_{\mathrm{tense}}$$

where D_{lax} = duration of the lax vowel, $V / \bar{V}$ = the ratio of short (V) to long ($\bar{V}$) vowel and D_{tense} = duration of the tense vowel. Similarly, Port (1981) found that tense and lax vowels were affected almost equally by the postvocalic consonant effect (cf. 1981: 265). However, Gopal (1990) argued against the existence of a fixed $\mathrm{V}/\ \bar{\mathrm{V}}$ ratio. Instead, his research showed that compression is context dependent: Before /t/ and /s/, the tense-lax vowel pairs maintain "a constant absolute duration across different rates" whereas the compression of tense vowels preceding /d/ and /z/ significantly differs from their lax counterparts, "so that the vowels maintained neither an absolute duration difference nor a constant proportional relationship" (Gopal 1990: 497).[23]

In 1985, Ostry & Munhall proposed a new way to compute "changes in movement duration associated with differences in back vowel height, consonant, and speech rate" with an equation of the form:

$$T = \frac{(c \bullet D)}{V_{\mathrm{max}}}$$

---

[22] The last factor mentioned here does not relate to the study at hand, since all except two tokens were monosyllabic. Cf. also footnote 31.

[23] Note that the difference in results may be due to various methodologies being employed by the different studies. Cf. Gopal 1990: 512 f. for a discussion.

where T = average duration, D = movement amplitude or displacement, c = a “constant whose value is characteristic of the system’s velocity pattern over time” and V_{max} = maximum velocity (Ostry & Munhall 1985: 646).[24] The researchers conclude that the effects of speech rate are relatively small because of “the variability in the way speakers alter rate in speech production experiments; speaking rate is not well controlled experimentally or, characteristically, in natural speech” (Ostry & Munhall 1985: 646; cf. also Miller *et al.* (1984), Eefting (1988)).

Three different models which have attempted to explain durational compression effects caused by an increase in speech rate will be briefly discussed here: The incompressibility model (Klatt (1973)) versus the multiplicative model (Lindblom (1968), Klatt (1973) and Port (1981)), and the additive model (Lehiste (1975)).

Klatt (1973) suggested the incompressibility model, which is based on the fact that vowel duration is influenced by a combination of two factors, namely postvocalic consonant voicing and the number of syllables in a word. The basic hypothesis is that “each segment can shorten only up to a certain point after which it cannot be abbreviated without its ‘destruction’ as a phone” (Port 1981: 271). He found that the combination of two factors which influence vowel duration leads to smaller duration compression than the sum of the individual factors. Vowels were shortened to 66% of their inherent duration when a syllable was added. Likewise, vowels in monosyllabic words were shortened to 66% when they appeared before a voiced consonant. If the two factors are simply multiplied, their combination should result in a shortening of the vowel to 44% of its inherent duration (= 66% • 66%). However, Klatt found the actual duration of vowels in this combined situation to be 54% of their inherent length (Klatt 1973: 1103; note, however, that Port argues 61% to be a more reasonable estimate, cf. 1981: 264). Therefore, Klatt concluded that there was an incompressibility effect which made vowels resist further shortening caused by a second factor. As such, he proposes that each factor added to the equation causes vowels to asymptotically approach their minimum duration, which is about 45% of their inherent length (1973: 1103). Mathematically, Klatt’s model can be expressed as (cf. Gopal 1996: 2):

$$D_o = K \bullet (D_i - D_{min}) + D_{min}$$

---

[24] Ostry & Munhall’s (1985) study only includes back vowels, although it is hard to imagine that this equation could not also be used for front vowels.

where D_o = the output duration, K = the compression factor, D_i = inherent vowel duration or input duration, D_{min} = minimum duration (i.e. 45% of the inherent vowel length).

Lehiste (1975), Port (1981) and Fourakis (1991), among others, have provided support for Klatt's model. However, Port (1981) argues for a purely multiplicative model rather than a combination of multiplicative and additive model as exemplified above. This is because he examined different factor combinations (speaking rate and postvocalic consonant voicing, vowel identity and postvocalic consonant voicing) and found that factors combine independently of each other. This means that Port's 1981 model is of the following form (cf. Gopal 1996: 2):

$$D_o = K \bullet D_i$$[25]

Port (1981: 267) suggests the following three rules as a predictor of vowel duration:

- *Rule 1*: Mean vowel duration is a function of the number of syllables of a word.
- *Rule 2*: The mean duration of the vowel in a VC syllable depends on the voicing of said consonant and the number of syllables. The ratio of plosive closure to vowel duration is .96 if the plosive is voiceless, it is .57 if the plosive is voiced and the word is monosyllabic, and .65 if the consonant is voiced and the word is monosyllabic.
- *Rule 3*: Vowel duration is a function of vowel tensity. Vowels are lengthened by an additional 14% if they are tense, and shortened by the same amount if they are lax.

Port's research shows that his model can account quite well for the data he accumulated, and can even cope with interspeaker variation (cf. Port 1981: 267). Moreover, while Klatt's incompressibility model failed to predict the effect of the two factors of postvocalic consonant voicing and vowel tensity, Port's constant ratio explanation managed to account for this combination as well (cf. 1981: 270). It must be mentioned, however, that the multiplicative model only works with phonological factors, so that non-phonological factors such as speaking rate are exempt from it. Nevertheless, Port's model has been successfully tested using a combination of speaking rate and a subphonemic feature such as voice onset time (VOT; cf. Kessinger & Blumstein (1998)).

---

[25] D_o, K and D_i are defined as before.

Lehiste's 1975 additive model proposes that the combination of independent factors will have an additive rather than multiplicative effect on vowel duration. This means that "when two factors that influence vowel duration combine, their effects will be independent and […] each of their effects on vowel duration will be constant in terms of an absolute amount of duration change" (Gopal 1990: 516). This can be expressed as:

$$V_{\text{tense}} = K + V_{\text{lax}}$$[26]

Lehiste (1975) concurs with Klatt (1973) somewhat, however, in that KIT + /k/ is the least compressible syllable. She ascribes this to the fact that KIT has the shortest intrinsic duration of the vowels and adds the postvocalic consonant as a second factor (1975: 84).

In conclusion, we have seen that all three models name factors beyond just speech rate as factors which influence vowel duration: intrinsic duration based on vowel identity/tensity, postvocalic consonant voicing, number of syllables and position of syllable within the word, as well as stress. We have already discussed intrinsic durations and the postvocalic consonant voicing effect. In addition, all but two tokens in the study are monosyllabic. This leaves stress, which will be discussed in section 2.2.3. Furthermore, one additional factor – namely word prominence/frequency – will be discussed directly below in section 2.2.2.

Coming back to the issue of speech rate and the three models described above, further studies undertaken for instance by Crystal & House (1988a), van Santen & Olive (1990) and Gopal (1996) have shown that none of the three models described above are able to account for all combinations of factors.[27] Each of the models only managed to capture part of the factor combinations and there were combinations which none of the models could explain (cf. Gopal 1996: 23). This shows that speaking rate is highly difficult to define and even more difficult to measure or control for. Gopal concludes that

> it would be extremely difficult (if not impossible) for subjects to produce comparable vowel durations directly, rather than controlling it indirectly by manipulating nominal rate categories. Moreover, if they could control durations directly, it is not clear how speech produced in this manner would relate to more natural productions variable of rate. (Gopal 1990: 513 f.)

This is one of the reasons why the study at hand did not carefully control for speech rate. Participants were encouraged to read at a normal speech tempo and enunciate clearly. In addition, participants read the first few words of the first reading task (cf. section 3.2) while

---

[26] *K* is defined as before, *V* = vowel duration.

[27] cf. Gopal (1996) for an in-depth discussion and critique of the models mentioned above.

the researcher pretended to adjust the volume of the recording. When participants re-read this passage after having been told the volume was now set correctly, they were less nervous and had settled into a comfortable reading routine. In the tasks that followed, speedy reading was rarely a problem.

Statistically speaking, the pool of data of 86 participants is large enough to cancel out individual differences in speech rate. In addition, the purpose of this study is not to state a millisecond value for the inherent length of different vowels, nor is it to propose a certain percentage of shortening of a vowel preceding a voiceless consonant. The purpose is to compare German learners to native speakers of English and to identify why some learners resemble their model more closely than others. Therefore, speech rate was not controlled in any other way than has been described above.

2.2.2 Word Prominence / Frequency

Word prominence is an important factor which influences vowel length. The term relates to “the information load the word carries in the message” (Umeda 1975: 436). As an example, nouns carry a higher (lexical) load in a sentence than do (grammatical) prepositions. Nouns are “hard to guess when they are missed in the flow of speech” (Umeda 1975: 436). Moreover, they are quite unpredictable from context and more infrequent, since they form a large open class, whereas prepositions are more limited in number and thus are more frequent. In general, one can say that infrequent words take longer to recall and may be enunciated with more exaggerated acoustic features. Therefore, the vowels in more infrequent words may be pronounced longer.[28] Bell *et al.* (2009) suggest that the duration of content words is more strongly influenced by word frequency than the duration of function words: “Content words are shorter when more frequent, and shorter when repeated, while function words are not so affected” (2009: 92). In the present study, all tokens are content words.

Many researchers have shown that segments which are less predictable due to linguistic context, and therefore occur in this syntactic position less frequently than others, are articulated more slowly. This leads to longer durations of said segments (cf. Lieberman (1963), Fidelholtz (1975), Wright (1979), Beckmann & Edwards (1990), Campbell & Isard (1991), Bard & Aylett (1999), Aylett & Turk (2004), Bell *et al.* (2009)). For instance, Lieberman showed that the vowel in the word *nine* was longer in “The number that you will

---

[28] For a discussion of how common words structurally differ from rare words, cf. Landauer & Streeter (1973).

hear is nine" than in the proverb "a stitch in time saves nine" (cf. 1963: 180). In both instances, the numeral occurs at a sentence boundary, but in the proverb, its occurrence is predictable due to its high frequency in this context, whereas any numeral apart from *nine* might appear in the first frame.

A second interesting result worth mentioning is reported by Wright (1979), who found that vowels in infrequent words (< 3 occurrences per million) were spoken up to 24% more slowly than vowels in frequent words (> 100 occurrences per million), "even when the memory and lexical access demands of the task were minimized" (1979: 411). This relates to the idea that words which are more frequent are stored differently in the brain, which also affects the way in which they can be recalled. Bell *et al.* say that "frequency or repetition leads to shorter or longer word durations by causing faster or slower lexical access, mediated by a general mechanism that coordinates the pace of higher-level planning and the execution of the articulatory plan" (Bell *et al.* 2009: 92). In their study, Bell *et al.* find strong correlations between duration and whether the token is a content or function word ($r = .61$). However, this effect is exceeded by the correlation of duration and word frequency ($r = -.70$; cf. Bell *et al.* 2009: 98). Frequency effects also occur when words are repeated several times in close succession (Bell *et al.* give a correlation coefficient of $r = -.48$; 2009: 98). The time to recall a repeated word is reduced drastically, and so does vowel length (cf. Oldfield & Wingfield (1965), Umeda (1975), Griffin & Bock (1998)). Last, frequency also influences perception: listeners perceive common words more quickly and more accurately (cf. Savin (1963)).

The question remains where this frequency effect originates from. Three main hypotheses have been suggested thus far:

- Influence of frequency on articulation: repetition results in automatization of the articulation processes (Bybee & Hopper (2001))
- Frequency affects words at several stages in the articulatory process:
    - More frequent words are articulated with less effort to ensure constant information density (Pluymaekers *et al.* (2005))
    - Exemplars are stored as locations in a mental map; frequency effects activate the central distribution of the words' exemplars[29] (Pierrehumbert (2002))

---

[29] Exemplars are "detailed long-term memories of particular percepts" (Pierrehumbert 2002: 8).

    - Sequences of frequent words are lexicalized and their articulatory plans routinized, so that their articulation becomes reduced (Bush (2001))
- Lexical access (Pierrehumbert (2002), Munson (2007), Bell *et al.* (2009))

Since lexical access seems to be the most widely accepted hypothesis, this is the only one which will be briefly discussed here. Generally speaking, the production process of a word consists of several stages:

> [C]onceptulization, retrieval of syntactic and semantic information from the mental lexicon, retrieval of phonological form, assembly of sounds into syllables (syllabification), and finally implementation of speech motor plan in terms of commands to specific muscles to execute the articulation.
> (Kang 2013: 205)

Many word production models assume that lexical access begins well before a person starts to speak, so that one or all of the abovementioned stages might be part of this process. We have already seen that more frequent words are accessed faster due to a higher level of activation. In addition, Bell *et al.*'s (2009) study shows that frequent content words are reduced more drastically than function words, again suggesting that the frequency effect takes place at the lexical access level. As such, the researchers argue that the frequency effect is

> implemented by a short-term coordination that moderates the pace of articulation when the progress of phonological encoding is slowed. A general motivation for a mechanism like this comes from the need for the production system to maintain temporal coordination between the conceptual/lexical and articulatory temporal streams of speech. (Bell *et al.* 2009: 105)

The frequency effect is not unlimited, however. In contrast, it depends on syntactic complexity and the lexical activation levels of the words in question (cf. Bell *et al.* 2009: 106). This means that speakers need to control the rate of speaking in different styles of speech so as to ensure that the listener understands. When conversation is at risk of breaking down, speakers will counteract the frequency effect by including pauses, fillers or repetitions (cf. Bell *et al.* 2009: 106). Bell *et al.* conclude that "[w]ord frequency is not the only factor affecting lexical access time, lexical access time is not the only factor affecting phonological and prosodic encoding, and slower access may not always lead to longer planning times" (2009: 106).

We can conclude that recent research has shown that word frequency is a strong factor in segment duration, but that "probabilistic information about words, phrases, and other linguistic structure is represented in the minds of language users", which also has a strong influence on segment duration (Jurafsky *et al.* 2000: 229). This has been shown to be a decisive factor in production (cf. Jurafsky *et al.* (2000), Gahl & Garnsey (2004) and Tily *et al.* (2009)), in perception (cf. Jurafsky (1996), Narayanan & Jurafsky (1998) as well as Frisch *et al.* (2000)) and even language learning (cf. Saffran *et al.* (1996), Seidenberg & MacDonald (1999), Xu & Tenenbaum (2007)).

On the subject of language learning: Word frequency is a major factor in the analysis section of this paper. The question to be answered is whether German learners can match their native speaker model more closely in perception and production in frequent words than in infrequent ones. This would suggest that German non-native speakers learn vowel length variation largely by mimicking native speakers. As more frequent words have a higher likelihood of being heard by learners, they will be reproduced with correct vowel length earlier. Frequency might also help learners store the correct pronunciation (including accurate vowel length) more easily and recall them when perceptually confronted with a word.

2.2.3 Environment / Position

"It is the nature of language that the probability of occurrences of phonemes and of conditions they occur is very uneven" (Umeda 1975: 434). Therefore, it is very difficult to compare vowel duration in different words, as the phonetic environment they occur in will vary. One environment which considerably affects vowel duration has already been mentioned: the nature of the final consonant. Vowels are longer before fricatives than stops, longer before velars than labiodentals and longer before voiced than voiceless consonants. The latter can be schematically expressed as

V → [+long] / _____ [+voice] [+consonant].

This is called the *postvocalic consonant voicing* effect and has been mentioned several times already. The effect results in vowel length being one of the cues to the voicing of the following consonant.

The situation is not as clear for the preceding consonant. Researchers who have studied this phenomenon have come to opposite conclusions. For instance, Oller (1973) reports that vowels which follow word-initial consonants were lengthened about 20-30 ms (cf. 1973: 1237). Larger values are mentioned by Crystal & House (26-71 ms, depending on stress; cf. 1988c: 1583). In contrast, Umeda found "[n]o consistent effect of preceding consonant on vowel duration […] except for /h/ preceding the vowel /æ/; in this case the vowel is the shortest of all" (1975: 436). Likewise, Port says that "initial consonant duration was scarcely affected by any of the variables" he tested (number of syllables, vocalic identity; Port 1981: 266). The preceding consonants vary in the study at hand; however, the following consonant is always either a voiced or a voiceless plosive.

A second factor which can affect the duration of a vowel in a word is the number of syllables of said word. It is clear that vowels in monosyllabic words can be studied most easily, since they are the only stressed vowel in the syllable and do not stand in competition with neighboring ones. Crystal & House report that the mean duration of the vowels in their corpus was 102 ms, with stressed vowels being 30 ms longer and unstressed vowels being 40 ms shorter, on average (cf. 1988c: 1575). Interestingly, Umeda found that stressed vowels in polysyllabic words behave quite similarly to vowels in monosyllabic words (1975: 434). Lehiste also found no notable difference between stressed syllables in di- and trisyllabic words (cf. 1975: 85). Again, however, there is counter-evidence: Oller's research shows that final-syllable stressed vowels are consistently longer, about 100 ms so (cf. 1973: 1236).[30] In any case, all vowel tokens except two in the present study occur in monosyllabic words, so that this factor cannot have an effect on the duration measurements.[31]

A third positional condition described as *prepausal* refers to vowels which occur before a consonant preceding a pause such as a phrase or sentence boundary. This can be schematically expressed as VC#. Crystal & House state that vowels preceding a final consonant and a pause are lengthened by 40% (cf. 1988c: 1577). This is a higher value than the 25% suggested by Mattingly (1968; cited in Oller 1973: 1244). Likewise, Jeng & Weismer note that vowels in utterance-final syllables were longer than those in non-final syllables (cf. 2004: 9). However, their results come from research on di-, tri- and four-syllabic words, which is not the exact condition as the vowels in the present study. Here, all vowels in the perception experiment occur in prepausal condition (cf. section 3.3). In the production experiment, the position of the tokens in the sentence and text reading tasks varies.

Generally, stressed words, which have more prominence in the sentence, will have more clearly enunciated vowels than unstressed words, where vowels are often reduced. This occurs in imperative, interrogative and declarative sentences (cf. Oller 1973: 1238; but note that Stack (1993) did not find durational differences in stressed vs. unstressed words). Crystal

---

[30] Word-initial consonant lengthening is also mentioned by Lehiste (1960) and Hoard (1966).

[31] Both polysyllabic words are compounds in which both members are stressed equally. In the sentence reading task, the sequence LOT + /b/ occurs the word *flashmob*. In the text reading task, the word *cobweb* contains the sequence DRESS + /b/.

& House state that “the absolute effect of stress is greater than that of tempo” and report 10 ms of lengthening in the case of tempo and 60 ms lengthening caused by stress (cf. 1988c: 1585; however, cf. Gay (1978) for counterevidence).[32] Support comes from Fourakis (1991), who claims that

> the effect of tempo is global, involving all subunits of an utterance. On the other hand, the shifting of stress from one syllable of a compound to the other is a local effect, and it should be expected that individual segments should be affected more than they would be by a change in tempo. (Fourakis 1991: 1825)

In the present study, the first two tasks presented participants with vowels in isolation (minimal pairs vs. a word list). In the sentence and text reading tasks, words were inserted into different positions in the sentences in order to receive a realistic mixture of stressed and unstressed vowels. Nevertheless, due to individual reading styles of the participants, there is no guarantee that every speaker stressed the same words in the sentence.

In 1994, van Santen proposed a sums-of-products model, which includes all of the factors mentioned above. It addresses, among other things, the effects on vowel duration exercised by the following factors (cf. van Santen 1994: 97):

- Syllabic stress
- Surrounding phonemes
- Position within the word
- Position within the utterance

Regarding the present study, the vast majority of the tokens were naturally stressed because they were monosyllabic.[33] Among the surrounding phonemes, only the consonants following play a role – here, voiced and voiceless plosives. Since vowels are usually the nucleus of a syllable and all tokens of the present study are of the type (C)CVC, the position of the vowels within the word in the present study is quite evident. The position within the utterance, however, can vary, and we have observed that this will have a strong effect on duration. Van Santen claims that “the interactions are often quite regular. When two factors interact, it is typically not the case that the effects of one factor are permuted by the other factor; instead, they are amplified or attenuated” (van Santen 1994: 96). This is clearly a departure from the claims made by Klatt (1973) in his incompressibility model and more closely matches the

---

[32] These different results could again be due to differential methodologies: Gay (1978) used nonsense words, Fourakis (1991) employed nonsense words which were more similar to real words. Crystal & House’s (1988c) results come from connected speech.

[33] Exceptions cf. footnote 31.

model proposed by Port (1981). However, van Santen's sums-of-products model allows both additive and multiplicative factors to operate at the same time by incorporating the respective product terms. In turn, van Santen's model assumes speech rate to be constant, which was not the case with the other two models.

Within the model, the factors which influence vowel duration are ordered in the following way (cf. van Santen 1994: 101):

1. Vowel identity (TRAP → FLEECE → DRESS)
2. Position within the word (word-final → other)
3. Postvocalic consonant voicing (voiced stops → voiceless stops)

This list is to be read as: If the vowel is TRAP, it will have a longer duration than if it is FLEECE or DRESS. When the vowel in question is in word-final position, it will be longer than in other positions. A postvocalic voiced stop will result in a longer vowel than a postvocalic voiceless stop. *Position within the word* involves both the position of the vowel within the syllable and the syllable type (open or closed). Unfortunately, van Santen makes no clear statement on how syllabic stress affects duration (cf. 1994: 114).

2.2.4 Summary

This chapter has shown that vowel length and vowel length variation largely depend on four factors: speech rate, word prominence / frequency, openness of the vowel and environment / position. Speech rate was not carefully controlled for in this study; word frequency as a factor in the production of vowel length variation will be discussed in section 4.2, and in section 4.3 regarding perception. The openness of the vowel will become important when we investigate the route German learners take in the acquisition of vowel length production (section 4.2) and when the effect of phonological proficiency on the acquisition vowel length variation is examined (section 4.5) by looking at how the merger of DRESS and TRAP and final devoicing influence the production and perception of vowel length variation. Since this dissertation studies the so-called postvocalic consonant voicing effect, the environment of the vowels in question is of course a factor in the analysis. The position of the word in the sentence was taken into account by inserting the tokens in different positions in the sentences, while in the perception experiment the tokens always appeared in sentence-final ("prepausal") position. The reasons for this are explained in section 3.3.

2.3 Production and Perception

Learners of a second language have been shown to perceive and produce L2 sounds differently from monolingual speakers of that language. They might also use different cues to categorize L2 sounds (cf. Bohn (1995), Flege *et al.* (1997), McAllister *et al.* (2002)). This effect is caused by previous linguistic experience in the establishing of the L1 (cf. Rochet (1995), Strange (1995), Iverson *et al.* (2003)). A critical period has been established, which proposes that the ability to perceive new sounds and create novel categories for them becomes more difficult the older the age of onset (the age when the speaker is first exposed to the L2, cf. Flege (1995), Munro *et al.* (1996)). Studies have shown that this ability may improve with increased exposure to the L2 (cf. Best & Strange (1992), Flege (1992), Best (1995)). Most studies also stress the fact that the combination of languages will have an effect on the difficulty of acquisition and the ultimate achievement. Regarding the present dissertation, the study compares German learners who have had exposure to their L2, English, in an English-speaking country (either the UK or the USA) for at least 4 months, to native speakers of English. Two further German learner groups who have had exposure to British or American English only in a non-naturalistic context (i.e. through instruction at school and university) will serve as a control group.

The exact role of experience is unclear. Some studies have shown that speakers who enter the L2 environment early are more successful in perceiving contrasts in a native-like way (cf. Flege (1991), Munro *et al.* (1996), Flege *et al.* (1997), Baker *et al.* (2002), Piske *et al.* (2001)). Others, however, have published results which indicate that early learners, too, differ from native speakers (cf. Munro (1993), Sebastián-Gallés & Soto-Faraco (1999), Pallier *et al.* (2003)). Oyama (1976) is prominently cited as the one study which found no relationship at all between LOR and pronunciation accuracy. However, there seem to be at least some late learners who are able to perceive L2 vowels accurately and in a native-like manner (cf. Flege & MacKay (2004)). Bohn & Flege (1992) showed that the relationship of the vowel in L1 and L2 plays a significant role (cf. also section 2.3.2.2). Bongaerts (1999) reports that some L2 learners manage to acquire an accent which is indistinguishable from that of a native speaker. It is important to note, however, that all the studies mentioned above define *experience* as 'age of first exposure to the L2'. This is unlike the definition used in this dissertation, in which *experience* equals 'time spent abroad in an English-speaking country'. Regarding age of first exposure to the L2, the vast majority of the participants in this study started learning

English at age 10 or 11.[34] This should be well before a critical period for L2 perception, if it exists, would have been passed.

It is also important to stress the fact that all of the studies cited above include speakers immersed in the L2 setting. However, experience has also been shown to be a major factor when the L2 is taught to students as a foreign language. In this context, Simon & D'Hulster define experience as "[the] amount of formal instruction and/or naturalistic or non-naturalistic setting" (2012: 269). This distinction between L2 learning in a Foreign Language Acquisition (FLA) setting and in an immersion context (SLA, called *L2 listeners* in the following quote) is quite important:

> FLA listeners, just like L2 listeners, but unlike monolinguals, have exposure to the target language. Yet unlike L2 listeners or monolinguals, FLA listeners have L2 exposure primarily through *formal instruction* in a *restricted setting*, with *little* or *unsystematic conversational experience* with *native speakers*.
> (Best & Tyler 2007: 19, original italics).

Regarding the present study, the control group, whose members have no experience abroad (C-UK and C-US, cf. section 3.1), can be said to exemplify FLA listeners, whereas the other two German learner groups (G-UK and G-UK) have at least 4 months of experience in either the UK or the USA, and are therefore L2 listeners in the sense of Best & Tyler (2007). It is interesting to see how FLA learners – in contrast to L2 listeners – acquire an L2 contrast (vowel length variation in the present paper) with limited "conversational experience with native speakers" (Best & Tyler 2007: 19).[35] This is also the topic of Simon & D'Hulster's research: they studied how native speakers of Dutch acquire the English vowel contrast between TRAP and DRESS – TRAP being absent from the Dutch vowel inventory just as in German. They found an asymmetry between production and perception: although all learners seemed to have created a new category for TRAP, some failed to produce this distinction in a native-like way (cf. Simon & D'Hulster 2012: 269). This may serve as evidence that perception precedes production. Other studies also focusing on FLA learners failed to establish a link between formal instruction and pronunciation accuracy (cf. Thompson (1991), Elliott (1995), Flege *et al.* (1995), Flege *et al.* (1999)). Cebrian (2006) tested the perception of English vowels by Catalan speakers. Her study consists of one group of speakers from

---

[34] The five bilingual speakers are the exception to the rule.

[35] All of the speakers in the control groups have some (limited) experience with native speakers, as they all completed an introductory course on speaking English in either British or American English. These classes are taught by native speakers of English at the University of Freiburg. However, their questionnaire data affirms that they do not have any additional contact to native speakers of English.

Barcelona with little knowledge of English (BC) and one group of late learners living in Toronto, Canada (TC). She concludes that

> motivational factors may also play a role given that the BC group is composed of students wishing to become English teachers and who may have more of a linguistic interest in English, as opposed to a more communicative interest on the part of the immigrants in Canada. An effect of limited experience may arise with less advanced and less motivated learners. (Cebrian 2006: 383)

Motivation and affect have been shown to be a crucial factor in the ultimate attainment of a native-like accent by other researchers as well (cf. Suter (1976), Purcell & Suter (1980), Krashen & Terrell (1988), Ellis (1994), Elliott (1995), Dörnyei & Skehan (2003)). Krashen & Terrell state that "[p]erformers with certain types of motivation, usually, but not always 'integrative,' and with good self-images do better in second language acquisition" (1988: 38). Schumann (1975; 1978) found affect to be more important than age in ultimate attainment of an L2. Other studies, however, have not been able to reproduce this effect. For instance, Flege *et al.* report that age accounts for 68% of the variance in their data, while integrative and instrumental motivation explain less than 3% (1999: 93). Similarly, Elliott (1995), Bongaerts *et al.* (1997) and Moyer (1999) assert that motivation does not automatically lead to accent-free pronunciation. Motivation and affect are only two parts of a holistic view of why some learners appear to learn foreign languages more easily than others. Motivation is part of conation – the way humans use their free will "to make choices that result in new behaviors" (Ortega 2009: 146). Affect, in contrast, "encompasses issues of temperament, emotions and how humans feel towards information, people, objects, actions and thought" (Ortega 2009: 146).

A third crucial factor is language learning aptitude. Some people appear to possess a special talent for pronouncing non-native sounds. These learners might be among the few who have been shown to attain a native-like accent even though they were late learners (cf. Bongaerts *et al.* (1997), Moyer (1999)). It has been claimed that this talent might stem from ability for mimicry or a high degree of musicality. Flege *et al.* included a variable called *sound processing ability* in their study, which is a mixture of "ability to remember how English words are pronounced", "ability to imitate foreign accents and dialects" and musical ability (1999: 92). They found that this factor did influence both L2 pronunciation accuracy and grammaticality judgments, although only to a small extent. In general, mimicry ability seems to positively correlate with ultimate attainment in L2 pronunciation. For instance, Suter (1976) had his participants imitate sequences of unfamiliar sounds and noticed that this influenced the degree of their foreign accent. Similarly, Thompson (1991) states that the selfreported mimicry ability of her participants explained 5% of the variance in their foreign accent (cf. 1991: 192).

Divergent findings have been reported regarding the influence of gender on foreign accent. Asher & García found that females received higher ratings than men matched for age, but this effect vanished as length of residence increased (cf. 1969: 339). Similarly, Tahta *et al.* (1981) report gender to be a significant factor for L2 pronunciation, with females scoring better than males.[36] In Thompson's study (cf. 1991: 192), gender accounted for 11% of the variation. Again, females performed better than males. However, there are also many studies which did not find any gender effects (e.g. Suter (1976), Purcell & Suter (1980), Flege & Fletcher (1992), Elliot (1995), Piske *et al.* (2001)). Flege *et al.* (1995) noticed that the gender effect was a function of AOL: Females with less than 12 years of AOL were rated higher than males, but males with more than 16 years of AOL scored higher than females. The researchers were unable to explain this effect (cf. Flege *et al.* 1995: 3129 f.).

Apart from AOL and LOR, the percentage of daily L1/L2 use has been named as a major factor determining the accent of a non-native speaker. Purcell & Suter reported this to be the third most important predictor for a foreign accent (cf. 1980: 282). Tahta *et al.* showed that the amount of English spoken at home explained 25.7% of the variance in accent in early bilinguals (cf. 1981: 269). Interestingly, Thompson (1991) observed that language use was correlated with degree of foreign accent, but was not a significant predictor, because it was confounded with AOL. Still, she writes that "a difference must be noted between subjects who have maintained their mother tongue and those who have lost it when it comes to estimating accent retention in the second language" (Thompson 1991: 200). In any case, language use is clearly only relevant for L2 speakers in immersion settings. Since all speakers in the present study learn English in an FLA setting, they are vastly German-dominant. In fact, they were asked to indicate on a questionnaire how often they use English actively and passively – both variables were found not to be statistically significant ($p = .37$ and $p = .56$, respectively; cf. also section 0). For this reason, this factor will not be discussed further. Instead, after this overview of factors which may influence the degree of foreign accent, let us look in more detail at how speakers learn to produce and perceive non-native (subphonemic) features. This is the topic of the next two sections.

---

[36] This effect was rather small, however (2.2%; cf. Tahta *et al.* 1981: 268).

2.3.1 Production

Most, if not all, adult L2 learners will initially have difficulty pronouncing non-native sounds. Although some learners seem to be able to become quite competent speakers, many (even advanced) learners retain an audible accent. This raises the question of whether adult speakers are capable of learning to pronounce new sounds or modifying their pre-existing L1 sound categories so that they match the L2.

Some sounds seem to be more difficult to pronounce than others. New sounds, for instance, appear to be more difficult to learn than similar sounds. This might be due to motor control issues – learners cannot use an already established articulatory pattern from their L1 and therefore simply *do not know* how to produce the sound. However, motor control issues might also be so serious in nature that adult learners *are not able to* articulate new sounds, simply because they do not have the same motor plasticity as children do. In contrast, problems might also stem from perception – if learners cannot perceive a new sound, they might be unable to produce it.

Similar sounds, too, might be pronounced differently by L2 learners and native speakers. Bohn & Flege (1992) note that "adult learners are able to establish phonetic categories for new L2 sounds, and thus eventually produce them authentically, whereas similar sounds will remain foreign-accented even after lengthy exposure to the L2" (1992: 132). However, the differences might sometimes be so small that they can be measured but are inaudible in conversation. Valdman summarizes the situation as follows: "The student must learn to make new responses to stimuli which are interpreted as identical to native language stimuli" (1976: 38). Flege & Hillenbrand state that the extent of this interlingual identification depends on auditory and articulatory similarity of the L1 and L2 phones (cf. 1984: 708). Therefore, it makes sense to relate the four vowels on which the study of the present dissertation is based – KIT, DRESS, TRAP and LOT – to the corresponding German vowels.

Bohn & Flege (1992) compared German speakers' production of KIT, FLEECE, DRESS and TRAP in (American) English and German. They showed that the German vowels KIT, FLEECE and DRESS have similar counterparts in English, but that German lacks TRAP. TRAP is therefore identified as a new phoneme for German learners. German /ɛː/ neither reaches the duration of English /æ/ nor are the two vowels close in vowel space (cf. 1992: 139). Regarding duration, Bohn & Flege (1992: 136 f.) list the following average values:

Table 2.1-3. Vowel length of four German and four English vowels. Source: Bohn & Flege 1992: 136 f.

	German	**English**
KIT	54 ms	144 ms
FLEECE	112 ms	175 ms
DRESS	/ɛ/ 76 ms /ɛː/ 163 ms	181 ms
TRAP	---	238 ms

The table clearly illustrates that all four English vowels in question are longer than their (similar) German counterpart. In turn, the German vowels are higher in vowel space, but not significantly so (cf. Bohn & Flege 1992: 137). Regarding DRESS, both German /ɛ/ and /ɛː/ match English /ɛ/ - German /ɛ/ spectrally, and /ɛː/ in durational terms (cf. Bohn & Flege 1992: 139). When German and English speakers were then asked to produce tokens containing these vowels, the speaker groups differed significantly in their production of vowel length and height, but not front-/backness (cf. Bohn & Flege 1992: 140). Bohn & Flege's study also showed a clear difference between experienced and inexperienced speakers, with experienced speakers producing more native-like vowels (both regarding spectral and durational characteristics). This is particularly salient in the separation for TRAP and DRESS, which only the experienced learners showed. The inexperienced learners had an almost complete overlap of these two vowels (cf. Bohn & Flege 1992: 143 f.). This, in turn, influenced the intelligibility test, where native speakers of American English judged vowels spoken by the experienced and inexperienced learner groups. The TRAP tokens of the experienced group were recognized correctly in 65.9% of all instances, compared to 50.9% of the tokens for the inexperienced group (Bohn & Flege 1992: 148).

Unfortunately, Bohn & Flege's study did not include the LOT vowel. In British English, LOT is pronounced as a low, rounded back vowel. It is relatively shorter in duration than the American English realization of it, which approximates /ɑː/ – a low (lower than British English), unrounded back vowel. The German vowel inventory does not include LOT. In fact, German does not have a back vowel as low as British and American English. The closest German vowel is the mid-low THOUGHT vowel /ɒ/, which appears in words such as *rotten* 'to rot'. However, this vowel is short in German, whereas it is long in British and American counterpart of /ɔ/ is /oː/ (a high-mid rounded back vowel), which does not exist in standard British or American English. The following three vowel trapezoids chart the articulation of the vowels on which the present study is based, as well as the vowels discussed above, in German, British and American English:

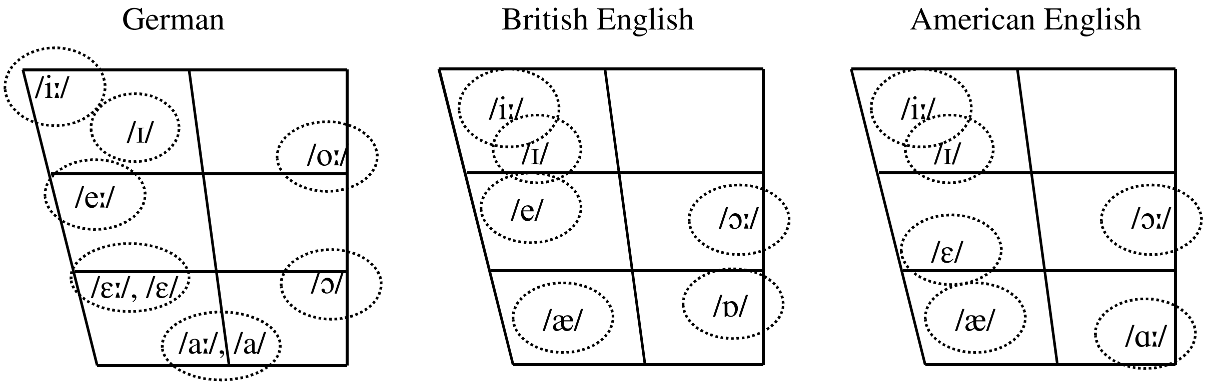


Figure 2.1-3. Position of KIT, DRESS, TRAP and LOT in German, British English and American English. FLEECE and THOUGHT as well as German /oː/ are given as reference vowels. Adapted from Furhop & Peters 2013: 23, Eckert & Barry 2002: 111 and Canepàri 2010: 44, 50.

Figure 2.1-3 illustrates that German FLEECE is more peripheral and higher in German than in English, whereas KIT is slightly more peripheral in German than in English. In addition, German has two realizations of DRESS – a mid-high long realization as well as a mid-low long and short realization. However, as has been mentioned before, most younger speakers have merged /eː/ and /ɛː/ under /eː/, so that a longer, mid-high and a shorter, mid-low realization of DRESS remain. The German inventory lacks TRAP and LOT but has a short mid-low THOUGHT vowel and /oː/. In contrast, FLEECE and KIT are slightly lower and not as peripheral in British and American English. British English DRESS is pronounced as a mid-high front vowel, whereas it is mid-low in American English. Both varieties have TRAP and LOT. TRAP is identical in British and American English, LOT is not. In conclusion, this means that KIT and DRESS can be categorized as similar vowels for German learners, whereas TRAP and LOT are new. Regarding vowel length, LOT pronounced by the American native speakers of the present study was 174 ms long; the British native speakers' realization of LOT was 108 ms (averaged over all tasks, cf. section 3.2). The other vowels pronounced by the native speakers also show shorter lengths than the ones named by Bohn & Flege (1992).

A question which has been posed and studied for a long time, but for which no satisfactory answer has been found so far, is whether perception precedes production in SLA, or vice versa. Many studies have addressed this question, but the results have been mixed at best. Studies by Trubetzkoy (1939), Flege (1987; 1991), Neufeld (1988), Llisterri (1995) and Flege & MacKay (2004) suggest that perception is a prerequisite for production. Other researchers claim that in some cases, production precedes perception (cf. Gass (1984), Strange (1995), Kluge *et al.* (2007)). In order to answer this question, several scholars have explored the possibility of phonetic training. After training perception of certain sounds, subjects managed a more target-like production. Such positive effects have been shown for instance by Rochet (1995) and Bradlow *et al.* (1999). A few studies have also investigated production-based training, with the effect that articulatory training improves perception (cf. Catford & Pisoni (1970), Sheldon & Strange (1982), Weiss (1992)). In summary, it seems that there is a mutually favorable relationship between perception and production. The question of how non-native speakers learn to perceive L2 sounds is the topic of the next chapter.

2.3.2 Perception

Understanding how L2 perception proceeds is very important for a number of reasons. Firstly, differences in the perception of native and non-native speakers may make the processing of an L2 more difficult (cf. Munro & Derwing (1995), Schmid & Yeni-Komshian (1999)). It might also lead to difficulties in word recognition (cf. Bradlow & Pisoni (1999), Mayo *et al.* (1997)), a topic which is addressed by the perception experiment of this dissertation. Last, Rochet (1995) has shown that perception has a strong influence on production, because speakers who have difficulty perceiving L2 sounds accurately also have a hard time producing these vowels in a native-like fashion.

The general consensus among researchers is that the perception of L2 vowels depends (at least initially) on the relationship to the L1 vowels.[37] This factor has been discussed by e.g. Best (1995), Kuhl (2000) and Flege & MacKay (2004). For instance, it seems easier for learners to perceive an L2 vowel which is assimilated to two L1 vowels. However, if two separate L2 vowels are perceived to be a single L1 vowel, this can cause problems (cf. Flege & MacKay 2004: 6). This explains why many German learners find it difficult at first to distinguish between TRAP and DRESS – they map both vowels onto the same vowel in German

---

[37] Cf. Flege *et al.* 1997: 440 ff. for an overview of how to assess the relationship between L1 and L2 vowels and possible problems related to different methodologies.

(DRESS). However, many speakers learn to produce and perceive this distinction with phonetic training.[38]

Chapter 4 will illustrate how the German learners in the present study managed to deal with these two vowels in production (4.2) and perception (4.3), whether perception and production are linked (4.4), and whether a merger of these two vowels is a factor in the acquisition of vowel length variation (4.5). First, however, we will discuss the effect of vowel duration on vowel recognition, because this is directly relevant to the study at hand. Hillenbrand *et al.* (2000) tested vowel recognition in CVC syllables in 300 utterances. Fifteen phonetically trained subjects served as listeners. Tokens were synthesized in four different versions (Hillebrand *et al.* 2000: 3013):

- with original vowel length
- with fixed vowel length at 272 ms (their calculated mean across all vowels)
- with fixed vowel length at 144 ms (two *sd* below the mean)
- with fixed vowel length at 400 ms (two *sd* above the mean)

The researchers found that length had a small overall effect in most vowels, but that TRAP and DRESS were greatly affected by differences in duration. The two were systematically confused with each other when duration was altered. Shortened TRAP was misheard as DRESS; lengthened DRESS was judged to be TRAP (cf. Hillebrand *et al.* 2000: 3017).

Other studies, such as the one conducted by Stevens (1959), support the view that duration is a strong factor in vowel recognition. In his study, vowels which were less than 100 ms long but showed spectral characteristics of TRAP were nevertheless identified as DRESS. Huang (1986) showed that vowels at 40 and 90 ms were identified as KIT, vowels at 140 and 235 ms were identified at FLEECE, even if the spectral information was kept constant. The results are far from uniform, however. There is a considerable number of studies that challenge the importance of duration in vowel recognition. For instance, Strange *et al.* (1983) asked listeners to identify tokens with three different characteristics: a) durational information retained, b) neutralized durational information by mapping all lengths to the shortest vowel and c) neutralized durational information by mapping all lengths to the longest vowel. The authors found that shortening the length to match the shortest vowel did not increase the

---

[38] Note, however, that studies have shown mixed results: Yule & Macdonald (1995) report that training effectiveness is highly dependent on the individual. Similarly, Suter (1976) found phonetic training had a rather limited effect, and that motivation and self-confidence as well as contact with native speakers were more important factors. In contrast, Logan *et al.* (1991) and Bradlow *et al.* (1997) showed that intensive training improved perception accuracy considerably.

number of mistakes listeners made in relation to the naturalistic tokens, but lengthening the tokens considerably increased the error rate.

In conclusion, there is considerable uncertainty as to what role duration plays in vowel recognition. Most researchers agree that there is a modest but measurable effect (cf. Hillebrand *et al.* 2000: 3015). It certainly plays a large role in L2 perception. The sub-chapters that follow will present four perception models which relate to L2 acquisition: Kuhl's Native Language Magnet Model (NLM), Flege's Speech Learning Model (SLM), Best's Perceptual Assimilation Model (PAM) and Escudero's Second Language Linguistic Perception Model (L2LP). All models make different hypotheses and proposals about how L2 perception develops, which is why a comparison of all models can be found at the end of the chapter (section 2.3.2.5).

2.3.2.1 Kuhl's Native Language Magnet Model (NLM)

Kuhl's Native Language Magnet model (NLM) says that "infants perceptually 'map' critical aspects of ambient language" and that "linguistic experience alters infants' perception of speech, warping perception in the service of language" (Kuhl 2000: 11850). She argues against nativist theoretical assumptions postulated e.g. by Noam Chomsky, which state that every baby is born with an innate language faculty and a universal grammar which is set to the ambient language through specified parameters and constraints triggered by language input. Instead, Kuhl follows the emergent hypothesis which posits that language input is mapped by infants' brains. Evidence for this comes from neurological studies performed on young infants (c.f. Dehaene-Lambertz & Gliga (2004), Mills *et al.* (2005), Conboy & Mills (2006), Imada *et al.* (2006)). These showed, among other things, that infants were able to parse speech phonetically correctly. This universal ability among infants is called *phonetic feature detector* by Eimas (1975: 346). Kuhl (2000: 11852) proposes three features of children's language learning:

- "[I]nfants detect patterns in language input."
- "[I]nfants exploit the statistical properties of the input, enabling them to detect and use distributional and probabilistic information contained in ambient language to identify higher-order units."
- "[I]nfant perception is altered – literally warped – by experience to enhance language perception."

The first two features are not directly relevant to the NLM and will therefore only be briefly explained. That infants can detect pattern in language is attested in a number of studies which show that children can perceptually sort syllables which vary across speakers and show different intonation contours (cf. Kuhl (1979)). Moreover, infants at birth prefer their mother's voice to other female voices; at an age of 9 months, they show a listening preference to their native language over other languages (cf. Kuhl (1983)).

Mattys *et al.* (1999) studied 9-month-old infants and showed that children make use of statistical properties of language by presenting them with CVCCVC strings. The idea behind this was that certain consonant combinations are statistically more likely to appear in certain positions. For instance, the sequence /ft/ is more likely to appear word-finally, while /vt/ appears more often between words. In the sequence above, the researchers inserted CC sequences which were either frequent or infrequent in English. Infants showed a clear preference and listened significantly longer to words which contained frequent CC sequences.

The last feature, namely that language experience "warps" infants' perception, is directly related to the NLM. Kuhl explains that "[l]anguage experience not only produces a change in infants' discriminative abilities and listening preferences, it results in a 'mapping' that alters perception" (2000: 11853). This can be shown by the magnet effect. This effect appears when some phonetic representations are mapped in the brain as prototypes and subsequently compared to other, possibly less prototypical, tokens. Brain studies have shown that prototypical tokens excite a special response in the brain, one which is not observed when non-prototypical tokens are used (cf. Aaltonen *et al.* (1997); Sharma & Dorman (1998)). Kuhl summarizes that the prototype functions like a magnet attracting other similar but less prototypical stimuli (cf. 2000: 11853). The perceptual magnet is language-specific and has been shown in children as young as 6 months of age (cf. Kuhl *et al.* (1992)).

The model was also tested on speakers of different languages. Japanese and English native speakers were presented with a continuum of syllables starting either with /□/ or /□/ - two English sounds which map onto a single Japanese sound. The results revealed that Japanese speakers demonstrate no magnet effect whatsoever, while the English native speakers show a clear map of more prototypical instances of /r/ and /l/ in the center and less similar ones on the boundaries. Kuhl concludes that "linguistic experience produces mental maps for speech that differ substantially for speakers of different languages" (2000: 11853; cf. also Kuhl & Iverson (1995), Iverson & Kuhl (1996)). What is important is that at first, the "contouring of the perceptual space" is universal (Kuhl 2000: 11853). Linguistic experience then alters perception, "completely revising the perceptual space underlying speech processing" (Kuhl 2000: 11853). This produces a filter through which language is perceived and makes L2 learning so difficult, because the mapping of the acoustic space has already been completed for the L1. However, the L2 might require a completely different map, the establishing of which might be constrained by the completed mapping of the L1.

Bosch *et al.* (2000) found that children learning an L2 after age 6 did not demonstrate a magnet effect for the L2, but did for the L1. Since the vast majority of speakers in the present study started learning English at age 10 or 11, they will have passed this "critical period" proposed by the NLM. However, according to the NLM, perceptual learning remains possible in adults. "Adult L2 learners might circumvent L1 interference effects if they can recapitulate infants' experience of L1 speech – that is, if they manage to receive 'exaggerated acoustic cues, multiple instances by many talkers, and massed listening experience'" (Flege & MacKay 2004: 5; cf. also McCandliss *et al.* (2002), McClelland *et al.* (1999)).

With regard to perception and production, Kuhl's model argues that there is a strong dependency between the two. Perceptual representations of words are stored in the brain and guide the development of production (cf. Kuhl 2000: 11854; Kuhl *et al.* 2008: 984). According to her theory, perception precedes production: "[S]ensory learning occurs first, based on experience with language, and this guides the development of motor patterns" (Kuhl *et al.* 2008: 985). Infant-directed speech with its exaggerated stress, increased pitch and slower tempo is instructive in children learning an L1; thus, it might be helpful for L2 learners as well. This would mean that teaching should confront learners with multiple speakers who exaggerate the linguistic cues of the L2. In any case, a "mass listening experience" is necessary for learners to perceive and subsequently reproduce features of the L2 which differ from the L1. This is where the NLM builds a bridge to the effect of experience, which was shown to be crucial in section 2.3. However, this dissertation will show that *experience*, defined as 'length of residence in an English-speaking country', is not the sole factor in how well learners can perceive and produce vowel length variation, but that the reason for their stay abroad must also be taken into account (cf. section 4.2.5)

In summary, Kuhl's NLM posits the following hypotheses:

- Infants perceptually 'map' critical aspects of ambient language" (Kuhl 2000: 11850).
- "[L]inguistic experience alters infants' perception of speech, warping perception in the service of language" (Kuhl 2000: 11850).
- There is a language-specific magnet effect which centers on prototypical realizations of sounds.
- L2 learning is difficult because it requires a re-mapping of the pre-existing L1 map.
- Teaching an L2 should include exaggerated cues similar to motherese.
- Perceptual learning is necessary for accurate production.

2.3.2.2 Flege's Speech Learning Model (SLM)

Flege's model focuses mainly on the ultimate attainment of learners. He divides L2 sounds into three different categories according to the way in which they relate to L1 sounds. According to him, L2 sounds can be *identical*, *similar* or *new* (cf. Flege 1987: 48). Sounds are identical when there is no acoustic difference between them and positive transfer from the L1 results in the L2 sound being produced correctly. Bohn & Flege (1990) characterize English and German KIT and DRESS as similar or nearly identical (1990: 303). New sounds have no counterpart in the L1 inventory. An example would be TRAP and LOT for German learners of English. Similar L2 phonemes differ acoustically from a counterpart in the L1 which is easily identified by the learners. For instance, both the German and English vowel inventory make use of GOOSE, but the vowel shows more formant movement in English. Catford (1965), Weinreich (1968), Valdman (1976) and others have shown that learners identify L2 sounds in terms of their L1 sounds and use the same articulatory patterns to produce them. Since there is no articulatory or acoustic model for new L2 phonemes, these can pose great trouble for learners.

However, similar phonemes, too, may be difficult, because even advanced speakers may retain a small but audible difference in the pronunciation of such phonemes. For instance, James (1985) reported that almost all errors involve new phonemes or phonemes which occur in both inventories but are realized differently (cited in Flege 1987: 48). Flege suggests that "learners may have difficulty establishing the articulatory patterns needed to produce both new and similar L2 phones authentically; or […] they may continue to identify both new and similar L2 phones in terms of an L1 category" (1987: 48).[39] The result of such a merged L1/L2 category is that the L2 sound is produced with an accent. In summary, the extent of the accent of a speaker relates to the perceived phonetic distance between L1 and L2 and the learner's (in)ability to establish L2-adequate categories.

The second important postulate by Flege relates to age of learning (AOL). According to Flege, earlier is better, because "the earlier L2 learning commences, the smaller the perceived phonetic distance needed to trigger the process of category formation" (1995: 264). Flege introduces the term *equivalence classification* to refer to a phenomenon which prevents adult learners from producing L2 phonemes in a native-like way. He defines this concept as follows: "Equivalence classification is a basic cognitive mechanism which permits humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category" (Flege 1987: 49). He maintains that this is a crucial mechanism for children to recognize that phonemes spoken by different speakers, in different pitches or with different intonational contours are still instantiations of a single category. Adult learners are at a disadvantage, however, since they already have well established categories for their L1 phonemes. This means that "they are likely to judge L2 phones (even those which differ auditorily from phones in L1) as being the realization of an L1 category" (Flege 1987: 50). Support for this hypothesis comes from a study by Baker *et al.* (2002), who showed that Korean adult learners assimilated English vowels to Korean vowels more strongly than child learners. However, with experience learners may establish a new category for the L2 sound which is independent or prior representations of the closest L1 sound. Still, learners may only approximate but never fully achieve the phonetic norm of the L2 phoneme (cf. Flege 1987: 50). It is important to note, here, that equivalence classification is more likely to occur when the L2 phoneme is similar to the L1 phoneme. New phonemes have may be too dissimilar for equivalence classification to occur (cf. Flege 1987: 50).

A third innovative point Flege discusses in his model is what happens to the pre-existing phonetic categories in L1 when learners establish categories for L2 phonemes. Previous research had focused on a unidirectional link whereby L1 influences L2, but Flege hypothesizes that the two systems are interdependent because they share a common phonological space. He hypothesizes that when L2 vowels are close in vowel space to L1

---

[39] Flege defines phonetic category as "the perceptual ability to 1) identify a wide range of phones as being 'the same,' despite auditorily detectable differences between them along dimensions that are not phonetically relevant; and 2) the ability to distinguish the multiple exemplars of a category from realizations of other categories, even in the face of noncriterial commonalities" (1995: 244).

vowels, the two vowels will dissimilate. For instance, Flege *et al.* (2003) examined the production of the diphthong FACE by Italian-English bilinguals differing in their age of arrival in Canada. Learners often identify this diphthong as instances of the Italian vowel DRESS. The researchers found that early learners produced FACE with significantly more tongue movement adequate for a diphthong than English native speakers. In contrast, the late learners produced the diphthong with much less tongue movement than the native speakers. Flege *et al.* conclude that the early learners produced more movement than native speakers because of phonetic dissimilation between the Italian L1 and the English L2 sound; the late learners were hypothesized to produce less movement than the native speakers due to failure to establish an adequate L2 category (cf. Flege *et al.* 2003: 467). This means that successful early learners might still show subtle differences in perception and feature weighting compared to native speakers (cf. Flege & MacKay 2004: 26). This might be due to motor control issues. Flege reasons that "[a]dults might be generally less able than young children to develop new articulatory patterns or to translate the sensory information associated with the new L2 phones into stable motor control patterns" (1987: 49). Another explanation might be social and/or psychological factors (cf. the discussion in the introduction).

Regarding the question of whether perception precedes production or vice versa, Flege's SLM clearly hypothesizes that perception is the key to production. As mentioned above, due to their pre-existing L1 categories, adult learners interpret an L2 vowel as an instance of an L1 vowel and produce the L2 vowel accordingly. Since perception can be trained, however, "the perceived relation between L1 and L2 vowels may change during L2 acquisition. Such changes in perception may, in turn, engender changes in vowel production" (Flege *et al.* 1997: 440). One question which remains unanswered by the SLM is the degree to which a foreign accent is influenced by the specific L1/L2 pairing.

In summary, the SLM includes the following important points:

- There are new, similar and identical phonemes in the L2 which learners must acquire.
- Category formation for a new or similar sound is influenced by the perceived distance of said sound to the closest L1 phoneme.
- Since categories are formed for the L1 through childhood and adolescence, learners who have passed these stages are at a disadvantage, because their pre-existing categories might prevent L2 categories from being formed in a native-like manner.

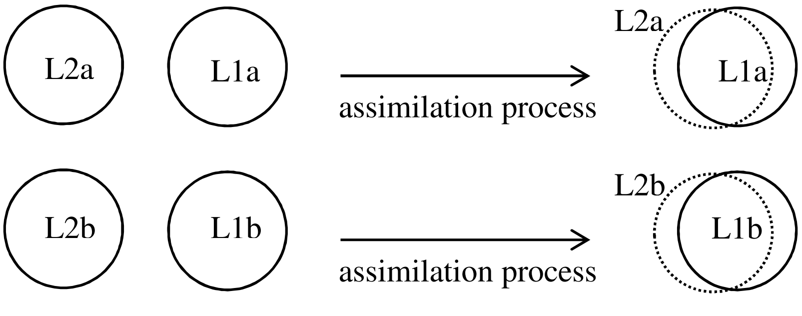
- Even successful early learners might show subtle differences compared to native speakers.
- The production of a sound reflects the representation in the phonetic category, which is established through perception.

2.3.2.3 Best's Perceptual Assimilation Model (PAM)

In contrast to Flege's SLM, which focuses on experienced learners, Best's PAM centers around naïve listeners. It takes the "ecological perspective that it is primarily the evidence about articulatory gestures in the speech signal that informs the perceiver" (Best 1994: 179). Similarly to Kuhl's NLM, this model also posits that the structure of the L1 can influence the perception of the L2: listeners perceive sounds according to their (dis)similarity to L1 sounds. When there are more similarities than dissimilarities between the sounds, the model predicts that the L2 sound is perceptually assimilated to the L1 sound.

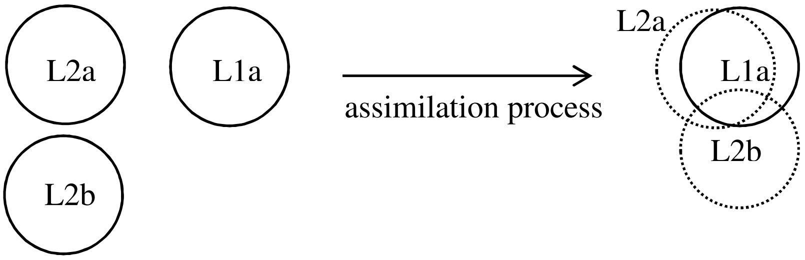
Best calls this *Single Category Assimilation* (cf. Best 1994: 180). In this case, the perceptual discrimination as well as the production of the L2 sound will be foreign-accented.

A second category mentioned in PAM is *Two Category Assimilation*, where two distinct L2 sounds are mapped onto two different L1 sounds (cf. Best 1994: 180).



According to PAM, Two Category assimilation will lead to excellent discrimination by learners (cf. Best & Tyler 2007: 23), because they can rely on the same number of distinctions between sounds in their L1 and L2.

The final category, *Category Goodness Difference*, explains that two L2 sounds may be mapped onto a single L1 category (cf. Best 1994: 180). However, one of these sounds is considered a better fit of the category than the other. For instance, German learners might map both English DRESS and TRAP to German DRESS, but since English DRESS is closer in vowel space to German DRESS than TRAP, speakers might consider DRESS tokens as more suitable instances than TRAP.



In the case of category goodness, it might also occur that one of L2 phonemes is not categorized because it fails to sufficiently match an L1 phoneme. This contrast between categorized and uncategorized phonemes is said to be discriminated well by the speaker because "it reflects a phonological distinction between an exemplar of a known phoneme versus something that is not an exemplar of that phoneme" (Best & Tyler 2007: 23). If both non-native phonemes are uncategorized, they will be moderately to poorly distinguished, depending on their proximity to the same or different native phonemes.

Finally, L2 sounds might also not be perceived as linguistic sounds at all, a rare phenomenon which the PAM terms *Non-Assimilable* (cf. Best 1994: 180). Discrimination is said to be good or even excellent, again depending on the perceived similarity of the sounds to non-speech sounds (cf. Best & Tyler 2007: 23). Here, Best's PAM is quite similar to Kuhl's NLM; both posit that there is a certain prototype of a category and sounds are judged according to their perceived similarity to this exemplar. In summary, the discrimination performance for late learners the PAM predicts is the following (from highest to lowest, cf. Best 1994: 181):

Two Category → (Non-Assimilable ↔ Category Goodness) → Single Category

In 2007, Best & Tyler revised their model, now called PAM-L2, in order to integrate more experienced learners. According to their model, learners differ within the following three criteria: "1) L1 acquisition at onset of L2 learning; 2) ratio of L1/L2 usage on an average daily basis, and; 3) ratio of L1/L2 in the language environment" (Best & Tyler 2007: 15). The researchers admit, however, that these criteria are difficult to measure objectively (cf. Best & Tyler 2007: 16). In any case, their two language systems interact: “[A] listener’s perception of phonetic information generally behaves as a coherent dynamical system in which changes in any subregion may affect the lay of the remaining landscape” (Best & Tyler 2007: 18; emphasis omitted). This results in perceptual changes caused by contact between L2 and L1. Similarly to Flege’s SLM, the researchers believe that perceptual learning remains possible over the entire lifetime (cf. Best & Tyler 2007: 24). Regarding the question of why children learn differently than adults, the PAM argues that “the environment […] changes, including the response of others to the individual’s appearance as a physical, cognitive and social being, and particularly as a language-user” (Best & Tyler 2007: 24). This supports the hypothesis proposed by the NLM, namely that feedback from the outside is crucial and that children have an advantage because they receive more high-quality input through motherese.

What is new and unique about the PAM is that it rejects the notion of listeners establishing categories through phonetic input. Instead, the PAM posits that experienced learners detect “higher-order articulatory invariants in speech stimuli” (Best & Tyler 2007: 25). Mental representations are not required for perceptual learning. Moreover, the PAM claims that the perception of learners depends on the learner’s focus and goal:

> [Some] contexts of speech perception might require focus at a phonetic level, or at a phonological level. We suggest that perceptual objects/events that are relevant to L2 speech learning are not merely phonetic. Language-relevant speech properties are differentiated not only at the phonetic level but also at the higher-order phonological level, as well as at the lower-order gestural level. L1-L2 differences at a gestural, phonetic, or phonological level may each influence the L2 learner’s discrimination abilities, separately or together, depending on the context or the perceiver’s goals. (Best & Tyler 2007: 25)

Regarding the present study, participants therefore need to learn to focus on what the PAM calls the *phonetic level* – “invariant gestural relationships that are sub-lexical yet still systematic and potentially perceptible to attuned listeners, for example, positional allophones” (Best & Tyler 2007: 25) – in our case, the allophonic variation of vowel length variation before voiced and voiceless plosives.

The following list is a summary of the predominant hypotheses of the PAM (cf. Best 1994: 167 f.):

- Non-native sounds are perceived through “an adjustment of selective attention rather than a permanent revision of the initial state of sensory-neural mechanisms.”

- This mechanism "is neither absolute in extent nor irremediable in adulthood, and it varies in degree among specific types of non-native contrasts and among individuals."
- Adult learners perceptually assimilate "non-native phones to the native phoneme categories with which they share the greatest similarity in phonetic characteristics."
- Non-native sounds may be processed through *Single Category Assimilation*, *Two Category Assimilation* or *Category Goodness Fit*. Less often, they will not be categorized as speech sounds (*Non-Assimilation*).

2.3.2.4 Escudero's Second Language Linguistic Perception Model (L2LP)

Escudero's L2LP model proposes a theoretical framework according to which methodological testing of second language perception can be performed. Escudero herself explains that the model "aims at describing, explaining, and predicting L2 sound perception at the initial, developmental, and end states" (cf. 2009: 167). It focuses on the acquisition of new, similar and subset L2 phonemes by learners.[40] As in Flege's SLM, the L2LP defines similar sounds as L1 and L2 sounds which are phonologically identical but phonetically different. The model suggests that learners start out with optimal L1 perception and unconsciously use cues of their L1 to perceive sounds of the L2. With time, they gradually adjust their perception using the same mechanisms as monolingual L1 speakers (cf. Escudero 2005: 111). Phonological representations need not be adjusted. Thus, "speech perception is a language-specific phenomenon that involves linguistic knowledge" (Escudero 2009: 152). The L2LP is based on the *Linguistic Perception Model* (LP), which explains that perception involves two sequential mappings: perception and recognition (cf. Boersma (1998), Boersma *et al.* (2003)). This means that listeners perceive an auditory form and map it to its phonological form, e.g. its place or manner of articulation. Subsequently, the phonological form is recognized and mapped to its underlying form (cf. Escudero 2009: 155 f.). The LP model makes the following two assumptions:

> First, it assumes that speech comprehension is a two-step process that involves two mappings, namely speech perception and speech recognition. Second, it assumes that speech perception is a pre-lexical and bottom-up process. […]. In addition, the model incorporates the idea that both speech perception and recognition are handled by linguistic grammars. (Escudero 2009: 155 f.)

---

[40] The perception of subset L2 sounds will not be discussed here, since this is not relevant to the present paper.

The LP also proposes that the way listeners perform this mapping is dependent on the particular L1/L2 combination. Escudero & Boersma (2003) formulate an *Optimal Perception Hypothesis*, which says that optimal listeners will a) "prefer auditory dimensions that reliably differentiate sounds in the production of her language" and "identify auditory inputs as the vowels or consonants that are most likely to have been intended by the speaker" (cf. Escudero 2009: 156). This also means that production and perception are linked: When the L1 and L2 in question differ concerning the way acoustic dimensions are integrated into perception and production, listeners will use different means in the perception of the L1 and L2 (cf. Escudero 2009: 156). For instance, Escudero & Polka (2003) studied the perception of DRESS and TRAP by three different learner groups: 1) monolingual Canadian English speakers (CE), 2) monolingual Canadian French speakers (CF) and 3) beginning Canadian English learners of Canadian French. The results showed that CF speakers mainly distinguished DRESS and TRAP by their F1 values, whereas CE speakers used a combination of F1 and durational cues. Moreover, the perceptual category boundary of these learners matched their production extremely closely (cf. Escudero 2009: 159). This is exemplified as follows:

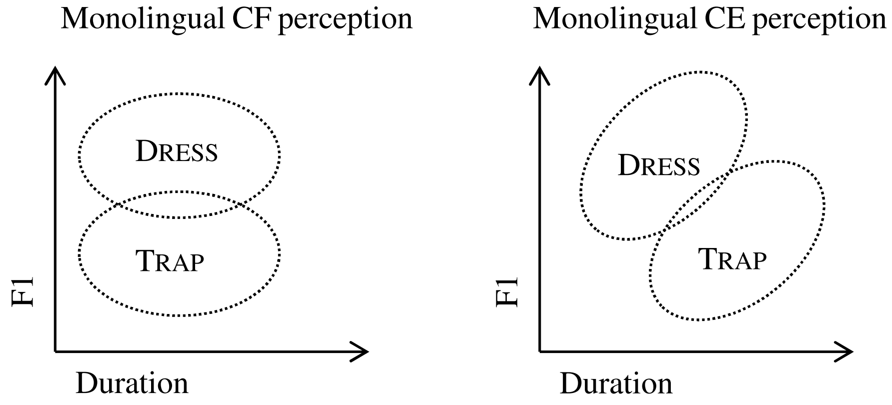


Figure 2.1-4. Canadian French (CF) and Canadian English (CE) speakers' realization and perception of DRESS and TRAP. Schematic figure adapted from Escudero 2005: 159.

Regarding the acquisition of similar phonemes by L2 learners, the L2LP predicts that learners will perceive and associate L2 sounds with L1 sounds when they possess auditory similarities. This can result in mismappings between the auditory form and the phonological representation, because the sounds differ in phonetic features. In the study mentioned above and exemplified in Figure 2.1-4, CF speakers of English might miscategorize CE TRAP with high F1 as an instance of DRESS, even if TRAP was realized with relatively long duration. With increasing exposure to the L2, the CF speakers are predicted to change their perception to include the cues relevant for correct L2 perception. Escudero describes this process in the following way:

> When faced with this situation, the learners' GLA [Gradual Learning Algorithm, cf. Boersma & Hayes (2001)], which in this situation acts as an error-driven constraint re-ranking mechanism triggered by mismatches between the output of perception and the lexicon, will change their perception grammars by small steps in order to decrease the probability of semantic mismatches. Finally, an optimal L2 perception will be attained when such mismatches no longer occur. (Escudero 2009: 177)

In the empirical research mentioned above, Escudero noticed that learners needed only little exposure to categorize an increasing number of TRAP tokens correctly (10 months; cf. 2005: 281).

For German learners of English, TRAP is a new sound which is not part of the German vowel inventory. Whereas learning similar sounds was not hypothesized to be a problem in the L2LP, learning a new sound is categorized as highly problematic, because the learner will use (inadequate) L1 cues to cope with the new sound (cf. Escudero 2005: 155). In this context, the L2LP proposes that the learner will perceive the foreign sound as the nearest L1 sound:

> In the NEW L2 sound perception scenario, the equation of two L2 categories with a single phonologically equivalent L1 category, which occurs at the abstract phonemic level, is accompanied by the perceptual mapping of the majority of the tokens of both L2 categories onto the same L1 category. In other words, most phonetic realizations or auditory events of the two L2 phonological representations will be perceived as a single L1 category. (Escudero 2005: 155, original emphasis)

This means that German learners will categorize instances of both TRAP and DRESS as realizations of DRESS. In order to attain native-like perception, learners have to form a separate category for the new sound. According to the L2LP, this happens through perceptual and representational learning. In the perceptual learning task, the learner must "[c]reate or split perceptual mappings" and "[i]ntegrate auditory cues" (Escudero 2005: 159, emphasis omitted). Representational learning involves the creation or splitting of categories and vowel segments (cf. Escudero 2005: 159). As such, learners must create new phonetic categories and phonological representations, which will then be copied to the L2 grammar. This enables optimal perception of the new L2 sound as well as integration of multiple auditory cues (cf. Escudero 2005: 161). For instance, this would allow German learners to integrate an additional cue – namely duration – into their L2 grammar in order to distinguish DRESS and TRAP. The following figure is a summary and representation of the individual steps from beginning learner to optimal L2 perceiver:

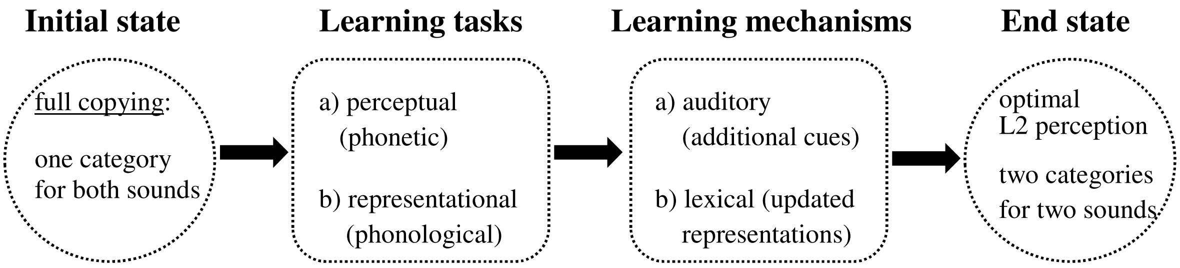


Figure 2.1-5. Developmental path predicted by the L2LP in the perception of new sounds. Adapted from Escudero 2005: 175.

According to the L2LP, learners have reached the end state of the learning process when they have attained optimal perceptional boundaries for the L2 while maintaining optimal boundaries for their L1. Both systems are "handled by two different grammars with the same constraints but different rankings" (Escudero 2009: 182). The model posits that any of the two grammars can be activated at any time, while the other is inhibited. When the auditory input is ambiguous and can belong to both grammars, the choice of grammar to be activated is purely coincidental (cf. Escudero 2009: 182).

The end state of different learners can differ. Escudero sees two types of explanations for this: loss of neural plasticity and insufficient input, where the latter outweighs the former (cf. Escudero 2005: 180). Similarly to Kuhl, Escudero questions whether adult learners ever receive sufficient high-quality input and suggests this as the reason why many adults do not achieve optimal L2 perception. She maintains that "very few learners have access to the enhancement of acoustic properties in their L2 environment, whereas such enhancement is a core feature of infant-directed speech or motherese" (Escudero 2005: 280). It is important to note here, however, that the L2LP makes a difference between the perception of new and similar L2 sounds: it hypothesizes that ambient input may be sufficient for learners to perform the boundary shift necessary for native-like perception of similar L2 sounds (cf. Escudero 2005: 280). Since the perception of new sounds also includes representational learning, ambient input will not suffice in this case.

In summary, Escudero's L2LP includes the following hypotheses:

- Listeners are optimal perceivers of their native language.
- L2 learners start with a copy of their optimal L1 perception and gradually adjust their perception using the same mechanisms as monolingual L1 speakers.
- Both L1 and L2 can be perceived in a native-like way, because they are based on different grammars.

- Learning similar sounds is not as difficult as learning new sounds, since it only includes adjusting pre-existing perceptual mappings and category boundaries.
- Adults need sufficient high-quality input to attain the optimal perceptual end state.

2.3.2.5 Comparison of the Perception Models – NLM, SLM, PAM, L2LP

The previous sections have introduced four different perception models. The NLM discusses and compares the development of speech perception from infancy to adulthood. In contrast, the SLM is primarily focused on explaining the ultimate attainment of L2 perception. The PAM tries to describe the perception of diverse forms of non-native contrasts. Finally, the L2LP models perceptual learning from the initial to the end state. Since the four models make a number of similar claims yet also many diverse ones, it is of value to compare them and identify points of (dis)similarity. One of the key divergences between the models is the question of whether L1 and L2 are interrelated or separated. In order to evaluate this question, Cook's concept of how language systems can be related in the mind will be used as a starting point (cf. 2002: 11). He proposes the following three possible interactions between the L1 and L2 system:

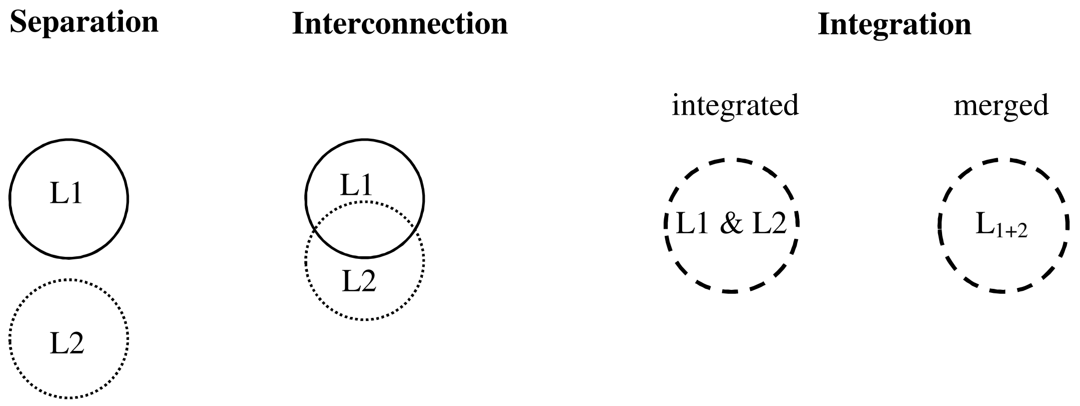


Figure 2.1-6. Possible representational statuses of L1 and L2 in the brain of learners or bilinguals. Adapted from Cook 2002: 11 and Escudero 2005: 116.

Naturally, all three possibilities entail certain assumptions and predictions. The separated systems hypothesis proposes that the two languages are stored as fully autonomous systems in the brain of a learner. A completely opposite view is exemplified by the mixed systems hypothesis. This hypothesis comprises two possible representations: The merged view posits that there is no differentiation of any kind in the two language systems of the learner. In contrast, the integrated view holds that even though the two languages share a common system, there is some specification between the two. An intermediate position is assumed by the connected systems hypothesis. Here, languages are thought to be stored separately with some limited interaction between the two systems.

Regarding the four perception models discussed in the previous sections, it is evident that they differ as to which of Cook's representations they support. Escudero's L2LP is a strong supporter of the separate hypothesis: L1 and L2 are said to be fully autonomous systems. Likewise, Kuhl's NLM and the magnet effect show that the two languages of a speaker exist in separate spaces. In contrast, Flege's SLM suggests that the two language systems exist in and share a common phonological space (integrated systems). This view is shared by Best's PAM, which states that changes in any region of one language can change the complete linguistic landscape. Two different predictions follow from these opposing views: The L2LP proposes that optimal L1 and L2 perception is possible, whereas the SLM posits that changes in the L2 will affect the L1 and vice versa. Kuhl's model is slightly more pessimistic than the L2LP. Although native-like attainment is believed to be theoretically possible, an ideal environment with sufficient high-quality input is necessary, which is rarely ever present in adult learners. The PAM, like the SLM, believes perceptual learning to be possible over the entire lifespan, but is skeptical as to whether adult learners can achieve native-like performance. The model gives the changing environment and socio-physical changes in the speaker as a reason for this.

A second large contrast emerges when one considers how these models view the perceptual learning of similar and new sounds. As a reminder, English KIT and DRESS were categorized as similar sounds, whereas TRAP and LOT are new sounds for German learners. Two of the models – the NLM and the PAM – consider L2 similar sounds as no special learning challenge. They reason that when L1 sounds or features similar to the L2 sounds/features to be acquired are present, the learner will have no difficulty using them in the L2.[41] In contrast, the SLM considers the acquisition of similar phonemes the greatest learning challenge. This is due to equivalence classification, which may block the establishing of categories in the L2. This, in turn, will make native-like perception impossible. Moreover, confrontation with a similar L2 sound might have an influence on the L1 as well, since the two sounds might dissimilate. The L2LP assumes an intermediate position. It suggests that perceptual learning of similar L2 sounds is a challenge for learners, because perceptual

---

[41] Note that the PAM does not agree with the establishment of categories in the first place and claims that learners have no perceptual mappings of sounds. Instead, they directly extract the invariants of articulatory gestures from speech. In contrast, the NLM argues that perceptual representations are stored in the brain.

mappings of the L1 might have to be adjusted to match the L2 category boundaries of a monolingual speaker. However, the learner does not need to create a new category as claimed by the SLM. Instead, perceptual learning of new sounds is thought to be the greatest challenge a learner can face. This view is supported by the NLM and PAM. Perceiving a new sound requires the learner to create completely new perceptual mapping and integrate and use cues which are not part of the L1 system. Contrary to this opinion, the SLM posits that the formation of a category for a new sound is relatively easy, because the category can be established in the phonological space without influencing or disturbing any pre-existing L1 sound.

Regarding the way perception works, the SLM does not explain how learners move from perceptual mappings to sound representations. Instead, the model assumes that the learner’s phonological system extracts phonetic information from the speech signal by means of sound categories. The PAM proposes that learners do not establish sound categories at all, but that the articulatory invariants detected in the speech signal are directly mapped onto higher-level units (cf. Articulatory Phonology; Browman & Goldstein (1989)). The NLM and the L2LP are similar in that both models suggest that there is a special mapping procedure between the recognition of abstract phonetic features and the mental representations of sounds. In the case of the NLM, this mapping procedure is said to take place via neural networks; the L2LP argues that neural, phonetic and psycholinguistic modeling are part of the process. Concerning the development of L2 perceptual learning, the L2LP refers to the mechanisms of category formation and boundary shifts, and the NLM states that categorization, distribution and warping are part of perceptual development. In this respect, the SLM is similar to the NLM, explaining that category establishment is a first step, which might in some cases be blocked by equivalence classification. The PAM does not give an explicit explanation of the learning mechanisms of the model. In addition, it is not clear how speakers recall the “higher-level invariants” when they are not represented or mapped onto a category in their brains. This lack of information is a weak point in the PAM.

Despite the differences between the models, there are also a number of similarities. Firstly, all models say that perception is language-specific. This includes the idea that perception and possible problems faced by learners depend to some degree on the similarity between the L1 and L2. Secondly, all models support the hypothesis that perception precedes production. Kuhl’s NLM argues that perception warps the acoustic space, which is then reflected in production. Flege’s SLM suggests that learners assimilate L2 sounds to L1 sounds, and only when this perception changes is accent-free production possible. Similarly, Best's PAM asserts that changes in the perception may alter the complete layout of the phonological space, which will be represented in production. Escudero's L2LP states that mismatches between the intended meaning of the speaker and the perceived meaning of the listener cause perception to change, which is, in turn, reflected in production.

The four models also agree that the learners' initial states are characterized by strong influence of their L1. The PAM claims that learners assimilate L2 sounds to their pre-existing L1 categories. The outcome of this assimilation process can lead to poor or good discrimination based on the match between the L1 and L2 categories. Similarly, the SLM posits that learners start out with their L1 system, which acts as a perceptual filter for L2 sounds. Therefore, sounds or features of the L2 may be filtered out by the L1, which hinders their perception. The NLM states that L2 learning is constrained by perceptual mappings established for the L1 which interfere with the maps required for the L2. According to the L2LP, learners use their entire L1 grammar and L1 categories to cope with L2 sounds.

Another similarity between the models can be seen in the way they describe the development of L2 learners. All of the models agree that there is no abrupt CP, but that learning is possible over the entire lifespan. In addition, all models refute that there is complete loss of plasticity in adults, although plasticity might be reduced compared to its extent in children. The individual models make slightly different proposals regarding the manner of development, however. The PAM suggests that learners advance by splitting and reorganizing their L1 categories. The SLM states that learners can create new categories only when there is sufficient phonetic dissimilarity between the L1 and L2 sounds, otherwise category formation will fail due to equivalence classification. Both models also stress that the acquisition of L2 sounds will alter the layout of the L1 phonological map. The NLM does not provide a description of the exact mechanisms in which new L2 maps and categories are formed, but claims that these mechanisms differ from L1 acquisition. The L2LP proposes that learners will establish a separate category when the L2 sound is new, i.e. has no counterpart in the L1, and adjust their mappings when the L2 sound is similar. Moreover, this process first happens on an auditory basis and then through representational learning.

The four models differ in the way they describe the end state of the perceptual learning process. The PAM does not clearly describe the end state beyond stating that L2 learning can influence L1. The NLM suggests that any pre-existing mappings from the L1 will constrain the establishment of native-like L2 categories, which is why adult learners can never become fully competent. These constraints are said to be low in early learners, which is why infants can learn multiple languages at once and become native speakers of several languages. However, when a language is learned after the L1 mapping process has been completed, the L2 will automatically be treated as a separate system. Since the SLM states that learners immediately map L2 sounds to L1 categories, it is crucial that these L1 categories not be too firmly established for L2 sounds to be able to be successfully incorporated into the phonological space. This is why the SLM posits that AOL is the most important success factor for ultimate attainment. A second important factor named by the SLM is the ratio of L1/L2 usage. This factor is also named by the PAM, which also incorporates high-quality input. Input, in turn, is one of the major factors mentioned in both the NLM and the L2LP, the latter of which even claims that input trumps AOL. The NLM and the L2LP both claim that adults hardly ever receive sufficient and adequate input, which is why they will not be able to attain native-like competence. However, while the NLM excludes native-like competence in L1 and L2 from the start, L2LP explicitly states that both systems can be optimal, because they are handled by two separate grammars.

We have observed a number of differences and similarities between and across the four models. As a summary, the following table presents an overview over the four models and their framework and propositions for how perceptual learning happens as well as predictions for the individual learning stages.

Table 2.1-4. Overview of the framework, perceptual learning process and learning stages described by the four perception models. NLM = Native Language Model (Kuhl), SLM = Speech Learning Model (Flege), PAM = Perceptual Assimilation Model (Best), L2LP = Second Language Linguistic Perception Model (Escudero). Adapted from Escudero 2005: 133, 138.

L2 Model	**Framework**			**Perceptual learning**			**Learning stages**		
	Type of model	*Elements of perception*	*Prerequisites*	*Development*	*Interaction between L1 & L2*	*Perception & production*	*Initial state*	*Mid-state*	*End state*
NLM	Language-specific	Neural maps, phonetic categories	Sufficient high-quality input similar to motherese	Language experience warps perception	Separate systems	Perception guides production	L1 maps, L1 categories	Creation of L2 maps and categories	Separate L1 and L2 maps and categories
SLM	Language-specific	Phonetic categories	Early AOL	Establishment of new category	Integrated systems	Perception precedes production	L1 categories	Category formation or merging	Early learners more but never fully native-like
PAM	Language-specific	Articulatory gestures	Early AOL, ratio of L1/L2 usage, high-quality input	Detection of higher-order articulatory invariants	Integrated systems	Perception of gestures influences production	L1 assimilation	Category split, reorganization	Not clear
L2LP	Linguistic, Acoustic-to-phonological	Linguistic mappings, phonological categories	Neural plasticity, sufficient input	Auditory perception	Separate systems	Perception is reflected in production	L1 grammar, L1 categories	Category formation, mapping adjustment	Both L1 and L2 can be optimal, separate grammars

2.4 Cues to Vowel Identity and Postvocalic Consonant Voicing

2.4.1 Durational and Spectral Cues in English

The sounds which speakers perceive and produce and which form part of their phonological inventory can be distinguished by many sub-properties of said sounds. These sub-properties are called phonetic *cues*. Speakers of a language must learn which cues signal contrasts in their language and are thus relevant, and which ones are redundant. L2 learners have the additional task of finding out if there might be cues in their L2 which are not relevant in their L1. Failure to do so might result in in erroneous categorization, which in turn might make perception and production very difficult for learners and may result in a foreign accent. However, new cues, which are particularly salient because they are not part of the L1 inventory, may be easier for learners to perceive and produce than similar cues.

In English, different types of information signal a phonemic contrast (cf. Lisker (1978), Nearey (1989)) – the two major cues for the tense/lax distinction are durational differences and spectral information (F1/F2). However, the weighting of these two cues differs among varieties. Speakers of North American English mostly rely on spectral cues in distinguishing vowels, particularly in the case of high vowels (cf. Hillenbrand *et al.* (2000), Kondaurova & Francis (2008)). Scottish English distinguishes KIT and FLEECE almost exclusively through spectral differences, whereas the same vowels have a considerable durational, but only a small spectral difference in Southern English (cf. Escudero (2001)). Since “production data to which a child is exposed influence[s] the preference for particular perceptual weightings”, children learning English will pay a different degree of attention to different cues (Escudero 2001: 250). For instance, a child learning Scottish English will pay more attention to spectral than durational cues, because adults will produce a larger spectral than durational difference (cf. Nittrouer (2000)). This suggests that the production and the perceptual weighting of cues is related (cf. Raphael (1972)). Moreover, this explains why native speakers have been shown to pay different amounts of attention to different cues (cf. Nearey (1989), Nittrouer (2000)). For instance, adult Scottish speakers are perceptually more sensitive to spectral than to durational cues (cf. Escudero 2000; 2002)).

Regarding the L2, research has shown that learners use different auditory information than native listeners when perceiving sounds. While – as noted above – American English listeners prefer spectral information in distinguishing KIT and FLEECE, L2 learners of different languages prefer durational cues (cf. Bohn (1995), Flege *et al.* (1997)). What is particularly interesting, however, is that durational cues are preferred both by learners whose L1 has durational contrasts (e.g. German, cf. Bohn & Flege (1990); Hungarian, cf. Altenberg & Vago (2006); Arabic, cf. Munro (1993); and Japanese, cf. Minnick-Fox & Maeda (1999)), and learners whose L1 does not (e.g. Mandarin, Korean, Spanish, cf. Bohn (1995), Flege *et al.* (1997), Wang & Munro (1999), Escudero & Boersma (2004), Escudero (2006); Portuguese, cf. Rauber *et al.* (2005); Catalan, cf. Cebrian (2006); Russian, cf. Kondaurova & Francis (2008)). Bohn claims that this is because duration is universally salient (cf. 1995: 300).

Learners have been found to be understood more easily by native speakers if they master both spectral and durational cues (cf. Flege *et al.* (1995), Kewley-Port *et al.* (1996)). Several studies have shown that L2 learners are capable of learning new cues, as well as adapting existing cues in their L1 to the L2, given sufficient experience. They can also change their cue weighting, i.e. the degree of attention they pay to cues available in the speech stream. For instance, Flege (1987) found that highly experienced English learners of French managed to pronounce French /y/ authentically. In their 2012 study, Simon & D'Hulster noted that experienced Dutch learners of English produced DRESS and TRAP with a clear contrast, whereas inexperienced learners did not, even though both groups were able to distinguish both sounds in a perception task. Bohn & Flege (1992) examined experienced and inexperienced German L2 learners of English in their production of the similar KIT, FLEECE and DRESS and the new sound TRAP. Their results showed that both the experienced and the inexperienced speakers produced intelligible tokens of KIT, FLEECE and DRESS (which were still acoustically different from native speakers). However, the groups differed in the new sound – the experienced group produced much more intelligible TRAP tokens than the inexperienced group. This means that the experienced German learners were more easily able to detect the spectral cues of TRAP and form a new category.

With regard to cue weighting, Escudero *et al.* (2009) studied the perceptual behavior of native Dutch and German speakers and Spanish learners of Dutch with respect to the Dutch phonemic contrast between /aː/ and /ɑ/. Their results show that cue weighting is dependent on the L1 of the speaker: native Dutch and German listeners weighted spectral information more heavily than vowel duration, whereas the Spanish learners of Dutch relied more on vowel duration. Morrison (2002) examined Japanese and Spanish learners' perception of KIT and FLEECE one and six months after their arrival in Canada. While Spanish only has one high front vowel /i/, the Japanese inventory includes phonemic vowel duration differences, so that there are two high front vowels /i/ and /iː/. Morrison showed that the Japanese learners failed to change their cue weighting and largely relied on durational differences to distinguish KIT and FLEECE both one month and six months after their arrival. However, the Spanish listeners, who were unable to perceive a difference between the two vowels at the initial stage, managed to establish category boundaries for KIT and FLEECE either based on spectral or durational information, but not both (cf. Morrison 2002: 29). Morrison's findings may be explained by the layout of the L1 versus the L2: It should be recalled that the two models SLM and L2LP discussed in the previous section both see the acquisition of similar sounds as a great challenge. This would explain why the Japanese learners continued to rely on durational cues exclusively even after lengthy exposure to their L2: Their two L1 phonemes /i/ and /iː/ are quite similar to English KIT and FLEECE, except for the spectral difference, which Japanese lacks. This similarity might have made it very difficult for Japanese learners to perceive the English spectral cue. In contrast, the Spanish learners were confronted with two cues which were completely foreign to them: spectral information and durational differences. The establishing of categories for new information is hypothesized to be comparably easy by both SLM and L2LP. Therefore, it is not surprising that the Spanish participants managed to establish a new category for KIT.

Turning to the issue of vowel duration as a cue to postvocalic consonant voicing, results have been mixed at best. Halle *et al.* (1957), Raphael (1972), Flege & Bohn (1989) and Crowther & Mann (1992) report that preceding vowel duration suffices as a cue to postvocalic consonant voicing. Their perception tests show that listeners perceive coda consonants following longer vowels as voiced. In contrast, Lisker (1957) proposes that closure duration is the main cue to consonant voicing. A third stance is taken by Denes (1955) and Port (1981), who argue that postvocalic consonant voicing is cued by the ratio of vowel to consonant duration. However, Crystal & House (1982) and Luce & Luce (1985) challenge the notion that closure duration or the vowel to consonant duration may reliably cue consonant voicing. Instead, F1 offset frequencies were identified as cues, with low F1 offsets being identified as vowels preceding a voiced consonant (cf. Walsh & Parker (1983), Hillenbrand *et al.* (1984), Fischer & Ohde (1990)). Interestingly, F1 offset was found to be a more reliable cue in low vowels than in high vowels, since F1 in high vowels is already quite low. Therefore, duration seems to be a more helpful cue to postvocalic consonant voicing in high vowels (cf. Hogan & Rozsypal (1980), Walsh & Parker (1983), Halle *et al.* (1957), Hillenbrand *et al.* (1984)). This means that postvocalic consonant voicing may be cued both by differences in duration and spectral quality of the preceding vowel.

Naturally, researchers have been trying to find out *why* learners struggle to perceive and/or produce non-native cues. As we have seen, cues differ in their relative salience in different languages (cf. Walley & Carrell (1983), Burnham (1986)), which in turn influences how easily they are learned by non-native speakers (cf. Francis *et al.* (2000), Francis & Nusbaum (2002)). The fact that learners of a language are more sensitive to some cues than others – for instance durational differences, as noted above – might be an innate characteristic (cf. Holt & Lotto (2006)). However, this might also result from systematic differences between the L1 and L2. This view is proposed by the feature hypothesis.

Feature hypothesis claims that learners have difficulty detecting L2 cues which are not distinctive in their L1. Consequently, they will fail to use these cues in production as well. Moreover, when a cue is used in a similar way in the L1 and the L2, speakers will be at an advantage compared to learners for whom the cue is completely new (cf. Flege (1995; 1998), McAllister *et al.* (2002)). For instance, Italian and Spanish learners have been shown to perform similarly to English native speakers in the use of closure voicing in final obstruent voicing. In contrast, Korean and Mandarin learners do not attend to this cue. An explanation can be found in differences between the L1 and L2: although none of the languages makes use of a voicing contrast in word-final position, closure voicing is a more predominant cue in Spanish and Italian than in Korean and Mandarin (cf. Flege *et al.* (1992), Flege *et al.* (1995), Flege (1998)). This has also been shown regarding vowel length difference as a cue to postvocalic consonant voicing: In studying English, Japanese and Mandarin speakers, Crowther & Mann (1992) found that Japanese speakers, whose language includes a phonemic length distinction, were perceptually more sensitive to vowel length variation than Mandarin speakers, whose language does not have intrinsic vowel length distinctions (1992: 711). Neither language has final consonants. The researchers conclude that “native experience with long/short vowels might be relevant to the use of vocalic duration as a cue to final consonant voicing” (Crowther & Mann 1992: 711).

A slightly different approach within this area is taken by Kondaurova & Francis (2008), who maintain that Spanish and Russian learners make use of their native allophonic experience with vowel duration in identifying L2 vowel contrasts. This claim is based on the fact that Spanish and Russian have allophonic duration differences with regard to stress and postvocalic consonant voicing. Unfortunately, the researchers do not explain how learners can directly transfer an acoustic cue from a suprasegmental domain, such as stress, to a phonological level (cf. Escudero *et al.* 2009: 643). Morrison's (2008a) study might give a tentative answer to this question, at least regarding the Spanish listeners. He finds that Spanish learners of English go through an initial stage where they rely on both durational and spectral cues in the identification of vowel identity, before reverting to durational cues only. He links this to a type of feedback process: Because Spanish learners' production of durational differences is more similar to that of native speakers' than their use of spectral characteristics, Spanish learners undergo a perceptual development where they abandon spectral cues in favor of durational ones.

Feature hypothesis has been somewhat discredited because it has been shown that cross-linguistic differences between the L1 and L2 do not account for the behavior of monolingual speakers. Morrison (2008b) found that neither speakers of Mexican Spanish nor Peninsular Spanish attend to durational cues in vowel identification tasks. Likewise, Benders & Escudero (to appear; cited in Escudero (2009)) report that Peruvian listeners do not make use of duration in categorizing Spanish vowels, either. These results seem to indicate that there must be universal or developmental experience. Both views are briefly outlined below.

A second, contrasting hypothesis has been put forward by Bohn (1995). He argues that native language is not always the major influence on why learners manage or fail to perceive and produce non-native cues. In contrast, he proposes the desensitization hypothesis, which states that non-native cues are in principle available to L2 learners, namely "whenever [L1] spectral differences are insufficient to differentiate [L2] vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences" (Bohn 1995: 294 f.). Support comes from Bohn & Flege's (1990; 1992) research. They studied the production and perception of FLEECE, KIT, DRESS and TRAP by German, Spanish, Mandarin and Korean learners of English. The German listeners relied more on duration to differentiate DRESS and TRAP than KIT and FLEECE. Because TRAP is a new sound for German learners, while the others all have similar counterparts, German learners do not have available spectral cues to differentiate TRAP.

In a similar study, Spanish listeners relied extensively on temporal cues to distinguish KIT and FLEECE, (KIT is an unfamiliar sound), but not to differentiate DRESS and TRAP, which might be assimilated to Spanish /e/ and /a/ (Flege & Bohn 1989: 85). Last, in Kondaurova &

Francis' study, Russian and Spanish speakers relied more on durational cues than spectral ones to differentiate English KIT and FLEECE. The researchers argue that due to the smaller vowel inventory of Spanish and Russian compared to English, speakers of these languages are less sensitive to spectral differences (cf. Kondaurova & Francis 2008: 3969; cf. also Hacquard *et al.* (2006)). All these findings supports Bohn's hypothesis that duration can be used as a cue to distinguish vowels when the L1 does not offer a spectral cue.

Escudero & Boersma (2004) give an explanation similar to Bohn (1995), but different in one aspect: Whereas Bohn (1995) claims that duration is universally salient, i.e. that all speakers have an innate category for durational differences, Escudero & Boersma (2004) claim that learners start out with *no* category. With growing input, learners then use distributional learning, an L1 acquisition strategy (cf. Maye *et al.* (2002)) to create a category for durational cues. Since the creation of a new category is said to be relatively easy in their model, it also becomes clear why Spanish learners weigh durational cues more heavily than spectral ones when identifying English vowels: Since their L1 already includes spectral cues for Spanish vowels, they would have to split this category to allow for English spectral cues. This, however, is more difficult than establishing a completely new category as in the case of duration.

2.4.2 Acquisition of Vowel Length Variation as a Cue to Final Consonant Voicing

This dissertation aims to answer the question of whether and how German L2 learners of English acquire vowel length variation before voiced and voiceless plosives. Examining the L1 acquisition process might help develop a strategy to make L2 learning easier. In addition, it might give hints as to the deviant behavior of L2 learners compared to native speakers. Therefore, this section is devoted to examining the acquisition of vowel length variation in L1.

Much is known about the processes of acquisition – both in L1 and L2 – in almost any linguistic subfield. The speech development of children has been extensively researched with regards to phonetics, phonology, morphology, syntax, semantics and pragmatics. Scholars have found that in comparison with adults, children have higher formant frequencies and pitch, longer segment durations and greater temporal and spectral variability (cf. Smith (1978; 1992), Smith *et al.* (1996), Hillenbrand *et al.* (1995), Lee *et al.* (1999), Gerosa *et al.* (2006), Koenig *et al.* (2008)). The acquisition of vowel length variation by children has also received continuous attention since around the 1970s. Researchers have studied both production and perception of vowel length variation before voiced and voiceless plosives by children. Results of studies on each of these abilities will be presented in turn.

The earliest study of infants' acquisition of vowel length variation seems to be Naeser's 1970 doctoral dissertation. Naeser reports that children aged 22 months and older show "correct" vowel length variation (1970: 16). Many studies have corroborated this finding for children of different ages. For instance, DiSimoni (1974), Raphael *et al.* (1980) and Krause (1982) found durational differences in vowels for children at ages 3, 4, 6 and 9. Stoel-Gammon *et al.* (1995), Stoel-Gammon & Buder (1998) and Buder & Stoel-Gammon (2002) reported vowel lengthening in 2- and 3-year-olds. More recently, vowel length variation has been shown to exist in children even before the age of 2 (cf. Ko 2007: 1882 f., Tauberer 2010: 82). In addition, children have been reported to use longer vowels before voiced codas compared to voiceless codas even when they do not produce the coda (cf. Weismer *et al.* (1981), Kehoe & Stoel-Gammon (2001)).

Conflicting results have been published as regards a developmental trend in children's production of vowel length variation. What is more, some studies have found a developmental trend for vowels preceding voiced consonants, but not voiceless ones. For instance, DiSimoni (1974), Krause (1982), Ko (2007) and Song *et al.* (2012) all found that the duration of vowels preceding unvoiced consonants remained stable for their participants (aged 3 to 9, 3 to adulthood, 11 months to 4 years and 1;6 to 2;6, respectively).[42] However, while DiSimoni (1974) described an increase in duration in vowels preceding voiced consonants, Krause (1982), Ko (2007) and Song *et al.* (2012) all report a decrease of vowel length in this environment with increasing age of the children. In Naeser's study, vowel length increased before voiceless consonants – at 22 months, the duration of vowels before voiceless consonants was 49% of that before voiced consonants, and at 34 months it was 63% (cf. 1970: 16).

---

[42] This finding seems to be quite surprising in light of the fact that overall vowel length seems to decrease with increasing age of the child (as proposed by Lee *et al.* (1999)). However, Song *et al.* (2012)) did not find overall durational changes, and Smith *et al.* (1996) found substantial differences in some children, but none in others, which considerably complicates the situation. The conflicting findings might be due to the different ages of the children studied: Lee *et al.* (1999) reported a decrease of overall duration between the ages of 9 and 12. The children observed in Song *et al.*'s (2012) study were only 1;6 to 2;6 years old, whereas Smith *et al.* (1996) examined children aged 7 to 11. Therefore, no final estimation can be given and the issue remains unresolved.

In Ko's (2007) study, the earliest age at which vowel length variation was encountered was 1;4.[43] None of the children examined showed a developmental pattern in the manner that they made "mistakes" in vowel duration, which Ko takes as evidence that vowel length variation is an automatic rather than a learned process (cf. 2007: 1884; cf. also Kehoe & Stoel-Gammon 2001: 424). Ko admits, however, that children might have learned a rule before using it actively in speech, i.e. that "children learn the appropriate control regimes for articulating the vowel in the segmental context so thoroughly that the phonetic implementation process appears to be an automatic process" (2007: 1884). A very recent study by Song *et al.* (2012) has repeated many of Ko's findings. Examining mother-child conversations, the researchers found vowel length variation as a function of coda voicing to appear at around the age of 1;6, although the use was not yet completely adult-like (cf. Song *et al.* 2012: 3041). For instance, children used exaggerated duration differences to signal coda voicing.

If a developmental trend was present, it might also be imaginable that it is language-specific, i.e. that all children start with vowel length variation and lose this distinction when their native language does not employ this differentiation. This hypothesis was successfully tested by Buder & Stoel-Gammon (2002), who compared Swedish and English children at 24 and 30 months of age. Like German, Swedish has phonemic vowel length variation. The researchers showed that Swedish children showed vowel length ratios of 1.7 at 24 months and 1.1 at 30 months, while the English children had ratios of 1.4 and 1.7 at those ages (2002: 1862). This supports their claim that Swedish children's sensitivity to vowel length variation decreased with time, while the English children become more attuned to this difference. However, counterevidence has been presented e.g. by Tauberer, who shows that the magnitude of the vowel length variation increases with the age of the children (cf. 2010: 83). Tauberer concludes from this that vowel length variation is indeed a learned phenomenon, and that children are not born with a universal vowel length effect, which they need to unlearn if their native language does not use this distinction (cf. Tauberer 2010: 83). The two conflicting findings of Buder & Stoel-Gammon (2002) and Tauberer (2010) are a clear sign that even though there is much data to analyze, considerable confusion remains as to the exact developmental status of vowel length variation in children.

---

[43] Note that Ko (2007) worked with a corpus not collected by her, so that the ages she mentions correspond to the earliest date at which a suitable minimal pair was found. It might thus be the case that vowel length variation appears in children's speech even earlier, as soon as they produce codas. In fact, in Tauberer's study, children aged 1;3 already showed vowel length variation (2010: 82).

We must also bear in mind that production of vowel length variation does not mean that children have stored this phonological category in a lexical entry. Van der Feest reports that infants at age 1;8 do not notice when they are shown a picture of a familiar object and the name of the object is mispronounced in the onset (2007: 149). For instance, she showed children a picture of a cat and played the stimulus *poes* [pus] (correct) or [bus] (incorrect). By age 2, children detected when a voiceless consonant was mispronounced as voiced, but not the other way around (van der Feest 2007: 149). This seems to suggest that “the ability to make a perceptual distinction does not translate immediately into the ability to associate the distinction with a word in the lexicon” (Tauberer 2010: 33). Likewise, Dietrich *et al.* (2007) demonstrated that English-speaking children at age 1;6 did not notice considerable vowel duration differences (near halving or near doubling) when the auditory stimuli were paired with pictures. However, when there were no images, they were able to discriminate the vowel duration differences (cf. Mugitani *et al.* (2009)). These finding is slightly at odds with the results reported by Swingley (2009), who tested 14- to 21-month-old infants in a visual fixation task using auditory stimuli mismatched both in onset and coda position. He presented the children with minimal pairs such as *boat*, *poat* ([voice] mismatch in the onset) and *boad* ([voice] mismatch in the coda). In this study, infants detected both onset and coda mismatches, even though “[w]ord-final consonants are, in general, less clearly articulated; they are heard only after perception of the initial parts of the word has led children to consider an interpretation; and they enjoy less of the (hypothesized) benefit of membership in dense phonological neighborhoods” (Swingley 2009: 266). It is also quite surprising that Ko’s (2007) and Song *et al.*’s (2012) studies should find children younger than the age of 2 to be able to produce a vowel length contrast when van der Feest showed that children only begin to specify this distinction in the lexicon at that age. This might be due to the difference between studying actual phonetic inventories and abstract phonological knowledge, however (cf. Tauberer 2010: 33).

Several studies have examined infants’ perception of vowel length variation. Eilers (1977) and Eilers *et al.* (1984) have shown that children aged 5 to 11 months use vowel duration as a cue to postvocalic consonant voicing. However, compared to adults, children require larger durational differences to perceive contrasts (cf. Eilers *et al.* 1984: 1217). Dietrich *et al.* (2007) studied Dutch- and English-speaking 18-month-old children. Recall that Dutch, like German, uses vowel length on the phonemic level to distinguish e.g. [a] and [aː]. Dietrich *et al.* found that the children’s perception matched that of adults of their native language: Dutch-speaking children perceived vowel duration as phonologically contrastive, whereas English-speaking children did not (Dietrich *et al.* 2007: 16027). This supports Crowther & Mann, who claim that the acquisition of vowel length variation is influenced by experience with the native language (cf. 1992: 721).

Last, Ko *et al.* (2009) examined children between 8 and 14 months and presented them with matched (long vowel + voiced coda / short vowel + voiceless coda) and mismatched tokens (long vowel + voiceless coda / short vowel + voiced coda). They report that 14-month-old children are sensitive to the mismatching of vowel duration and consonant voicing only in the short condition, whereas 8-month-olds did not react to either condition.[44] This warrants two conclusions: Perceptual sensitivity to vowel length variation seems to develop between 8 and 14 months, and – if Ko's mention of earliest production at 16 months holds – perception seems to precede production. Furthermore, it is quite noteworthy that the older children in Ko *et al.*'s (2009) study only showed sensitivity in the short condition (i.e. vowel + voiceless consonant). Ko *et al.* explain this as follows:

> [It] may well be a consequence of infants' familiarity with vowel lengthening effects such as phrase-final lengthening and vowel elongation in infant-directed speech. Vowels are lengthened due to a variety of causes, and thus long vowels appear in variable contexts in the input. Therefore, infants may treat shortening as a more relevant cue for the phoneme boundary than lengthening or treat lengthening as more acceptable than shortening. (Ko *et al.* 2009: 138)

The same findings have been reported for Japanese 18-month-olds (Mugitani *et al.* (2009)) and Dutch 21-month-olds (van der Feest & Swingley (2011)) as well as for adults (Hogan & Rozsypal (1980)). This perceptual behavior thus appears to be established in the early acquisitional stages and remains in adulthood. Examining the native speakers' performance in the perception experiment of the present dissertation also supports this finding – native speakers perform better when the coda consonant is voiceless (cf. sections 4.3.1.1 and 4.3.2.1).

Interestingly, however, there seems to a development in the use of cues over the course of childhood after all. In a study by Wardrip-Fruin & Peach, 3-year-old children used only durational cues to final consonant voicing, 6-year-old children used spectral cues exclusively, and adults relied on both cues alike depending on their availability (1984: 167; cf. also DiSimoni (1974), Smith (1978), Eilers *et al.* (1984)). Krause (1982) found that children aged 3 and 6 both used durational cues to distinguish between voiced and voiceless coda

---

[44] Cf. Tauberer 2010: 32 f. for a critique of these findings and the ensuing interpretation.

consonants, but that they needed increased durational differences compared to adults in order to shift their perception. Moreover, the 3-year-olds needed an even greater increase in duration than the 6-year-olds. This would provide some evidence as to a developmental pattern, which Ko *et al.* (2009) also found in infants aged 8 to 14 months. One might therefore summarize that very young infants already show sensitivity to durational cues, while cue weighting and the acquisition of additional cues might take several years. Tauberer concludes as follows:

> In the first stage in the first few months after birth, acoustic dimensions that take part in [voice] become perceptually accessible, including VOT and only later preceding vowel duration. During the second stage around 8–14 months, associations between phonetic cues to [voice] begin to form, as seen by infants' ability to notice a mismatch between vowel duration and the other cues to post-vocalic voicing. But at this stage, the infants are not storing this information into their lexical entries. Only between 20 and 24 months does it appear that voicing information is stored in lexical entries.
> (Tauberer 2010: 34)

We are left with the question of what reasons there are for children to show vowel length variation in the first place. Tauberer (2010: 30) identifies the following three possibilities:

- Adult-like representation: The process is phonological, i.e. vowel duration depends on the voicing of the coda consonant.
- Phonemic vowel length: The duration difference is stored as a property of the vowel.
- Performance: The duration difference is due to physiological processes.

Phonemic vowel length can be discredited due to the results published by Dietrich *et al.* (2007) and Mugitani *et al.* (2009), which show that children do not notice a change in vowel duration in a word learning task. Mugitani *et al.* reason that

> the inability of English infants to link a word with an object in a word-learning task […] can be attributed to their failure to use the vowel length cue to link novel words to novel objects, presumably because of the emergence by 18 months of more interpretative phonological categories, rather than simply acoustic–phonetic categories. (Mugitani *et al.* 2009: 240)

Likewise, performance is ruled out by the findings of Weismer *et al.* (1981), namely that children show vowel length variation even if the coda consonant is omitted, and that children notice mismatches in vowel duration and coda voicing (Ko *et al.* (2009)). This leaves us with the conclusion that children have an adult-like phonological representation of vowel length variation.

2.5 Research Questions

The previous sections have introduced many diverse theoretical issues. We focused on a description of vowel length and vowel length variation in English and German and saw which factors influence vowel length variation in English. Afterwards, we examined how learners perceive and produce an L2 differently from native speakers of that language. Four major perception models were discussed and compared. The final section was dedicated to cues. We considered durational and spectral cues in English, established vowel length variation as a cue to postvocalic consonant voicing and found out how children acquire vowel length variation. This section will outline the major research questions this dissertation sets out to answer and which have emerged from the theoretical discussion.[45]

Starting from the last point in line, the obvious question on which this dissertation centers is how vowel length variation is acquired by adult second language learners. Since this has not been previously studied, it is not clear how their learning process will compare to child L1 learners. Production and perception of vowel length variation will be examined in turn.

Regarding L2 production, it has been shown that learners are more easily understood by native speakers if they master both spectral and durational cues. To this end, the production experiment studies whether German L2 learners can use durational cues and produce vowel length variation in a native-like manner. The production experiment confronted participants with four reading tasks of varying formality – minimal pairs, a word list, sentences and a text. Only participants who show solid vowel length variation both in the formal and the informal tasks can be said to have mastered it.

The major research questions pertaining to production are:

- In which contexts can German learners successfully reproduce vowel length variation?
- Are learners more successful in formal tasks such as minimal pair and word list reading, where they pay closer attention to their pronunciation?
- Which factors make learners successful?
- In which vowels (KIT, DRESS, TRAP, LOT) and coda plosives (/b/, /p/, /d/, /t/, /k/, /g/) can learners most easily match their native speaker target?
- What is the role of the stay abroad of the learners and what is the role of their reason for it?

---

[45] This section is designed as an overview of the major research questions. For the exact nature of the experiments described here as well as the factors mentioned please see section 3 – Methodology.

The perception experiment confronted learners with sentences in which minimal pair tokens which only differ in the voicing of the final plosive, are inserted (e.g. *bat* / *bad*). The final plosive was cut off, so that the only remaining difference between the tokens was preceding vowel length. Participants had to listen to these tokens and identify which token they heard. Research has shown that learners prefer durational cues in identifying phonemes, but this experiment will show whether durational cues are sufficient for learners to identify words when the spectral cues of the final plosive are missing. This would mean that they can perceive vowel length variation as a distinctive cue. We will also examine which factors make learners successful in perceiving vowel length variation. The following is a list of the major research questions concerning perception:

- Can German learners successfully perceive vowel length variation as a cue?
- Which factors make learners successful?
- Do learners use the same cues as native speakers in perceiving vowel length variation?
- Is VLV in similar vowels (KIT, DRESS) easier to perceive than in new ones (TRAP, LOT)?
- Which tokens including which (missing) coda plosives (/b/, /p/, /d/, /t/, /k/, /g/) are easiest to identify?
- What is the role of the stay abroad of the learners and what is the role of their reason for it?

Naturally, the topic of production and perception also includes the issue of which precedes which. Research has produced arguments in both directions, so that no clear hypothesis can be formulated as yet.

A third large topic in this dissertation is the relationship between phonological proficiency and success at producing and perceiving subphonemic vowel length variation. Phonological proficiency was assessed twofold, namely by measuring whether participants a) had successfully established two distinct categories for TRAP and DRESS, and b) did or did not exhibit devoicing. Concerning a possible merger of TRAP and DRESS, this is relevant regarding the representation of sounds in a learner's mind. Are sounds represented as several distinct cues (e.g. spectral vs. durational cues), or as a bundle of cues which together form a category? If the production and perception of vowel length variation goes hand-in-hand with separate categories for TRAP and DRESS, we may assume that learners form bundles of cues to represent a category. Regarding devoicing, two hypotheses seem plausible: Learners who exhibit strong devoicing might be unable to produce accurate vowel length variation as well. However, learners might also use exaggerated vowel length variation in order to compensate for devoicing. In summary, the major research questions in this field are the following:

- Is success at producing phonemic contrasts an indicator for success at producing and perceiving subphonemic variation?
- Are speakers who merge TRAP and DRESS less able to produce and/or perceive vowel length variation? To put it another way, is vowel length variation learned as part of a bundle of cues which make up a phoneme category?
- Is there a link between vowel length variation and devoicing? Do German learners use vowel length variation in order to increase the contrast between final voiced and voiceless plosives? Or are learners who show strong devoicing also unable to produce vowel length variation accurately?

The topics touched upon here will be discussed in this order in the Results & Analysis chapter (4). After a general note on statistics (4.1), we will consider the results of the production experiment in section 4.2. The following section 4.3 will answer the research questions pertaining to the perception experiment. Perception and production will be linked in section 4.4, before we turn to the question of how phonological proficiency influences the acquisition of vowel length variation (4.5).

First of all, however, chapter 3 – Methodology.

3. Methodology

This chapter introduces the participants in the study (section 3.1) and describes the design of the production (3.2) and perception experiment (3.3). Section 3.4 explains the data obtained for this dissertation. In particular, this part will discuss how vowel length and vowel length variation, the merging of DRESS and TRAP, devoicing and word frequency have been measured, as this has some implications for the statistical analyses performed within the scope of this dissertation.

3.1 Participants

This study works with two types of participants: native speakers of English and German learners of English. The German participants were mainly approached by e-mailing higher semester students of English at the Universities of Freiburg and Tübingen, who were likely to have advanced far enough in their studies to have completed a stay abroad in an English-speaking country (this is a compulsory requirement for many degrees). Some also learned about the study through leaflets posted on bulletin boards at the university. The native speakers were approached in a number of different ways: apart from the leaflets mentioned above, some native speakers were contacted on an online British expat forum; one was a guest researcher at the University of Freiburg. Most of the native speakers asked friends to join the research project. All participants agreed to participate in the study without pay, but had the chance to win one of ten small prizes. In total, 86 participants aged between 20 and 30 years completed the experiment. They are grouped as follows:

Table 3-1. Overview of participants.

	American group	**British group**
Native speakers	10	10
German learners	30	30
Control group	3	3

Henceforth, the groups of speakers will be referred to as:

Americans: 10 native speakers of American English, born in different states all over the United States. 9 of them were students on exchange at the University of Freiburg; one was a guest lecturer at the University of Freiburg.

British: 10 native speakers of British English, 7 born in England, 1 born in Wales, 1 born in Scotland (but raised in different cities in England) and one born in Stuttgart and raised in an English-only household until age 6.

G-US (German non-native speakers of English – US): 30 native speakers of German who have completed a stay abroad in the United States for at least four months.

G-UK (German non-native speakers of English – UK): 30 native speakers of German who have completed a stay abroad in the United Kingdom for at least four months. The vast majority of participants in this group stayed in England; three participants were included in the study even though they had stayed in Scotland[46] due to the fact that they had lived in dorms with exclusively English students and were taught in Received Pronunciation at university. Thus, their exposure to Scottish English was rather limited.

The two groups *G-US* and *G-UK* include five bilingual speakers. All of these speakers have one native English-speaking parent with whom they sometimes converse in English. However, none of these speakers considers him-/herself as bilingual. Instead, all of them indicate that German is their native language, because they grew up in Germany and only have German-speaking friends. Their life is German-dominant except for the fact that they sometimes (not always) use English to speak to one of their parents. Therefore, these speakers were included in the German groups.

In addition to these four major groups, two control groups were recorded:

C-US (German control group – US): 3 native speakers of German with no stay abroad who had completed one semester of English studies at the University of Freiburg at the time of the recording, including an introductory speaking course in American English held by a native speaker of this variety.

C-UK (German control group – UK): 3 native speakers of German with no stay abroad who had completed one semester of English studies at the University of Freiburg at the time of the recording, including an introductory speaking course in British English held by a native speaker of this variety.

The exposure to English of the participants of the control group was carefully controlled for. Apart from them not having spent any time abroad in an English-speaking country,

---

[46] Due to the Scottish vowel length rule, participants with a stay abroad in Scotland were largely excluded from the study, except those named above. Cf. Aitken (1981; 1984).

participants also had to indicate that they had not been in a relationship or lived with a native speaker of English in order to be considered for the study. Moreover, they were only selected if none of their teachers or private tutors had been a native speaker of English. Their exposure to music or movies in English was the same as for the participants in groups G-UK or G-US; however, this passive exposure cannot be considered to have a strong impact on the development of phonological proficiency in English.[47]

Most of the German participants started learning English at school at age 10 or 11. Before the start of the experiment, participants signed a consent form and filled out a questionnaire relating to their background and their use of English (cf. section 3.4.1). A copy of the questionnaire, the consent form and an overview of the participants can be found in Appendix A and B.

In order not to reveal the true purpose of the study, participants were told that it dealt with their production and perception of vowels in English. All participants received identical instructions in their native language. The order of the experiment – reading tasks before listening task – was chosen in a way so that participants were unlikely to learn the true objective of the study. Instead, by having to read minimal pairs first participants were likely to think that the study was about final devoicing, which is a prominent feature of German learners' English.

In the analysis, results will be reported for the two major varieties British English and American English. Thus, groups *British*, *G-UK* and *C-UK* will form the *British group*; groups *Americans*, *G-US* and *C-US* will form the *American group*. Individual speakers will be addressed by their ID numbers. Speaker numbers 001 – 060 refer to groups G-UK and G-US. Numbers 101 – 110 are the American native speakers, 201 – 210 the British ones. The control groups C-UK and C-US are numbered 301 – 306.

3.2 Production Experiment

Participants completed four reading tasks with decreasing formality (cf. Labov (1972)) in the following order: minimal pairs, a word list, sentences and a text (see Appendix C for a copy of the tasks). The reading tasks were chosen to include different levels of formality because researchers have often been concerned about the validity of results coming from artificial tasks. Moyer (2007), for instance, states that "[i]mitation tasks and reading aloud isolated

---

[47] This factor was not statistically significant in the analysis.

words (or even phones/phonemes) may elicit closer-to-target production than free speaking, possibly due to a greater focus on accuracy in decontextualized tasks" (2007: 113). Similar objections are voiced e.g. by Dickerson (1975), Oyama (1976) Tarone (1982) and Sato (1985). As such, two of the tasks in the present study, namely the minimal pair and the word list reading, may be subject to the influence of decontextualization. However, the two tasks were included on purpose in order to observe whether learners could produce vowel length variation both in more reduced contexts, where they were expected to pay close attention to pronunciation, and contextualized tasks such as reading sentences or a text. The latter two tasks were incorporated into the study in order to provide insight into more of an "authentic" pronunciation of the learners. A drawback of these types of tasks is that "phonological skills are difficult to isolate from other language skills once the participant moves beyond the word-level" (Moyer 2007: 114).

Each vowel and consonant combination appeared in the minimal pairs three times, except for TRAP followed by /b/ (e.g. *tab*), which is included in the minimal pairs only twice due to an unfortunate design flaw. From a statistical viewpoint, this is not a problem, however. It should also be mentioned here that there are no distractor tokens in the minimal pairs list, which might not seem advisable at first. However, this turned out not to be a problem thanks to the fact that the first few tokens each participant read were not recorded but used to gauge and adjust the volume of the recording. Although some participants seemed slightly nervous at the beginning of the first recording, their reading speed had decreased considerably when they read the list the second time.

In the word list, sentences and text, the test tokens were inserted into a number of distractor words.[48] Each vowel and consonant combination appeared twice in the word list, three times in the sentences and once in the text. Participants were asked to read the text twice, so that each token was recorded twice in the text as well.[49] The reason for this was that even with only one instance of vowel and consonant combination in the text, the text was already rather long and participants might have felt overwhelmed to read an even longer text.

After reading the minimal pairs and the word list, participants were told that the following sentence reading task would be much easier in order to encourage the change from a more formal reading style to a more natural one. The text for the text reading task was a

---

[48] The test tokens are marked in bold print in Appendix C, while they were not marked in any way in the copy given to the participants of the study.

[49] Cf. Moyer (2007) and Long (2005) for relevant points regarding practice effects which might have appeared due to participants' reading the text a second time.

story written by the researcher about Little Bear and Little Tiger, two famous characters by the German children's book author Janosch. Next to the text were two pictures of Little Bear and Little Tiger. This was supposed to invite the most informal reading style of the participants. It seemed to work, too, since many participants were very enthusiastic about reading the story and even used different voices for the characters.

Data was recorded using a portable Edirol R 09 recorder at 96 kHz, 24 bit and a Sony ECM-MS907 microphone. The recordings took place in a quiet office at the University of Freiburg. Vowel length was then measured for all speakers and all tokens using Praat (Boersma & Weenink (2012); see also section 3.4.2). In total, 20,468 tokens from 86 speakers were evaluated, equaling 238 tokens for each speaker. Out of those 238 tokens, 70 are from the minimal pair reading task, 48 from the word list reading, 72 from the sentences, and 48 from the text reading task.

3.3 Perception Experiment

In the second part of the experiment, participants were asked to listen to a set of 48 sentences and decide which of two words they heard last. The sentences included minimal pair tokens in which every vowel and consonant combination appeared twice. Most of the sentences were relatively short and semantically empty, in the form of "I can say …", "There is no…" or "This is a …", as exemplified in the following example:

(1) I can say *lib*.
(2) I can say *lip*.

Three sentence pairs were designed to be longer with more semantic meaning in order to counter a fatigue effect which can set in when participants become bored with a repetitious task. An example of such a longer sentence pair is the following:

(3) Can you put it in the *back*?
(4) Can you put it in the *bag*?

The final consonants of the tokens in question were cut off using Praat (Boersma & Weenink (2012)) so that the only deciding factor between the minimal pairs was vowel length.[50] Since the consonant that follows is already triggered towards the end of the vowel by e.g. the vowel formants falling, this part of the vowel was cut off as well.[51] The vowel was then lengthened synthetically (again using Praat) in order to reach original length. The minimal pair tokens were inserted into the same frame, so that suprasegmentals such as intonation and pitch were identical. Before the tokens were accepted to be used in the experiment, they were played to a native speaker of British English and a native speaker of American English, respectively, in order to verify that they sounded natural even after partial synthetic lengthening.

The sentences were recorded by a male and female native speaker of either British or American English; participants were played half of the sentences spoken by the male native speaker, the other half spoken by a female native speaker. This was done in order to minimize the risk of the speaker's gender influencing the outcome of the experiment. All four native speakers were university-educated and used a standard accent. Like the participants in the study, they were between 20 and 30 years old. The two American speakers come from Nebraska and Virginia; the British speakers both grew up in London. Groups G-US and C-US listened to the sentences spoken by the two American native speakers; groups G-UK and C-UK were confronted with the sentences spoken by the two British native speakers.

It should be mentioned here that the prepausal condition (cf. section 2.2.3) of the tokens might not seem ideal due to its influence on vowel duration. However, this position was necessary in order to allow a clean cut-off of the final consonant. If another word had followed the token, listeners would have perceived the missing consonant much more clearly, which might have hindered them from making an unbiased decision. Also, since all tokens appeared in prepausal condition, the influence of this position on vowel duration was equal in all instances. Therefore, this methodology seemed to be the best suited to the purpose of the study.

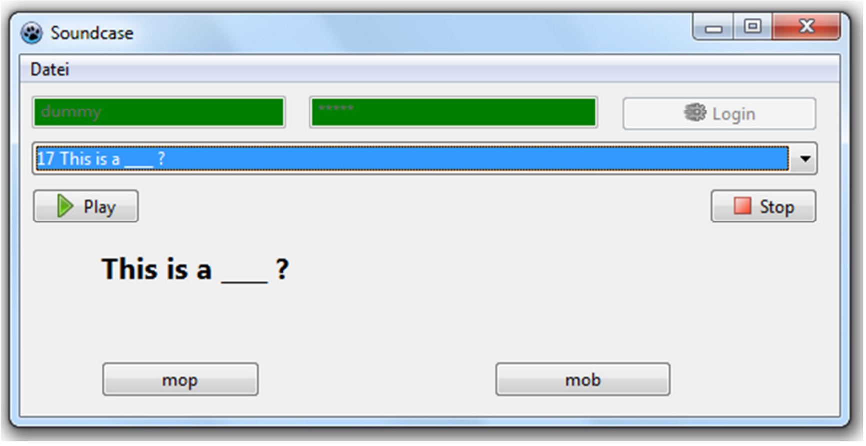
The perception experiment was run using software designed specifically for this purpose (*Soundcase*). The interface of Soundcase presents the subject with a play button and a stop button, and it displays the current sentence on the screen. Two buttons at the bottom show the two possible answer tokens (cf. Figure 3-1). The order of tokens was chosen so that two items

---

[50] Although cutting off the final consonant of a token might seem to render the tokens slightly unnatural, the same methodology has previously been successfully employed by e.g. Wardrip-Fruin (1982) to study the perception of vowel length variation.

[51] F1 offset of the vowel was found to be a cue to consonant voicing by Halle *et al.* (1957), Walsh & Parker (1983) and Hillenbrand *et al.* (1984). Cf. chapter 2.4.1.

of a pair did not occur directly in sequence. Every participant completed the experiment in the same order.



dummy

*****

17 This is a ___ ?

This is a ___ ?

Figure 3-1. Interface of the Soundcase software used in the perception experiment.

The participants received headphones and were told before the beginning of the experiment that they would be confronted with 48 sentences and would have to decide which of two words they heard last in the sentence. Participants were informed that they were allowed to listen to each token several times. However, once they had made their choice and moved on, they were not allowed to go back. This was necessary to ensure that in case of participants figuring out that they had to cue in to vowel length, they would not change their previous answers. Instructions were identical for all participants. Interestingly, about 90% of German participants remarked after the first token that it seemed “cut off prematurely,” while only one out of twenty native speakers commented on this.

Participants’ answers were automatically recorded by Soundcase and compared to the correct answers. Each participant’s incorrect answers were counted, leading to a final test score for each participant. The higher the score, the more mistakes a participant made; the highest score to be reached was of course 48, which equaled no correct answer. In addition, the total number of incorrect answers of all participants was calculated for each token in order to gain an understanding of which tokens were misheard most often.

3.4 Data

3.4.1 Questionnaire Data

Before the beginning of the experiment, participants filled out a questionnaire which asked them for personal information. The information gathered was of the following type:

a) Questions on background information such as gender, age, birthplace and nationality of the participants and the participants' parents
b) Questions related to their stay abroad (location, length, reason)
c) Questions concerning the participants' "history" with English (when they started learning English, which variety of English their teacher at school spoke, if they ever lived with or were in a relationship with a native speaker of English)
d) Questions designed to gain a deeper understanding of the participants' current use of English (e.g. "How often do you use English actively?" or "Do you use English mostly with native or non-native speakers of English?")
e) Questions related to the variety the participants preferred in order to ensure that they showed a clear preference for either one of the two varieties, had completed a stay abroad in the country of their choice and also listed this variety as the one they had most contact with through music or movies
f) Two questions in which participants had to indicate on a scale how important a native-like accent was to them and how they judged their own accent
g) Knowledge of other languages besides English and German
h) An empty box where participants were asked to enter any other information related to language learning (bilingual upbringing, possible impairments, etc.)

A copy of the questionnaire can be found in Appendix B.1.

Some of the data gained from the questionnaire proved not to be significant in the statistical analysis and will therefore not be mentioned in the results (section 4).

3.4.2 Measuring Vowel Length and Vowel Length Variation

In the present study, vowel length was measured using Praat (Boersma & Weenink (2012)). On a spectrogram, "a phoneme boundary is usually determined at a discontinuity in excitation, formant structure, or both" (Umeda 1975: 434).Vowels can be identified by their formants, i.e. vertical bars of a darker shade. The main task is to extract the vowel from the neighboring consonants.[52] Most researchers agree that the consonant closure and release of the following consonant do not belong to the vowel. Indeed, with a research question such as that of the present study, including the closure phase of the following consonant would be counterintuitive. Since the closure phase in voiceless consonants is longer than in voiced

---

[52] For an extensive discussion on segmentation and possible problems cf. Peterson (1955) and Peterson & Lehiste (1960).

ones, the results of the measurements would not be very descriptive of vowel length. However, should the consonant release of the preceding consonant be included? Many researchers include the release burst and aspiration in the measurement of the vowel's duration (e.g. Peterson & Lehiste (1960)), because it represents the transition from consonant closure to vowel. However, this is not the case when the consonant in question is a fricative. Fricative noise is never included in the measurement of vowel length, since friction turbulence it is only generated during the closure phase of the consonant. In order to be consistent and introduce as little bias as possible, friction noise, consonant release and aspiration of the preceding consonant were ignored in the measurements. This means that in the present study, vowel length was measured from the onset of energy in the higher vowel formants (F1, F2, F3, sometimes F4) to the sudden drop in intensity of said formants. Figure 3-2 can be considered as an example:

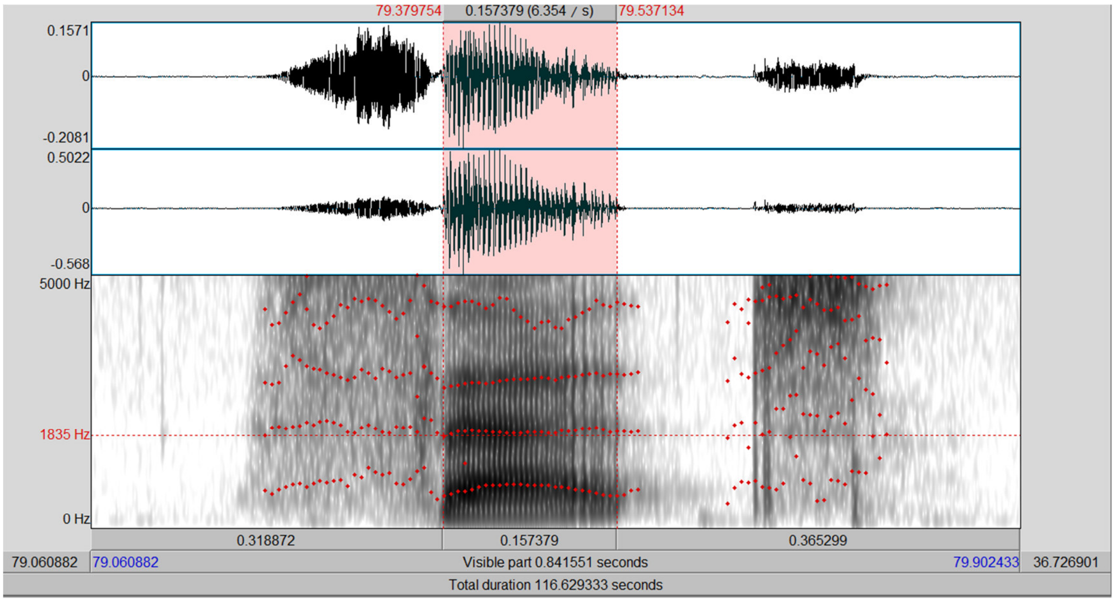


Figure 3-2. Vowel length of DRESS measured in the word *set*. Speaker #108, female native speaker of American English, word list reading.

The picture is a little more complicated when the preceding consonant is a liquid or a glide, since these groups of consonants ("semi-vowels") also show formants. Consider Figure 3-3 below. The word in question is *slob*, where the consonant preceding LOT is /l/. As can be seen from Figure 3-3, /l/ shows darker formant bars similar to the vowel's in F1 and F2, but has lighter formants than the vowel in F3 and F4. Thus, the sudden increase in intensity in F3 and F4 is where the vowel is easiest to separate from the preceding approximant. This is the selected starting point used to measure vowel duration in the present study.

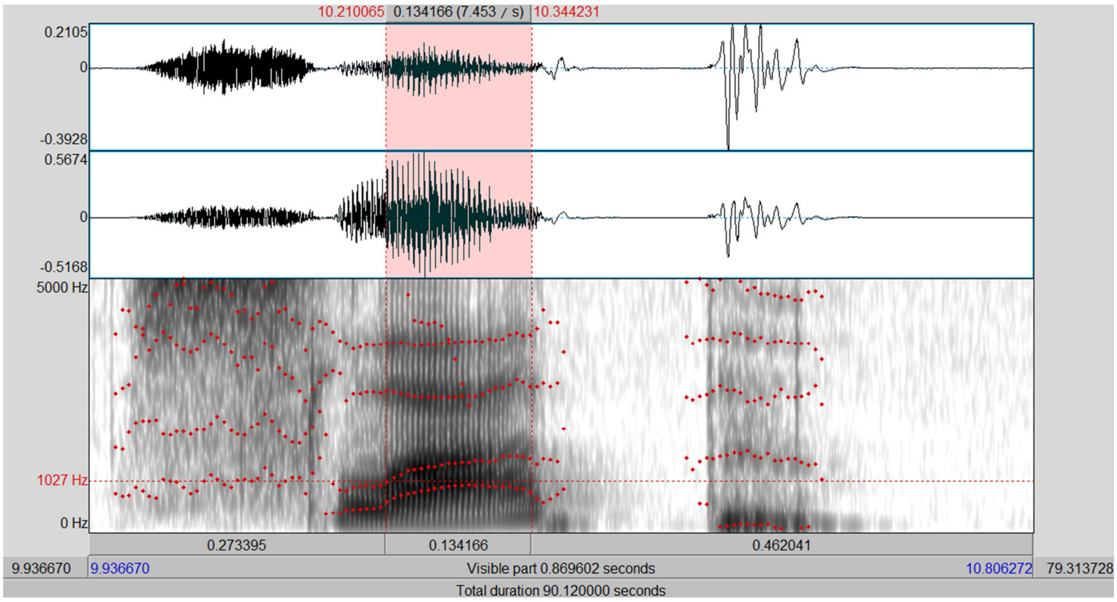


Figure 3-3. Vowel length of LOT measured in the word *slob*. Speaker #108, female native speaker of American English, minimal pair reading.

Once the length of all vowel tokens is measured, the next step is to calculate vowel length variation. In this dissertation, vowel length variation is expressed as a percentage. As explained in section 3.2, the vowel duration of a total of 20,468 tokens was measured. Words were paired together so that they were distinctive by their final consonants /b/ – /p/, /d/ – /t/ and /g/ – /k/, respectively. In the minimal pairs task, the words were of course minimal pairs, while this was not the case for the other tasks, in which the position of the word within the sentence was used to determine word pairs. For instance, in the text, *frog* and *shock* were such a pair for LOT. Both words occur sentence-finally and are therefore a good match.

Subsequently, the percentage deviation of the lengths of the paired words was calculated by first assuming that one vowel, say *A*, was the “truth” and comparing it to the other vowel (*B*). This yields the following mathematical formula:

$$\left[\left(\frac{A_{\text{Length}}}{B_{\text{Length}}}\right) - 1\right] \bullet 100$$

In turn, the deviation of vowel B from vowel A was calculated by swapping the lengths of vowels A and B in the formula above. The reason why this calculation was preferred over first calculating the mean value and then calculating how much each vowel deviated from the mean was that

a) the above formula allows for a more easily accessible way of referring to the results, e.g. “vowel A is 20% shorter than vowel B” instead of “vowel A deviates from the mean by 10%, while vowel B deviates from the mean by 20%”. This means that statements on the degree of variation each participant exhibits are possible, and

b) the statistics program R calculates the deviation from the mean individually before performing the statistical analyses used in this dissertation.

After having obtained the two deviation values for vowel A (A_σ) and vowel B (B_σ), the absolute values of the two deviations can be added together and divided by two in order to arrive at a final percentage score, which expresses the vowel length variation of this word pair:

$$\frac{(|A_\sigma| + |B_\sigma|)}{2}$$

Let us consider two examples of such a calculation to exemplify this further. Imagine vowel A is 144.385 ms long, while vowel B is 151.387 ms long. The deviation of vowel A from vowel B is

$$\left[\left(\frac{144.385}{151.387}\right) - 1\right] \bullet 100 = -4.625232\%$$

This means that vowel A is -4.625232% “longer” than vowel B. Swapping the values for A and B yields a result of +4.849534%, which is the percentage by which B is longer than A. Using the formula above we can calculate the variation percentage for the two vowels:

$$\frac{(|-4.625232| + |4.849534|)}{2} = 4.737383$$

A percentage variation of 4.74 is rather small, which indicates that the vowel lengths of these two example tokens do not vary much. Another example will illustrate greater variation. Let vowel C be 144.359 ms and vowel D 230.981 ms. The two deviation values are -37.50179 and 60.00457. This results in a final variation percentage of 48.75318, which is ten times greater than in the first example.

3.4.3 Measuring Merger of TRAP and DRESS

Part of this dissertation deals with the question of whether phonemic success is an indicator for success at acquiring subphonemic vowel length variation. German lacks the TRAP vowel, which leads to inexperienced learners' merging TRAP and DRESS under DRESS, which is the nearest available phoneme in the German inventory. Thus, the question arises whether those learners who have successfully acquired TRAP and keep it distinct from DRESS will be more successful at reproducing and perceiving the vowel length variation investigated within this study. It is therefore important to find a way of quantifying the degree of merging of TRAP and DRESS. To this end, Pillai scores were calculated for each speaker. The Pillai score was first used in linguistics by Hay *et al.* (2006) as a statistical method of determining the degree of distinctness of two sets of data. When it comes to mergers, Hay *et al.* argue that the Pillai score is "superior to taking Euclidean distances between means […] because it takes account of the degree of overlap of the entire distribution" (Hay *et al.* 2006: 467). The Pillai score is a value between 0 and 1, with 0 meaning that the two distributions fully overlap and 1 equaling two completely distinct distributions. In this experiment, the two distributions are the F1 and F2 values of DRESS and TRAP.

In order to measure the Pillai scores for the merger between TRAP and DRESS, eleven tokens were chosen from the word list and measured for each speaker. The word list was chosen because it is semi-formal, thus ensuring that participants pay attention to their pronunciation. If the merger persists at this semi-formal level, it is assumed to be present in the subject's speech. Also, participants read single words in the word list, which means that each word is articulated carefully and does not become subject to vowel reduction. The words were chosen randomly from the word list, but the same words were measured for each speaker. In order to have a visual representation of each speaker's vowel space, vowel plots were created for each speaker as well. Two plots and their respective Pillai scores will serve as an example here; a full list of all plots and Pillai scores can be found in Appendix D.

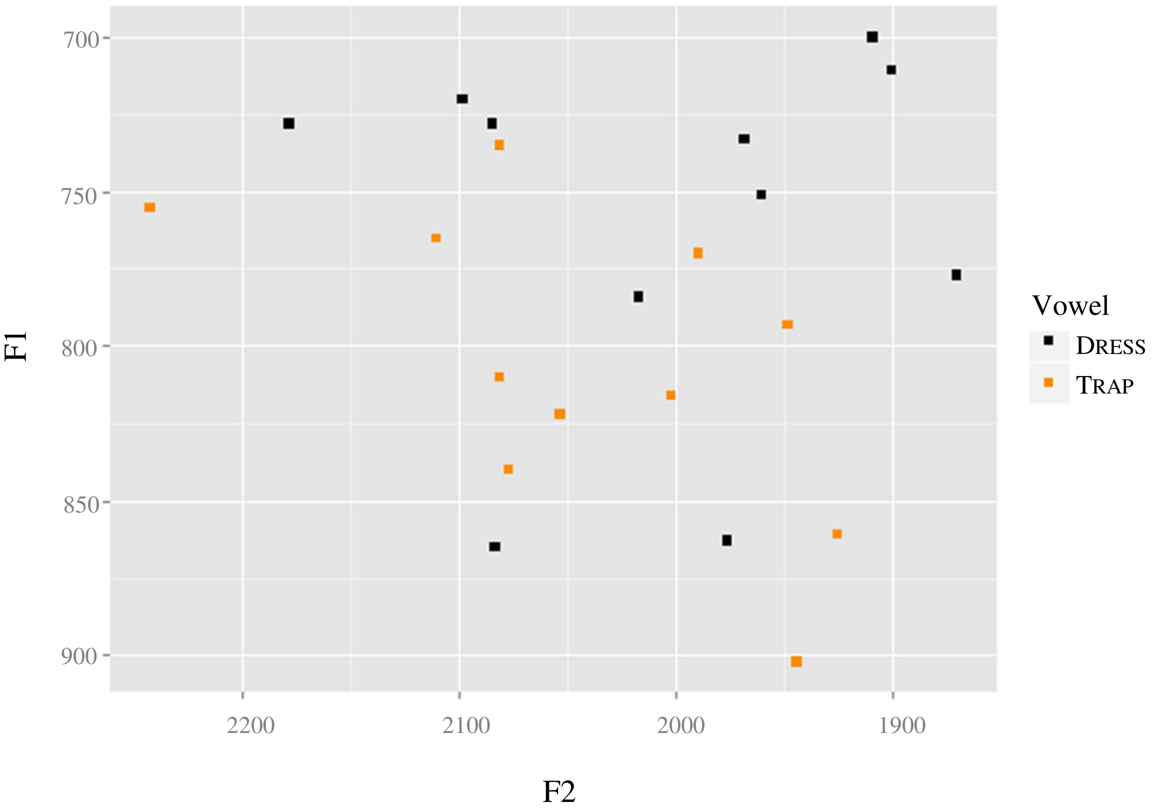


Figure 3-4. Speaker #025 with overlapping distributions of TRAP and DRESS, corresponding to a Pillai score of .23.

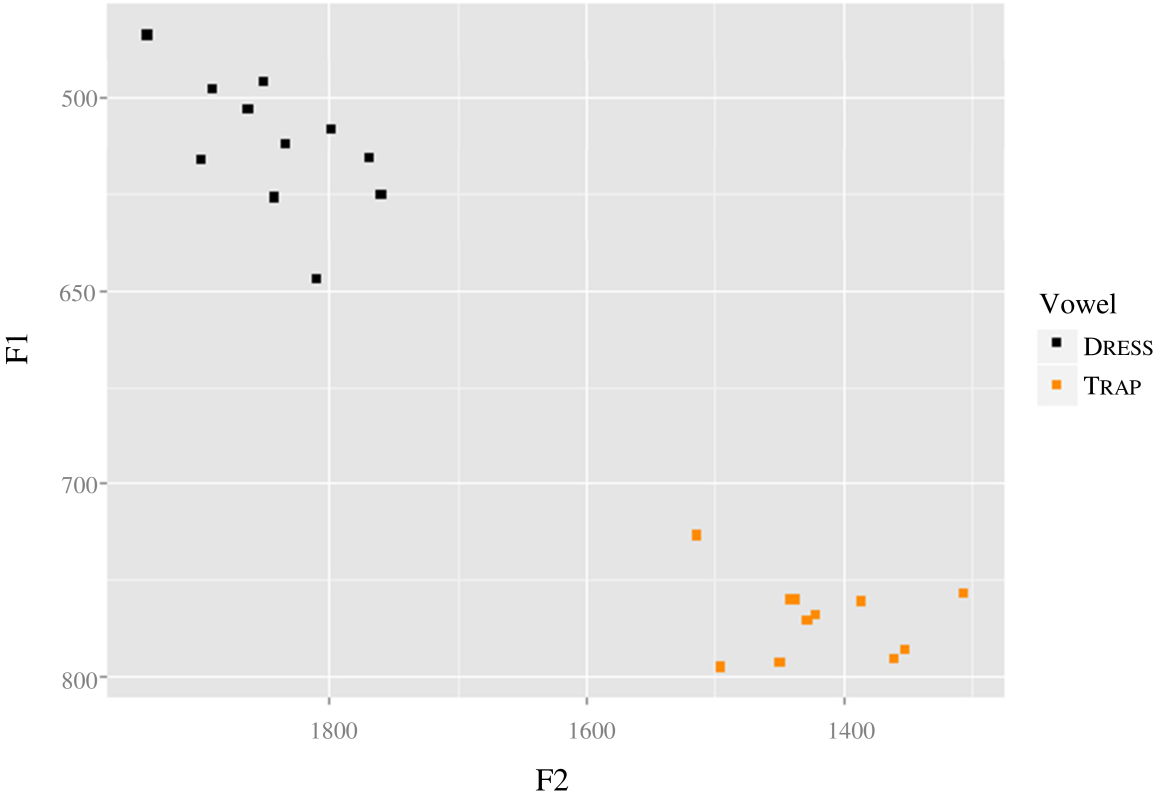


Figure 3-5.Vowel plot of speaker #029 with distinct distributions of TRAP and DRESS, corresponding to a Pillai score of .96.

3.4.4 Measuring Devoicing

Devoicing was also measured by using Pillai scores. In this case, the two distributions used to determine the score were the lengths of the closure phase and aspiration of the voiced and voiceless plosives, respectively. Both break and aspiration are longer in voiceless plosives. The data comes from the minimal pair reading. Since participants are likely to assume that the minimal pair reading focuses on final devoicing, they are supposed to pay close attention to their pronunciation. This means that if devoicing persists at this very formal level, subjects are assumed to exhibit final devoicing. 24 tokens were chosen from the minimal pairs list, one pair for each vowel + consonant combination. The same tokens were measured for each speaker; however, if a speaker did not release one or both of the final consonants in a pair, a different pair (with equal vowel + consonant combination) was chosen from the list.

After obtaining all the measurements, Pillai scores were calculated for each speaker. Again, a Pillai score closer to 1 suggests that the speaker releases voiced and voiceless plosives distinctly, while a score closer to 0 indicates that the two distributions overlap, which means that the speaker exhibits final devoicing. The plots and Pillai scores of each speaker can be found in Appendix E; two scatterplots will again serve as examples:

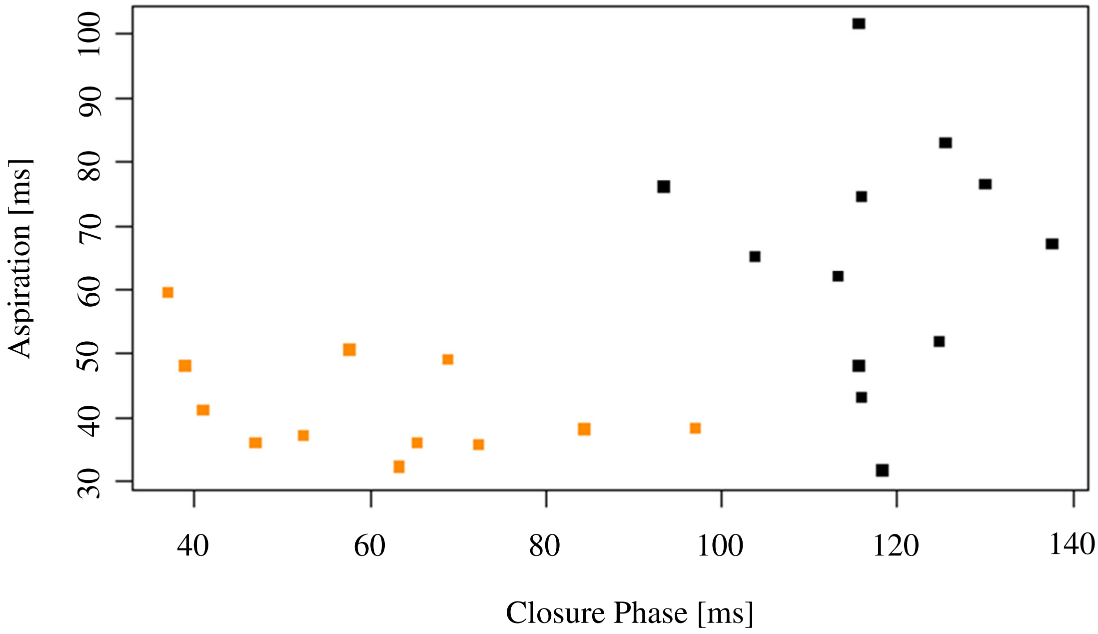


Figure 3-6. Scatterplot of distinct subject (#024), corresponding to a Pillai score of .83. Orange squares = voiced tokens, black squares = voiceless tokens.

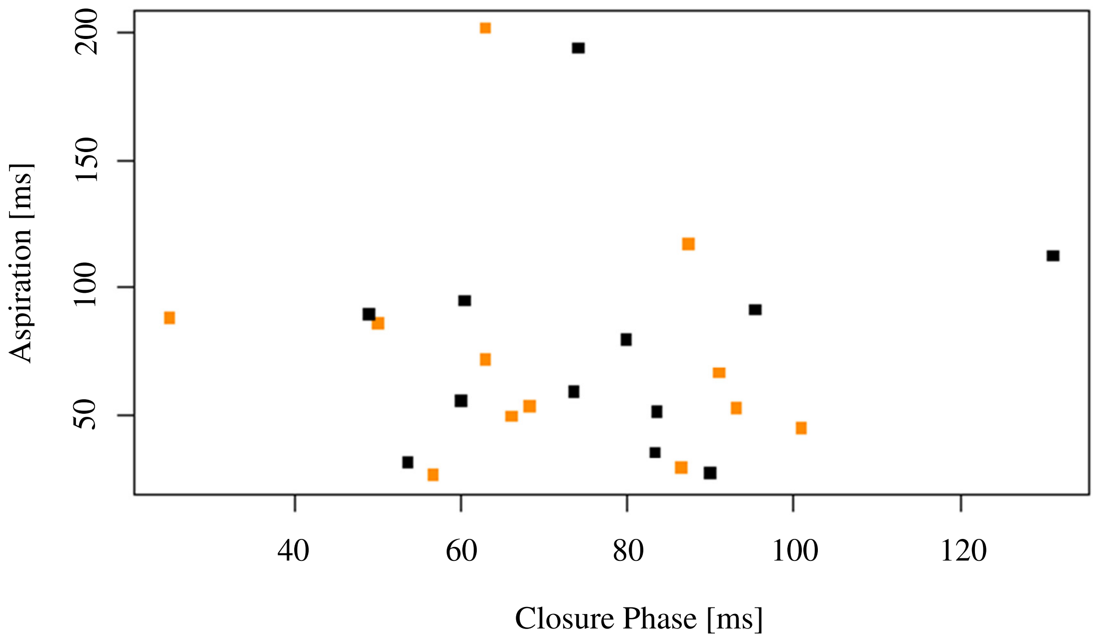


Figure 3-7. Scatterplot of merged subject (#025), corresponding to a Pillai score of .03. Orange squares = voiced tokens, black squares = voiceless tokens.

3.4.5 Measuring Word Frequency

Word frequency was measured in order to determine whether the vowel length variation in more frequent words was reproduced with higher accuracy by the German learners than the variation in less frequent words. This would suggest that learners acquire vowel length variation largely by mimicking native speakers, since words of higher frequency have a larger probability of being heard and used by learners. A second question is whether the less frequent words in the perception experiment will be heard correctly less often than the higher frequency words.

In order to determine the frequency of each word, two corpora were used. For the American group, the words in question were searched for in the *Corpus of Contemporary American English* (COCA). The *British National Corpus* (BNC) was used for the British group.

Two possible issues should be mentioned here. Firstly, the occurrences in the corpus were not examined in detail in order to see if all findings matched the grammatical form of the word in question (e.g. *lack* as verb or noun). In many instances in the experiment, the syntactic class of the word was ambiguous in any case. Moreover, word class is unlikely to be a major influence on perception and production as long as the phonological sequence is correct.[53] The only difference might be the frequency with which a word occurs in one syntactic class or another. Secondly, the sizes of the two corpora differ considerably: Whereas COCA is a corpus of roughly 450 million words the BNC only lists about 100 million words. Since speakers of the two varieties will not be compared with speakers of the respective other variety, however, this does not pose a problem.

Out of the 100 million words from the BNC, 10% are from spoken data. In the COCA, spoken data amounts to roughly 95.5 million words. Given that learners need to be confronted with native speakers' speech in order to pick up on vowel length variation, the best methodology would be to only take into account the frequency of the tokens in the spoken sample of the two corpora. Unfortunately, the BNC does not allow the separation of spoken data from written data, while this is possible in the COCA. In order to remain consistent in the methodology, the analysis therefore takes into account the occurrences of the tokens in the entire corpus.

---

[53] Note that Yeni-Komshian *et al.* (2001: 283 f.) found that grammatical category (verb vs. noun) *did* have an effect in how accurately words were pronounced. Korean-English bilinguals produced verbs more accurately than nouns, and they detected grammatical errors more easily in verb constructions than in noun ones. However, they argue that this is a reflection of the linguistic structure of Korean and can therefore not be transferred to the speakers of the present study, who are native Germans. The latter should show similar behavior to native speakers of English, because both languages are predominantly SVO languages (Yeni-Komshian *et al*. 2001: 293). In addition, as mentioned above, the question whether the words were verbs or nouns was left unanswered in the present study.

4. Results & Analysis

4.1 A Note on Statistics

Chapter 4 presents the results of the study, divided into the British group (participants in groups *British*, *G-UK* and *C-UK*) and the American group (participants in groups *Americans*, *G-US* and *C-US*). All statistical analyses were performed using the open-source statistics program R and are based on a confidence interval of 95%.

The results of the production experiments will be visualized using boxplots. Boxplots are ideal to visualize the data collected in the present study, because they can be used for any statistical distribution due to their non-parametric nature. In addition, generalized linear models (GLM) were employed to evaluate the relationship between a dependent and one or more explanatory variables. The strength of the relationship between the variables can be expressed using the coefficient of determination R^2. The value of R^2 lies between 0 and 1, with values near 0 expressing poor fit of the model, and values near 1 equaling an almost perfect fit (which makes strong predictions possible). One problem with R^2, however, is its high sensitivity to outliers (lack of robustness), which are most certainly present in the data collected for this study. This results in R^2 often having quite a low value. This is not surprising in the context of this study, which deals with humans, after all. The performance of participants – whether in the production or perception experiment – is rarely ever a result of just one factor. Rather, it is a combination of factors, some of which may be highly individual and hard to measure (e.g. internal factors such as motivation, IQ, talent, psychology of the learner, etc.). Therefore, R^2 might not adequately represent the strength of the association between the variables studied here.

A second issue with GLMs is that there is no standard test to assess the degree of interaction between the two variables. Parametric tests, such as correlation, yield Spearman's rho or Kendall's tau, but these tests are not adequate for models which work with non-parametric data. However, with the kind of data this study has gathered, generalized linear regression is the only suitable statistical test to work with. Baayen (2008) suggests using the measure *C* as an indication of goodness of fit of the model in generalized linear regression. C is "an index of concordance between the predicted probability and the observed response. […] When C takes the value 0.5, the predictions are random, when it is 1, prediction is perfect. A value above 0.8 indicates that the model may have some real predictive capacity" (2008: 223).

The graphic plots which are based on (generalized) linear models all include a regression line. This is a non-parametric locally weighted scatterplot smoothing curve as proposed by Cleveland (1979). “Robust locally weighted regression is a method for smoothing a scatterplot […] in which the fitted value […] is the value of a polynomial fit to the data using weighted least squares” (Cleveland 1979: 829). As such, the regression line exemplifies the ideal curve for the data points.

The Wilcoxon rank-sum test (also often called *Mann-Whitney U test*) assesses whether two samples of independent values come from the same data set. More than two samples of data require the Kruskal-Wallis one-way analysis of variance. This study uses the Wilcoxon test e.g. in the analysis of the production experiment in order to assess whether the production of vowel length variation in the four different vowels and tasks differs significantly between the native speakers and the German learners. The Kruskal-Wallis test is employed when the production of vowel length variation of one group before the three different voiced and unvoiced consonants is studied.

The chi² test is a statistical test of significance used when data comes in the form of contingency tables, i.e. when it is of a categorical nature. This is the case when the role of the female or male speaker, the vowel or the consonant in the perception experiment is examined. For smaller sets of data, Fisher’s exact test yields more stable statistics. Since it is debatable what *smaller* means (fewer than 10 values in each cell of a 2 x 2 table has been proposed), both tests were employed in a pilot analysis. The results from the Chi² test and Fisher’s exact test did not vary more than .001; therefore, the Chi² test was established as a safe method to use. The strength of the association between the two binary variables can be expressed by the phi coefficient (φ). The coefficient ranges from -1 to +1, where -1 and +1 indicate complete (dis)agreement and 0 expresses that there is no relationship between the variables. The effect size can be categorized threefold:

Table 4-1. Effect size of the phi coefficient according to Cohen (1992: 156 f.).

Effect size	φ
small	± .1
medium	± .3
large	± .5

Statistical notations such as *sample mean*, *probability value, standard deviation* and *standard error* are reported so frequently that they are assumed to be known to the reader. It is important to note, however, that whenever this dissertation speaks of the sample mean ($\bar{x}$), it refers to the *arithmetical mean*. When there is a large probability that the mean may be influenced by outliers, the more robust median will be stated. There is no fixed statistical notation for the median; this study uses $\tilde{x}$. The significance level used in this dissertation is .05, the confidence interval which was chosen is 95%.

The remainder of this chapter is structured as follows:

Section 4.2 will report on the production experiment. It will be shown to what extent the German learners can match their native speaker model in the production of minimal pairs, words from a word list as well as in sentence and text reading. The results of the perception experiment will be discussed in section 4.3. We will then examine whether and in which way perception and production are linked (section 4.4). Afterwards, we will deal with the question of whether the acquisition of vowel length variation is influenced by phonological proficiency (4.5). The last section, 4.6, is designed as a summary and overview of the findings of this chapter.

4.2 Production Experiment

This section reports on the results gained from the production experiment. The questions to be answered here are: Which learners can successfully reproduce vowel length variation, and which factors lead to some participants' being able to match their native speaker model more closely than others?

The data gathered from the production experiment (i.e. the length of the individual vowels) is Gaussian-distributed; therefore, linear models were employed to perform the statistical analyses. The results of the analysis will be illustrated using boxplots. The gender of the participants ($p = .91$),[54] the emphasis they place on a native-like accent ($p = .19$) and language use (how often they use English actively or passively) was not found to be statistically significant and will therefore be ignored in the analysis ($p = .37$ and $p = .56$, respectively).[55]

The two sub-chapters which follow will report on the results from the British and the American group. Afterwards, a more generalized section will investigate the production of vowel length variation before the three different consonant pairs /b, p/, /d, t/ and /g, k/. A comparison of the performance of the German learners in group G-UK and G-US will show differences and similarities in the production of vowel length variation regarding the four vowels and three consonant pairs. Last, the effect of other factors such as the length of and the reason for the stay abroad, word frequency and accent self-rating will be analyzed.

4.2.1 British Group

4.2.1.1 Vowel KIT

The following boxplots show a comparison of the vowel length variation exhibited by the British native speakers, the German learners (G-UK) and the control group (C-UK) in the different reading tasks. KIT, DRESS, TRAP and LOT will be considered in this order. For each vowel, we will examine vowel length variation in the minimal pairs, the word list, the sentences and the text.

---

[54] Many other L2 accent studies also did not find a significant effect of gender; cf. Suter (1976), Purcell & Suter (1980), Flege & Fletcher (1992), Elliott (1995), Piske *et al.* (2001). Flege (1995) suggests that gender effects might be related to AOL and L2 experience.

[55] Suter (1976) and Purcell & Suter (1980) did not find a statistical effect of amount of conversation in English at work or at school in their accent studies. Flege & Fletcher (1992), Thompson (1991) and Elliot (1995) likewise did not find that daily use of the L2 influenced accent in a significant manner. Flege *et al.* (1997; 1999) and Piske *et al.* (2001) suggest that the accent in the L2 improves when the L1 is only spoken rarely. Since the participants in the present study have all learned English as a foreign language and not in an immersion setting, it is not surprising that this study comes to the conclusion that language use is not a significant factor.

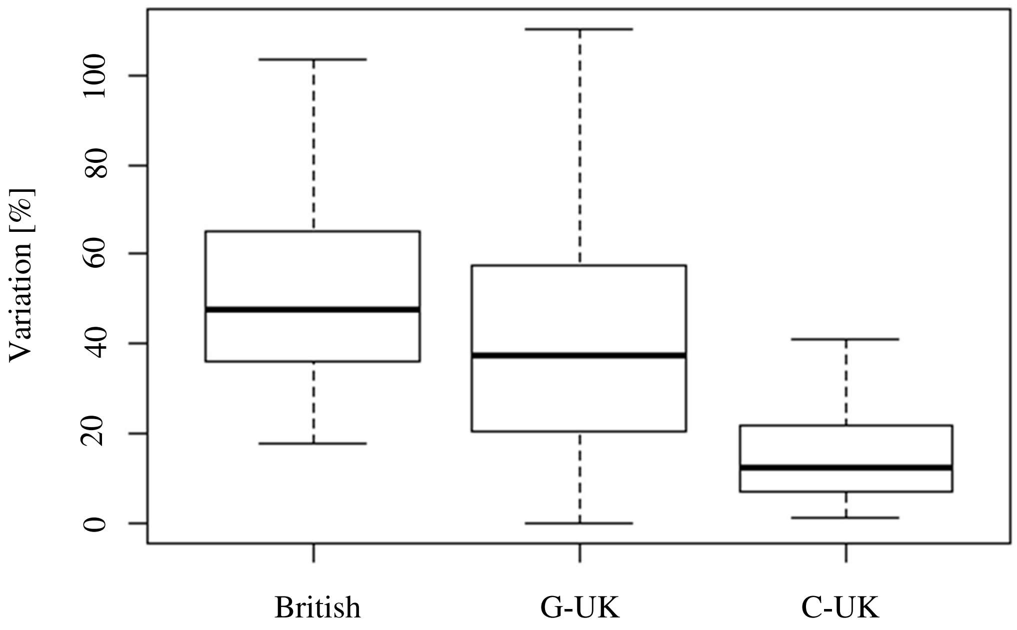


Figure 4-1. Boxplots of the vowel length variation in KIT, British group, minimal pairs reading.

Group G-UK matches the native speaker model quite closely (*p* = .06) in the minimal pairs reading, even though the median $\tilde{x}$ lies slightly lower and there is larger spread in the data. The learner group C-UK, however, displays only little variation and is significantly different from both groups British (*p* = .007) and G-UK (*p* = .01). This first result already shows that exposure to the target language in a native setting has a strong influence on the acquisition of vowel length variation. The speakers in group G-UK, who have been abroad, show more native-like variation than the control group, whose speakers have not spent any time in an English-speaking country. This stark contrast shows with a control group as small as three subjects.

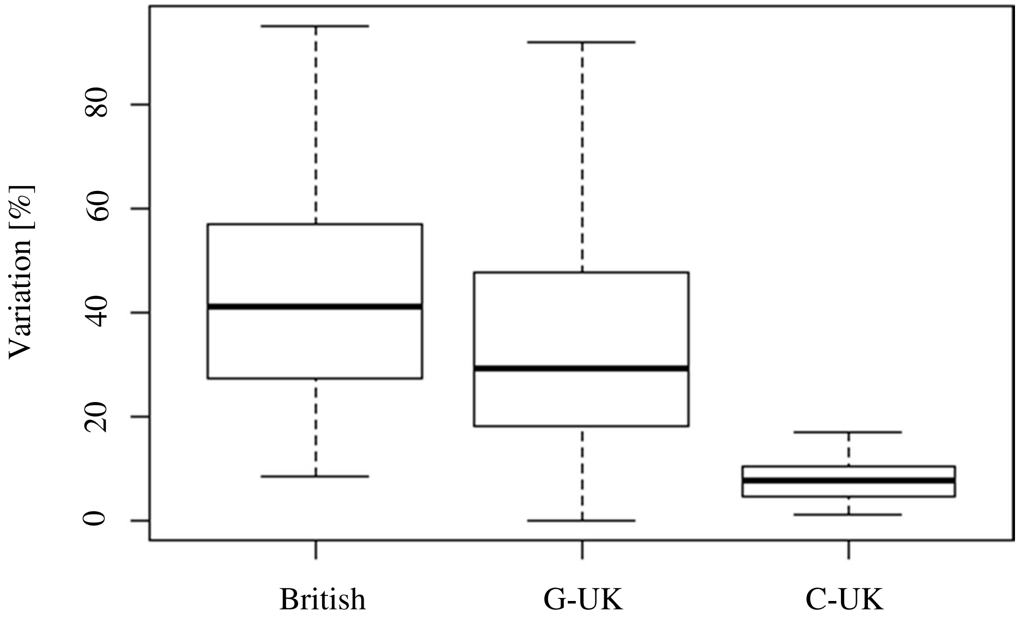


Figure 4-2. Boxplots of the vowel length variation in Kɪᴛ, British group, word list reading.

In the word list reading, group G-UK matches the British native speakers closely again ($p = .11$). The control group shows hardly any variation and significantly differs from the other two groups (British $p = .007$, G-UK $p = .0004$). This sudden lack of variation in the control group might stem from the fact that the word list is only semi-formal, which might have led speakers in this group to pay less attention to their pronunciation. In addition, all 86 speakers show significantly less variation in the word list than in the minimal pairs ($p < .0000000001$).[56] Since the control group is less experienced than the speakers in group G-UK, this might have affected them more.

---

[56] The levels of variation exhibited in the text reading are also significantly lower than in the minimal pairs ($p = .02$). The levels of variation in the text reading are higher than in the word list reading, however.

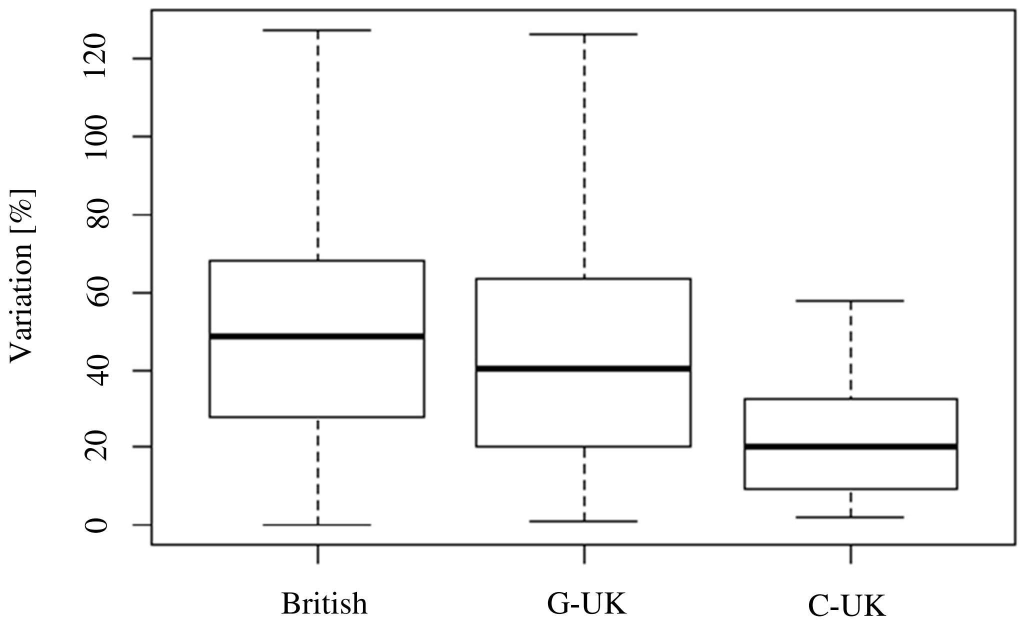


Figure 4-3. Boxplots of the vowel length variation in Kɪᴛ, British group, sentence reading.

Group G-UK is not significantly different from the native speakers in the sentence reading task (*p* = .30). The two boxplots are quite similar in their extremes and the median $\tilde{x}$. In contrast, group C-UK differs significantly from both the native speakers (*p* = .007) and group G-UK (*p* = .002, respectively).

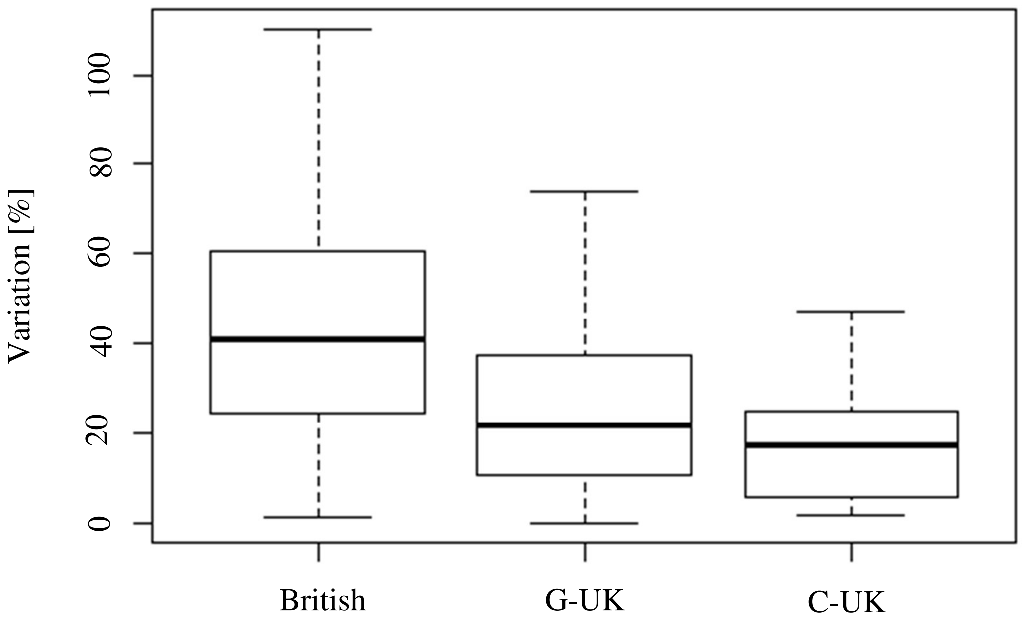


Figure 4-4. Boxplots of the vowel length variation in Kɪᴛ, British group, text reading.

In the text reading task, both German groups are significantly different from the British model (G-UK *p* = .02; C-UK *p* = .03). However, they do not differ significantly from each other

($p = .12$). This is an interesting finding. The level of variation exhibited by all speakers in the text reading is significantly lower than the variation in the minimal pairs reading ($p = .02$). We saw before that this is also the case for the word list reading, where only the control group C-UK seemed to be affected. The text reading is the most informal task, in which all speakers are assumed to focus more on content than pronunciation. Therefore, it is quite possible that even the more advanced learners of group G-UK were affected in this task and show lower variation values than before.

One might argue that the lower variation values come from the fact that participants read more quickly in the text reading than in the minimal pairs, which would lead to less variation. While this is certainly true, it does not explain why the reading of the word list rather than the sentences shows significantly lower levels of variation. If variation was this strongly influenced by speech rate, the values should decrease linearly along the following route: Minimal pairs → Word list → Sentences → Text. However, they do not.

Table 4-2 below gives an overview of the mean percentage of vowel length variation in KIT in the four different production tasks.[57]

Table 4-2. Mean percentage of variation of the British group in four different production tasks. Vowel KIT. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
British	56.12	47.67	57.57	48.48
sd	(31.15)	(30.40)	(45.04)	(35.66)
G-UK	47.57	36.39	49.33	31.48
sd	(51.65)	(29.97)	(39.86)	(42.71)
C-UK	15.84	8.33	24.45	19.01
sd	(12.34)	(5.0)	(19.32)	(14.66)

It is quite clear from the table that the British subjects of the study are the model toward which group G-UK orientates itself. In the first three tasks, group G-UK displays (on average) 10% less variation than the native speakers. The text reading task seems to be the most difficult one, since the difference between the native speaker model and group G-UK amounts to 18% in this task. Group C-UK does not manage to match either the native speakers or

---

[57] Note that *sd* is quite large because the table includes measurements before voiced as well as voiceless consonants.

group G-UK. This difference is particularly evident in the word list reading, where group C-UK only produces about one fifth of the variation exhibited by the native speakers.

Two things have become clear regarding the British group and the vowel KIT: Group G-UK shows more progress in matching their variation to the model, while group C-UK shows much less variation in all of the tasks. A second factor is that group G-UK has quite a large spread in the data. Some speakers even outperform the British model in the minimal pairs task (cf. Figure 4-1). This suggests that this group is quite inhomogeneous, as can be expected from a group of learners.

4.2.1.2 Vowel DRESS

This section deals with the vowel DRESS. As before, the boxplots show a comparison of the vowel length variation exhibited by the British native speakers, the German learners (G-UK) and the control group (C-UK) in the different reading tasks.

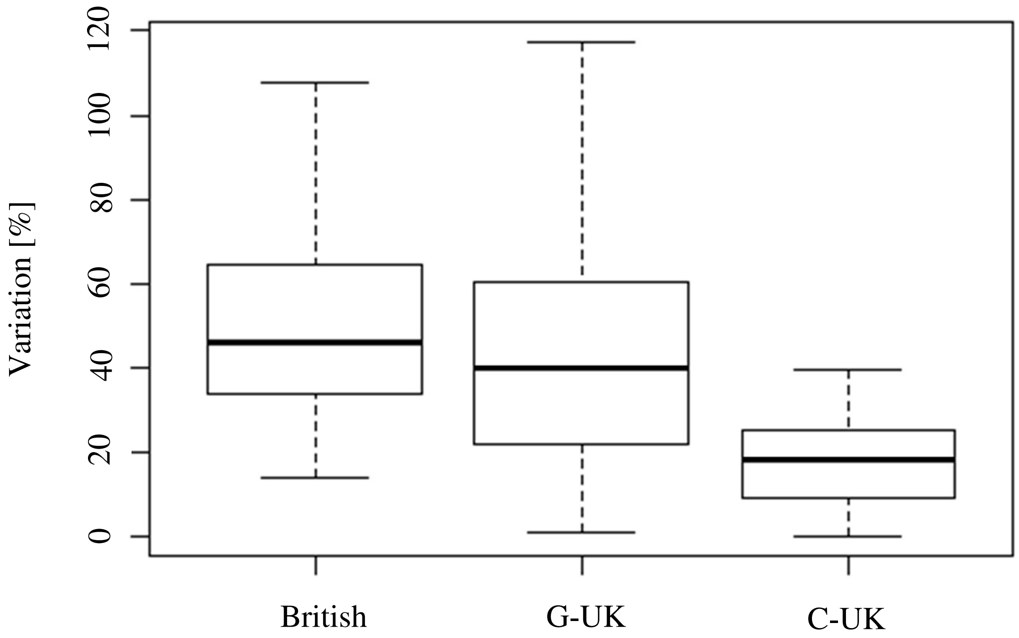


Figure 4-5. Boxplots of the vowel length variation in DRESS, British group, minimal pairs reading.

Table 4-5 shows quite clearly that group G-UK matches the British native speakers well, with some speakers outperforming the model. Thus, it comes as no surprise that the difference is not statistically significant (*p* = .15). In contrast, the control group C-UK is statistically different from both the native speakers (*p* = .007) and the German learners (*p* = .004).

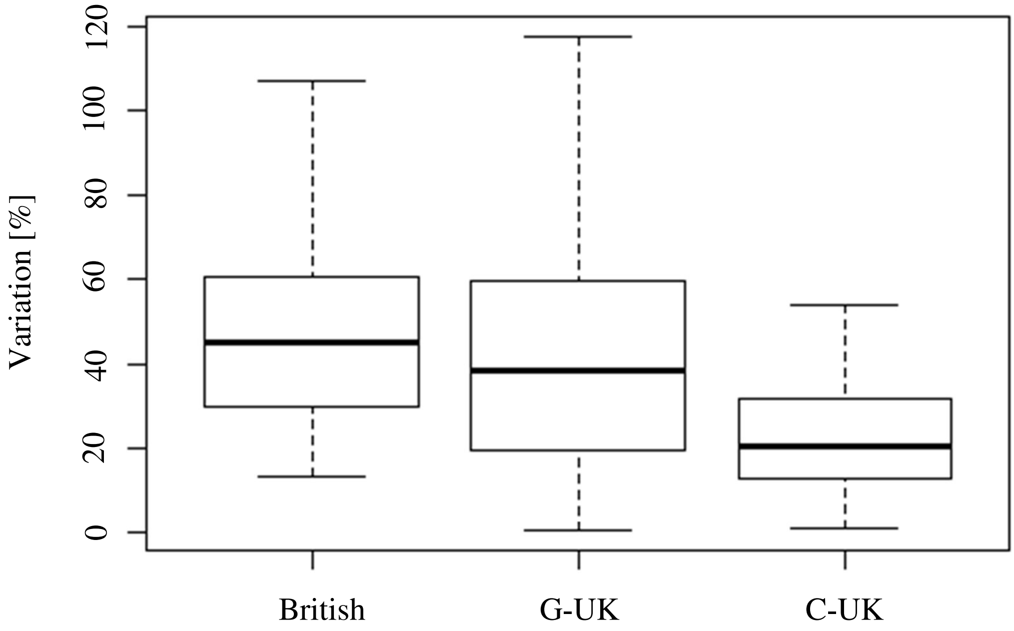


Figure 4-6. Boxplots of the vowel length variation in DRESS, British group, word list reading.

The word list shows an almost identical picture to the minimal pairs reading. Group G-UK behaves very similarly to the native speakers (*p* = .10), whereas group C-UK is significantly different from both the British group (*p* = .007) and group G-UK (*p* = .04).

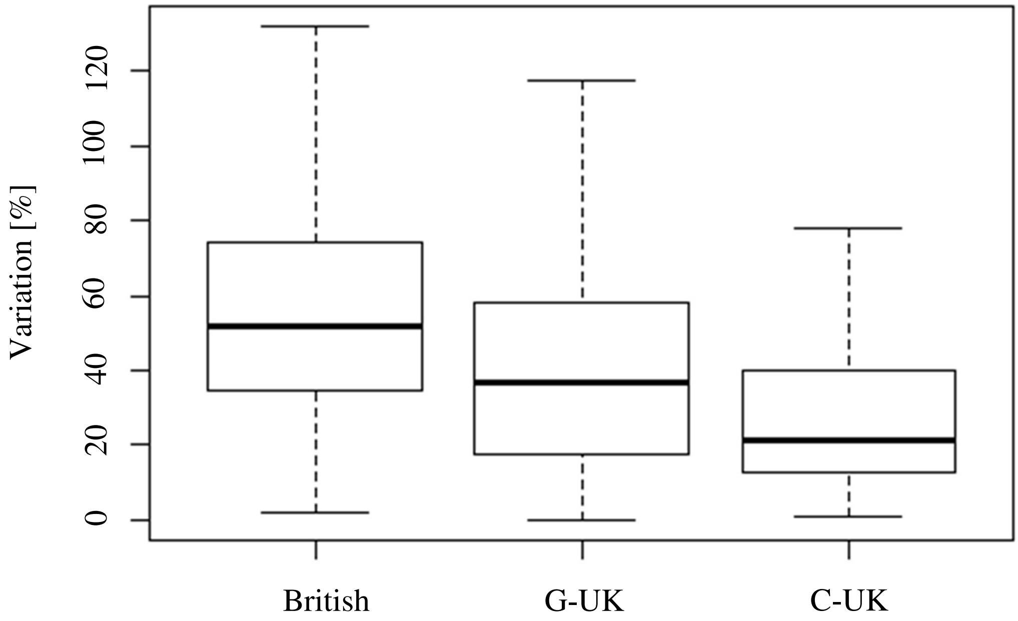


Figure 4-7. Boxplots of the vowel length variation in DRESS, British group, sentence reading.

The sentence reading task is the first task for DRESS in which the variation exhibited by both German groups significantly differs from the British native speakers (G-UK *p* = .05, C-UK *p* = .007). The control group is also significantly different from group G-UK (*p* = .01).

Moreover, this is the only task for DRESS in which group C-UK shows notable vowel length variation, with the upper extreme at about 80%.

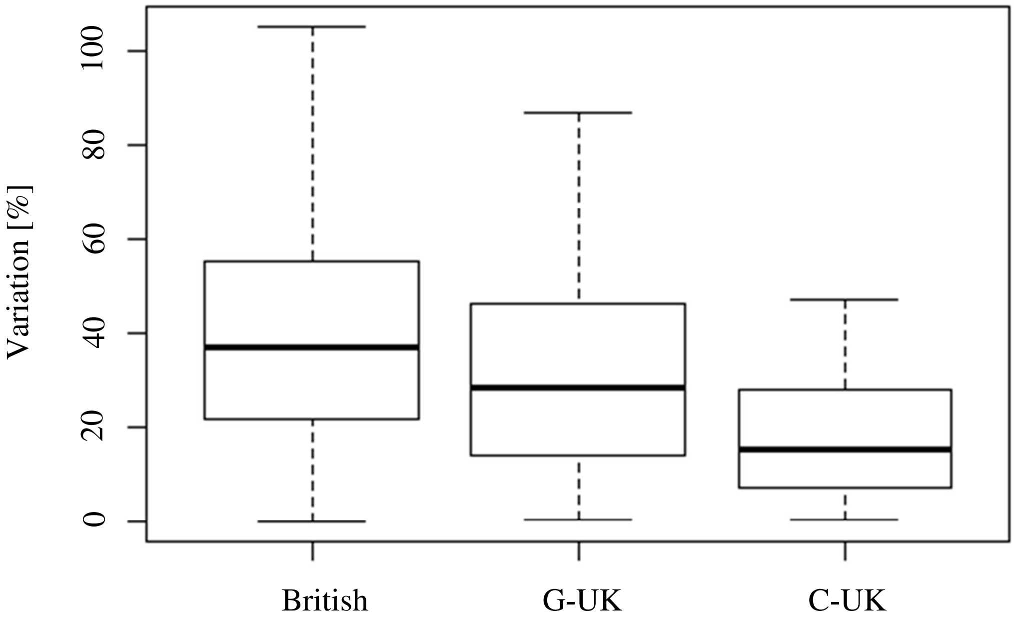


Figure 4-8. Boxplots of the vowel length variation in DRESS, British group, text reading.

In the text reading, group G-UK behaves differently to the British group again (*p* = .04). The control group differs significantly from the native speakers (*p* = .02) and group G-UK (*p* = .006). It is quite striking that the variation of about 80% exhibited by group C-UK in the previous task decreases again, almost by half, in the text reading.

A summary of the mean percentage of variation in DRESS in the four different production tasks is given in Table 4-3 below.

Table 4-3. Mean percentage of variation of the British group in four different production tasks. Vowel DRESS. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
British	54.81	54.34	66.19	42.84
sd	(31.23)	(34.69)	(60.48)	(31.65)
G-UK	49.96	47.12	47.40	36.06
sd	(43.26)	(41.30)	(47.41)	(34.31)
C-UK	18.06	25.74	30.76	20.21
sd	(11.86)	(21.79)	(31.89)	(19.03)

DRESS is one of the most interesting vowels in the British group. As with KIT, group G-UK assumes a medium position between the British model and the control group. There is again large spread in the G-UK, with some German subjects outperforming the model in the minimal pairs and the word list. Interestingly, group G-UK is not significantly different from the British model in two of the four tasks: Minimal pairs (*p* = .15) and word list (*p* = .10). They are highly significantly different in the sentence reading (*p* = .05) and the text reading task (*p* = .04). This means that the learner group can match the British model only in the formal tasks. With increasing informality of the tasks, however, group G-UK pays more attention to the content and shows less vowel length variation than the native speakers. Group C-UK shows more variation than before in all tasks. Still, the speakers of the control group only reach about half the variation of the native speakers, again showing that they lack input generated through interaction with native speakers during a stay abroad.

4.2.1.3 Vowel TRAP

The following boxplots show a comparison of the vowel length variation in TRAP again exhibited by the British native speakers, the German learners (G-UK) and the control group (C-UK) in the four different reading tasks.

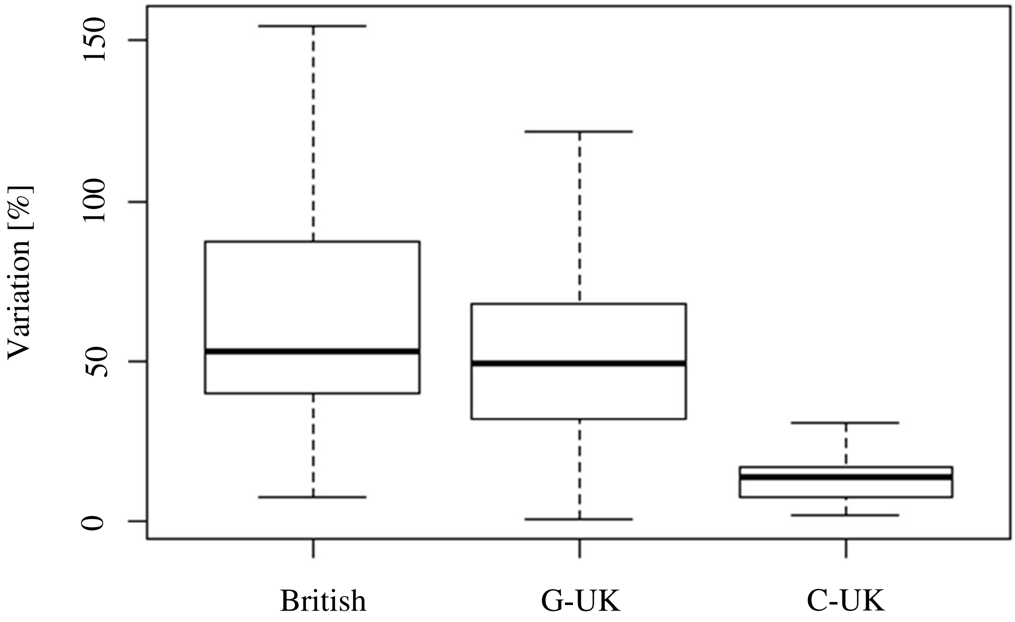


Figure 4-9. Boxplots of the vowel length variation in TRAP, British group, minimal pairs reading.

It is quite clear from the plot that group G-UK behaves similarly to the British native speakers (*p* = .17), since the medians of both groups are almost identical. Group C-UK, however, does not exhibit much variation to speak of and is significantly different from both the native speakers (*p* = .007) and the German learners with experience abroad (*p* = .0004).

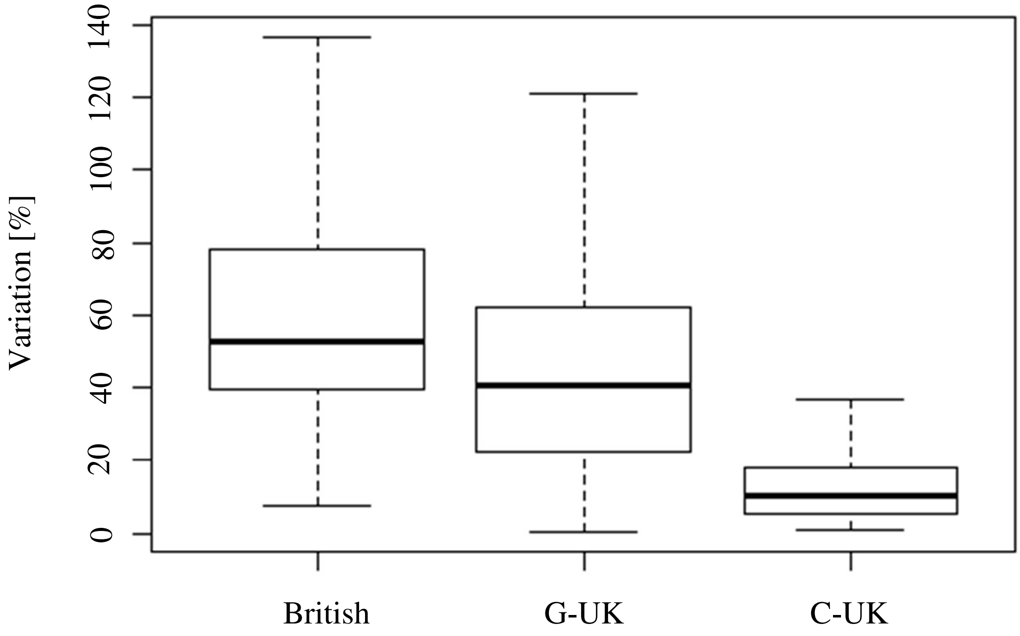


Figure 4-10. Boxplots of the vowel length variation in TRAP, British group, word list reading.

In the word list, the control group C-UK shows nearly the same low level of variation as it did in the minimal pair reading task before. Therefore, it is not surprising that group C-UK differs significantly from both the British native speakers (*p* = .01) and group G-UK (*p* = .04) again. The German learners with experience abroad, group G-UK, match their native speaker model more closely (*p* = .16). However, group G-UK's minimum is the lowest of all three groups.

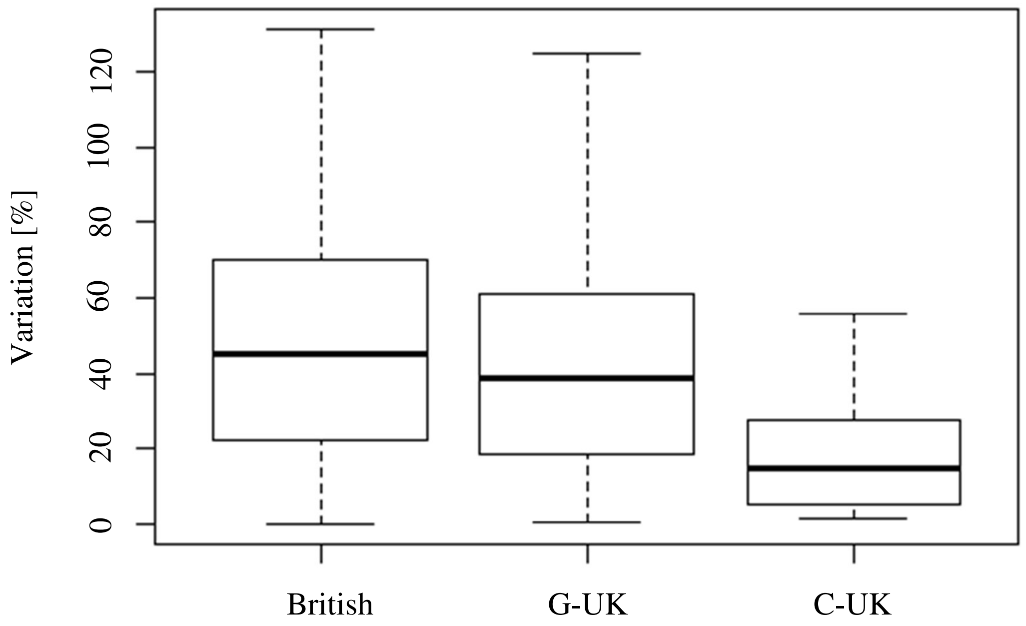


Figure 4-11. Boxplots of the vowel length variation in TRAP, British group, sentence reading.

In the sentence reading task, group G-UK matches the British native speakers closely again (*p* = .22). Although the control group's variation has increased, it is still significantly different from the native speakers' (*p* = .008) and that of group G-UK (*p* = .01).

TRAP is the first vowel in which the German learner group G-UK matches the native speakers not only in the two formal tasks (minimal pair and word list reading), but also in the semi-informal sentence reading. This suggests that the high level of variation exhibited by the native speakers in TRAP (cf. Table 4-4), which is the greatest vowel length variation of all four vowels, makes it easier for learners to perceive and reproduce. Since TRAP is also the lowest vowel of the four in the British vowel space, it is intrinsically longer than the other vowels, which makes vowel length variation in this vowel particularly salient. Interestingly, this seems not to be the case with the control group, however, whose members show a much higher degree of variation in DRESS than in TRAP.

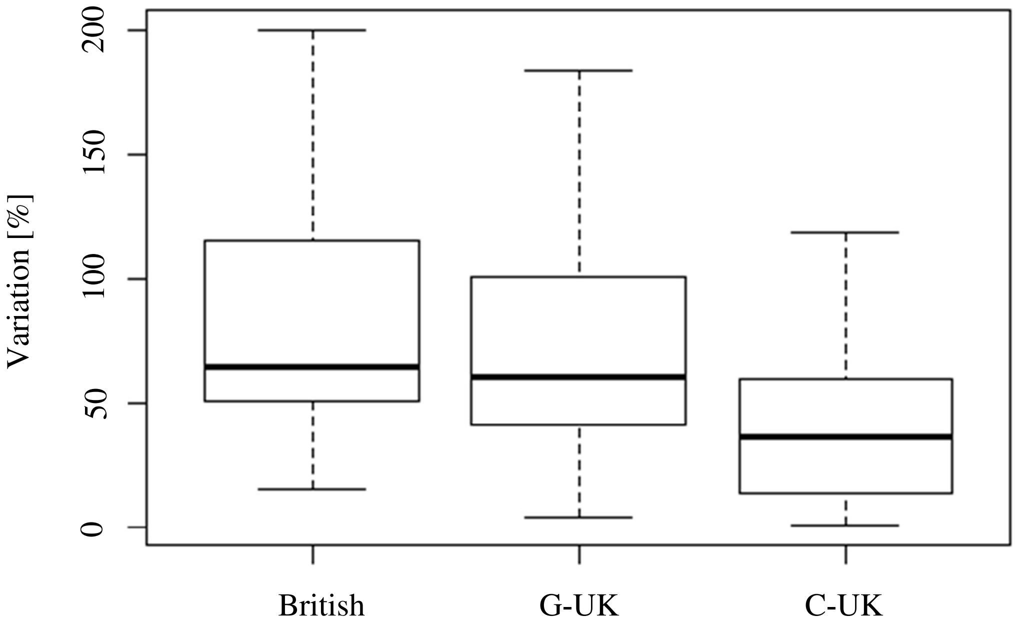


Figure 4-12. Boxplots of the vowel length variation in TRAP, British group, text reading.

All groups show the greatest variation in the text reading. Group G-UK behaves very similarly to the British native speakers (*p* = .29), as in all tasks before. Thus, the hypothesis proposed above seems to hold true. The control group is significantly different from both the native speakers (*p* = .007) and group G-UK (*p* = .001).

Table 4-4 summarizes the mean percentage of variation in TRAP in the four different production tasks.

Table 4-4. Mean percentage of variation of the British group in four different production tasks. Vowel TRAP. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
British	65.37	65.22	57.48	89.47
sd	(41.24)	(40.10)	(53.63)	(59.39)
G-UK	60.13	50.64	48.23	80.69
sd	(46.14)	(44.27)	(45.55)	(67.49)
C-UK	14.55	15.87	22.62	45.28
sd	(8.81)	(17.07)	(23.78)	(41.14)

TRAP is the vowel where group G-UK matches the British model group most closely, possibly because the native speakers show the greatest variation in this vowel. Therefore, it is more easily perceptible by learners. This hypothesis is supported by the fact that group C-UK with no experience abroad also shows considerable variation in comparison with other vowels, except in the minimal pairs. It thus appears that the variation in this vowel is most easily matched by learners. The text reading tasks seems most interesting, because all groups show the highest degree of variation in this task, even though it is the most informal one. This cannot be explained by the position of the tokens in the text, either, since only two out of the six TRAP tokens appear sentence-finally, which might have an additional lengthening effect. However, it is quite possible that the two sentence-final tokens were pronounced very carefully due to the semantic load they carry in the story:

1) "Oh, that's too **bad**. I don't like frogs."
2) "I think you just like to **brag**."

Since many participants read the story quite animatedly and gave life to the characters, they might have articulated TRAP in these instances overly long. It is still hard to imagine, however, that two out of six instances of a vowel could have such a strong influence.

4.2.1.4 Vowel LOT

The final vowel to be discussed for the British group is LOT. The boxplots will again present the vowel length variation exhibited by the British native speakers, the German learners (G-UK) and the control group (C-UK) in the different reading tasks.

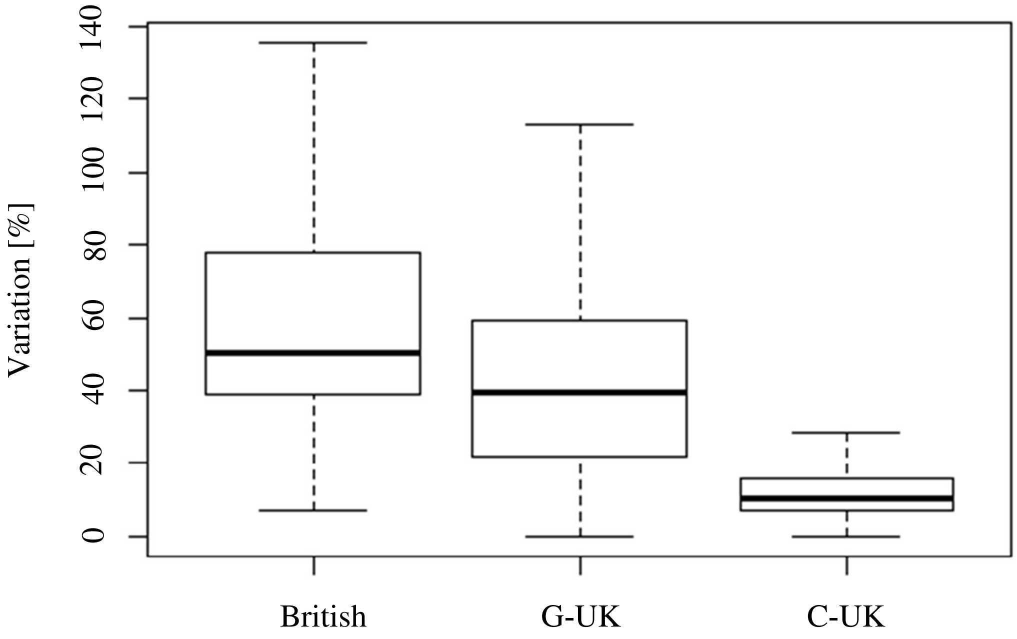


Figure 4-13. Boxplots of the vowel length variation in LOT, British group, minimal pairs reading.

The British native speakers show the highest degree of variation in the minimal pairs reading, even though the minimum value lies below 10%, which is very low for native speakers. This leads to the fact that the minima of all groups lie very close together. The other values of groups G-UK and C-UK (medians, quartiles, maxima), however, are very different from those of the native speakers. Therefore, it comes as no surprise that group G-UK differs significantly from the native speakers (*p* = .02) and the control group (*p* = .05). Group C-UK only exhibits little variation overall and therefore cannot match the British model (*p* = .01).

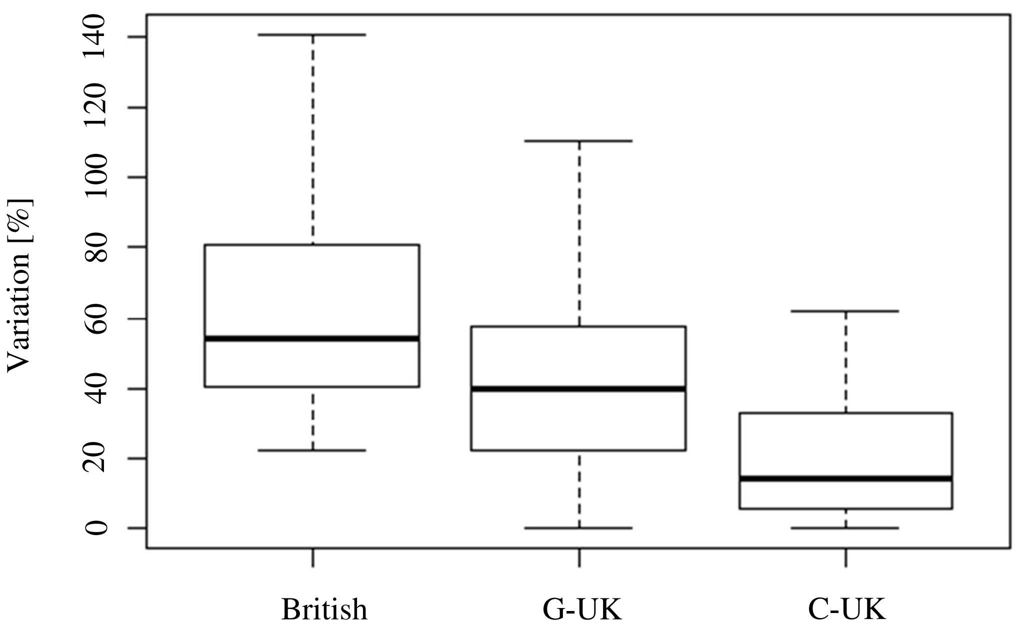


Figure 4-14. Boxplots of the vowel length variation in Lot, British group, word list reading.

The minimum variation of the British native speakers in the wordlist is greater than in the minimal pairs and exceeds the minima exhibited by groups G-UK and C-UK. Interestingly, group C-UK shows much more variation in this semi-formal task than in the minimal pairs. This leads to the two German groups performing similarly (*p* = .21). Both groups G-UK and C-UK are significantly different from the British model (*p* = .03 and *p* = .003, respectively).

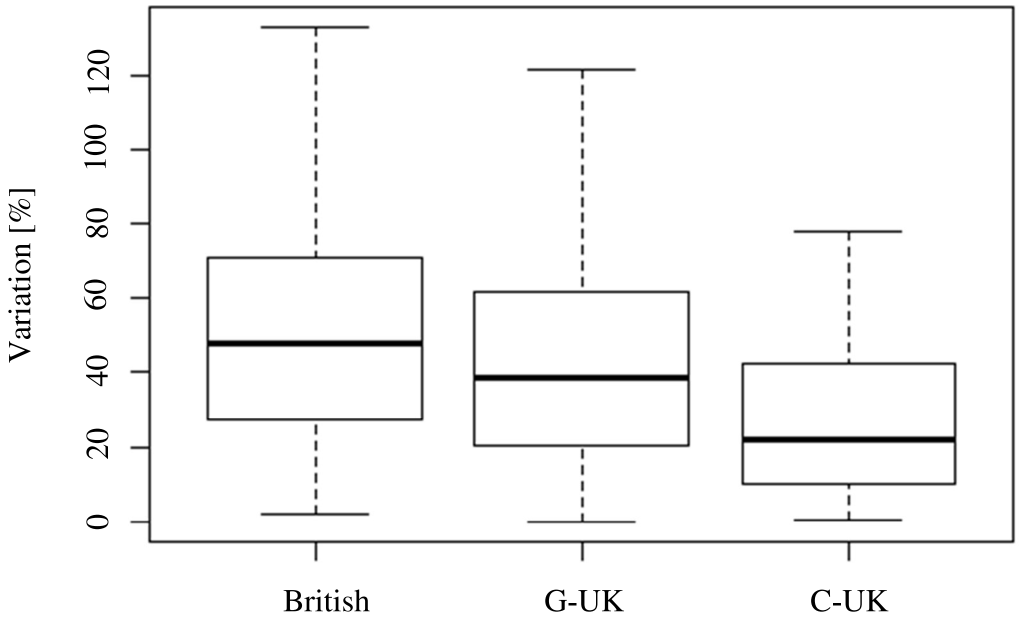


Figure 4-15. Boxplots of the vowel length variation in Lot, British group, sentence reading.

All groups differ significantly from each other in the sentence reading task. Again, there is a large spread in the British native speakers' data, and the minimum has decreased again. Group G-UK significantly differs from the model group ($p = .02$), as does the control group ($p = .007$). Both German groups also behave significantly differently from each other ($p = .02$), which illustrates again that group G-UK assumes an intermediate position between the native speaker model and the control group. Group C-UK shows a considerable amount of variation again with the maximum at around 80%.

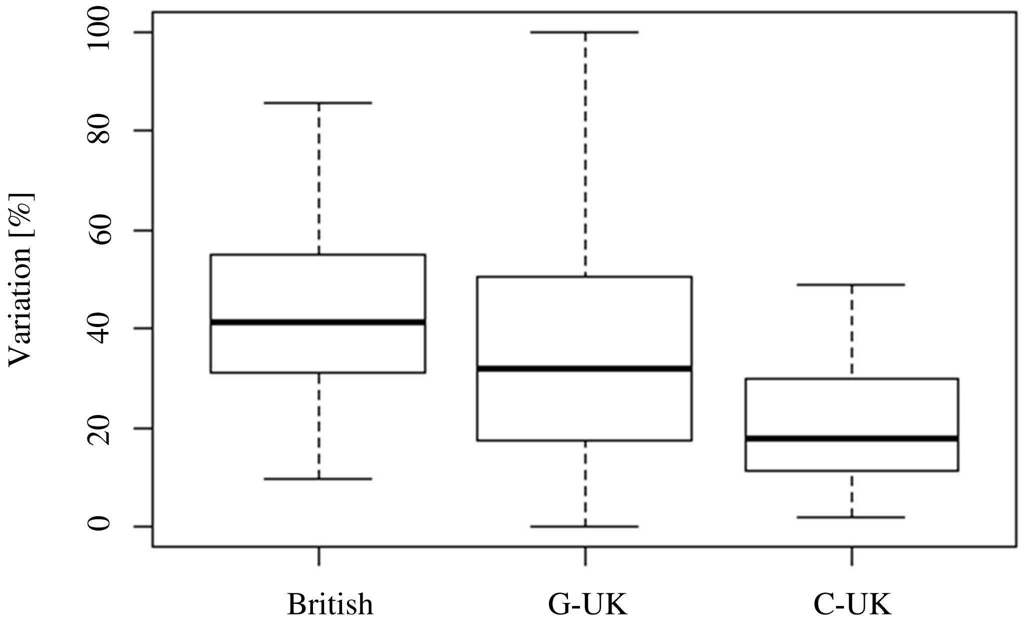


Figure 4-16. Boxplots of the vowel length variation in LOT, British group, text reading.

The text reading is very interesting, since some participants in group G-UK outperform the British native speakers. However, on closer inspection, this is not a result of group G-UK exhibiting stronger variation than before. Instead, the British native speakers' variation decreases in the text reading and approaches the German learners' degree of variation (cf. Table 4-5). This leads to group G-UK not being significantly different from the British native speakers (*p* = .16). Although group C-UK also shows some variation, the group is significantly different both from group G-UK (*p* = .03) and the native speakers (*p* = .007).

A summary of the mean percentage of variation in LOT in the four different production tasks can be found in Table 4-5 below.

Table 4-5. Mean percentage of variation for the British group in four different production tasks. Vowel LOT. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
British	59.58	65.74	62.95	47.61
sd	(31.32)	(37.03)	(57.93)	(24.97)
G-UK	48.47	45.49	51.11	40.88
sd	(39.14)	(32.08)	(49.35)	(40.50)
C-UK	13.95	22.16	29.33	23.28
sd	(12.21)	(21.05)	(24.48)	(19.53)

As was the case with TRAP, LOT is a vowel where a large amount of variation can be observed, even in the group which has no experience abroad (C-UK). Yet, in contrast to TRAP, group G-UK behaves significantly differently from the British model group in three out of four tasks: Minimal pairs ($p = .02$), word list ($p = .03$) and sentence reading ($p = .02$). This suggests that even though variation is easily perceptible in this vowel, learners have not yet managed to reproduce it. There is large spread in group G-UK; some subjects even outperform their model.

4.2.1.5 Summary

The analysis of the vowel length variation produced by the British group (native speakers, G-UK and C-UK) has shown that the British native speakers have the strongest variation in TRAP ($\bar{x} = 69.39\%$), followed by LOT ($\bar{x} = 58.97\%$), DRESS ($\bar{x} = 54.55\%$) and KIT ($\bar{x} = 52.46\%$).[58] This sequence is matched by group G-UK, which exhibits 59.92% of variation in TRAP, 46.49% in LOT, 45.14% in DRESS and 41.19% in KIT. Again, this can be seen as evidence that German learners acquire vowel length variation largely by mimicking native speakers. The stronger the vowel length variation exhibited by the native speakers, the easier it is to perceive and reproduce by German learners.

A look at group C-UK reveals another interesting factor. Group C-UK does not match the sequence proposed above, since its members exhibit more variation in DRESS than in LOT (23.69% vs. 22.18%). However, TRAP is still the vowel in which most variation can be observed (24.58%), and KIT the one with least variation (16.91%). A hypothesis to explain the behavior of group C-UK might be that L1 influences the acquisition of vowel length variation to a certain extent before a native-like sequence emerges. Since British English DRESS is much closer to German DRESS than British English LOT to a similar German vowel, this might make it easier for learners to reproduce this length variation in DRESS. In the end, however, the difference between the variation in DRESS and LOT amounts to only 1% in both groups G-UK and C-UK, while it is greater among the native speakers. Therefore, this might also mean that vowel length variation in DRESS and LOT emerges almost simultaneously. However, the results from the American group (cf. section 4.2.2.2) seem to favor the hypothesis that L1 influences the acquisition of vowel length variation. In the American group, the control group C-US also manages to match the model group quite closely and shows considerably more variation in DRESS than in LOT.

---

[58] Numbers averaged over all tasks.

4.2.2 American Group

4.2.2.1 Vowel KIT

The following section illustrates the vowel length variation exhibited by the American native speakers, the German learners G-US and the control group C-US in the different reading tasks. KIT, DRESS, TRAP and LOT will be considered in this order. For each vowel, we will examine vowel length variation in the minimal pairs, the word list, the sentences and the text.

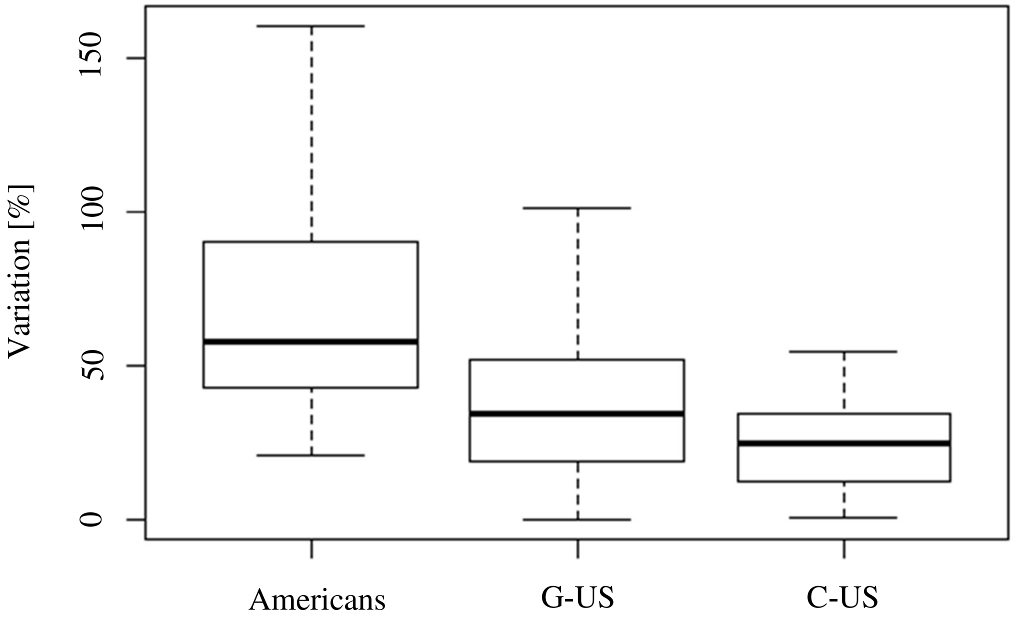


Figure 4-17. Boxplots of the vowel length variation in KIT, American group, minimal pairs reading.

It is quite evident from this first boxplot that both groups G-US and C-US show a much lower degree of variation than the American native speakers. While both German groups' minima approach 0, this is not the case for the native speakers. Thus, it comes as no surprise that both groups G-US and C-US are significantly different from the native speakers ($p = .001$ and $p = .007$, respectively). Even though the spread is much smaller in group C-US than in G-US, the two learner groups do not differ significantly from each other ($p = .10$). In comparison to the British group, this is an interesting finding. The German learner group G-UK was able to match the British native speakers in three out of four tasks in KIT. However, the variation exhibited by the British native speakers in this vowel is much smaller than that which we can observe from the American native speakers. Therefore, group G-US would need to show much greater variation than group G-UK in order to match the American native speakers.

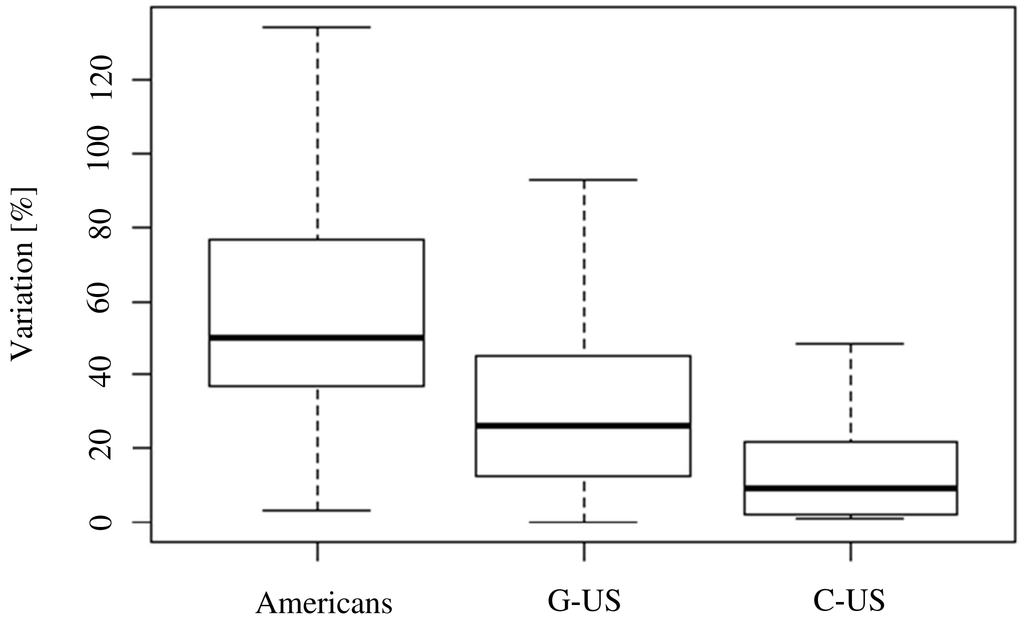


Figure 4-18. Boxplots of the vowel length variation in Kit, American group, word list reading.

The trend which we saw before continues in the minimal pairs, although some differences are visible. For instance, the American native speaker's minimum also approximates 0, which was not the case before. Also, group C-US now differs significantly from group G-US ($p = .05$). This shows that even though neither group was able to match the American speakers in the most formal minimal pairs reading task, at least the control group did not differ from the German learners with experience abroad. The semi-formal word list reading seems to capture the control group's attention more, so that they focus less on pronunciation and therefore perform significantly worse than group G-US ($p = .05$) and the native speakers ($p = .007$). Group G-US lies in the middle field again, but still cannot match the native speakers ($p = .001$).

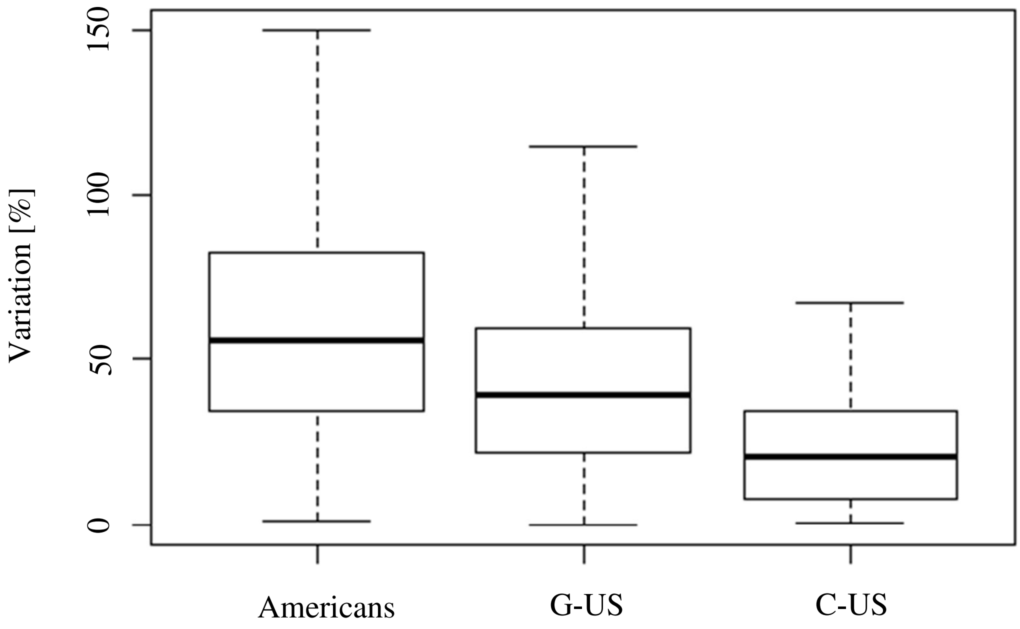


Figure 4-19. Boxplots of the vowel length variation in Kɪᴛ, American group, sentence reading.

The results from the sentence reading task are very similar to the two tasks before. The American native speakers represent the model towards which groups G-US and G-US orientate themselves. G-US is more successful, although the speakers in this group differ significantly from the native speakers (*p* = .004). Group C-US exhibits significantly less variation than both the native speakers (*p* = .008) and group G-US (*p* = .007).

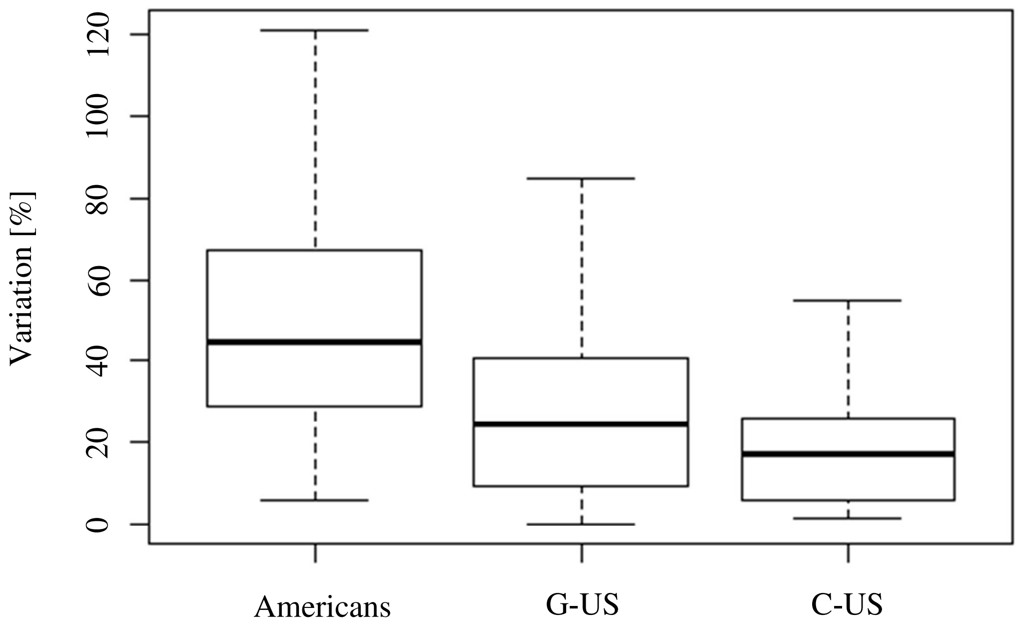


Figure 4-20. Boxplots of the vowel length variation in Kɪᴛ, American group, text reading.

The text reading task is very interesting. All three groups show less variation than before, which is to be expected in informal reading. The text is often read quite fluently, which reduces vowel length variation even further. However, group G-US reduces their variation the most and shows the lowest level of variation out of all tasks, resulting in this group being very similar to the control group (*p* = .12). By comparison, group C-US has greater variation in the text reading than in the word list. Still, the overall level of variation is quite low, so that C-US ends up being statistically different from the native speakers (*p* = .007). G-US cannot match the native speakers, either (*p* = .005).

As before, the following table lists the mean percentages of variation for the three groups *Americans*, *G-US* and *C-US* in the four production tasks.

Table 4-6. Mean percentage of variation of the American group in four different production tasks. Vowel KIT. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
Americans	71.57	59.24	70.97	53.53
sd	(41.37)	(34.83)	(57.98)	(37.74)
G-US	40.44	33.09	47.91	29.10
sd	(38.54)	(30.88)	(41.19)	(26.15)
C-US	25.99	15.09	26.21	19.03
sd	(16.20)	(15.97)	(29.10)	(15.02)

Table 4-6 shows that both groups G-US and C-US show a much lower level of average vowel length variation than the American native speakers. Group G-US exhibits only roughly 40% to 50% of the variation shown by the native speakers; group C-US, 20% to 30%. This means that G-US and C-US differ significantly from their native speaker model in all tasks. The high variation values exhibited by the American native speakers in the minimal pairs and sentence reading task are quite surprising and inexplicable.

4.2.2.2 Vowel DRESS

The second vowel to be examined is DRESS. As before, the boxplots show a comparison of the vowel length variation exhibited by the American native speakers, the German learners (G-US) and the control group (C-US) in the different production tasks.

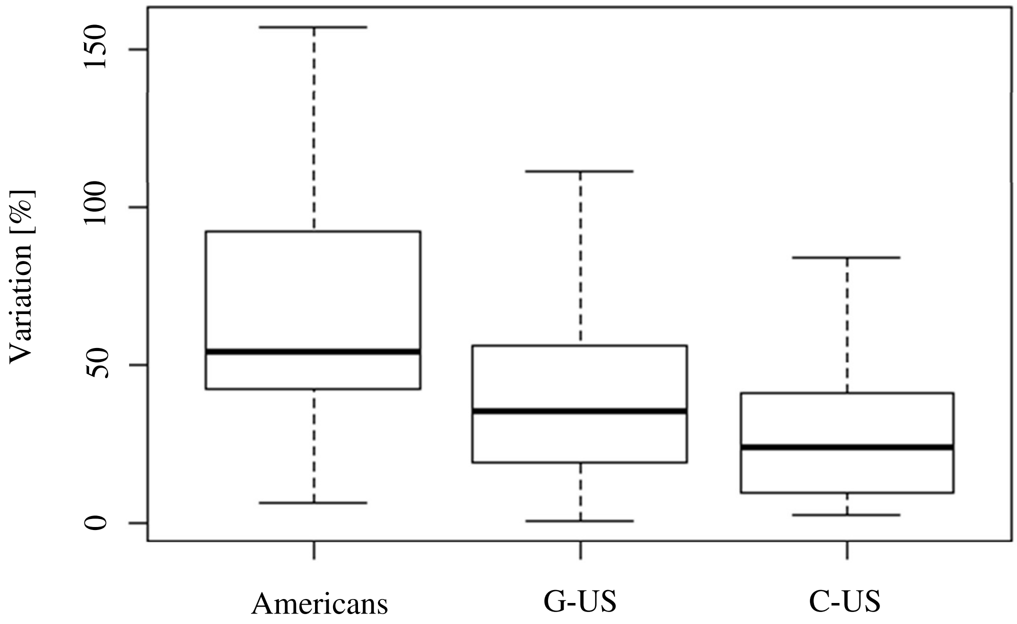


Figure 4-21. Boxplots of the vowel length variation in DRESS, American group, minimal pairs reading.

The first boxplot for DRESS foreshadows a general trend to be observed in all four tasks: Both German groups perform very similarly and also imitate their model quite closely. This is due to the fact that the control group shows larger variation in this vowel than in all the others, which suggests that the vowel length variation in DRESS is acquired more easily by German learners than the variation in other vowels. Since DRESS is pronounced lower in American English than in British, it lies closer to the German realization of it. In recent years, many German speakers have merged /eː/ (e.g. *Beeren* 'berries') and /ɛː/ (e.g. *Bären* 'bears') under /eː/. The two remaining phonemes now representing DRESS are /eː/ and /ɛ/. This might make the length contrast in DRESS particularly salient in German and therefore easier to reproduce in English.[59] A supporting factor of this hypothesis is that the control group is not statistically different from group G-US in any of the tasks. Thus, what makes learners successful in acquiring vowel length variation in DRESS might be transfer from L1 in addition to exposure to native speakers (which group C-US does not have). This interesting connection between group G-US and the vowel DRESS can also be observed in the perception experiment. For this reason, this issue will be revisited in section 4.3.2.2.

---

[59] Group G-UK was able to match the British native speakers in the two formal tasks in DRESS; group C-UK also showed considerable vowel length variation in this vowel.

What remains is the statistical analysis: Group G-US performs significantly differently from the American native speakers (*p* = .009), but is similar to the control group (*p* = .18). In turn, the control group is statistically different from the native speakers (*p* = .001).

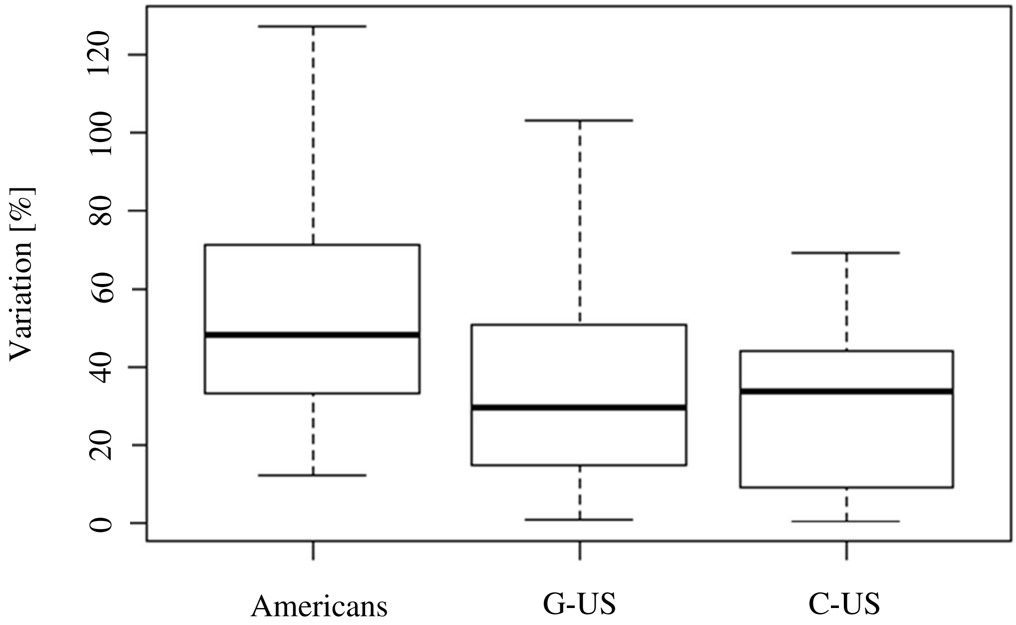


Figure 4-22. Boxplots of the vowel length variation in DRESS, American group, word list reading.

A fascinating finding in the word list reading is that the median of the control group C-US lies higher than that of group G-US. This is further evidence that the acquisition of vowel length variation might not be guided solely by input from native speakers, but also by the connection between L1 and L2. Considering the large spread in both group C-US and the American native speakers, however, the control group is still significantly different from the model group (*p* = .03). Likewise, group G-US cannot match the native speakers (*p* = .02). The two German learner groups perform similarly (*p* = .33).

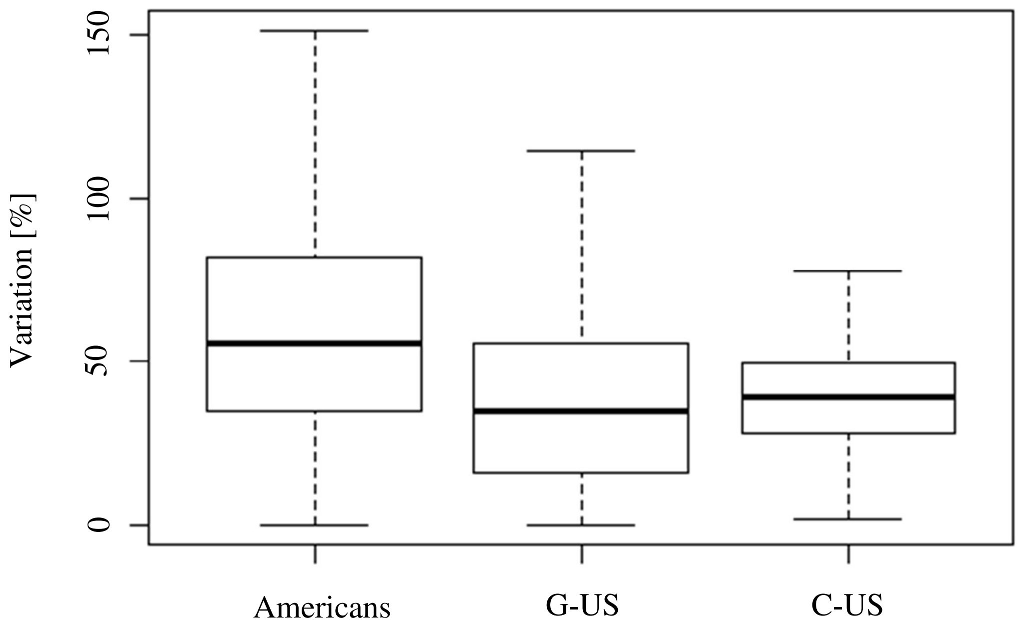


Figure 4-23. Boxplots of the vowel length variation in DRESS, American group, sentence reading.

In the sentence reading task, both the median and the mean of group C-US is higher than that of group G-US. Nevertheless, neither of the German learner groups can match the native speakers (G-US $p = .0006$, C-US $p = .007$). This is also due to the fact that the American speakers' variation is greatest in this task. Again, groups G-US and C-US are very similar ($p = .23$).

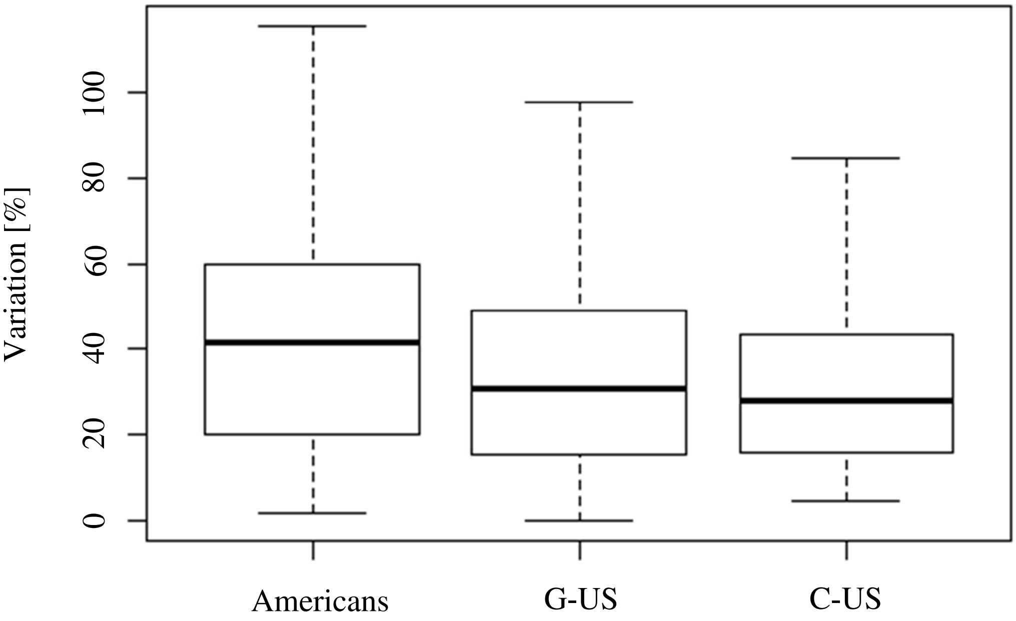


Figure 4-24. Boxplots of the vowel length variation in DRESS, American group, text reading.

At first sight, Figure 4-24 seems to suggest that groups G-US and C-US match the native speakers well in the final text reading task, even though they are still significantly different from the model ($p = .02$ and $p = .03$, respectively). This is a misinterpretation, however. Instead, the American speakers' average vowel length variation decreases drastically in this task compared to the others. Therefore, it is more accurate to say that the American native speakers come closer to the two German learner groups. Groups G-US and C-US perform similarly once more ($p = .19$).

The average vowel length variation of all three groups in the different production tasks is listed in the table below.

Table 4-7. Mean percentage of variation of the American group in four different production tasks. Vowel DRESS. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
Americans	68.71	60.60	75.72	46.52
sd	(41.69)	(45.34)	(68.33)	(36.28)
G-US	43.56	39.07	43.26	37.76
sd	(36.57)	(37.82)	(39.48)	(36.79)
C-US	36.93	30.56	43.82	30.67
sd	(30.65)	(20.22)	(31.66)	(20.88)

Table 4-7 illustrates what has become evident from the boxplots already: groups G-US and C-US perform quite similarly in all the tasks. Group C-US even outperforms group G-US in the sentence reading. This suggests that the acquisition of vowel length variation, at least where DRESS is concerned, is not solely based on input from native speakers and mimicking said input, but also on the relationship between L1 and L2. Since DRESS also stands out in the American group in the perception experiment, this issue will be further discussed in section 4.3.2.2.

4.2.2.3 Vowel TRAP

TRAP is the next vowel to be examined. The following boxplots will again display the vowel length variation exhibited by the American native speakers, the German learners (G-US) and the control group (C-US) in the four different production tasks.

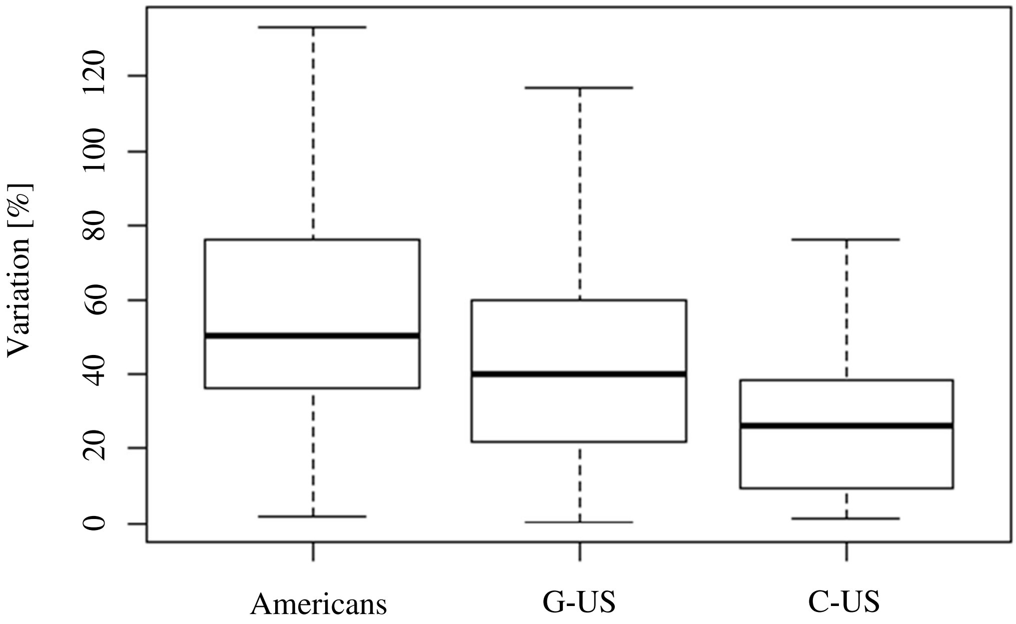


Figure 4-25. Boxplots of the vowel length variation in TRAP, American group, minimal pairs reading.

In the minimal pairs reading, groups G-US and C-US perform similarly (*p* = .06). Both groups are statistically different from the American native speakers (G-US *p* = .03, C-US *p* = .007), whose median and highest value lie well above the German learners'.

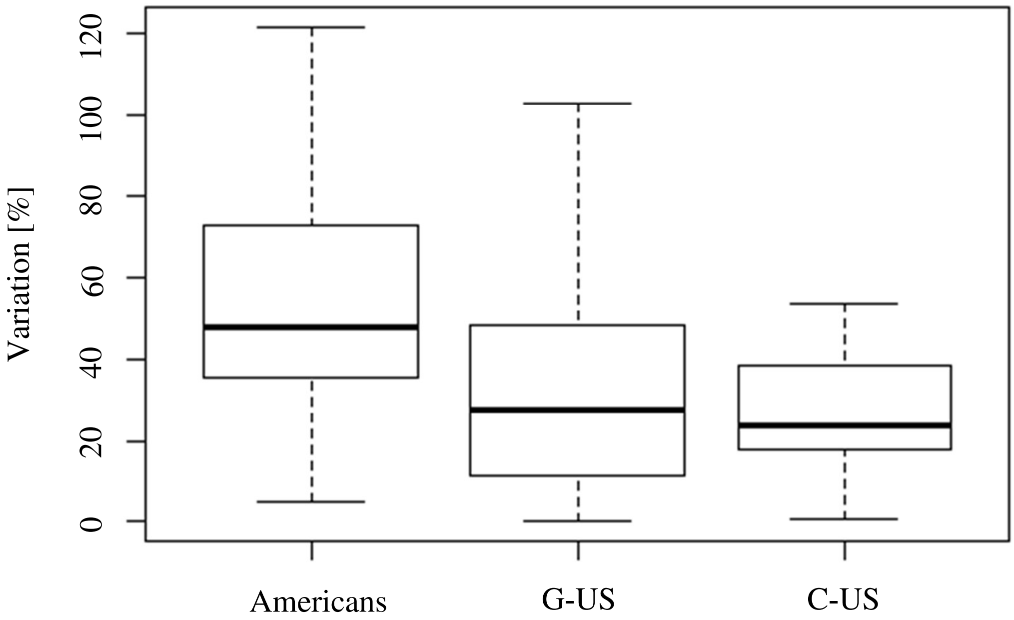


Figure 4-26. Boxplots of the vowel length variation in TRAP, American group, word list reading.

The situation in the word list reading is similar to the minimal pairs reading. Groups G-US and C-US are even more similar than before (*p* = .52). Although the spread in group G-US is quite large, this group differs significantly from the American native speakers (*p* = .01). Likewise, group C-US performs significantly differently from the native speakers (*p* = .0001).

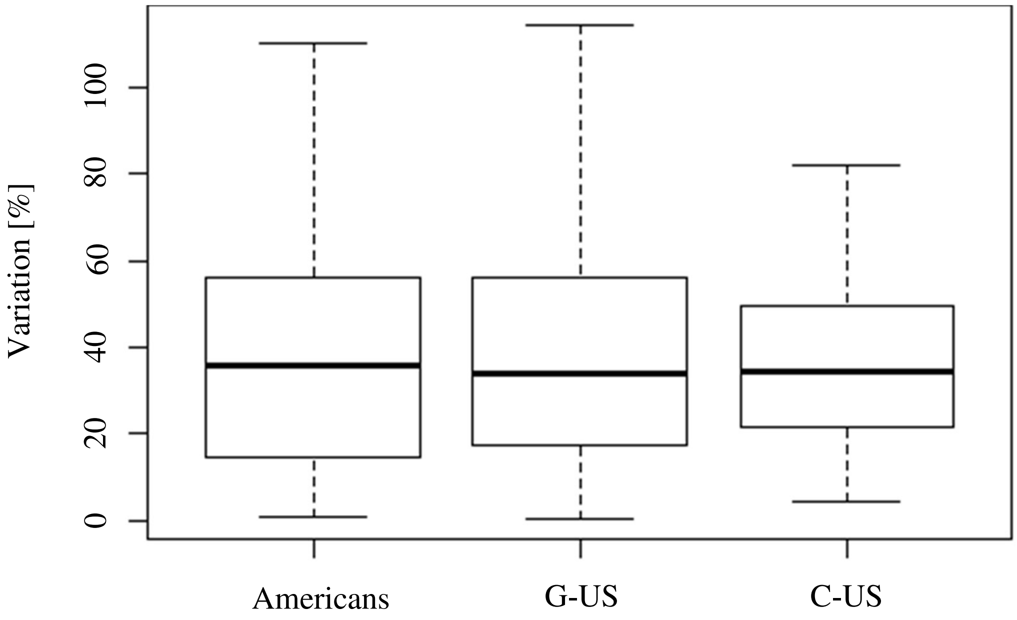


Figure 4-27. Boxplots of the vowel length variation in TRAP, American group, sentence reading.

The third reading task paints quite a different picture. Suddenly, all three groups show very similar variation values, with group G-US even outperforming the American native speakers. This is due to a) the American speakers' variation decreasing between the minimal pairs, word list and sentence reading tasks, and b) group C-US showing greater variation than before. Interestingly, although the spread in group C-US is quite small compared to the other groups, their median is the highest of all three groups. Group G-US's variation remains quite steady over the first three tasks. As a result, none of the groups are statistically different from each other. Group G-US matches the American group most closely (*p* = .47), and is also quite similar to the control group (*p* = .16). Group C-US does not statistically differ from the native speakers, either (*p* = .13).

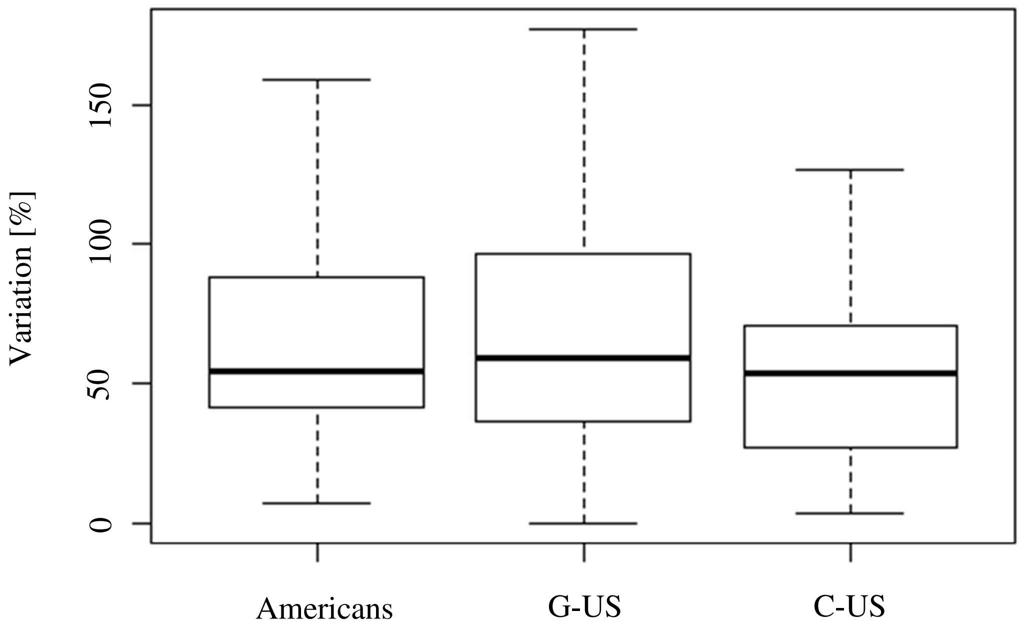


Figure 4-28. Boxplots of the vowel length variation in TRAP, American group, text reading.

The last reading task appears very similar to the one before. As in the sentence reading task, some speakers of group G-US outperform the native speakers again. Therefore, it comes as no surprise that group G-US matches the native speaker model well (*p* = .33), as does group C-US (*p* = .11). Groups G-US and C-US also show similar vowel length variation in this task (*p* = .08).

Table 4-8 summarizes the mean percentage of variation in TRAP in the four different production tasks.

Table 4-8. Mean percentage of variation of the American group in four different production tasks. Vowel TRAP. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
Americans	59.45	58.53	44.29	65.52
sd	(35.73)	(36.25)	(41.56)	(36.94)
G-US	49.81	35.32	44.13	81.76
sd	(44.72)	(32.42)	(39.73)	(74.77)
C-US	27.73	31.91	45.95	64.17
sd	(20.44)	(26.09)	(51.58)	(54.71)

TRAP is characterized by a distinction between the formal tasks (minimal pairs and word list), where the two learner groups cannot match the native speaker model, and the two informal tasks (sentence and text reading), where they manage to do so quite successfully. As was the case the British section, it is not easy to explain why the groups show the highest level of variation in the most informal task, namely text reading. It might have to do with the semantic load of the story after all. Even though TRAP is not the vowel with the highest vowel length variation exhibited by the native speakers, the two German groups G-US and C-US only manage to match the native speakers in two out of the for tasks.

4.2.2.4 Vowel LOT

The final vowel to be discussed for the American group is LOT. The boxplots will again present the vowel length variation exhibited by the American native speakers, the German learners (G-US) and the control group (C-US) in the different reading tasks.

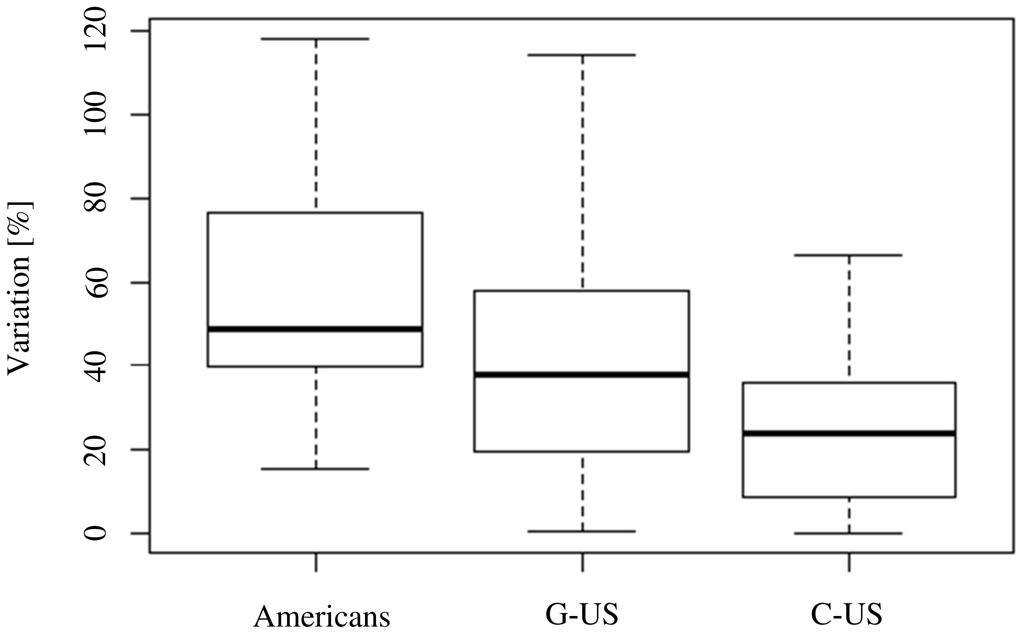


Figure 4-29. Boxplots of the vowel length variation in LOT, American group, minimal pairs reading.

In the minimal pairs reading, group G-US shows the largest spread of the three groups. However, G-US cannot match the American native speakers (*p* = .04), whose median lies well above the two learner groups'. The control group (C-US) differs significantly from the American native speakers (*p* = .007), but not from group G-US (*p* = .08).

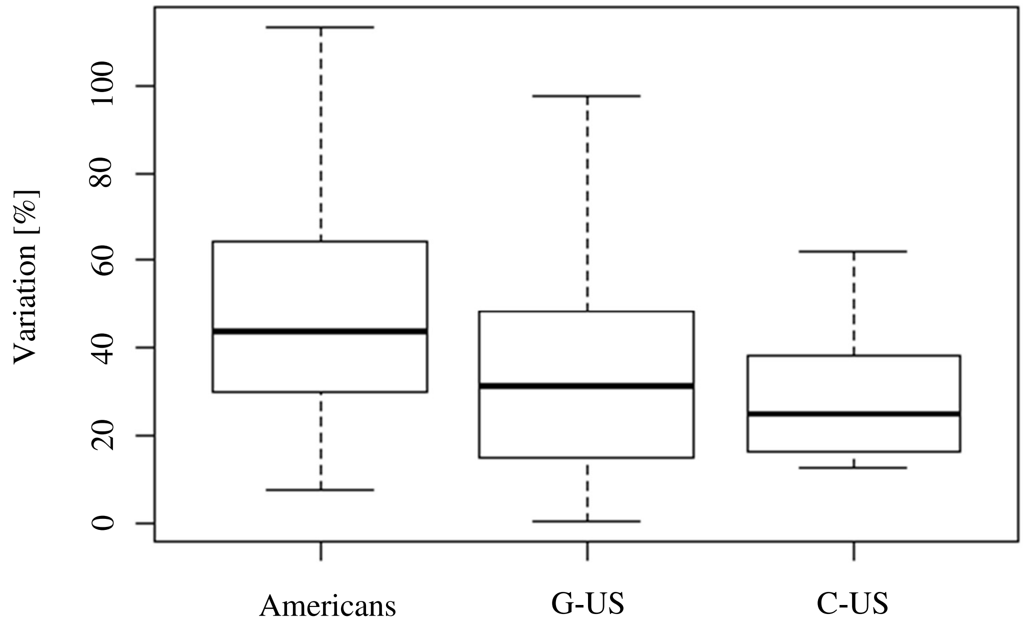


Figure 4-30. Boxplots of the vowel length variation in LOT, American group, word list reading.

The word list shows groups G-US and C-US to be even more similar than before (*p* = .39). It is also interesting that the minimal value of group G-US lies below that of the control group C-US. Neither of the two groups can match the native speaker model. However, group G-US is less statistically different from the Americans (*p* = .01) than group C-US (*p* = .008).

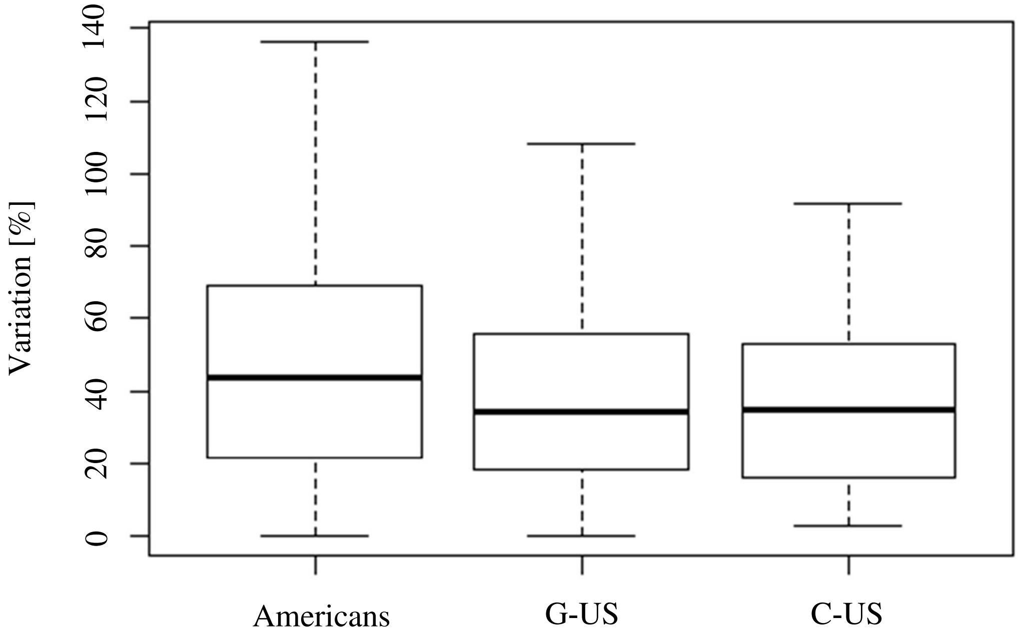


Figure 4-31. Boxplots of the vowel length variation in LOT, American group, sentence reading.

Figure 4-31 illustrates that the two German learner groups G-US and C-US perform almost equally (*p* = .83). This is the most similar performance displayed in any of the tasks or vowels. Neither group can match the native speaker model. Group G-US differs significantly from the American native speakers (p = .002), as does group C-US (p = .001).

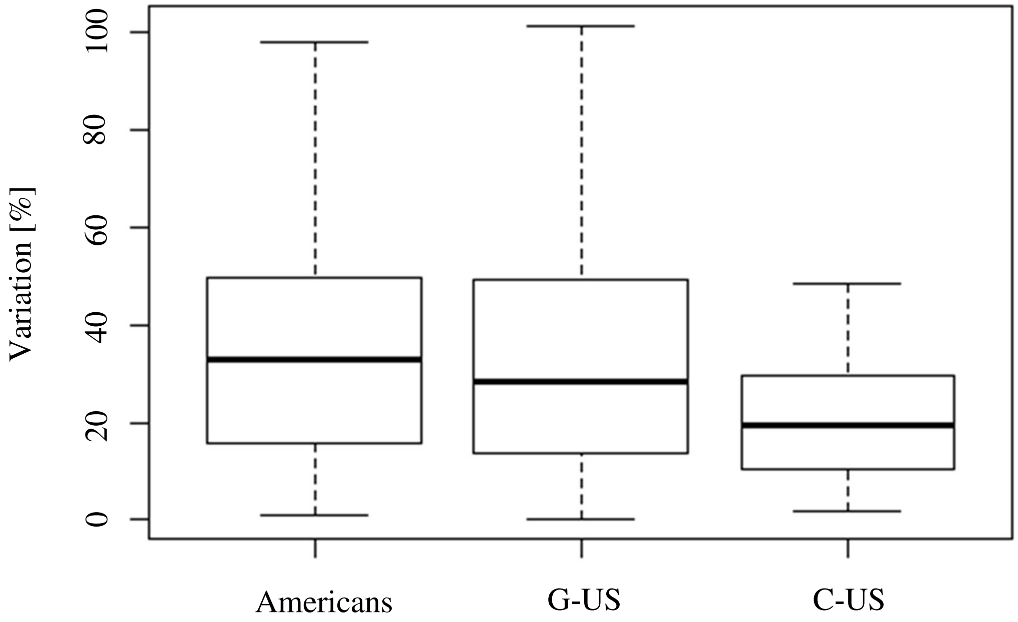


Figure 4-32. Boxplots of LOT, American group, text reading.

In the final reading task, it seems that group G-US performs very similarly to the American native speakers. This is indeed the case (p = .12); however, it is again due to the fact that the Americans' vowel length variation decreases in this task and thus comes closer to the variation exhibited by group G-US. Group C-US is statistically different from both the American native speakers (p = .003) and group G-US (p = .001).

A summary of the mean percentage of variation in LOT in the four different production tasks can be found in Table 4-9 below.

Table 4-9. Mean percentage of variation of the American group in four different production tasks. Vowel LOT. One standard deviation is given in the brackets.

	Minimal pairs	**Word list**	**Sentences**	**Text**
Americans	57.14	54.85	56.01	37.70
sd	(25.51)	(52.91)	(52.94)	(31.95)
G-US	45.58	36.56	44.51	36.89
sd	(36.16)	(29.17)	(40.61)	(32.67)
C-US	33.02	30.55	43.17	23.83
sd	(34.09)	(17.69)	(46.59)	(19.55)

Regarding LOT, group G-US manages to match the American speakers in only one task, namely the text reading. Moreover, this is only the case because the American speakers' variation decreases. Groups G-US and C-US perform very similarly in three out of four tasks (minimal pairs, word list and sentence reading). This is particularly interesting because the control group shows such a high degree of variation.

4.2.2.5 Summary

The analysis of the vowel length variation exhibited by the American group (native speakers, G-US and C-US) yields a number of very surprising results. The American speakers do not exhibit most variation in the lowest vowels, as the British native speakers do. Instead, they show an almost inverted sequence: 63.82% of variation in KIT, 62.89% in DRESS, 56.95% in TRAP and 51.43% in LOT. This is certainly unexpected and no explanation can be offered.

Interestingly, however, the German learners in group G-US do not imitate the sequence produced by the American native speakers, but the one we already know from group G-UK: They show the greatest variation in TRAP (52.76%), followed by LOT (40.89%), DRESS (40.91%) and KIT (37.64%). Group C-US almost follows this sequence, with an average of 42.44% of variation in TRAP, 35.50% in DRESS, 32.64% in LOT and 21.58% in KIT. Here we can see again that the similar vowel DRESS might make it easier for inexperienced learners to acquire vowel length variation here than for the unfamiliar LOT.

Since the British native speakers, groups G-UK and G-US all show the same route of acquisition (group C-UK and C-US are very similar except for LOT having lower levels of variation than DRESS), I suggest this sequence as the common route for the acquisition of vowel length variation by German learners of English. It seems that inexperienced learners initially favor vowel length variation in DRESS before a native-like sequence takes over. The final sequence goes in accordance with the average vowel length variation exhibited by the British native speakers (TRAP → LOT → DRESS → KIT), and the expected variation exhibited by the American native speakers. Why the American speakers in this dissertation do not follow the expected sequence is not clear; however, with a sample as small as ten speakers it is quite possible that individual speakers had a considerable effect on the group. Since the rule that lower vowels show higher vowel length variation holds and has been illustrated by many researchers, the American speakers in this group must be viewed as an exception.

4.2.3 Effect of the Consonant

This part of the dissertation examines the vowel length variation exhibited before the six different consonants. Since no statistical effect of the vowel was found, the variation of all vowels produced in all tasks was pooled together for a global overview. Section 4.2.3.1 will report on the British Group, while section 4.2.3.2 deals with the American Group.

4.2.3.1 British Group

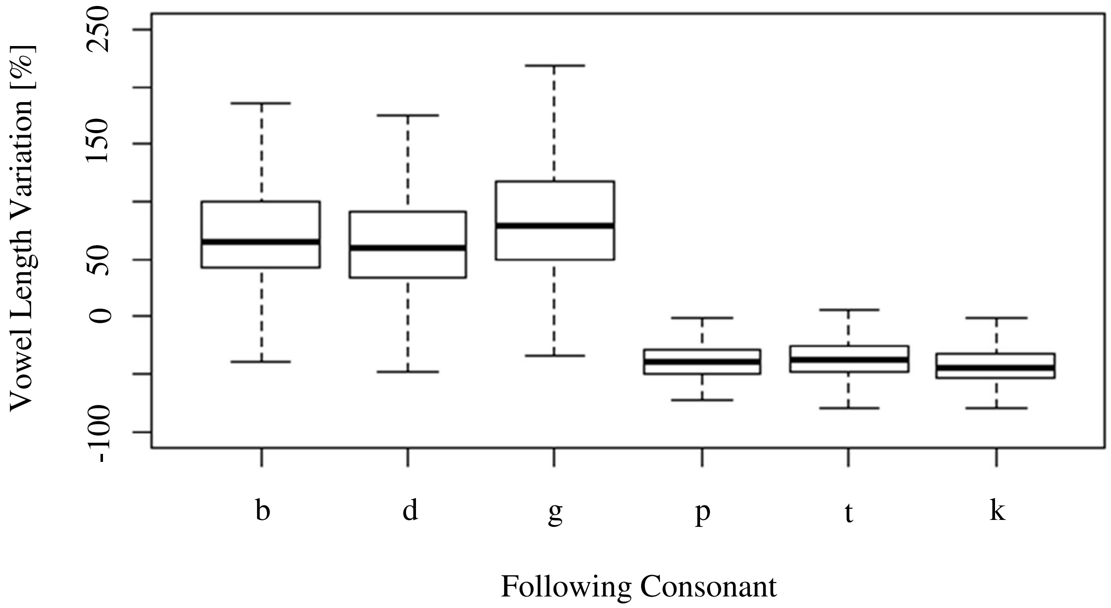


Figure 4-33. Mean vowel length variation exhibited by the British native speakers in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

The vowel length variation exhibited by the British native speakers is quite steady regarding all three voiced consonants and all three unvoiced ones. The only difference can be seen in vowels preceding /g/, which show slightly greater variation than before /b/ and /d/. However, the variation exhibited before the three voiced consonants is statistically different ($p < .001$). This is also the case for the vowel length variation in connection with the three unvoiced consonants ($p = .01$). What is also clearly visible from Figure 4-33 is that the vowel length variation is firmly divided between the voiced and voiceless consonants: vowels before voiced consonants are lengthened (meaning they show positive vowel length variation), whereas vowels preceding unvoiced consonants are shortened (that is, they show negative variation). In addition, one standard deviation, which includes 68.3% of the data, still lies within the positive range with respect to the voiced consonants, and in the negative range concerning the unvoiced consonants:

Table 4-10. Mean vowel length variation exhibited by the British native speakers, groups G-UK and C-UK in vowels preceding the six voiced and voiceless plosives. One standard deviation is given in the brackets.

	Following consonant					
	b	**d**	**g**	**k**	**p**	**t**
British	74.59	67.11	89.01	-41.71	-36.91	-34.84
sd	(51.17)	(51.34)	(59.53)	(19.93)	(33.75)	(19.52)
G-UK	46.41	61.48	61.86	-28.90	-20.84	-27.21
sd	(64.36)	(66.51)	(60.96)	(30.57)	(44.20)	(38.32)
C-UK	14.18	19.62	13.65	-6.87	-6.67	-10.04
sd	(30.71)	(35.95)	(30.12)	(21.87)	(22.86)	(25.61)

The table illustrates that vowels preceding voiced consonants are lengthened, while vowels preceding voiceless consonants are shortened, which is of course the expected result for native speakers. This differs strongly from the behavior of groups G-UK and C-UK. Consider Figure 4-34 below, which illustrates group G-UK's production of vowel length variation in relation to the different consonants:

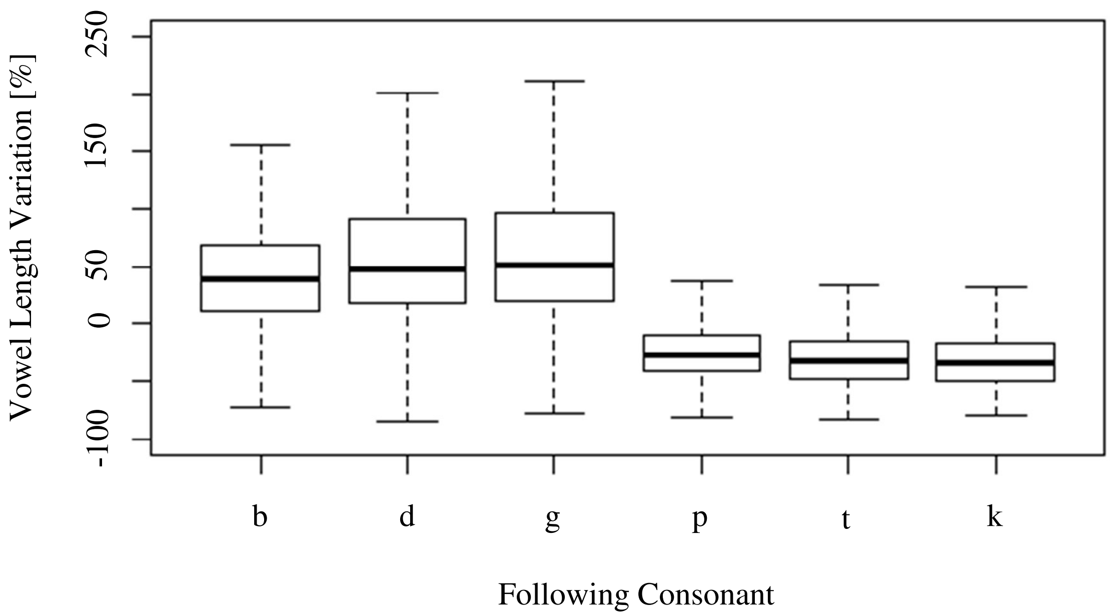


Figure 4-34. Mean vowel length variation exhibited by group G-UK in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

Although Figure 4-34 suggests that group G-UK's variation also remains quite steady within the groups of voiced and voiceless consonants, a number of differences can be observed. Firstly, the vowel length variation is not as strong as the British native speakers'. This is illustrated by the fact that it remains below that of the native speakers in all six cases (cf. also

Table 4-10). The only consonant pair where group G-UK behaves similarly to the British native speakers is /d, t/. However, group G-UK still performs significantly worse than the native speakers regarding both /t/ ($p = .0005$) and /d/ ($p = .0003$). Variation preceding the other four consonants is all highly significantly different from the British native speakers' (all in the range of $p < .000000000000001$), so that it is impossible to draw a conclusion as to which of the two remaining consonant pairs might make it easier to reproduce vowel length variation correctly.[60]

Secondly, the vowel length variation – or rather, vowel shortening – before the voiceless consonants is much weaker in group G-UK compared to the native speakers. All of the upper whiskers of the boxplots relating to unvoiced consonants in group G-UK extend into the positive range, which indicates that a number of speakers pronounce a vowel preceding a voiceless consonant *longer* than one preceding a voiced consonant. An example of this is Speaker #001, who pronounced DRESS in *peg* -12.94% shorter than in *peck* in the minimal pairs reading. Likewise, the lower whiskers of the boxplots for the voiced consonants extend into the negative range. This is also the case for the British native speakers; however, it is more extreme in group G-UK, where the most negative values lie almost on a straight line for all six consonants.

Another striking difference appears when we consider the standard deviation of group G-UK. Whereas the standard deviation of the British native speakers lay below the mean in all six cases, it exceeds the mean in five out of six cases in group G-UK. In the sixth consonant, /g/, the standard deviation almost equals the mean. This illustrates that within 68.3% of the data of group G-UK, there are already quite a number of tokens with inverted vowel length variation, i.e. where vowels are shorter before voiced consonants than before voiceless ones. This is of course quite a clear sign that the learners in group G-UK are not as proficient as the British native speakers.

Let us consider the control group C-UK now. Figure 4-35 below illustrates the results of group C-UK on the same scale as group G-UK and the native speakers.

---

[60] /t/ and /d/ also play an interesting role in the results of group G-UK in the perception experiment. Cf. section 4.3.1.2.

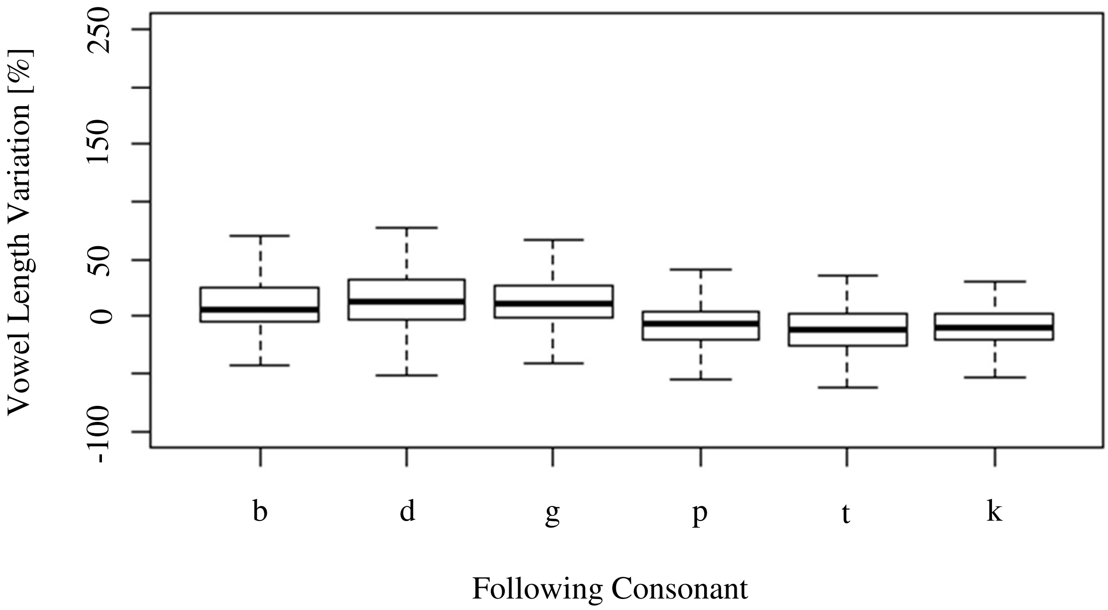


Figure 4-35. Mean vowel length variation exhibited by group C-UK in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

It is immediately visible that the range of variation is very small for all consonants, with the medians lying between -10% and +20%. Moreover, the variation preceding voiced and voiceless consonants does not differ much. This suggests that the speakers in group C-UK have not yet acquired a rule for vowel length variation. For better visibility and further discussion, the results of group C-UK were rescaled in a second graph:

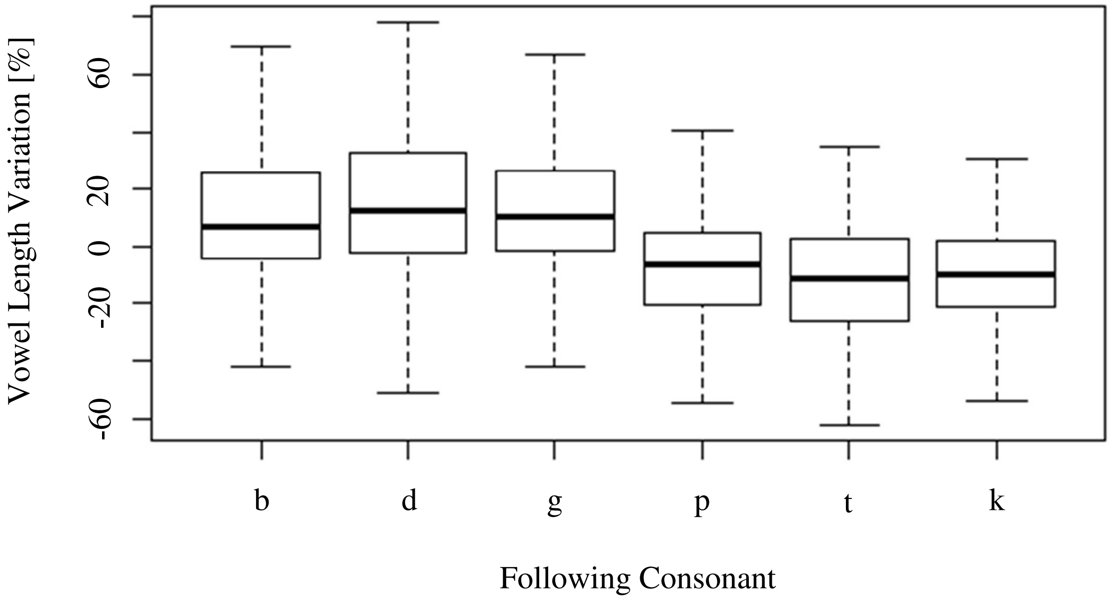


Figure 4-36. Rescaled figure illustrating the mean vowel length variation exhibited by group C-UK in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

The new scale in Figure 4-36 allows a more accessible discussion of the results. Some very small differences between voiced and voiceless consonants can be observed, though these do not amount to statistical significance (voiced consonants $p = .32$; voiceless consonants $p = .25$). Similar to group G-UK, some speakers in group C-UK produce tokens with inverted vowel length variation. A number of vowels followed by voiceless consonants were lengthened by up to 50%; some vowels followed by voiced consonants were shortened up to -55%. Since the overall variation exhibited by group C-UK is so small it is not surprising that it is statistically different from the British native speakers in all six cases (all in the range of $p < .000000000000001$). The standard deviation exceeds the mean by far in all cases, particularly regarding the voiceless consonants. Therefore, the control group can easily be called the least proficient group.

4.2.3.2 American Group

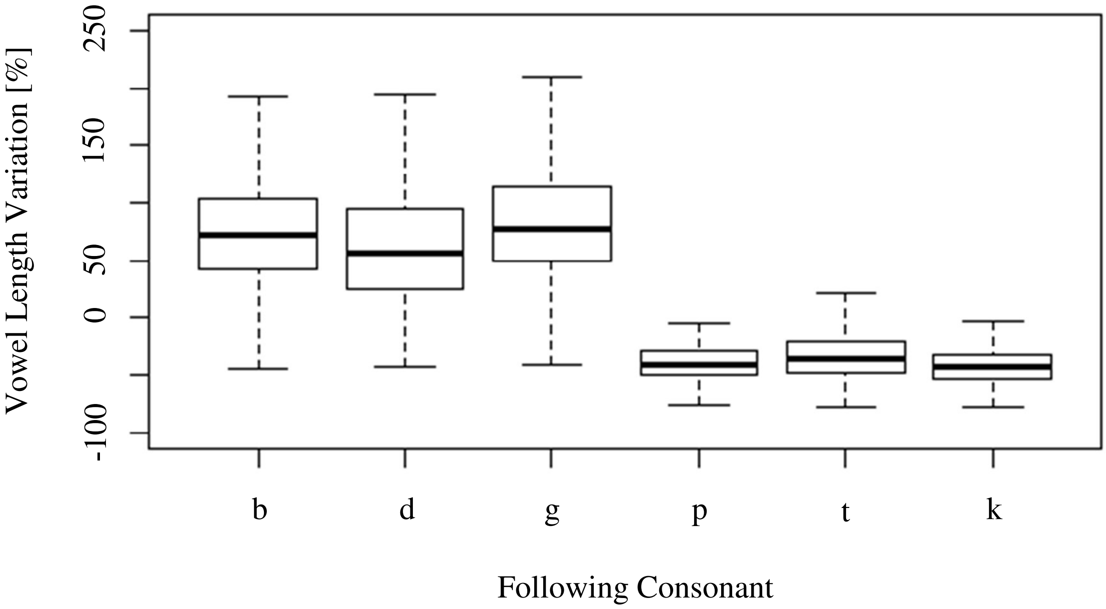


Figure 4-37. Mean vowel length variation exhibited by the American native speakers in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

The vowel length variation exhibited by the American native speakers is quite similar to that of the British speakers. It is also quite steady in association with all three voiced and all three unvoiced consonants. Again, the variation in the vowels preceding /g/ is slightly greater than that occurring before /d/ and /b/. The variation before the three voiced consonants is statistically different ($p = .0004$). However, this results from the variation in the vowels preceding /d/ being smaller. The variation in the vowels before /b/ and /g/ is actually similar ($p = .09$).

Regarding the voiceless consonants, the variation exhibited by the American speakers also differs significantly (*p* = .00005). Again, this is due to /t/, with the variation being similar in the vowels preceding /k/ and /g/ (*p* = .08). This illustrates that vowel length variation before /d/ and /t/ is somewhat different from that occurring before the remaining four plosives.[61] As was the case with the British native speakers, the vowel length variation exhibited by the American native speakers is also firmly divided between the voiced and voiceless consonants: vowels before voiced consonants are lengthened (meaning they show positive vowel length variation), vowels preceding unvoiced consonants are shortened (that is, they show negative variation). One standard deviation lies well within the positive range regarding the voiced consonants, and in the negative range concerning the unvoiced consonants. Consider Table 4-11 below:

Table 4-11. Mean vowel length variation exhibited by the American native speakers, groups G-US and C-US in vowels preceding the six voiced and unvoiced plosives. One standard deviation is given in the brackets.

	Following consonant					
	b	**d**	**g**	**k**	**p**	**t**
Americans	78.43	63.86	86.37	-39.78	-37.24	-31.78
sd	(54.25)	(55.45)	(59.02)	(34.59)	(29.89)	(23.81)
G-US	42.23	52.02	52.15	-24.72	-18.75	-21.92
sd	(59.42)	(74.44)	(59.17)	(29.09)	(35.48)	(47.06)
C-US	25.72	30.25	28.43	-12.05	-10.78	-5.14
sd	(48.20)	(53.46)	(38.73)	(46.34)	(34.11)	(80.97)

The results from the table are quite expected for native speakers. Vowels preceding voiced consonants are lengthened, while vowels preceding voiceless consonants are shortened. The table also gives a first indication that groups G-US and C-US are not as proficient. Figure 4-38 below illustrates group G-US's production of vowel length variation in relation to the different consonants:

---

[61] Following/t/ and /d/ sounds also had an interesting effect in the results of group G-US in the perception experiment, which will be discussed in section 4.3.2.2.

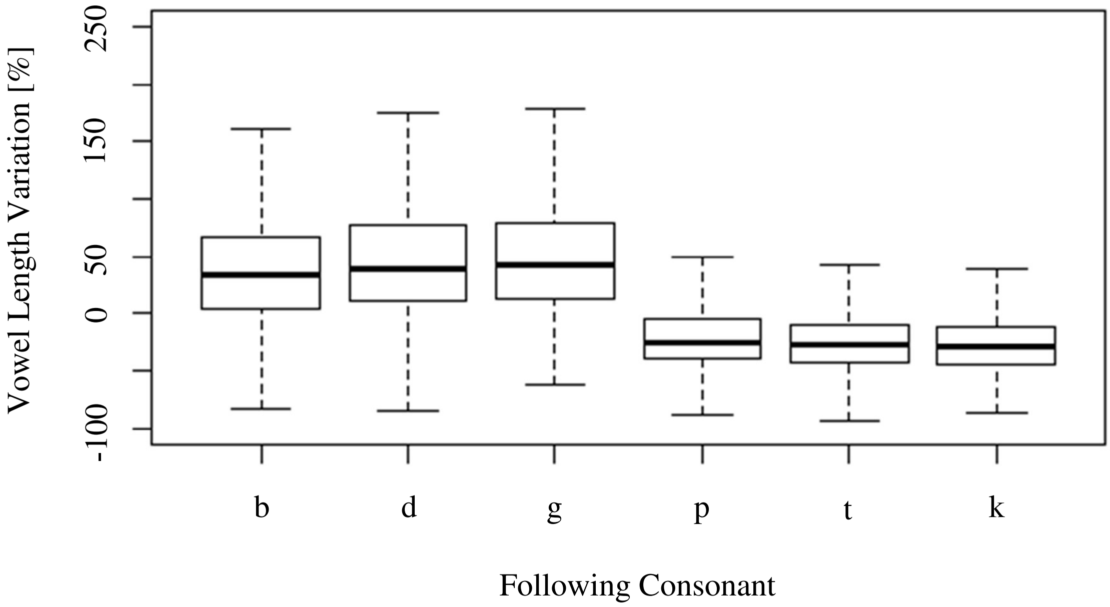


Figure 4-38. Mean vowel length variation exhibited by group G-US in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

Group G-US's variation is not unlike group G-UK's, although their variation is not as great as group G-UK's. This might be an indication that vowel length variation in production is more difficult to match for learners of American English than for those of British English. This is unexpected, because vowel length variation is usually considered to be more extreme in American English. However, from Table 4-10 and Table 4-11 (as well as from the graphs and tables in sections 4.2.1 and 4.2.2) we can see that this is not the case for the speakers in this study.

The degree of variation exhibited by group G-US is much lower than that of the American speakers. As was the case with the British speakers and group G-UK, group G-US matches the American speakers more closely in the vowel length variation preceding /d/ and /t/ than in the other consonants. However, the effect is statistically extremely small: $p < .0000001$ for /d/ and /t/ vs. $p < .000000000000001$ for the other four plosives.

The vowel length variation before the voiceless consonants is much weaker in group G-US compared to the native speakers. All of the upper whiskers of the boxplots relating to unvoiced consonants extend into the positive range, which indicates that in this group, too, a number of speakers pronounce a vowel preceding a voiceless consonant *longer* than one preceding a voiced consonant. For instance, Speaker #024 pronounced LOT in *cob* -50.04% shorter than in *cop* in the minimal pairs reading. In addition, the lower whiskers of the boxplots for the voiced consonants extend into the negative range. This is also the case for the American native speakers; however, it is more extreme in group G-US, where the most negative values lie almost on a straight line for five out of the six consonants. All these findings are very similar to the results from the British native speakers and group G-UK.

The standard deviations of group G-US exceed the mean in all six cases, which does not happen in the American native speakers' data. Thus, speakers in group G-US pronounce a number of tokens within the 68.3% limit of the data with inverted vowel length variation. Again, this is evidence that the learners in group G-US are not as proficient as the American native speakers. This effect is even stronger in the control group C-UK, as Figure 4-39 below illustrates:

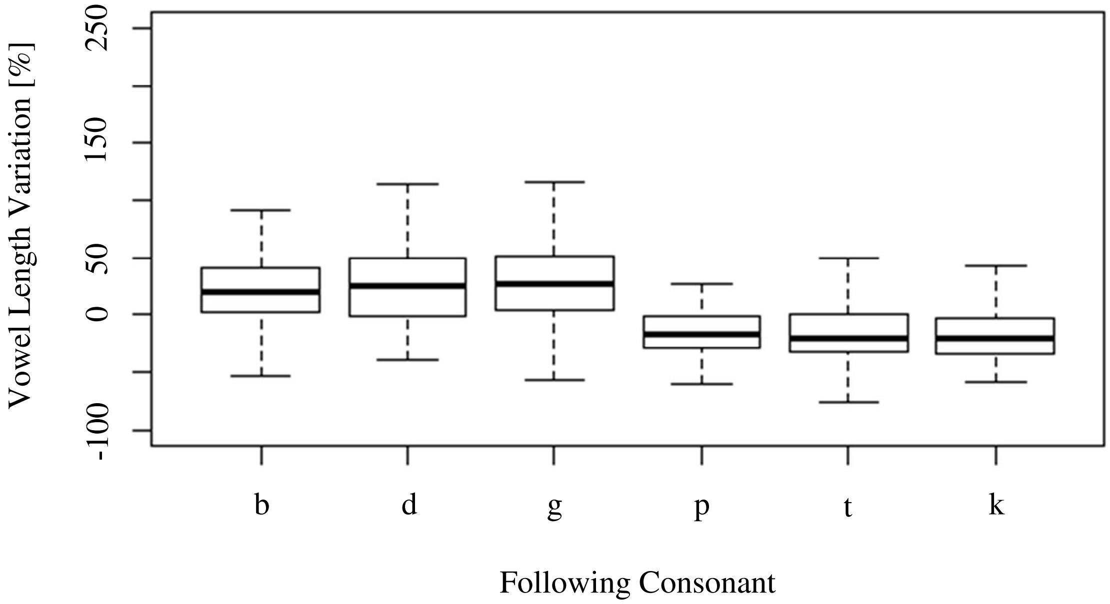


Figure 4-39. Mean vowel length variation exhibited by group C-US in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

The graph clearly shows the small range of variation in the six consonants. The medians lie between -12% and +30%, which is a bit higher than in group C-UK. As is the case for group C-UK, the variation preceding voiced and voiceless consonants does not differ much, which indicates that the speakers in both control groups lack a rule for vowel length variation. In order to be able to discuss group C-US in more detail, the results were again rescaled in a second graph:

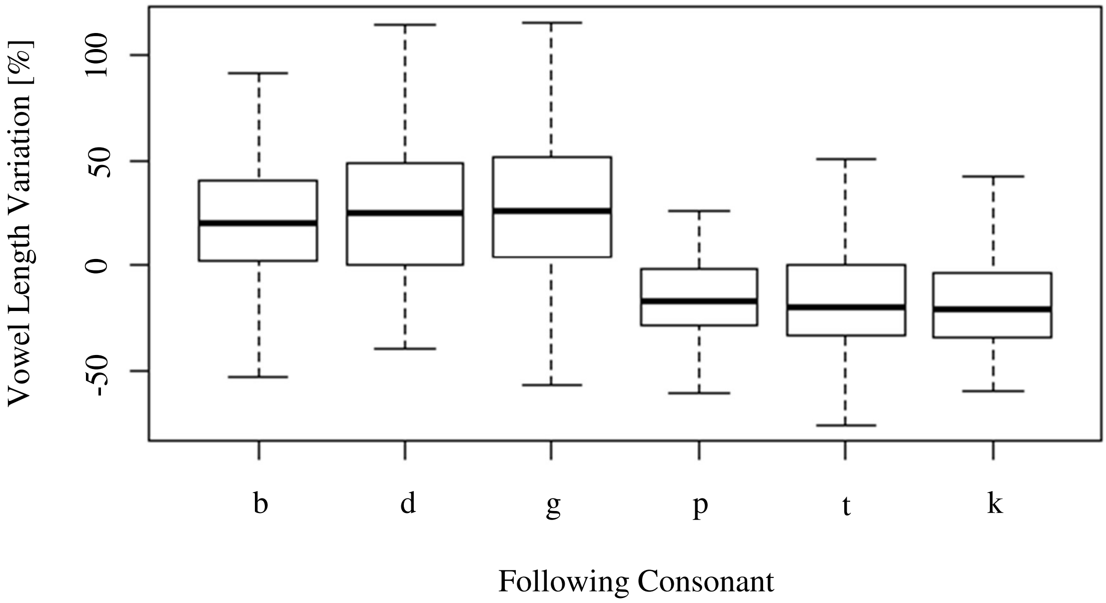


Figure 4-40. Rescaled figure illustrating the mean vowel length variation exhibited by group C-US in vowels preceding the three voiced and voiceless plosives. Numbers averaged over all vowels and tasks.

This new scaling shows that there are small differences in the vowel length variation preceding voiced and voiceless consonants. However, among the voiced consonants, variation does not differ significantly ($p = .48$). The difference in variation among the voiceless consonants does not amount to statistical significance, either ($p = .49$). In results similar to those for group G-US, the most extreme values of the voiceless consonants lie in the positive range and the most extreme values of the voiced consonants lie in the negative range. Thus, speakers produced tokens with inverted vowel length.

The overall weak variation of group C-US leads to this group's vowel length variation being significantly different from the American native speakers' variation for all six consonants (/d/ and /t/ $p < .0000000001$, all others $p < .000000000000001$). Although the probability values are very small, it is still noteworthy that the control group, too, is less statistically different in the variation in vowels preceding /t/ and /d/ than in the other four plosives. Thus, the hypothesis that vowel length variation is most easily acquired in vowels preceding /d/ and /t/ holds true. This is quite possibly due to the fact that /t/ is a very salient coda consonant due to its many allophonic realizations (aspirated, unreleased or glottalized). Since this effect also occurred in the perception experiment, it will be discussed in full detail later within that context (sections 4.3.1.2 and 4.3.2.2).

A last point to be addressed is the standard deviation, which exceeds the mean by far in all cases, particularly regarding the voiceless consonants (/t/ is the most extreme case). Once again, the control group is clearly the least proficient group.

4.2.4 Comparison of the British and the American Group

Comparing groups G-US and G-UK yields interesting results. With regard to KIT, G-US and G-UK show quite similar variation values; however, due to the variation values of the American group being much higher than that of the British group in all tasks, G-UK manages to match their model more closely, namely in three out of the four tasks (cf. Table 4-14).

Looking at DRESS reveals a similar picture to KIT. The American native speakers' variation is much greater than the British native speakers', except in the text reading task, where it is only slightly more (46.52% compared to 42.48%). Likewise, the variation exhibited by groups G-UK and G-US is similar, except in the word list (G-UK 47.12% vs. G-US 30.56%). This suggests that group G-UK is able to match their native speaker model more easily because the British native speakers do not exhibit as strong a variation as the American native speakers.

TRAP is a particularly fascinating vowel for several reasons. Firstly, the degree of variation exhibited by the American native speakers is unexpectedly low in this vowel, especially in the sentence and the text reading, where the British native speakers show significantly more variation. Likewise, group G-UK shows much greater variation that G-US. Thus, the German learner groups both imitate their native speaker model. Group G-US has much lower values than the American native speakers in the two formal tasks, but an almost identical value in the sentence reading (44.13%, compared to 44.29% of variation exhibited by the American native speakers). In the text reading, the German speakers even show stronger variation than their native speaker model (71.76% vs. 65.42%). This leads to group G-US not being significantly different from the native speakers in the two informal reading tasks. In contrast, group G-UK performs very similarly to the British native speakers, even in the text reading task, where the level of variation among the native speakers is particularly high (89.47% compared to 80.69% exhibited by group G-UK).

Both group G-UK and G-US differ significantly from their native speaker model in three out of the four tasks concerning LOT (the one task in which both groups manage to match the native speakers being text reading). In their production of this vowel, the British native speakers show more variation than the American native speakers in the two informal tasks again, which is quite surprising. Due to LOT being pronounced lower in American than in British English, one would expect the American native speakers to show greater variation than the British speakers. As before in TRAP, group G-UK shows a higher degree of variation in LOT than group G-US again. The following table is an overview of the performance of groups G-US and G-UK regarding the four different vowels and tasks. Significant values are in bold print and indicate that groups G-UK and G-US produced significantly different vowel length variation.

Table 4-12. Statistical significance of the difference between the vowel length variation exhibited by groups G-US and G-UK in the four different vowels and tasks. Calculated using Wilcox tests. Significant *p* values in bold print.

	KIT	**DRESS**	**TRAP**	**LOT**
Minimal pairs	**$p = .02$**	**$p = .01$**	**$p < .00001$**	$p = .31$
Word list	**$p = .02$**	**$p = .001$**	**$p < .0000001$**	**$p = .0001$**
Sentences	**$p = .04$**	$p = .19$	$p = .11$	**$p = .03$**
Text	$p = .06$	$p = .26$	$p = .52$	$p = .15$

We can see from the table that both groups differ mainly in the formal tasks. In addition, G-UK and G-US also differ significantly in the sentence reading task in KIT. In LOT, it is the word list and the sentences where the two groups perform differently. Taken together with the fact mentioned above – that British native speakers exhibit lower degrees of vowel length variation in almost all tasks and vowels – this explains why group G-UK matches their native speaker model in many more tasks than group G-US. The production of vowel length variation of the British versus the American native speakers is listed in the table below. Significant differences are again marked in bold.

Table 4-13. Statistical significance of the difference between the vowel length variation exhibited by the British and American native speakers in the four different vowels and tasks. Calculated using Wilcox tests. Significant *p* values in bold print.

	KIT	**DRESS**	**TRAP**	**LOT**
Minimal pairs	**$p = .00001$**	**$p = .00001$**	$p = .21$	$p = .72$
Word list	**$p = .001$**	$p = .35$	$p = .15$	**$p = .0008$**
Sentences	**$p = .03$**	$p = .36$	**$p = .02$**	$p = .17$
Text	$p = .24$	$p = .53$	**$p = .0002$**	**$p = .0001$**

In KIT and DRESS, the American native speakers produce more variation than the British native speakers in all tasks. However, in the case of TRAP, the British native speakers outperform the Americans in the two informal sentence and text reading tasks. In LOT, the British native speakers produce more variation in all tasks. The lower level of variation in TRAP and LOT among the American native speakers is still quite surprising, but might be one of the reasons why these are the only two vowels in which group G-US matches them in the production of vowel length variation.

Next, let us have a look at the effect of the vowel itself on the production of vowel length variation in the British group versus the American one.[62]

The vowel length variation exhibited by the American native speakers is not significantly different in the four vowels (KIT $p = .24$; DRESS $p = .18$; LOT $p = .73$). This is also the case for the British native speakers (KIT $p = .07$; DRESS $p = .08$; LOT $p = .23$). This means that vowel length variation is a fixed rule for native speakers, who have acquired it as children and subsequently use it for all vowels.

The picture looks quite different when we turn to the German learners. Group G-US shows significant effects in all three vowels (KIT $p = .02$; DRESS $p = .008$; LOT $p = .005$), as does group G-UK (KIT $p < .0001$; DRESS $p = .002$; LOT $p = .007$).

The two control groups are different again. Group C-US does not show any significant difference in the vowel length variation exhibited in the different vowels (KIT $p = .16$; DRESS $p = .68$; LOT $p = .93$), and neither does group C-UK (KIT $p = .46$; DRESS $p = .94$; LOT $p = .79$). This is certainly due to the fact that the control groups show only very low variation values and very similar ones across all tasks. Thus, it does not warrant the conclusion that the control groups are particularly proficient or match the native speakers well.

For a better overview, Table 4-14 shows a comparison of the British and American groups with regard to performance in the production experiment. It lists the tasks and vowels in which groups G-UK and G-US manage to match their native speaker model, that is to say, when they are *not* statistically different from the native speakers.[63]

---

[62] As the lowest vowel, TRAP was chosen to function as the intercept in the linear model. Of course, as the intercept it does not appear in the text above, because the other vowels are compared to the intercept, i.e. the vowel length variation exhibited in TRAP.

[63] C-UK and C-US are not listed since C-UK fails to match the native speaker model in any of the tasks and C-US only matches the American native speakers in the sentence and text reading task in TRAP.

Table 4-14. Overview of the tasks in which groups G-UK and G-US match their native speaker model, i.e. when they are *not* significantly different from the British and American native speakers, respectively.

	G-UK				**G-US**			
	KIT	DRESS	TRAP	LOT	KIT	DRESS	TRAP	LOT
Minimal pairs	✓	✓	✓					
Word list	✓	✓	✓					
Sentences	✓		✓				✓	
Text			✓	✓			✓	✓

The table illustrates that group G-UK seems to be more proficient over the spread of all vowels. Whereas group G-US only manages to match the American native speaker models in the two informal tasks in TRAP and the most informal one in LOT, group G-UK matches the British native speakers in all tasks in TRAP, three out of four in KIT, the two formal ones in DRESS and the least formal task in LOT. However, a straightforward assumption of greater proficiency would be a simplification of the facts. As was mentioned above, the British native speakers show less variation in KIT and DRESS, which might make it easier for group G-UK to match their native speaker model. However, in TRAP and LOT the variation exhibited by the British native speakers is greater than that of the American native speakers. This was completely unexpected. Moreover, it is quite surprising that even though the British native speakers show more variation in both vowels, group G-UK is able to match them in four out of four task in TRAP and one out of four tasks in LOT, whereas group G-US only matches the American native speakers in two out of four tasks in TRAP and one out of four tasks in LOT. Thus, the acquisition of vowel length variation seems to be aided by having both a native speaker model which shows strong, easily perceptible variation, and a model that does not exhibit too great a degree of variation for learners to match it. In addition, there seem to be other factors at play, for instance ego boundaries (learners are more set in their character and may be more hesitant to reproduce a particularly salient vowel length variation for fear of being ridiculed by native speakers if they overdo it), or even motor control issues.

A look at the performance of the control group might also help shed some light on this complicated situation. It should be remembered that the learners in the control group have had limited exposure to native speakers within their introductory pronunciation class, but have not stayed abroad in an English-speaking country for any amount of time. Therefore, they represent an earlier stage of learning than the speakers in groups G-UK and G-US.

With regard to all vowels and all tasks, group C-US exhibits more variation than group C-UK. There are only two instances where both groups perform similarly: In KIT (19.03% vs. 19.01%) and in LOT (23.83% vs. 23.28%), in both cases in the text reading. Moreover, group C-US manages to match the American native speakers in the sentence and text reading task in TRAP, while group G-UK fails to do so in any of the vowels and tasks. This seems to suggest that group C-US is more proficient than group C-UK. However, this might also just be an initial phase, seeing as the more advanced groups G-UK and G-US showed a completely opposite picture. It is also quite possible that there are other individual factors at play beyond a "simple" stay abroad. These will be discussed in the next section.

4.2.5 Other factors

It has become clear that a stay abroad is a significant factor in how well participants produce vowel length variation, but not the only one. The effect of the length of stay is not as easy to determine. In order to examine this factor more closely, participants were sorted into groups according to the following criteria:

- No stay abroad (*Home*)
- 4 – 6 months (*Half a year*)
- 7 – 12 months (*One year*)
- 13 –18 months (*One and a half years*)
- 19 – 24 months (*Two years*)
- 25 months + (*More than two years*)[64]

Subsequently, a linear model was applied, comparing the participants' production of vowel length variation to that of the native speakers. The effect is highly significant ($p = .0009$, $R^2 = .24$). The following order lists the performance of the different groups in comparison with the native speakers (from least similar to most similar):

> *Home ← One year ← Half a year ← One and a half years ← More than two years ← Two years*

This sequence seems slightly confusing at first. One would think that the performance should improve, and thus match the native speakers' more closely, the longer participants stayed abroad. Part of the sequence fits this hypothesis, too: Participants of the control groups C-UK

---

[64] The longest stay abroad amounted to 96 months.

and C-US perform worst and are highly significantly different from the native speakers (*p* < .00000001). Participants who stayed abroad for half a year perform worse than those who stayed one and a half years and also differ significantly from the native speakers (*p* = .05). In contrast, participants in group O*ne and a half years* are not statistically different from the native speakers (*p* = .24). The two groups *Half a year* and *One and a half years* are outperformed by participants who stayed abroad longer than two years. Participants in the latter group are not statistically different from the native speakers (*p* = .64).

Participants from groups *One year* and *Two years* do not seem to fit into the sequence proposed above, however, since participants from group *One year* performed worse in the production experiment than those with a stay abroad of only half a year. In addition, participants who stayed abroad for one year also differ significantly from the native speakers (*p* = .0003). Likewise, participants who stayed abroad for two years performed better in the experiment than those who stayed even longer than two years. Moreover, participants with a stay abroad of two years are the most similar to the native speakers (*p* = .98).

In summary, the sequence above suggests that generally, a longer stay abroad improves participants' performance in the production experiment. After a stay of one and a half years, participants did not perform statistically differently from the native speakers anymore. However, the effect is not linear, which indicates that another factor must be at play. In the case of the participants in this study, what needs to be taken into account is not just how long they stayed abroad, but also *why*. This was one of the questions posed in the questionnaire. The reasons listed by the participants can be grouped as follows:

- *Home*: Participants from groups C-UK and C-US who have no experience abroad
- *School*: Participants who went abroad during school, usually during grade 10
- *University*: Participants who completed a stay abroad during their studies at university
- *Family*: Participants whose families moved to or were stationed in the UK or USA[65]
- *Foreign language assistant* (*FLA*): Participants who went abroad during the course of their university studies in order to be a foreign language assistant at an English or American school[66]
- *Travel*: Participants who traveled through the UK or the USA

---

[65] Two out of the six participants in this group grew up in an English-speaking country, with German being the language spoken at home and within the family. The other participants were ten and seventeen years old when they moved to the UK/USA. All but one participant have one native English-speaking parent.

[66] Participants in this group taught German to English-speaking students.

As in previous cases, a linear model was employed to compare the individual groups' performance in the production experiment to that of the native speakers. The results are exemplified by Figure 4-41:[67]

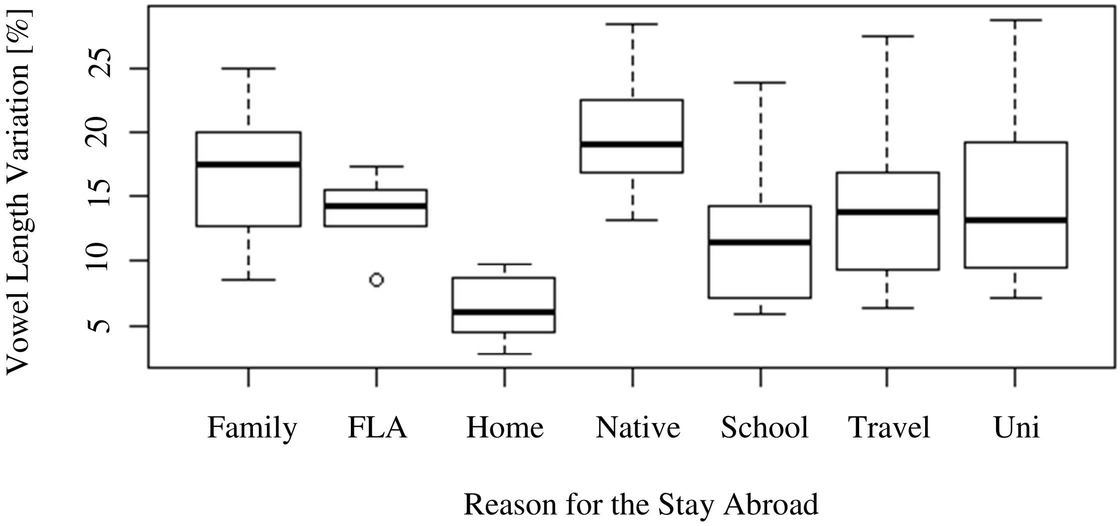


Figure 4-41. Vowel length variation exhibited in the production experiment in relation to the reason for which the speakers stayed abroad. *Home* indicates the control groups who did not stay in an English-speaking country; group *Native* comprises the native speakers. Numbers averaged over all vowels and tasks. $p = .0002$, $R^2 = .32$.

The results from Figure 4-41 can be summarized in the following sequence (groups ordered from least similar to most similar to the native speakers):

Home ← School ← Travel ← FLA ← University ← Family

This sequence is expected. It is not surprising that the participants in the control groups, who have no experience abroad, perform worst in the production experiment and are highly significantly different from the native speakers ($p < .0000000001$). The control groups are followed by participants who went abroad during school ($p < .00001$). The reason why this group performs so differently from the native speakers might be that the time of the stay abroad lies farther in the past for participants of this group than all the others. On average, their stay abroad was 6.5 years ago (*sd* = 2.4).

Participants who went to the UK or USA for travel performed only slightly better than those who went there during school, and they are also very different from the native speakers ($p < .0001$). The argument here may be that these participants came into contact with the

---

[67] Note that the comparison between the groups is not based on the median alone, but on an overall comparison of the performance of all speakers in each group. Therefore, although group *Travel* has a higher median than group *University*, it is less similar to group *Native*.

English language mostly for conversational purposes, for instance in restaurants, hotels or during sightseeing. This is quite different for the next two groups, whose members went abroad to be an FLA or to continue their studies at a partner university. Participants who worked as an FLA performed significantly worse than the native speakers ($p = .0009$), as did the speakers in group *University* ($p = .008$). Group *FLA* also showed less vowel length variation than group *University*. In contrast to the participants who merely traveled, participants who went abroad to be an FLA or to study at university also came into contact with academia, which requires a different type of English skills (more on this below). Therefore, they can be said to have “delved deeper” into English, which has evidently helped them to be more proficient in reproducing vowel length variation.

The most proficient group and the only group which is not statistically different from the native speakers is group *Family* ($p = .15$). Two speakers in this group spent their first few years in the UK/USA; three of them went there later in their life in order to visit family. All five participants have one native English-speaking parent. However, for all of them German is their dominant language and the language used within the family. The remaining sixth participant’s father was stationed in the UK for two years. Due to the fact that five out of the six participants in this group are bilingual (even though German is their dominant language), it is not surprising that this group matches the monolingual native speakers most closely in the production task.

Of course, the reason for the stay abroad might conceal highly individual factors such as motivation, which was not and cannot be easily measured. However, it stands to reason that participants who study English and want to become proficient both because they enjoy the language and because they want to be teachers or perhaps want to improve their chances on the job market have both intrinsic and extrinsic motivation, while participants who went abroad during school were looking more for a good time and an adventure. Moreover, five out of six participants who went abroad because of their family have a native English-speaking parent, which probably gives them a completely different level of identification with the English language than participants of other groups. Thus, while the reason for the stay abroad was shown to be a major factor in the acquisition of vowel length variation, we must bear in mind that other highly individual psychological factors will always also play a role.

Having examined the reasons why participants stayed abroad, we can now return to the effect of length of the stay. We should recall that the analysis showed that participants who stayed abroad for half a year performed better than those who stayed abroad for a year, and those who stayed longer than two years were outperformed by those who stayed two years. This can be explained when we look at the distribution of participants and their reason for the stay abroad in the individual groups:

Table 4-15. Distribution of participants in the five categories for the reason of their stay abroad.

	Reason for the stay abroad				
	Family	*University*	*FLA*	*Travel*	*School*
Half a year	1	12	5	2	3
One year		9	1	7	8
One and a half years		2		1	1
Two years	3	1			
More than two years	2	1			

Table 4-15 shows that the group whose members stayed abroad for half a year comprises 23 participants. Out of those, one went abroad to visit family, 12 to complete their studies at a partner university, and five worked as foreign language assistants. Thus, almost 80% this group consists of participants who were shown to be the three most proficient groups in comparison with the native speakers. The group who stayed abroad for one year comprises 25 participants. In contrast to the group who went abroad for only half a year, however, there are only nine participants who went abroad to a foreign university and one FLA in this group. In addition, seven participants went for travel; 8 during school. This means that only 40% of participants in this group belong to the more proficient groups.

Regarding the groups with stays abroad of two years and more than two years, the situation is similar. The first group consists of four participants; the second one has three participants. In the group of two years, three speakers went abroad to visit family; two of them spent their childhood years abroad. The fourth speaker attended university in England. The group of speakers who stayed abroad for more than two years shows a different layout: Only two participants went abroad to visit family (when they were in their late teens), one studied at different universities in the USA. Therefore, the group who stayed abroad for two years consists of three speakers from the most proficient group and one from the second most proficient group; the group who stayed abroad for more than two years comprises only two speakers from the most proficient and one speaker from the second most proficient group.

In summary, these results suggest that it is not purely the length of stay which makes a non-native speaker successful in reproducing vowel length variation accurately. Instead, the reason why the speaker went abroad and the ensuing depth of exposure to the target language must be considered. This section already alluded to the fact that the participants in groups *University* and *FLA* came into contact with academic English as well as conversational English, while participants from groups *Travel* and *School* possibly had more contact with conversational English. Although participants who went abroad during their school years obviously also attended school abroad, the exposure they had to academic English was most probably limited to when they wrote essays in English. At university, however, students are confronted with academic English every day and write many essays, term papers and exams in English. As an FLA, participants teach German to native-English speaking students at school, which also requires the ability to make use of abstract linguistic knowledge and possibly to deal with students' written work. A theory which considers this is the *BICS* and *CALP* model by Cummins (1979; 1991). Cummins argues that there are two different areas of language skills a learner needs to master: Basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP). BICS are needed for everyday conversation (oral proficiency). These skills can be acquired within the first two years of learning the L2. However, there is a large area beyond simple conversational skills, which takes much longer to master – about five to seven years. Cummins calls these literacy-related skills, which are necessary to deal with the cognitive processing of written texts. In 1991, Cummins re-named the two skill sets *Conversational* and *Academic Language Proficiency*.

BICS and CALP differ in the way they are used in L1 and and L2. BICS are highly language-specific, because they are used in content-embedded social interaction. They draw on vocabulary and grammar structures specific to the L2. In contrast, CALP can be transferred from the L1 to the L2, since these skills involve meta-linguistic knowledge such as evaluation, synthesizing, comparing and analyzing. However, although the skills per se can be transferred, the specific vocabulary needed to express thoughts and ideas in the L2 needs to be acquired by the learner, too. Figure 4-42 below illustrates the interaction between BICS and CALP:

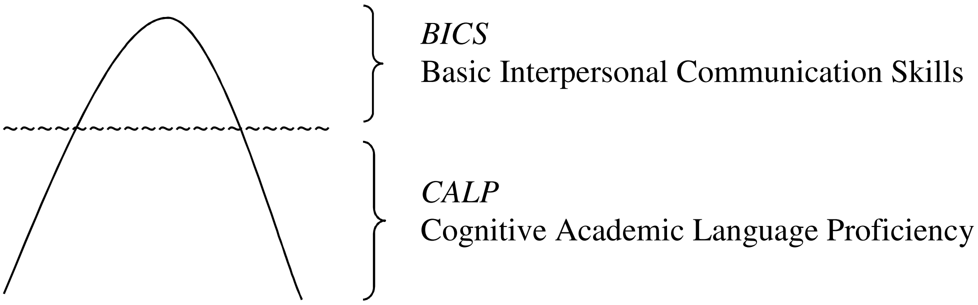


Figure 4-42. Cummins' *BICS* and *CALP* model.

With regard to the present study, Cummin's theory explains why participants in groups *University* and *FLA* manage to produce native-like vowel length variation more successfully than speakers in groups *School* or *Travel*. As mentioned before, speakers who traveled around the UK or USA used English predominantly for social, everyday interaction. Therefore, the skills they used are mostly BICS. Participants in group *School* might have used some CALP skills, too, although not to the same extent or on the same level as participants from groups *FLA* and *University*. The latter two groups dealt with academic situations almost every day during their stay abroad. Thus, a large part of their interaction drew on BICS as well as CALP. This means that they had extensive exposure to both context-embedded and academic English. This deeper contact with all levels of the L2 is very probably what gives them an advantage over the less experienced groups.[68]

Section 2.2.2 introduced how word frequency affects vowel length variation. The analysis of the production result shows that this is a factor for all groups – native speakers and German learners. However, the extent to which word frequency helps determine the length of the vowel produced differs among the groups. As such, it is a stronger factor for group G-US (p = .0005) than for the American native speakers (p = .004). Interestingly, the control group C-US remains completely unaffected (p = .26). In the British group, the German learners (G-UK) show stronger influence (p = .002) than the British native speakers (p = .02). Group C-UK is not influenced by word frequency (p = .13). These results might be due to the advanced learning stage of groups G-UK and G-US in comparison to C-UK and C-US. Since the speakers in the control group have had more limited interaction with native speakers of English, they might not have stored more frequent words so that they are easier and quicker to recall. In contrast, participants in groups G-UK and G-US who have had considerable exposure to native speakers of English might have experienced the factor of word frequency

---

[68] It is clear that speakers from group *Family* are most proficient, since they were exposed to English from infancy on, even if German is their dominant language.

during their stay abroad – they might have used some words much more often than others. As a result, frequent words might be stored where they are more easily accessible – similar to the way a native speaker would store them. Therefore, the native speakers as well as groups G-UK and G-US show sensitivity to word frequency, whereas groups C-UK and C-US do not. However, there is also a chance that this result was caused by the small group size of groups C-UK and C-US or their overall very low levels of variation.

An interesting question is, of course, whether speakers who rate themselves as more proficient actually perform better than those who doubt their abilities. This is indeed the case in groups G-UK ($p = .006$) and G-US ($p = .05$). This factor becomes even stronger when the two groups are pooled together ($p < .0000001$). In contrast, the importance participants ascribed to achieving a native-like accent did not influence their performance in a significant manner ($p = .19$). However, the two factors interact: speakers who found it highly important to attain a native-like accent also judged their own accent to be better $p < .007$, $C = .72$). In addition, participants who speak English predominantly with native speakers performed significantly better in the production experiment than those who speak English more often with other non-native speakers ($p = .007$).

Many results have been presented in connection with the production experiment. A summary can be found in the final section of this chapter (4.6). First, however, let us have a look at the findings of the perception experiment.

4.3 Perception Experiment

The question to be answered in this section is: Which learners successfully perceive vowel length variation?

A first interesting finding is that the gender of the participants did not have a significant influence on their performance in the perception experiment ($p = .26$). Numerous scholars (cf. Zaidi 2010: 40 f.; Alonso-Nanclares *et al.* (2008)) have argued that females perform better in perception tasks than men. This hypothesis is not supported by the present study. Figure 4-43 shows two boxplots comparing the performance of female and male participants in the perception experiment:

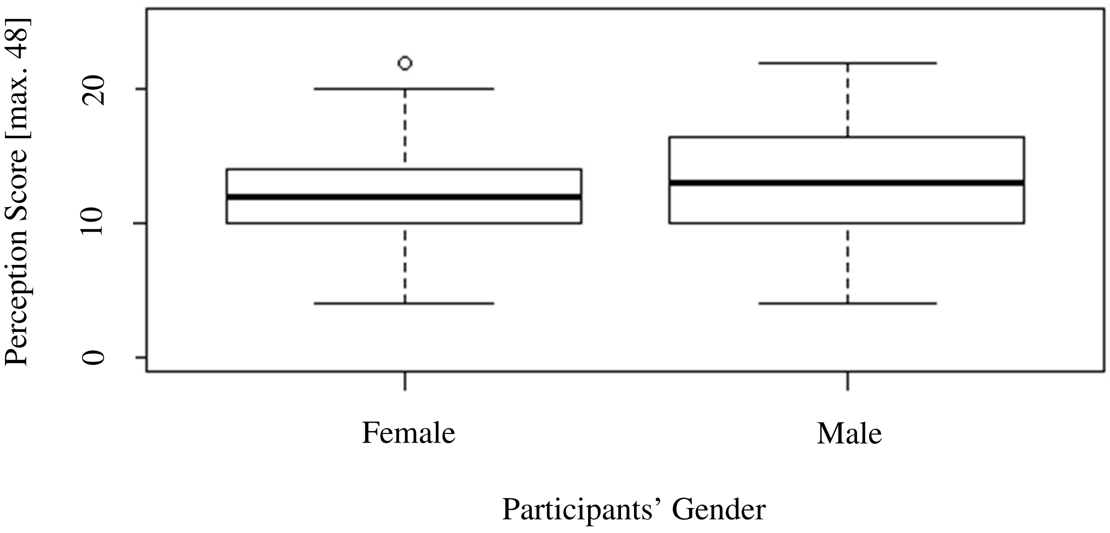


Figure 4-43. Perception score (number of incorrect answers in the perception experiment) in relation to the participants' gender. *p* = .26.

The boxplots show that females and males performed similarly. On average, females made 11.9 mistakes in the perception experiment (*sd* = 3.8), males, 12.5 (*sd* = 4.8).

The importance participants ascribed to having a native-like accent and the way they rated their own accent was not found to be significant (*p* = .09 and *p* = .15). There was also no effect of active and passive language use (*p* = .07 and *p* = .65, respectively). Likewise, learners who predominantly conversed with native speakers of English did not perform significantly better in the perception experiment than those who spoke more often with other non-native speakers (*p* = .29). Therefore, these factors will henceforth be ignored.

Interestingly, a stay abroad did not have as clear or as strong an effect on perception as it did on production. As before, participants were sorted into the following groups depending on the length of their stay abroad:

- No stay abroad (*Home*)
- 4 – 6 months (*Half a year*)
- 7 – 12 months (*One year*)
- 13 –18 months (*One and a half years*)
- 19 – 24 months (*Two years*)
- 25 months + (*More than two years*)

The result of the analysis lists the groups according to their similarity to the native speakers in the perception experiment (from least to most similar):

One year ← One and a half years ← Half a year ← Home ← More than two years ← Two years

It is easily visible that there is no true pattern, with the control groups without any stay abroad performing better than learners who stayed abroad for up to one and a half years. Moreover, only the group who stayed abroad for one year is significantly different from the native speakers ($p = .006$). However, the two most proficient groups are the ones who stayed longest.

Regarding the reason for the stay abroad, the picture is similarly unclear. Let us recall that participants were distributed into the following groups:

- *Home*: Participants from groups C-UK and C-US who have no experience abroad
- *School*: Participants who went abroad during school, usually during grade 10
- *University*: Participants who completed a stay abroad during their studies at university
- *Family*: Participants whose families moved to or were stationed in the UK or USA
- *Foreign language assistant* (*FLA*): Participants who went abroad during the course of their university studies in order to be a foreign language assistant at an English or American school
- *Travel*: Participants who traveled through the UK or the USA

The performance in the perception experiment of the different groups was again compared to that of the native speakers. The results are as follows (again sorted from least similar to most similar):

Travel ←University ← FLA ← Home ← School ← Family

Again, no clear picture emerges from the analysis. Only groups *Travel* and *University* performed significantly worse than the native speakers ($p = .0006$ and $p = .005$). It thus seems that perception is not significantly influenced by a stay abroad. Instead, it might be affected more strongly by age. As we can see, the most similar group to the native speakers is group *Family*, which comprises subjects who spent their first years abroad and were therefore confronted with English from early age on. The second most proficient group is group *School*. The participants in this group went abroad during their teenage years. The other groups either have not had long-term exposure to native speakers (*Home*) or only went abroad only after they were 20 years old (all other groups). This interpretation must be considered with caution, however, as it is based on little evidence. Moreover, it has often been suggested that phonological sensitivity decreases even in very young children, which might mean that the participants in group *School* have reached this level of proficiency by chance rather than their earlier stay abroad being the causal factor.

The next two sections will analyze the results more closely. Particular focus will lie on the question of which words in the perception experiment were identified correctly with greater or lesser frequency and which factors account for the differences of performance in the learner groups. Section 4.3.1 will report the results of the British Group; section 4.3.2 deals with the findings from the American group.

4.3.1 British Group

4.3.1.1 British Native Speakers

This section analyses the British native speakers' responses to the perception experiment in more detail. Since this group consists of native speakers, it might be surprising that not all of them performed equally well in the experiment (mean mistakes $\overline{x} = 12.5$, $sd = 4.8$). However, there are some results which apply for the group as a whole.

First of all, the frequency of the word does not make a statistical difference ($p = .21$), meaning that more frequent words were not identified correctly significantly more often than infrequent ones. This suggests that native speakers acquire vowel length naturally; it is not a learned feature. This stands in stark contrast to the results for the German learners, as will be discussed in the following section.

The vowel lengths of the words produced by the speakers who recorded the tokens for the perception experiment were also measured to calculate the variation value between the minimal pairs used in the perception experiment.[69] This was done to assess whether this variation was a factor in the number of tokens identified correctly by the participants in the experiment. Since ten native speakers of British English completed the experiment, a maximum of ten incorrect answers can occur for each token. As Figure 4-44 shows, variation does indeed help British native speakers identify the words in the perception experiment more easily. What is interesting, however, is that the shortening of the vowel leads to higher success rates than lengthening does:

---

[69] All vowel lengths and corresponding variation values can be found in Appendix F.

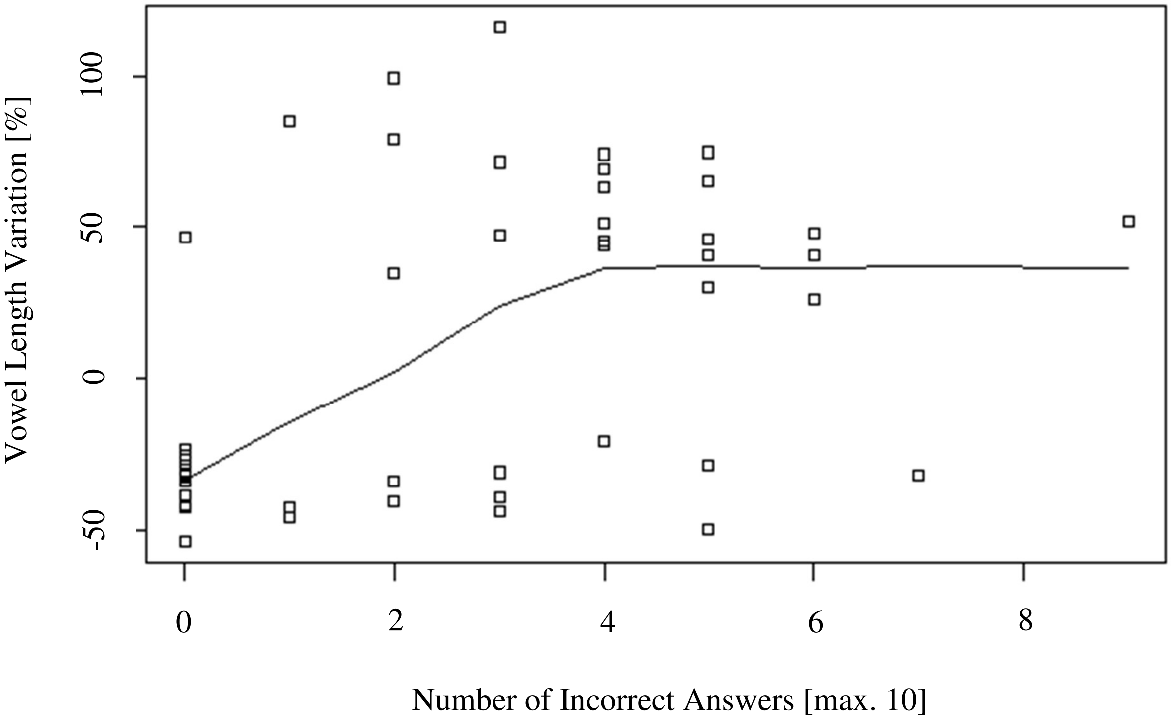


Figure 4-44. Number of British native speakers who identified tokens incorrectly in the perception experiment in relation to the vowel length variation exhibited by the speakers of the tokens. Since there are 10 British native speakers, this is the maximum number of incorrect answers which can occur. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .00001$, $C = .64$.

As can be seen from Figure 4-44, the strongest increase in incorrect answers occurs between a shortening of -50% and a lengthening of about +40%. This cutoff point at +40% is quite clear in comparison with the American native speakers, where no fixed cut-off point can be established (cf. Figure 4-50). The next section 4.3.1.2 will show that the German group behaves quite differently, too.

Coinciding with the result above is the fact that the British native speakers identified tokens correctly more often when a voiceless plosive followed. Most mistakes were made when /d/ was the coda consonant ($\overline{x} = 4.875$),[70] followed by /b/ ($\overline{x} = 4.375$), /g/ ($\overline{x} = 2.625$), /p/ ($\overline{x} = 2.125$), /k/ ($\overline{x} = 1.625$) and /t/ ($\overline{x} = .75$). It is very probable that the error rates for vowels preceding /t/ are so low because /t/ is often glottalized or unreleased. Therefore, speakers might be more used to cuing in to other factors (e.g. vowel length) when dealing with vowels preceding /t/. This also ties in with the result that most mistakes were made when /d/ followed. This consonant is not usually unreleased; it might therefore have been confusing for

---

[70] This has to be read as: In each of the 8 trials where the token ended in /d/ (4 vowels • 2 appearances), an average of 4.875 out of the 10 native speakers chose incorrectly. If the average should fall below zero, this means that the total number of mistakes in the 8 trials was below 8. In the German learner groups G-UK and G-US, the number of speakers who can choose incorrectly is of course 30; the 8 trials for each consonant remain.

the native speakers that the consonant was inaudible due to it having been cut off before the experiment.

Overall, the effect of the consonant is highly significant ($p < .000000001$) and of medium strength ($\varphi = .328$). This even holds true when looking at voiced and voiceless consonants separately. The fact that fewer mistakes were made before /g/ than before /d/ and /b/ is also highly significant ($p = .009$) but the effect is not as strong ($\varphi = .197$). The finding that /t/ has such a low incidence of mistakes compared to /p/ and /k/ is also significant ($p = .04$), but only to a small effect ($\varphi = .159$).

With regard to the different vowels, an average of 3.75 mistakes were made with KIT, 2.66 with DRESS, 2.83 with TRAP, and 1.66 with LOT.[71] The fact that fewer mistakes were made when LOT was involved is highly significant ($p = .004$), although the effect is rather small ($\varphi = .166$).

The sex of the speaker does not make a statistical difference for the British native speakers ($p = .15$). This is again a finding which differentiates the native speakers from the German learners in group G-UK (see below). Group C-UK's results will be mentioned in relation to group G-UK's, as C-UK's small group size does not lead to robust findings.

4.3.1.2 G-UK / C-UK

In comparison with the British native speakers, group G-UK behaves quite differently. The first obvious difference is that word frequency, which did not play a role for the native speakers, is a highly significant factor for group G-UK ($p < .0001$). This means that the more frequent the token is, the more likely it is identified correctly by the German listeners. Consider Figure 4-45 below:

---

[71] This has to be read as: In each of the 12 trials where the vowel appeared (twice before each of the six consonants), an average of *x* speakers out of 10 native speakers chose incorrectly. The rest remains as described in footnote 73.

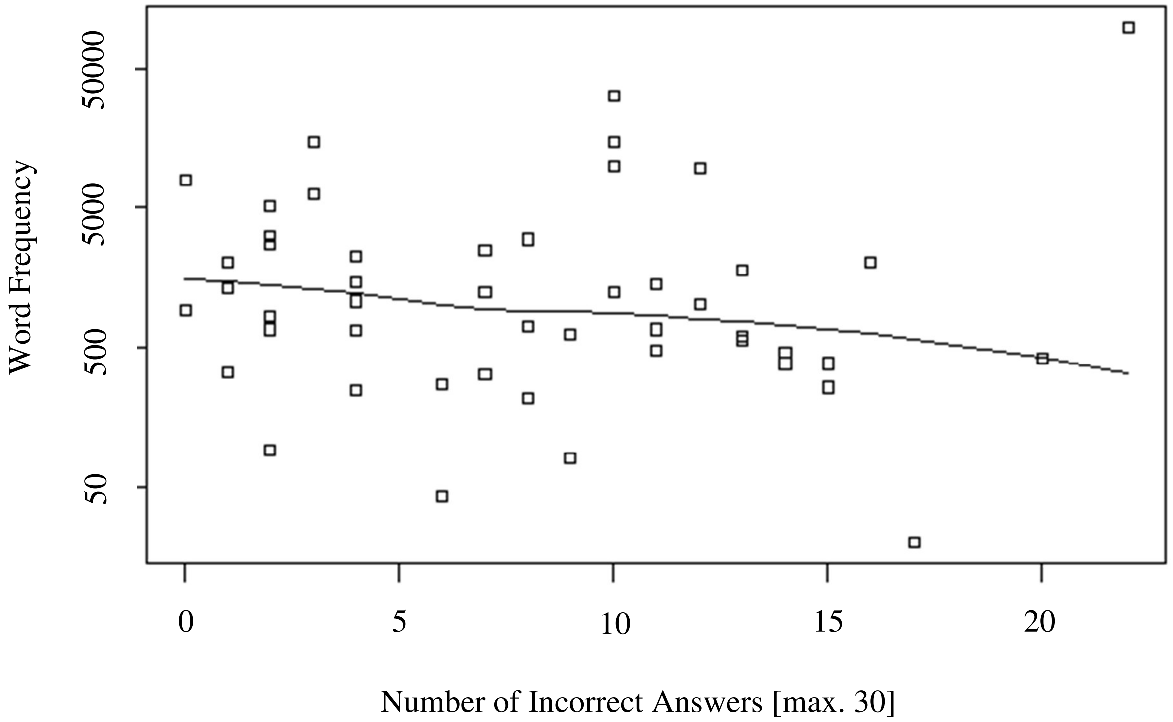


Figure 4-45. Number of tokens identified incorrectly by group G-UK in the perception experiment in relation to the frequency of the tokens in the BNC. Since there are 30 speakers in group G-UK, a maximum of 30 incorrect answers can occur. The y-axis is logarithmically scaled. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .0001$, $C = .62$.

Figure 4-45 shows quite clearly that group G-UK makes more mistakes with words that are less frequent. This is probably due to the fact that they mimic native English speakers. The more frequent a word is, the more likely it is that learners have heard it before. Naturally, the more often learners hear a certain word pronounced, the more easily they can store it in their mental lexicon. Thus, vowel length variation is a byproduct of the phonetic output of the native speaker. This means that German learners are able to not only reproduce the vowel length variation of high frequency tokens more easily, but also to identify it correctly more often when they hear it again. Unfortunately, due to small group size, which results in a maximum number of 3 incorrect answers, the statistics of group C-US are not robust in this case ($p = .96$).

There is one exception in group G-UK, which can be seen in the upper right hand corner of the graph. *Back* was the most frequent word in the BNC (96,989 occurrences), but it was identified incorrectly by 22 out of 30 participants. The vowel in this word is 49.78% shorter than the vowel in the minimal pair word *bag*. Other words with a lower variation value (24.68% in *back*/*bag* vs. 19.58% in *lack*/*lag*) were identified correctly more successfully: only 1 participant misheard *lag*, 10 misidentified *lack*. It is therefore very probable that the high error rate in *back* is due to semantic influence, since the word pair *back*/*bag* was inserted into a longer sentence frame in order to counteract a fatigue effect which might affect listeners after a longer stretch of listening (cf. section 3.3). Thus, the sentence frame “Can you put it in the ____?” might have semantically favored the word *bag*. This also happened with the British native speakers, 5 of whom (i.e. 50%) misheard *back* and 2 of whom (i.e 20%) misheard *bag*.

A second factor for the perception of group G-UK is the degree of variation exhibited by the speakers in the perception experiment. In contrast to the native speakers, in group G-UK vowel lengthening leads to fewer mistakes, as Figure 4-46 shows:

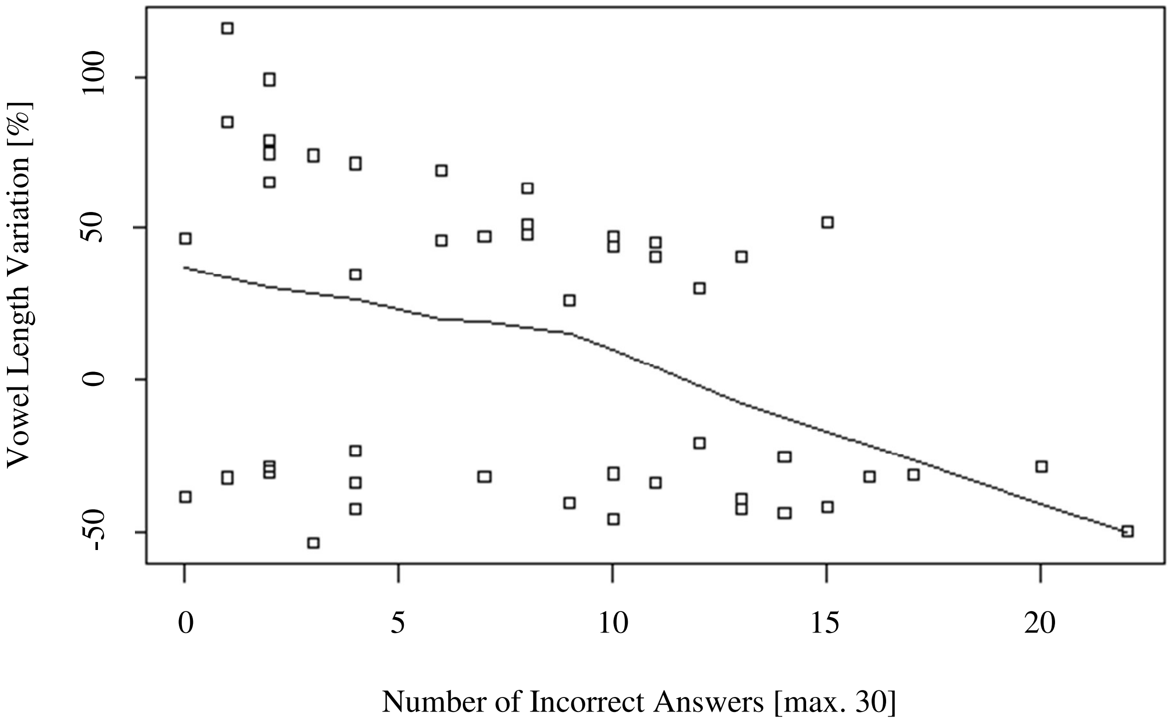


Figure 4-46. Number of tokens identified incorrectly by group G-UK in the perception experiment in relation to the frequency of the tokens in the BNC. Since there are 30 speakers in group G-UK, a maximum of 30 incorrect answers can occur. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .000001$, $C = .65$.

It is evident from Figure 4-46 that there is a strong link between successful perception and variation exhibited by the speakers of the tokens. This was also the case for the native speakers; however, group G-UK differs from the British native speakers in the way in which vowel length variation is processed. While more British native speakers identified words correctly in which the vowel was shortened, i.e. when a voiceless plosive followed, group G-UK made more mistakes in this environment. This finding matches the results proposed by van der Feest & Swingly (2011), who found that Dutch listeners, too, were more likely to make mistakes when perceptually presented with tokens whose vowel length was shortened due to a voiceless coda consonant (cf. 2011: 62). The same result occurs in group C-UK, with $p = .04$.

In the case of group G-UK, words ending in /p/ were misheard incorrectly most frequently ($\overline{x} = 13.125$), followed by /k/ ($\overline{x} = 11.125$). Interestingly, /t/ completely goes against this trend and has the second lowest error rate of all consonants ($\overline{x} = 4.5$). Fewer participants in group G-UK made mistakes when the coda consonant was voiced: On average, 8.875 participants misheard words ending in /b/, followed by 8.0 mistakes in words ending in /d/. Fewest mistakes were made when the word final consonant was /g/ ($\overline{x} = 2.5$).

Group C-UK replicates this finding to the letter: most mistakes were made in tokens ending in /p/ ($\overline{x} = 1.25$), followed by /k/ ($\overline{x} = 1.0$) and /t/ ($\overline{x} = .875$). Regarding tokens ending in voiced consonants, words ending in /b/ were misheard most often ($\overline{x} = .875$), followed by /d/ ($\overline{x} = .5$) and /g/ ($\overline{x} = .375$). Due to small group size, however, the overall effect of the following consonant is not statistically significant in group C-UK ($p = .22$, $\varphi = .22$)

The results of groups G-UK and C-UK with regard to the effect of the vowel directly relate to the vowel length variation exhibited by the speaker of the tokens. The case is particularly clear when a voiced consonant follows. The stronger the lengthening of the vowel before the voiced coda, the more likely German learners are to recognize the token correctly. Let us focus on group G-UK. The vowels were lengthened by 76.66% on average when /g/ followed, and an average of 2.5 learners misheard these words. Vowels preceding /b/ were lengthened by 50.66% on average; 8 learners clicked the wrong token. Finally, vowels preceding /d/ were lengthened by 47.11% on average, and an average of 8.875 participants answered incorrectly. This points to an almost linear relationship between vowel lengthening and success at perception, although three data points are certainly far too few to be conclusive.

Looking at vowel shortening paints a different picture. It might seem logical that the stronger the vowel shortening, the more participants would choose correctly, because the shorter length of the vowel would at some point trigger the perception of the learners in a way which prevents them from choosing a voiced coda consonant, simply because they tune in to the fact that a vowel preceding a voiced consonant should be longer. However, this formula *stronger vowel shortening = fewer mistakes* only holds true for vowels preceding /p/ and /k/, not /t/. Vowels preceding /p/ are shortened by 33.2% on average, and an average of 13.125 participants misheard these tokens. Vowels preceding /k/ were shortened slightly more

($\bar{x}$ = 42.48%), and fewer participants chose the incorrect token ($\bar{x}$ = 11.125). This ties in with the explanation offered above, namely that the more a vowel is shortened, the more participants favor the tokens with the unvoiced coda. Even though the difference in mistakes made by the participants is not significant ($p = .25$), the difference in vowel length variation is ($p = .02$). This suggests that German learners seem to be unable to cue in to vowel shortening in the same linear way as they did with vowel lengthening. The degree to which vowels are shortened (in the above example, roughly 30% vs. 40%) does not increase or decrease the likelihood of tokens being perceived correctly to a degree that is statistically significant.

It becomes interesting when we turn to /t/. Vowels preceding /t/ were shortened least ($\bar{x}$ = 31.34%), but only 4.5 participants on average clicked the incorrect answer. This probably stems from the fact that /t/ is a phoneme which has several allophones in coda position. It can be unreleased [t˺] or even glottalized [ʔ]. This renders /t/ in this position very salient. Learners are frequently confronted with these various allophonic realizations of /t/. Therefore, they might not have perceived /t/ as having been cut off at all. It is very likely that they perceived /t/ as being unreleased or glottalized, and thus found the choice between the voiced and the unvoiced option quite easy. This is supported by the fact that tokens which ended in /d/ were identified incorrectly most often out of the voiced options.

In summary, the effect of the consonant is highly significant in group G-UK ($p < .000000000000001$), and of medium strength ($\varphi = .274$). This also holds true when looking at voiced and voiceless consonants separately. Like the British native speakers, group G-UK made significantly fewer mistakes before /g/ than before /d/ and /b/ ($p < .00000001$). The strength of this effect is medium ($\varphi = .229$). In addition, group G-UK also performed better when the coda consonant was /t/ rather than /p/ or /k/. This effect is highly significant ($p < .0000000001$) and of medium strength ($\varphi = .264$).

With regard to vowels, most mistakes occurred with KIT ($\bar{x}$ = 8.83), followed by DRESS ($\bar{x}$ = 8.0), TRAP ($\bar{x}$ = 7.83) and LOT ($\bar{x}$ = 7.42). None of these vowels has a statistically significant lower or higher error rate, however ($p = .53$). Interestingly, group C-UK is different here: This group made most mistakes with LOT ($\bar{x}$ = 1.08), followed by KIT ($\bar{x}$ = 1.0), TRAP ($\bar{x}$ = .75) and DRESS ($\bar{x}$ = .42). The effect is not statistically significant, either ($p = .14$).

One very interesting finding is that group G-UK made significantly more mistakes when the tokens were spoken by the male speaker: On average, they made 9.71 mistakes, compared to 6.33 mistakes with the female speaker. Although this effect is small ($\varphi = .127$), it is highly statistically significant ($p < .00001$). This happened with neither the British native speakers ($p = .15$) nor the control group C-UK ($p = .26$). Interestingly, many German speakers mentioned after completing the perception experiment that the male speaker had been more difficult to understand, although they could not elaborate on this when asked for an explanation.

A last point to be discussed is the effect of the absolute length of the vowel. The methodology section explained that the tokens in the perception experiment were inserted into a sentence frame in order to make listening more natural. It is conceivable that if the tokens had been presented in isolation, participants would not have had enough time to perceive and process them. Instead, they knew that the token in question would always be presented at the end of the sentence. In addition, they were able to read the sentence from the screen. Furthermore, if they were not able to grasp the word the first time, they were allowed to listen to it several times. This methodology proved to be well chosen, since the analysis shows that group G-UK identified significantly more tokens correctly in cases where these were longer.[72] Consider Figure 4-47:

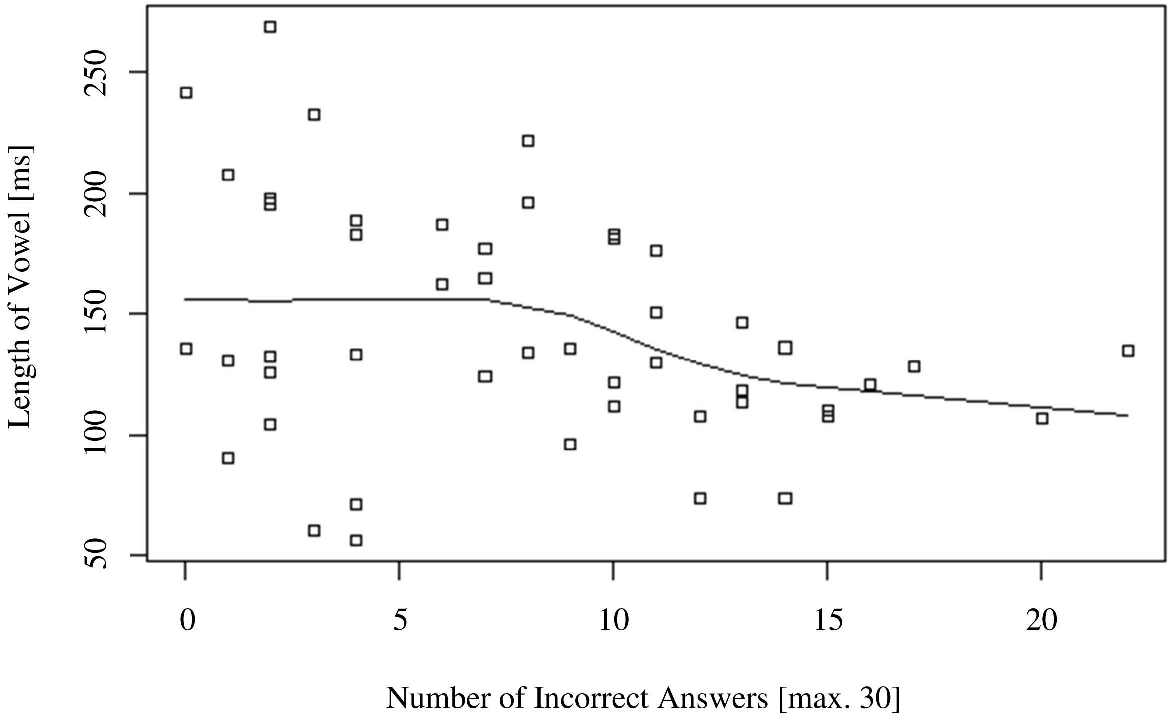


Figure 4-47. Number of incorrect answers given by group G-UK in the perception experiment in relation to the length of the vowels of the tokens (in ms). Since there are 30 speakers in group G-UK, this is the maximum number of incorrect answers which can occur. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .0001$, $C = .63$.

---

[72] This factor is not statistically significant for the British native speakers ($p = .11$) or group C-UK ($p = .21$).

Figure 4-47 clearly shows that group G-UK made significantly more mistakes with tokens whose absolute vowel length was longer. Thus, not only vowel length variation, but also the length of each separate vowel is an important factor which determines how successfully German learners identify tokens.

One might argue that this factor does not truly exist, but is a side-effect of the result that speakers of group G-UK identify more tokens correctly when the tokens show positive vowel length variation. This means that words which end in a voiced consonant have lengthened vowels, and are therefore naturally longer than a minimal pair word which ends in a voiceless consonant and has a shorter vowel. However, the length of the word is a separate factor, not a side-effect. In order to provide evidence for this, tokens were sorted into groups according to whether they ended in a voiced or voiceless consonant. Aftterwards, a generalized linear model was applied, first calculating whether words whose vowels were longer among the group of tokens ending in a voiced consonants were identified correctly more often than words with shorter vowels from the same group. Figure 4-48 presents the results:

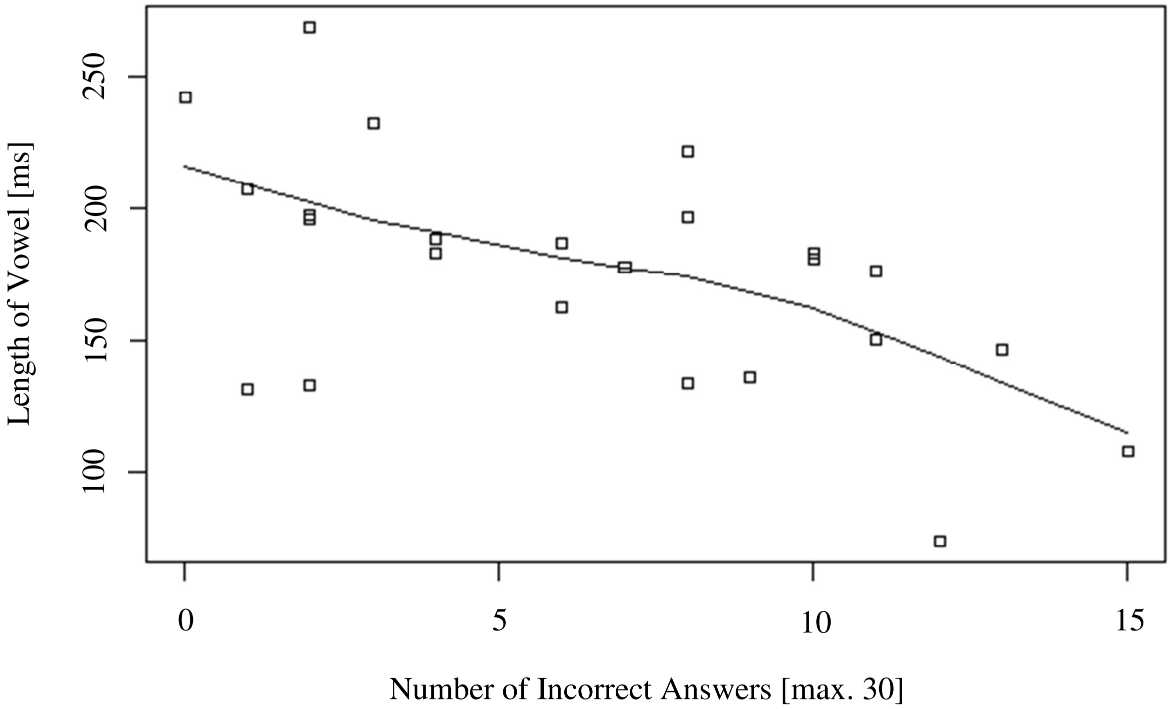


Figure 4-48. Length of the vowels of the tokens ending in voiced consonants in relation to the number of tokens which were identified incorrectly by group G-UK. Since there are 30 speakers in group G-UK, this is the maximum number of incorrect answers which can occur. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .00001$, $C = .71$.

As is easily visible from the graph, tokens which were longer in the group of tokens which end in a voiced consonant were identified correctly more often than those which were shorter.

The effect is highly significant ($p < .00001$) and quite strong ($C = .71$). This indicates that German learners indeed need more time to process vowels. Moreover, they seem to be quite sensitive to small changes in length, since the effect is almost linear. In contrast, word length is not a significant factor for the British native speakers ($p = .11$).

The opposite question, namely whether group G-UK also correctly identified more tokens which are shorter among the ones which end in a voiceless consonant was also considered. It might well be possible (and has been argued before) that tokens with very short vowels signal to participants that the vowel is too short to be considered a vowel preceding a voiced consonant. Therefore, the shorter the vowel, the more often participants should make the decision that the word in question ends in a voiceless consonant. Whether they truly did so is presented in Figure 4-49:

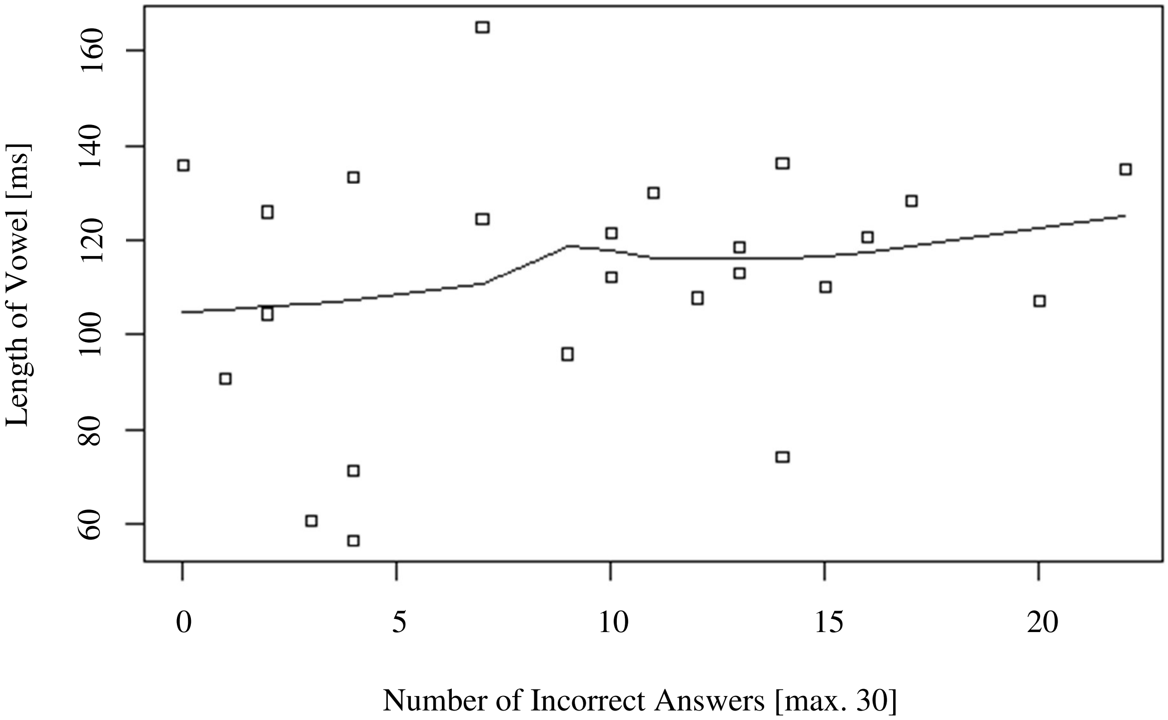


Figure 4-49. Length of the vowels of the tokens ending in voiceless consonants in relation to the number of tokens which were identified incorrectly by group G-UK. Since there are 30 speakers in group G-UK, this is the maximum number of incorrect answers which can occur. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p = .02$, $C = .56$.

The graph illustrates that the effect is not as significant and not as strong as with voiced consonants. However, a trend can still be observed. This indicates again that the German learners in group G-UK can deal more easily with longer words and with lengthened vowels. They cannot tune in as easily and are not as sensitive to small changes in shorter vowels. Although they increasingly chose the voiced consonant option as the vowels become longer, there is no true cut-off point for group G-UK. This stands in contrast to groups G-US, where a clearer cut-off point can be observed (cf. section 4.3.2.2), and group C-US, which shows the same pattern.[73]

4.3.2 American Group

4.3.2.1 American Native Speakers

This section reports on the American native speakers' performance in the perception experiment. With a mean mistakes score of 9.0 (*sd* = 3.9), the American native speakers performed better than the British native speakers. The following analysis will illustrate which factors account for the performance of the American speakers.

As with the British native speakers, word frequency is not a statistically significant factor for the American native speakers ($p = .21$). This again supports the hypothesis that the perception of vowel length variation is not a learned feature, but acquired naturally. Correspondingly, this stands in contrast to the behavior of group G-US (cf. section 4.3.2.2)

Regarding the vowel length variation exhibited by the speakers of the tokens, it is noteworthy that this factor is not as strong as with the British native speakers. Although larger or smaller variation leads to better performance, this effect is only moderately significant ($p = .05$). Overall, the American native speakers make significantly more mistakes when the vowel is lengthened ($p = .02$). Consider Figure 4-50 below:

---

[73] Neither factor is significant in group C-UK. Voiced consonants $p = .56$, voiceless consonants $p = .57$.

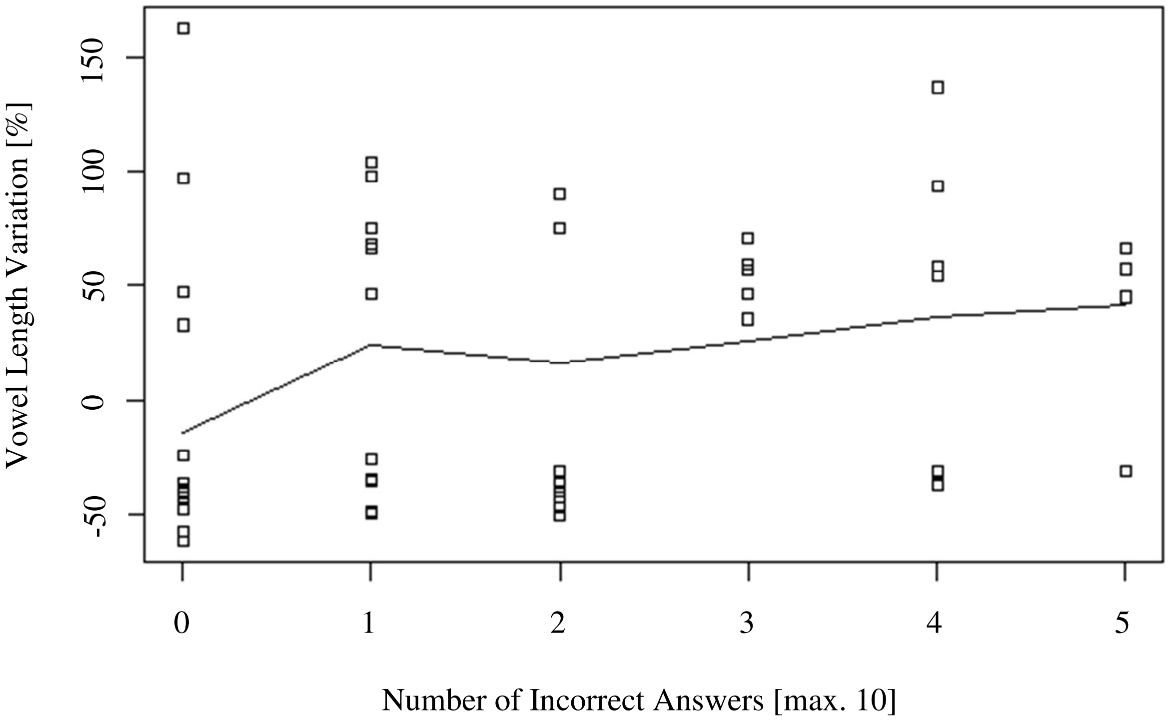


Figure 4-50. Number of American native speakers who identified tokens incorrectly in the perception experiment in relation to the vowel length variation exhibited by the speakers of the tokens. Since there are 10 American native speakers, this is the maximum number of incorrect answers. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .05$, $C = .61$.

The graph shows that a maximum of 5 out of 10 speakers identified tokens incorrectly. Moreover, it illustrates that the vowel length variation exhibited by the speakers of the tokens is only a moderately significant factor in how successfully words are perceived by the American native speakers. Although many identified tokens correctly when they were shortened by -50% (cf. bottom left hand corner of the graph), the same shortening led to 4 or 5 out of 10 speakers choosing the incorrect token of a different word pair (cf. bottom right hand corner). Likewise, a lengthening of +100% to +150% leads to 0 mistakes in some tokens (cf. upper left hand corner), but to 4 mistakes in another token. This finding might again be explained by semantics: In the latter case, i.e. where a lengthening of more than 130% results in four speakers identifying the word incorrectly, the token in question was *cob*, with the minimal pair token being *cop*. It is very probable that speakers are more familiar with the word *cop* (word frequency 10,042 vs. 503 for *cob*), and therefore opted to choose this token over the lesser known *cob*.

The second token in this category (4 incorrect answers, +93% lengthening) is *deb*, with the minimal pair token *dep*. These two tokens were suggested by a native speaker of American English as abbreviations for *debutant* and *department*, respectively. Both of these tokens might have struck speakers as uncommon; however, all ten speakers identified *dep* correctly. The word *deb* might still be less familiar, evidenced by a simple Google search. The term *deb* does not produce any familiar items – among the first hits is a clothes store, skin care products and a software program. *Dep*, on the other hand, lists a number of *State Departments of Environmental Protection*, which might be known to the speakers. This might have led them to choose *dep* instead of *deb* even though the vowel might have been considerably longer than expected. A last noteworthy factor which supports the semantics hypothesis is that word frequency is not a factor in this word pair, because *deb* is more frequent in the COCA (813 occurences) than *dep* (608).

Another very interesting result is that 3 out of 4 tokens in the bottom right hand corner, i.e. where shortening of up to -50% leads to 4 or 5 out of 10 speakers incorrectly identifying the token, involve TRAP: *back*, *lack*, *bat* (*rip* is the fourth token). This is particularly fascinating since TRAP was found to be the vowel where the American native speakers produce the second lowest degree of vowel length variation. No explanation can be offered for these results here; more research is necessary to ascertain whether this performance is unique to the group of American native speakers chosen for this study or whether it is part of a larger change in TRAP happening in American English.

Like the British native speakers, the American native speakers identified tokens correctly more often when a voiceless plosive followed. Most mistakes were made when /b/ followed ($\bar{x} = 2.875$), followed by /d/ ($\bar{x} = 2.5$), /k/ ($\bar{x} = 1.875$), /g/ ($\bar{x} = 1.625$), /t/ ($\bar{x} = 1.25$) and /p/ ($\bar{x} = 1.125$). In comparison to the British native speakers, the effect is not as strong ($\varphi = .162$), but significant ($p = .02$). It is also noteworthy that while the British group made the fewest mistakes with /t/, this is not the case for the American native speakers. Glottalized final /t/ is not a feature of American English, which means that it might be slightly less salient in American English than in British English. Still, it can be unreleased in American English, too, which distinguishes it from /d/. This also shows in the error rates, which are twice as high for /d/ than for /t/. The only consonant pair where more mistakes were made before the voiceless consonant than the voiced one is /k/ and /g/. This might be an influence from the two words *back* and *lack*, which were identified incorrectly most often.

Looking at voiced and voiceless consonants separately does not make a difference in this group. The fact that fewer mistakes were made before /g/ than before /d/ and /b/ is not significant ($p = .15$), nor is the difference in mistakes between /t/, /k/ and /p/ ($p = .34$).

With regard to the different vowels, an average of 1.9 mistakes were made with KIT, 1.75 with DRESS, 2.75 with TRAP, and 1.1 with LOT. The fact that fewer mistakes were made when LOT was involved is significant ($p = .01$), although the effect is rather small ($\varphi = .152$).

The sex of the speaker does not make a statistical difference for the American native speakers ($p = .55$). It was also not a factor in group G-US, whose results will be discussed below. As before, group C-US's results will be explained in relation to the findings of group G-US.

4.3.2.2 G-US / C-US

Group G-US behaves quite differently than the American native speakers in a number of respects. First, while word frequency was not significant for the American native speakers ($p = .21$), it is a highly significant factor for group G-US:[74]

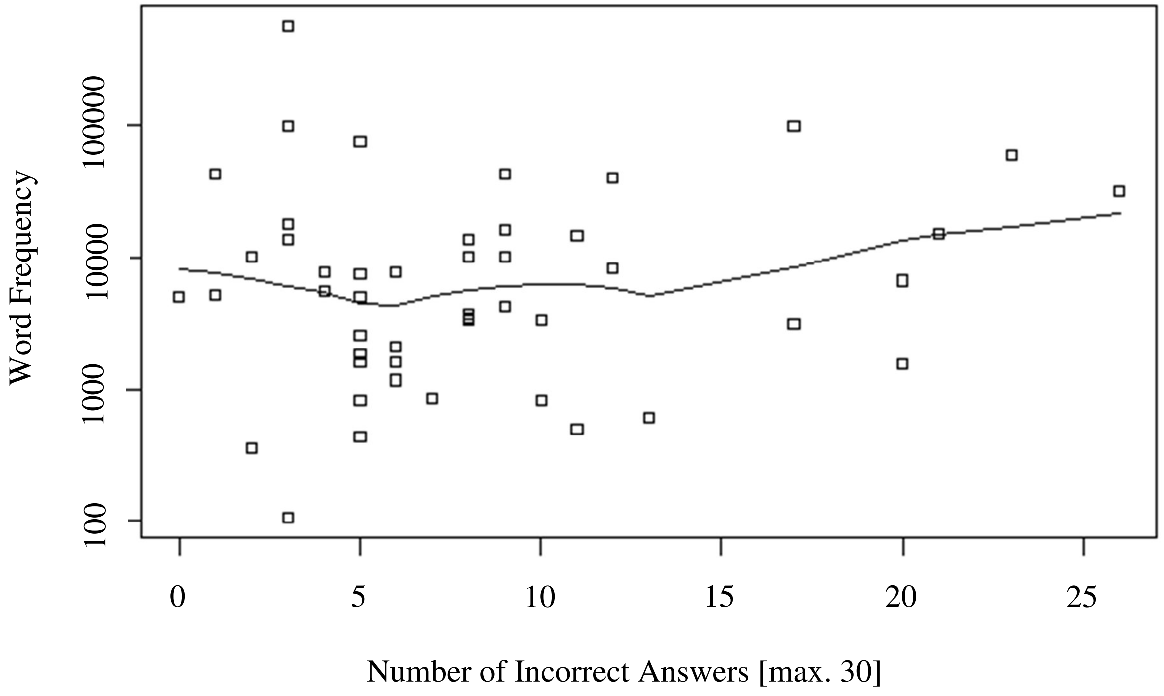


Figure 4-51. Number of tokens identified incorrectly by group G-US in the perception experiment in relation to the frequency of the tokens in the COCA. Since there are 30 speakers in group G-US, this is the maximum number of incorrect answers. The y-axis is logarithmically scaled. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .0000001$, $C = .60$.

As was the case with to group G-UK before, *back* was misheard most often (26 mistakes). This supports the hypothesis that this token was strongly influenced by the semantics of the surrounding frame sentence. Three other frequently incorrectly matched tokens all involve

---

[74] It is not significant for group C-US ($p = .43$).

LOT: *dock* (20 mistakes), *knot* (21 mistakes) and *pot* (20 mistakes). It is interesting that in all of these words, LOT is followed by a voiceless consonant. All of these tokens are shortened by -40% to -50% in comparison to their minimal pair counterpart. Other words which show a lesser degree of shortening (e.g. *hid*, -30%) were misheard much less often (8 out of 30). This suggests that the vowel LOT causes these high error rates among German learners. In American English, LOT is realized as a tense low back vowel resembling [ɑː] in both quality and quantity. This means that the realization of LOT before a voiceless plosive might be as long or even longer than the realization of KIT, DRESS or TRAP before a voiced plosive. The German learners might therefore have perceived LOT followed by a voiceless consonant as “unusually long” and opted for the voiced environment instead.

Overall, however, group G-US behaves quite similarly to group G-UK in that the speakers in both groups clearly idɑentify tokens more successfully when the words show higher frequency. This supports the hypothesis that German learners acquire vowel length variation largely by mimicking native speakers. When they hear a certain word pronounced very often, they can store it in their mental lexicon and are able to recognize and reproduce it correctly more easily.

Another result relating to the perception of group G-US is the influence of the degree of variation exhibited by the speakers in the perception experiment. In results similar to those for to groups G-UK and C-UK, fewer mistakes occur when the speakers of the tokens exhibit greater variation. Consider Figure 4-52 below:

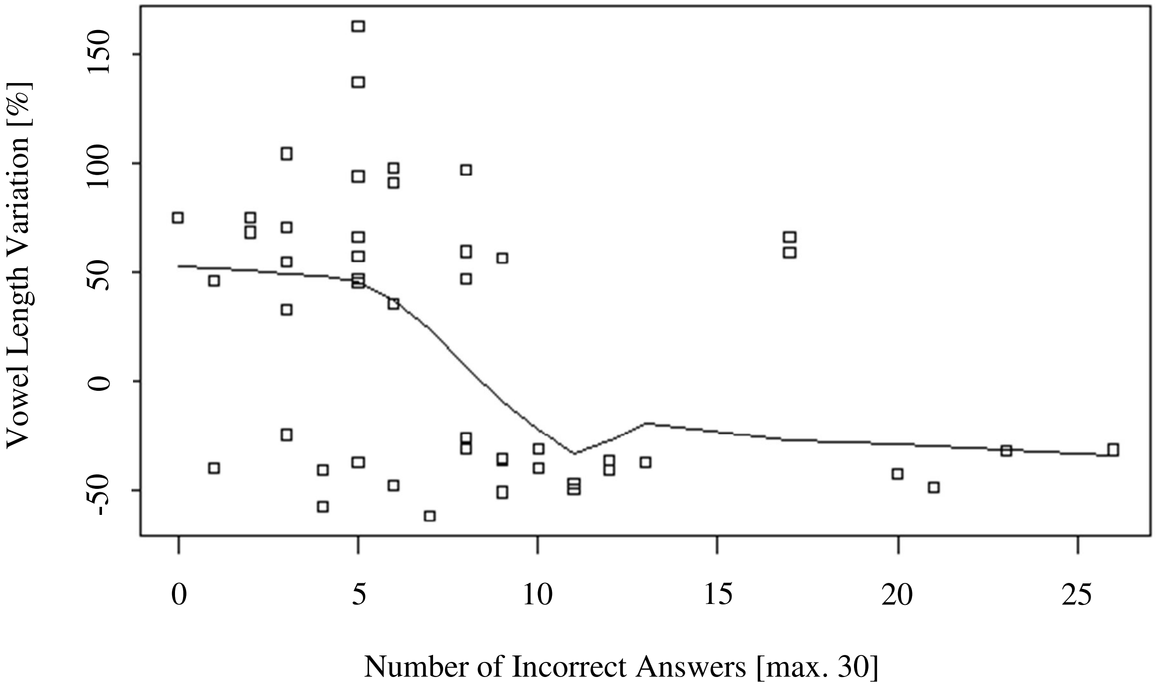


Figure 4-52. Number of tokens identified incorrectly by group G-US in the perception experiment in relation to the vowel length variation exhibited by the speakers of the tokens. Since there are 30 speakers in group G-US, this is the maximum number of incorrect answers. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .00000001$, $C = .65$.

Figure 4-52 shows that the German learners in group G-US make significantly fewer mistakes when the tokens are lengthened ($p < .00000001$). This is also the case for the control group C-US ($p = .007$).

Group G-US makes the fewest mistakes when the word is either shortened by -50% or lengthened by more than +50%. There is a strong tendency for words to be recognized correctly more often when they are lengthened; it is not quite as clear with shortening. Some words are recognized very easily when they are shortened by -50%, others are not. A lengthening of +50% leads to less than one third of participants making mistakes. There are two exceptions, which can be seen in the upper middle of the graph. Although both tokens are lengthened (+66% and +59%, respectively), 17 speakers misheard these tokens in each case. The tokens concerned are *brig* and *lib*. Other tokens involving KIT, which are lengthened less (*lid*, +35%; *hid*, +32%) were identified incorrectly only 6 and 3 times, respectively. Tokens which involve exactly the same sequence, i.e. KIT + /g/ and KIT + /b/ are difficult to compare. The other KIT + /g/ token (*pig*) is lengthened by +100%, which is very probably a substantial enough lengthening to dissuade speakers from choosing a voiceless ending (3 out of 30 speakers misheard this token). The second KIT + /b/ token is *rib*, which is lengthened by +60%, very similar to *lib*. Still, it was only misheard by 8 out of 30 participants. The frame, which was “I can say *lib* / *lip*” and “I can see a *brig* / *brick*”, cannot have caused this large number of errors in either case this time. This suggests that these two tokens were most probably influenced by the frequency of the tokens. *Brig*, ‘a sailing vessel’, occurs in the COCA 845 times; its minimal pair token *brick* has a frequency of 7,807. It is therefore highly likely that learners were not familiar with the word *brig* and opted to choose *brick* instead. *Lib*, an abbreviation for *liberation*, which often occurs in the expression *women’s lib*, is listed in the COCA only 353 times, compared to 8,398 occurrences for the minimal pair token *lip*. Again, it is highly likely that learners did not know the word *lib*, but were familiar with *lip*.

It is very interesting to note that overall, the relationship of group G-US and the American native speakers is the same as the one between group G-UK and the British native speakers with respect to the vowel length variation exhibited by the speakers of the tokens: While the German learners cue in to vowel lengthening and make significantly fewer mistakes when the vowel is lengthened, the American native speakers show the exact opposite effect. They cue in to vowel shortening and identify words correctly more often when the vowel is shortened. In comparison to the British native speakers, however, the effect is more gradual (cf. Figure 4-50). Ko *et al.* propose that vowel lengthening might be easier to process because it appears naturally in many languages, e.g. as an expression of emphatic or lexical stress or in child-directed language (2009: 138). Since vowel lengthening seems to be more common than vowel shortening, this might lead to perceptual asymmetry (cf. van der Feest & Swingley 2011: 62). This might be a reason why German learners, who do not have subphonemic but phonemic vowel length variation, cue in more to the more natural lengthening of vowels rather than to shortening.

The German learners in group G-US made most mistakes when the tokens end in /k/ ($\bar{x} = 13.625$), followed by /t/ ($\bar{x} = 10.5$) and /p/ ($\bar{x} = 8.625$). When it comes to voiced codas, most tokens were identified incorrectly when /b/ followed ($\bar{x} = 7.125$), then /g/ ($\bar{x} = 5.125$) and /d/ ($\bar{x} = 4.875$). The effect of this difference on the number of mistakes is of medium strength ($\varphi = .229$) and highly significant ($p < .00000000000001$). Unlike the case for group G-UK, this result cannot be easily explained by the vowel length variation exhibited by the speakers of the tokens within the group of voiced or voiceless consonants. Table 4-16 below lists the average vowel length variation of the tokens before the individual consonants and the corresponding average number of speakers who identified these tokens incorrectly:

Table 4-16. Average lengthening / shortening of the tokens in the perception experiment and corresponding coda consonant in relation to the average number of speakers who identified these tokens incorrectly. VLV = vowel length variation.

Average VLV and corresponding coda consonant (%)	**Average number of incorrect answers**	***sd* of shortening/ lengthening (%)**
/k/ - 38.5	13.125	(6.77)
/t/ - 35.9	10.5	(8.47)
/p/ - 47.0	8.625	(9.64)
/b/ + 94.8	7.125	(38.48)
/g/ + 65.1	5.125	(21.77)
/d/ + 58.5	4.875	(19.62)

The difference in the number of mistakes before /p/, /t/ and /k/ is highly significant ($p = .0006$) and moderately strong ($\varphi = .143$). The effect of the shortening works similarly to group G-UK in group G-US. Fewest mistakes occurred with the highest shortening (/p/, -47.0%). /k/ and /t/ show very similar shortening rates. The reason why more tokens were identified correctly when the final consonant was /t/ rather than /k/ might again be explained by the salient coda position of /t/. Unlike in British English, final /t/ is not glottalized in American English, but it can be unreleased. This might also be the reason why this effect of /t/ is not quite as strong in group G-US, but still noticeable.

The discrepancy in the number of mistakes between the voiced consonants can be neglected ($p = .07$, $\varphi = .086$), but it is noticeable that vowel lengthening is not as clearly a factor in group G-US as in group G-UK. Moreover, it seems to be distributed quite differently. The table above shows that most mistakes were made before /b/, which is the consonant before which the vowels were lengthened most (+94.8%). Fewest mistakes occurred with the coda consonant before which vowels were lengthened least, i.e. /g/, (+65.1%). It is also evident from the table that the standard deviation of the average lengthening of these tokens is quite high. Therefore, it is possible that listeners were confused by the individual lengthening differences of the tokens. This would explain why most tokens were identified incorrectly when they ended in /b/ – these tokens show the largest spread ($sd = 38.48$, $\bar{x} = 7.125$), followed by /g/ ($sd = 21.77$, $\bar{x} = 5.125$) and /d/ ($sd = 19.62$, $\bar{x} = 4.875$).

With regard to group C-US, it is noteworthy that this group behaves quite differently from group G-US and the native speakers. However, this might be expected from the fact that two out of three speakers in this group performed quite well (8 and 11 mistakes in the perception experiment), while one participant misheard almost half of the tokens (21 mistakes). Similarly to group G-UK, group C-US made more mistakes when a voiceless consonant followed. Within this group, the order is quite different, however: most mistakes occurred with /t/ ($\bar{x} = 1.375$), followed by /p/ ($\bar{x} = 1.25$) and /k/ ($\bar{x} = .875$). In the group of voiced consonants, speakers misheard tokens ending in /g/ most often ($\bar{x} = .75$), followed by /b/ ($\bar{x} = .375$) and /d/ ($\bar{x} = .25$). The effect of the consonant is significant ($p = .02$) and of medium strength ($\varphi = .31$). In group C-US, neither the number of mistakes among the group of voiced consonants nor that among the voiceless consonants is significant ($p = .25$ and $p = .47$, respectively).

When we look at vowels, most mistakes in group G-US occurred with TRAP ($\bar{x} = 9.83$), followed by LOT ($\bar{x} = 8.83$), KIT ($\bar{x} = 8.5$) and DRESS ($\bar{x} = 6.08$). The fact that fewest mistakes were made with DRESS is highly significant ($p = .002$), although the effect is small ($\varphi = .103$). This is an interesting link to the production experiment, in which group G-US also matched their native speaker model more closely in DRESS than any other vowel.

Group C-US is different again: This group made most mistakes with KIT ($\bar{x} = 1.0$), followed by TRAP ($\bar{x} = .75$) and LOT ($\bar{x} = .75$). Fewest mistakes occurred with DRESS ($\bar{x} = .67$). However, the effect is not statistically significant ($p = .70$). Unlike in group G-UK, the sex of the speaker of the tokens was not significant in either group G-US ($p = .22$) or C-US ($p = .35$).

A last point to be mentioned is the effect of the absolute length of the vowel. This is a strong factor for group G-US as well. The following figure illustrates this:

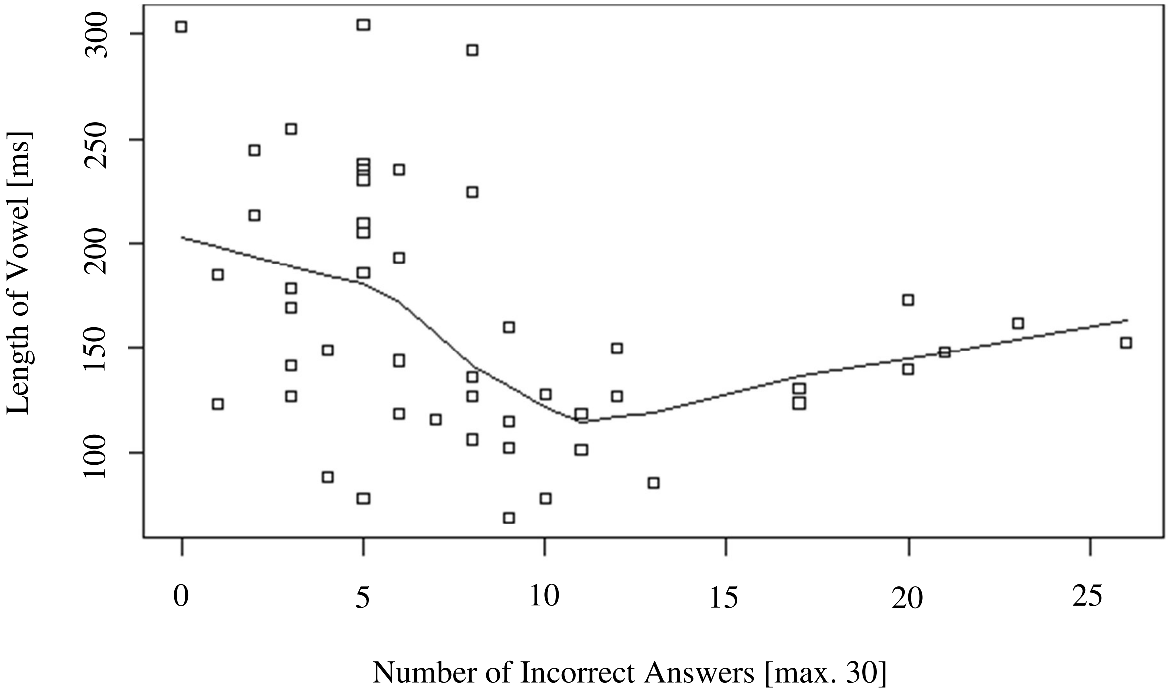


Figure 4-53. Number of incorrect answers given by group G-US in the perception experiment in relation to the length of the vowels of the tokens (in ms). Since there are 30 speakers in group G-US, this is the maximum number of incorrect answers. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p = .00001$, $C = .63$.

Figure 4-53 shows that the effect is not linear. The line falls between roughly 200 and 100 ms, indicating that participants in group G-US make very few mistakes at the upper boundary of 200 ms. The mistakes increase as the length of the word approaches 100 ms. Interestingly, some words which fall into this interval were identified correctly less often, as the right part of the graph indicates. However, the overall effect is highly significant ($p = .00001$) and of medium strength ($C = .63$). For group C-US, this factor is also highly significant ($p = .002$) and quite strong ($C = .71$). In contrast, it is not significant for the American native speakers ($p = .64$).

In order to assess whether this effect is not influenced by vowel length variation, tokens were sorted into groups according to whether they ended in a voiced or voiceless consonant, as was done for group G-UK. Subsequently, a generalized linear model was applied to calculate whether words which were longer among the group of tokens ending in a voiced consonants were identified correctly more often than shorter words. Figure 4-54 illustrates the results:

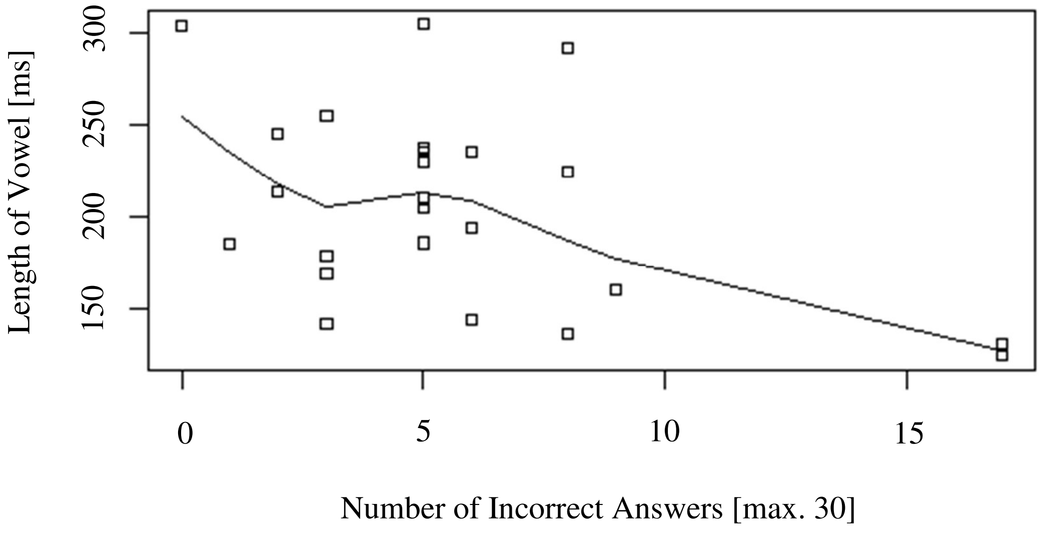


Figure 4-54. Length of the vowels of the tokens ending in voiced consonants in relation to the number of tokens which were identified incorrectly by group G-US. Since there are 30 speakers in group G-US, this is the maximum number of incorrect answers. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .0001$, $C = .65$.

The graph shows that tokens which were longer within the group of tokens ending in a voiced consonant were indeed identified correctly more often than those that were shorter. The effect is highly significant ($p < .0001$) and quite strong ($C = .65$). The same can be said for group C-US ($p = .006$, $C = .68$). This means that the effect in groups G-US and C-US is the same as in groups G-UK, which gives further evidence that German learners require time to process vowels. As was the case for group G-UK, the speakers in group G-US also seem very sensitive to small changes in length – the effect is almost linear.

In group C-US, the effect of the tokens which are shorter among those ending in a voiceless consonant is not significant ($p = .20$). However, it is significant in group G-US ($p < .000001$). Consider Figure 4-55:

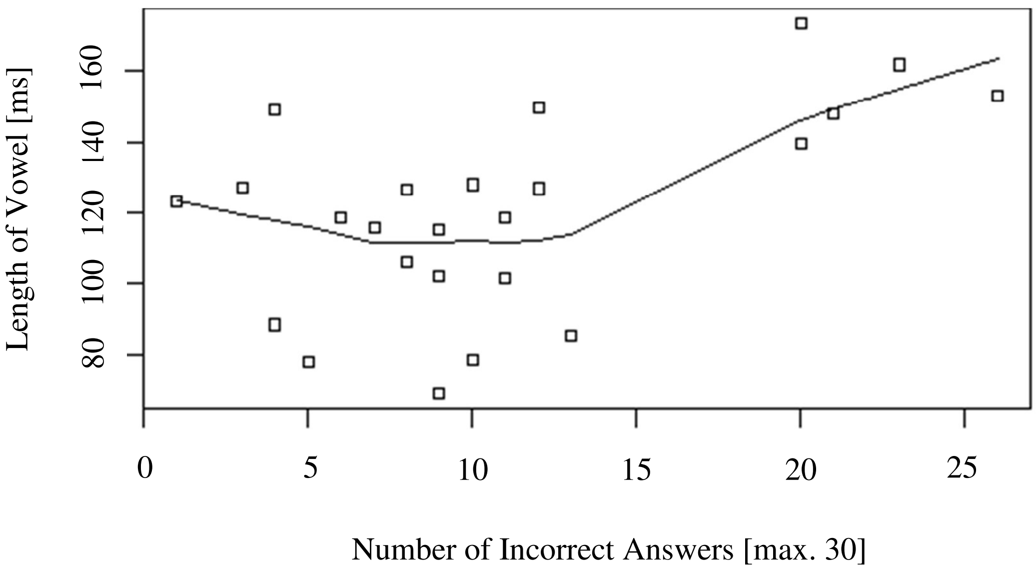


Figure 4-55. Length of the vowels of the tokens ending in voiceless consonants in relation to the number of tokens which were identified incorrectly by group G-US. Since there are 30 speakers in group G-US, this is the maximum number of incorrect answers. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). $p < .000001$, $C = .64$.

Figure 4-55 illustrates that the effect is not as strong as with the voiced consonants, but still highly significant. It is much stronger than in group G-UK, however. The upper right hand corner of the graph clearly shows that words which are longer than 140 ms are increasingly misidentified as having a voiced coda. This cut-off point cannot be easily established in group G-UK. The overarching trend is the same, however: The length of the word is a highly significant factor in how successfully German learners identify tokens.

4.4 Linking Perception and Production

The previous two sections have illustrated the native speakers' and German participants' performance in the production and perception experiment. The obvious question to ask now is of course whether production and perception are linked. Are learners who are more successful at perceiving vowel length variation also more successful at producing it? Many researchers have argued that perception is a prerequisite for production (cf. Trubetzkoy (1939), Flege (1987; 1991), Neufeld (1988) and Llisterri (1995)), or vice versa (cf. Gass (1984), Strange (1995), Kluge *et al.* (2007)). This chapter will shed light on this question with regard to vowel length variation. Since this is an issue which does not hinge on whether participants stayed abroad in the USA or UK, but draws on the common factor that they are all learners of English, the two German learner groups G-US and G-UK are pooled together. This also has the advantage that the increased data set of 60 subjects allows for a more robust statistical analysis.

Figure 4-56 illustrates that production and perception are closely linked. The x-axis shows the perception score of each speaker, that is, the number of tokens each speaker identified incorrectly in the perception experiment. Since participants listened to 48 sentences, this is the maximum score and equals no correct identification. This is plotted against the variation exhibited by each speaker, which can be found on the y-axis. The results clearly show that subjects who exhibit greater variation in the production experiment make significantly fewer mistakes in the perception experiment (*p* = .02).

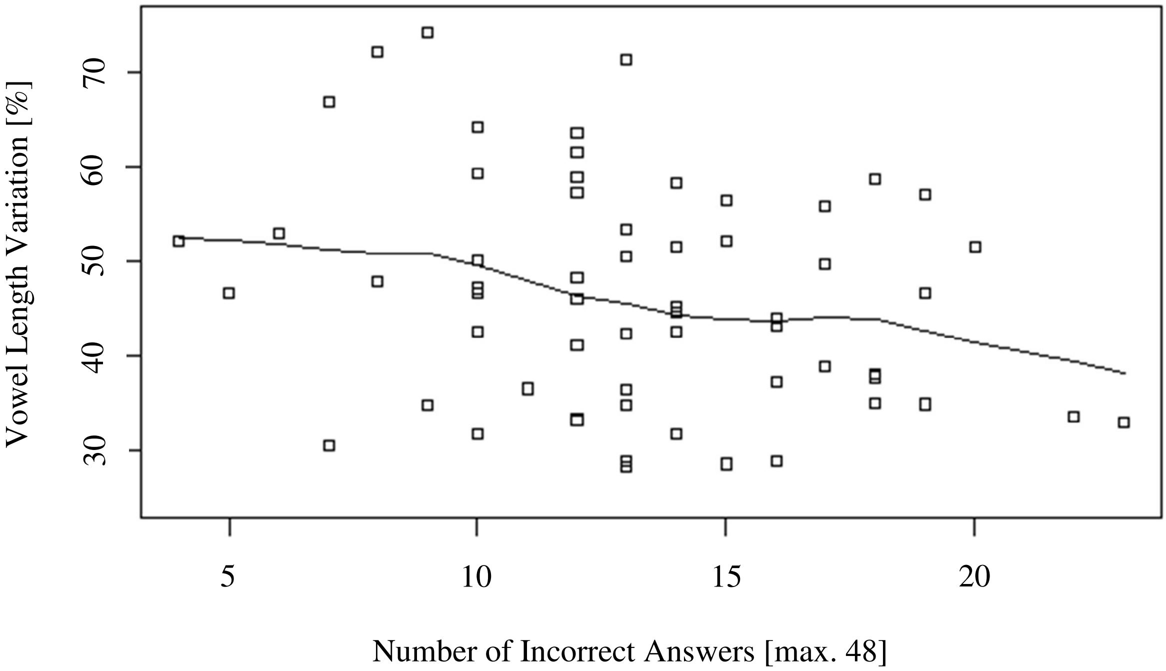


Figure 4-56. Mean value of variation of each speaker plotted against the perception score of each speaker (the number of incorrect answers out of 48 sentences). The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). *p* = .02, *C* = .60.

Figure 4-56 suggests that the success of a speaker in the perception experiment can be predicted by examining said speaker's production of vowel length variation. Interestingly, this is a mutual influence, since the opposite also holds: Participants who make fewer mistakes than average in the perception experiment exhibit significantly greater vowel length variation than those who perform below average (*p* = .02, cf. Figure 4-57 below). In group G-UK, the ratio of speakers in the two groups *above average* and *below average* is 15 to 15, while it is 16 to 14 in group G-US. The average perception result is 12.7 (*SE* = .7) in group G-UK and 13.3 (*SE* = .8) in group G-US. Taking into account the standard error, this means that the true average for group G-UK lies within the interval between 12.0 and 13.4; the true average for group G-US lies within the interval between 12.5 and 14.1. The two intervals overlap at 13, which is why this value was chosen as a cut-off point for the average number of mistakes.

This means that speakers who identified ≤ 13 tokens incorrectly performed better than average; speakers whose perception score is ≥ 14 performed worse than average. A boxplot of the production of vowel length variation for these two speaker groups is shown in Figure 4-57:

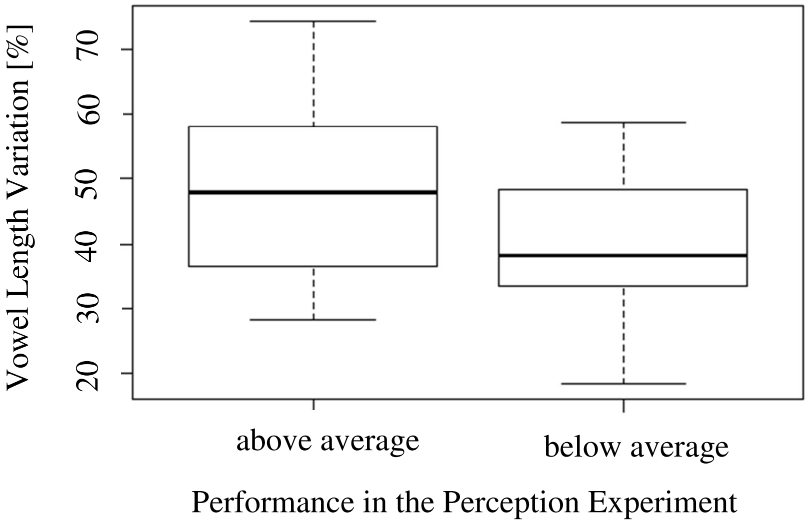


Figure 4-57. Variation exhibited in the production experiment in relation to the performance in the perception experiment. *Above average* indicates ≤ 13 mistakes, *below average*, ≥ 14 mistakes. $p = .02$.

Participants who performed above average in the experiment show an average of 48.68% (*sd* = 2.2) variation, while the figure for participants in the group *below average* is 40.30% (*sd* = 1.94). The spread in the groups is therefore roughly the same.

The results mentioned so far indicate that perception and production are closely linked. However, the direction of the link has not been inspected so far. Let us examine Figure 4-56 again. There are no speakers in the upper right hand corner, which suggests that there are no speakers in groups G-UK and G-US who have difficulty perceiving vowel length variation but produce it with ease. This is a first indication that perception might precede production. A second idea might be to look for speakers who can perceive vowel length variation, but do not produce it. These speakers can be found in the control groups C-UK and C-US. Consider Figure 4-58:

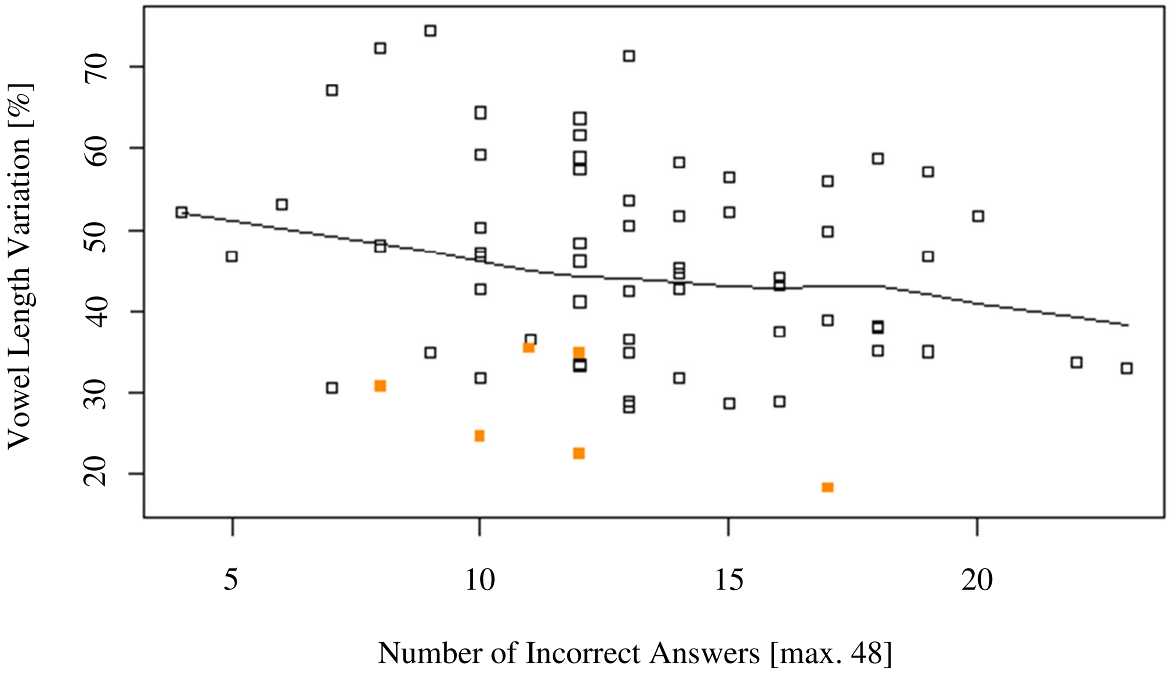


Figure 4-58. Mean value of variation of each speaker plotted against the perception score of each speaker (the number of incorrect answers out of 48 sentences). The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)) and takes into consideration data from all four groups. Empty black squares = groups G-UK and G-US; orange squares = groups C-UK and C-US.

Figure 4-58 shows that among the speakers from the control group charted in orange, all but one cluster in the lower left and middle. This indicates that they make fewer mistakes than average ($\bar{x} = 11.5$, *SE* = 1.23, *sd* = 3.0), but that their production of vowel length variation is quite low ($\bar{x} = 27.66$, *SE* = 2.84, *sd* = 6.96). Quite clearly, these speakers can hear differences in vowel length, even though they do not produce them. This strongly suggests that perception precedes production in the acquisition of vowel length variation. However, we must bear in mind that the experiment in which the perception of vowel length variation was examined was quite limited and somewhat decontextualized, so that a more elaborate and holistic study would be needed in order to verify this first hypothesis. In addition, it remains unclear why the speakers in the control groups do not produce vowel length variation when they can clearly perceive it.

4.5 Linking Phonemic Success and Subphonemic Variation

This section examines the link between phonemic success and success at acquiring vowel length variation. It stands to reason that learners who have trouble acquiring phonemic contrasts will also find it very difficult to acquire subphonemic contrasts, since these are hard to perceive and rarely ever mentioned or taught to learners. For German learners, however advanced, two phonemic issues often persist: The merger of TRAP and DRESS and devoicing.

Since the German phoneme inventory lacks TRAP, many learners merge TRAP and DRESS under DRESS, the nearest available vowel. Students of American English might find acquiring separate categories for these two vowels particularly challenging, because DRESS is pronounced lower in American than in British English. As such, it lies very close to TRAP, which makes it even more difficult for learners to keep these two vowels distinct. However, even though DRESS and TRAP are more distinct in vowel space in British English, many learners of this variety also struggle to keep them apart even at an advanced stage. The distinction between these two vowels is rarely mentioned at school, but it is certainly taught during speaking courses at university which all of the German participants in this study have attended. Thus, they are well aware of the difference between these two vowels, whether they produce this difference or not. Section 4.5.2 will deal with the question of how this influences the perception of vowel length variation.

Possibly the most difficult feature of a German accent to eradicate is the devoicing of word-final segments. Researchers have related devoicing to the phonetic conditions present in prepausal position (cf. Sievers 1901: 289 f., Bloomfield 1933: 373; Lindblom 1983: 237). Since the vocal folds do not vibrate during a pause, final devoicing can be taken as assimilation to the pause (cf. Lightner 1972: 332 f., Ingram 1989: 35). German does not pronounce voiced segments in final position, so that *Rad* 'bike' and *Rat* 'advice' sound the same: /ʁaːt/. In English, the difference between a voiced and a voiceless coda is meaning-distinguishing, which means that German learners have to train to produce this distinction if they want to communicate successfully. Devoicing is dealt with intensively in the speaking courses at university. Still, it remains a difficult feature to master even for advanced learners. Two ways in which this might influence the acquisition of vowel length variation are conceivable: One is that German learners use vowel length variation in order to counterbalance devoicing, maybe even outperforming their native speaker model in the process. The other hypothesis that may have credence is that speakers who are unsuccessful at learning this phonemic distinction are also unable to reproduce vowel length variation accurately. We have seen from the previous section that perception and production mutually influence each other. Thus, devoicing in pronunciation might make it very difficult for German learners to perceive vowel length variation. Instead, they might be more focused on listening to the coda consonant of the syllable, which is of course inaudible in the experimental setting. The next section will shed light on which of these two hypotheses actually holds true.

4.5.1 Devoicing

Neither the British nor the American native speakers show devoicing ($\bar{x} = .82$ and $\bar{x} = .78$, respectively). For the German learners, the mean Pillai score for devoicing is .48 (*sd* = .24), illustrating that the spread of the data is quite large. Thus, there are some speakers who exhibit devoicing, others do not. Since devoicing is a feature common to German learners of English, the two German learner groups G-US and G-UK are again pooled together to receive a general picture of the influence of devoicing. Figure 4-59 presents the results of the generalized linear model used to evaluate the relationship between devoicing and perception:

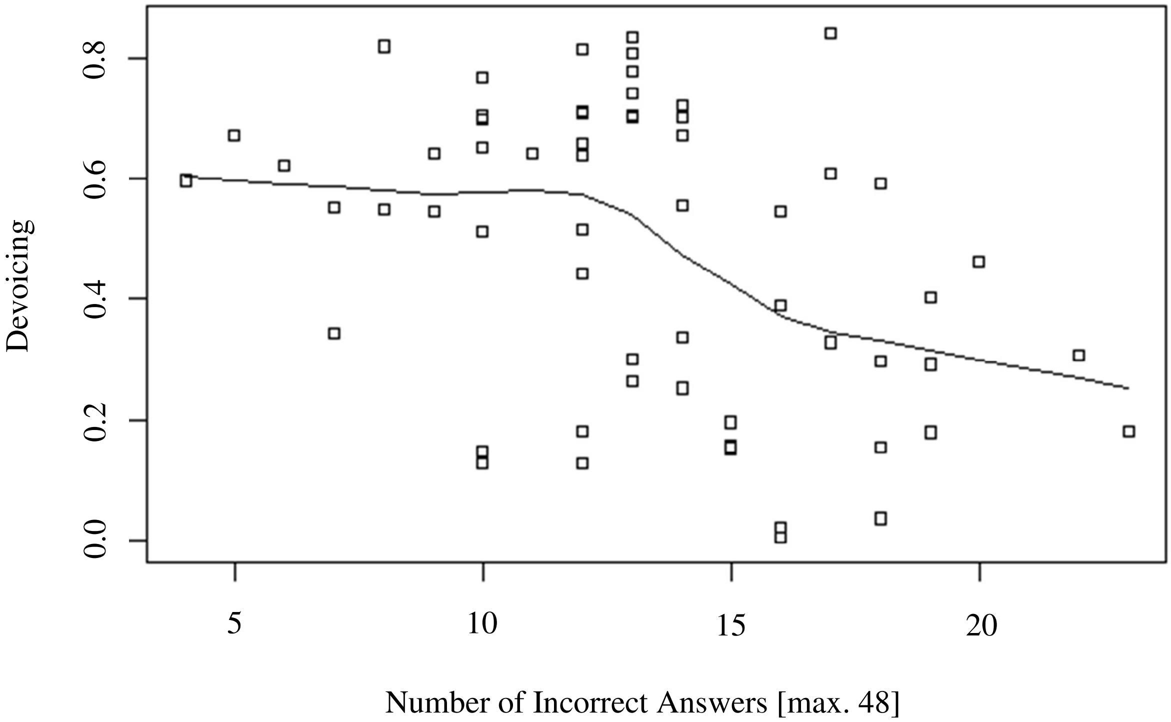


Figure 4-59. Devoicing plotted against the perception result of the participants in groups G-US and G-UK. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). Devoicing was measured using Pillai scores (see section 3.4.3). $p < .001$, $C = .62$.

Figure 4-59 illustrates that there is a definite link between devoicing and perception of vowel length variation. Participants with a higher Pillai score, i.e. low levels of devoicing, make significantly fewer mistakes in the perception experiment ($p = .001$). There is one exception in the lower left corner, a speaker who has a relatively low Pillai score of .34 and a perception score of 7. This subject (#028) seems to perceive vowel length variation quite well despite strong devoicing. Furthermore, there are a number of speakers in the top center who exhibit hardly any devoicing but made between 10 and 14 mistakes in the perception experiment. Remember that speakers with ≥ 14 mistakes were categorized in the group whose perception score was below average, so these speakers still performed better than average. Figure 4-59 illustrates this quite nicely, since this is the point where the scatterplot smoothing curve starts falling, i.e. when the link between devoicing and perception score becomes relatively strong. This means that the link between devoicing and the perception of vowel length variation is particularly strong for those speakers who do not perceive this variation well. In addition, devoicing is a stronger factor in group G-UK (*p* = .009, *C* = .60) than in group G-US (*p* = .04, *C* = .55).

Since the speakers in the control groups all made a similar number of mistakes and show strong devoicing, this factor is not statistically significant in either group C-UK (*p* = .39) or C-US (*p* = .61). Instead, the data of the six speakers from the control groups were added to the existing data pool of the German learners from groups G-UK and G-US in order to examine how the control group behaves in relation to the more advanced learner groups. The following Figure 4-60 illustrates this:

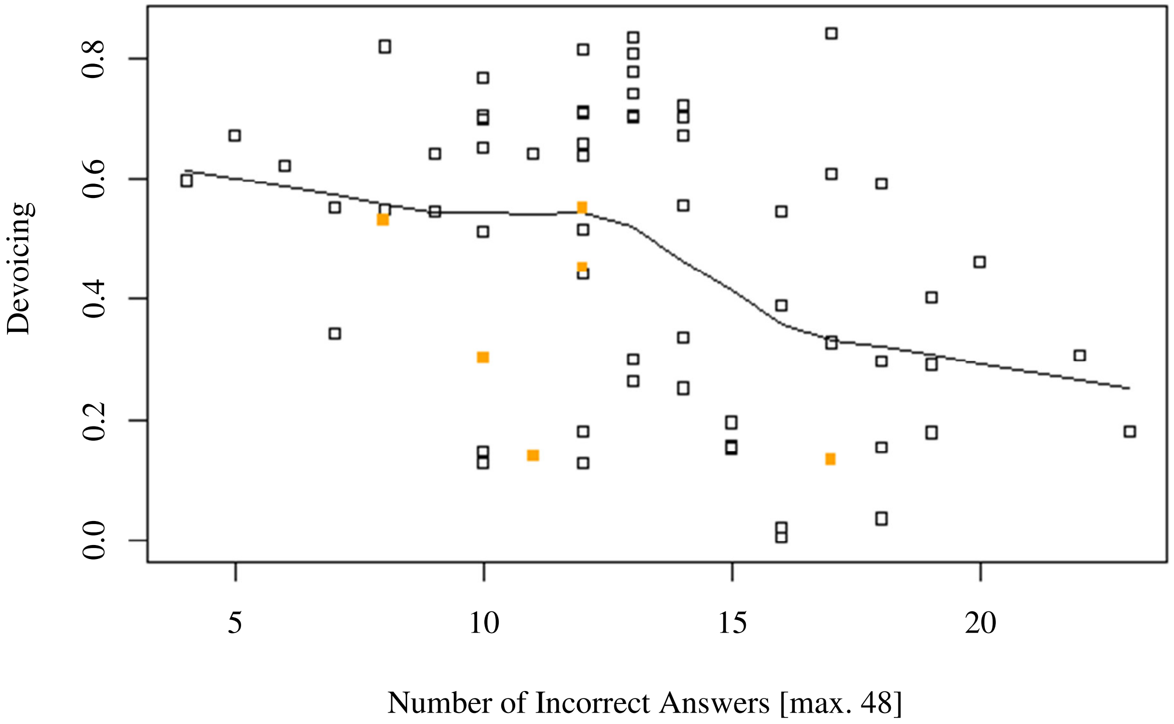


Figure 4-60. Devoicing plotted against the vowel length variation of the individual participants in groups G-US, G-UK, C-UK and C-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)) and takes into account data from all speaker groups. Devoicing was measured using Pillai scores (see section 3.4.3). *p* < .001, *C* = .61. Empty black squares = groups G-UK and G-US; orange squares = groups C-UK and C-US.

Figure 4-60 shows that the control groups C-US and C-UK actually pattern well with the other two learner groups. There is a horizontal 3-by-3 divide: Three speakers conform well to curve, while the other three can be found among the less proficient learners.

Devoicing is also a factor in the production of vowel length variation. Speakers who have strong devoicing produce significantly less vowel length variation, as Figure 4-61 below demonstrates:

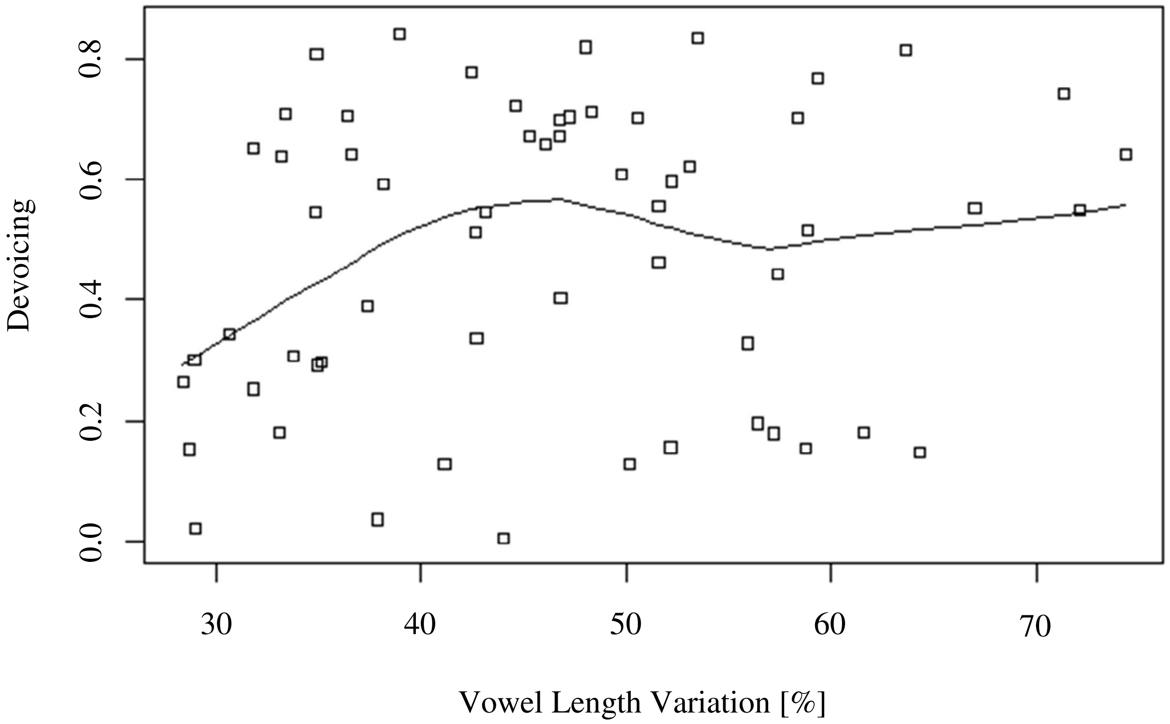


Figure 4-61. Devoicing plotted against vowel length variation of the individual participants in groups G-US and G-UK. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). Devoicing was measured using Pillai scores (see section 3.4.3). *p* = .01, *C* = .56.

Figure 4-61 allows interpretation in the direction of both hypotheses mentioned above. Although the statistics indicate that speakers who do not exhibit devoicing also successfully produce vowel length variation, there are a few individual speakers in the lower right. These speakers show robust vowel length variation, but also have strong devoicing. They might – consciously or unconsciously – use vowel length differences in order to counterbalance their devoicing. It is noteworthy that there are also speakers who pattern in the upper left of the graph; these speakers display no devoicing but do not produce much vowel length variation. It is unclear why this is so; in any case, these speakers should be well able to acquire variation if they were made aware of its existence.

The two control groups C-UK and C-US are an interesting addition to the existing picture. Taken by themselves, devoicing does not seem to be a significant factor for either group (C-UK p = .58; C-US p = .64). However, this might again be due to small group size. This is why the control groups were added to the pool of data from the two learner groups G-UK and G-US. Figure 4-62 colors the control groups as orange squares:

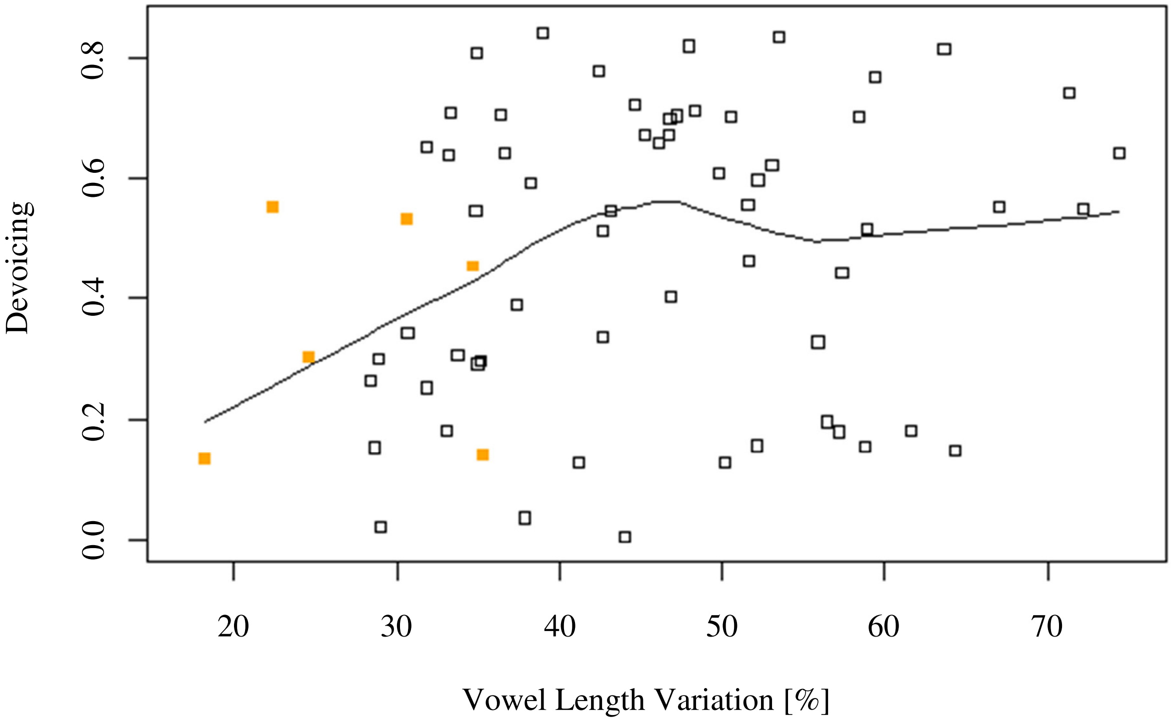


Figure 4-62. Devoicing plotted against vowel length variation of the individual participants in groups G-US, G-UK, C-UK and C-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)) and takes into account data from all four speaker groups. Devoicing was measured using Pillai scores (see section 3.4.3). p = .0005, C = .58. Empty black squares = groups G-UK and G-US; orange squares = groups C-UK and C-US.

Figure 4-62 now illustrates that the control groups are an extension of the learner groups G-UK and G-US. They cluster at the leftward end of the graph. Four of the speakers are quite close to the scatterplot smoothing curve, which has been altered to take into consideration the new data from the control groups. The addition of the six less experienced speakers also has a strong effect on the significance of devoicing on the production of vowel length variation, which is now at p = .0005.

In summary, we have seen that devoicing is a factor in both the perception and production of vowel length variation. Although this did not seem to be the case with the control groups, adding them to the existing pool of data has provided interesting insight. Groups C-UK and C-US do not improve the model which links perception and devoicing, which is doubtlessly influenced by the fact that 5 out of the 6 speakers in these groups performed above average in the perception experiment. In contrast, when we look at the production of vowel length variation and how this is influenced by devoicing, we can clearly see that the control groups, i.e. the less experienced speakers, are an extension at the lower end of the model. We can conclude that speakers who have a low level of devoicing show significantly better perception and production of vowel length variation.

4.5.2 Merger of DRESS and TRAP

This section deals with the question of whether and how strongly the perception of vowel length variation is influenced by the second issue affecting phonemic success: the merger of TRAP and DRESS. It has been suggested that perception precedes production in the acquisition of phonemes (e.g.by Rochet (1995)). Therefore, participants who show merged TRAP and DRESS might do so because they are unable to perceive a distinction between these two vowels. This lack of perception ability might extend to the subphonemic level and make it difficult for speakers to perceive vowel length variation adequately. Particularly since the degree of vowel length variation differs between the two vowels (lower vowels show stronger variation), having merged vowels might create considerable problems for learners.

The mean Pillai score for the merger of TRAP and DRESS in the two German learner groups is .57 (*sd* = .25), indicating that the spread of the data is quite large. The British and American native speakers do not exhibit this merger (Pillai scores of $\bar{x} = .88$ and $\bar{x} = .84$, respectively).

Figure 4-63 presents the result of the analysis:

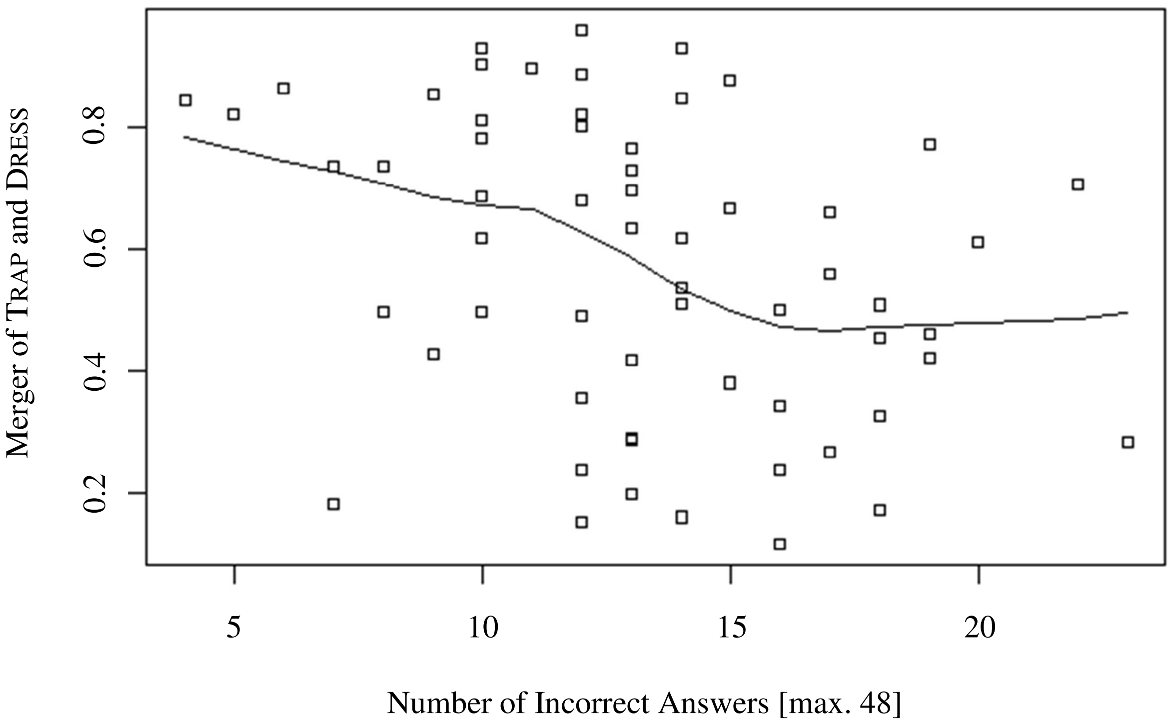


Figure 4-63. Merger of TRAP and DRESS plotted against the number of incorrect answers in the perception experiment of each speaker in groups G-UK and G-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). The merger of TRAP and DRESS was measured using Pillai scores (see section 3.4.3). *p* = .003, *C* = .64.

Figure 4-63 illustrates that participants who exhibit a merger of TRAP and DRESS make significantly more mistakes in the perception experiment than those who have two separate categories for the vowels (*p* = .003). Again, there are a number of exceptions: One speaker shows a low Pillai score of .17, but only misjudged seven tokens. This is the same speaker #028 who had a low Pillai score for devoicing. Therefore, this participant can be seen as a true exception. There are two participants in the top right corner who do not exhibit a merger (Pillai scores of .71 and .77), but identified 22 and 19 tokens incorrectly. Both speakers have low values for devoicing (.31 and .18, respectively), which might indicate that out of the two phonemic factors *merger of TRAP and DRESS* and *devoicing*, the latter is the stronger influence. This is also supported by a statistical analysis: When both factors are entered into a generalized linear model, devoicing is still as significant as before, with *p* < .001. However, the merger is then only significant at *p* = .04. On a related note, the merger of TRAP and DRESS is a stronger factor in group G-US (*p* = .04, *C* = .62) than in group G-UK (*p* = .05, *C* = .64).

Introducing the two control groups C-UK and C-US yields fascinating results. Two of the speakers converge with groups G-UK and G-US, while the other four seem to pattern with speaker #028. This is illustrated by the following Figure 4-64:

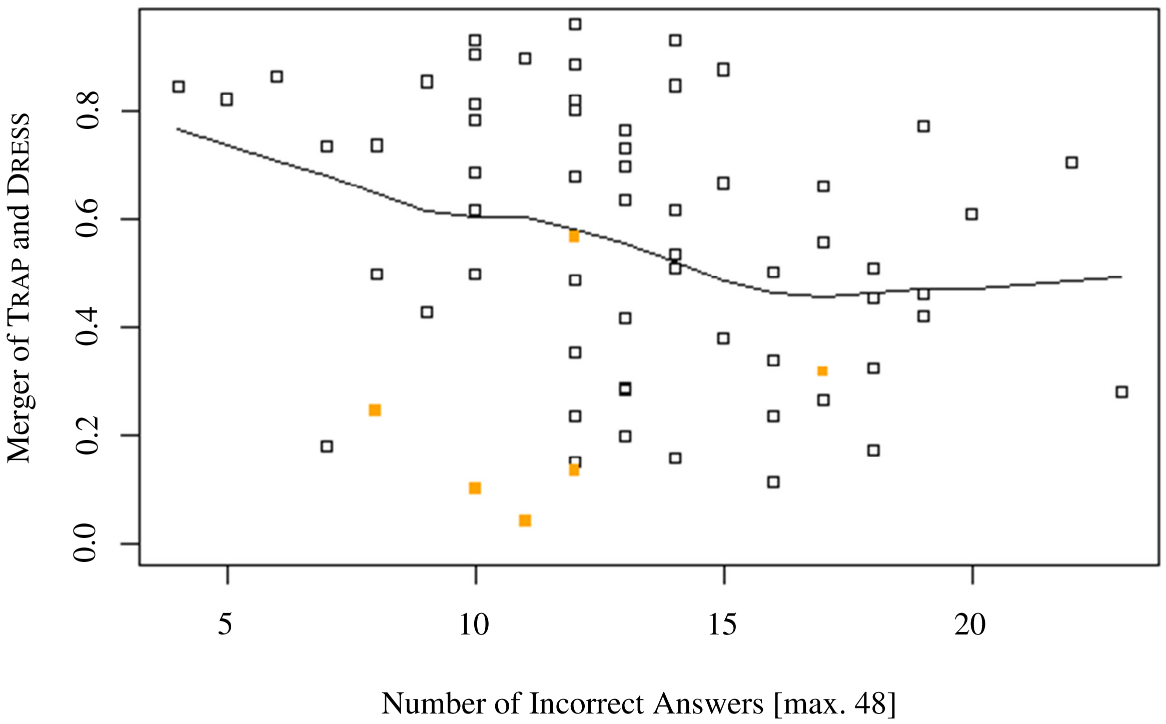


Figure 4-64. Merger of TRAP and DRESS plotted against the number of incorrect answers in the perception experiment by the participants in groups G-US, G-UK, C-UK and C-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)) and takes into account data from all four speaker groups. Merger of TRAP and DRESS was measured using Pillai scores (see section 3.4.3). *p* = .03, *C* = .59. Empty black squares = groups G-UK and G-US; orange squares = groups C-UK and C-US.

In Figure 4-64, the two learner groups C-US and C-UK are again illustrated in orange. The scatterplot smoothing curve has been edited to include this new data. We can see that the upper two orange dots conform to this curve quite well, while the lower four do not. This is due to the fact that these four speakers performed quite well in the perception experiment but strongly merge TRAP and DRESS. This merger did not affect them negatively in the experiment, because the two tokens between which the speakers had to choose were always either two tokens of TRAP or two tokens of DRESS. As long as the speakers were able to perceive the vowel length variation between the two tokens, it did not matter whether they classified them as instances of TRAP or DRESS.

Let us now examine the link between the merger of TRAP and DRESS and the production of vowel length variation. Figure 4-65 illustrates the result of the analysis of these two factors:

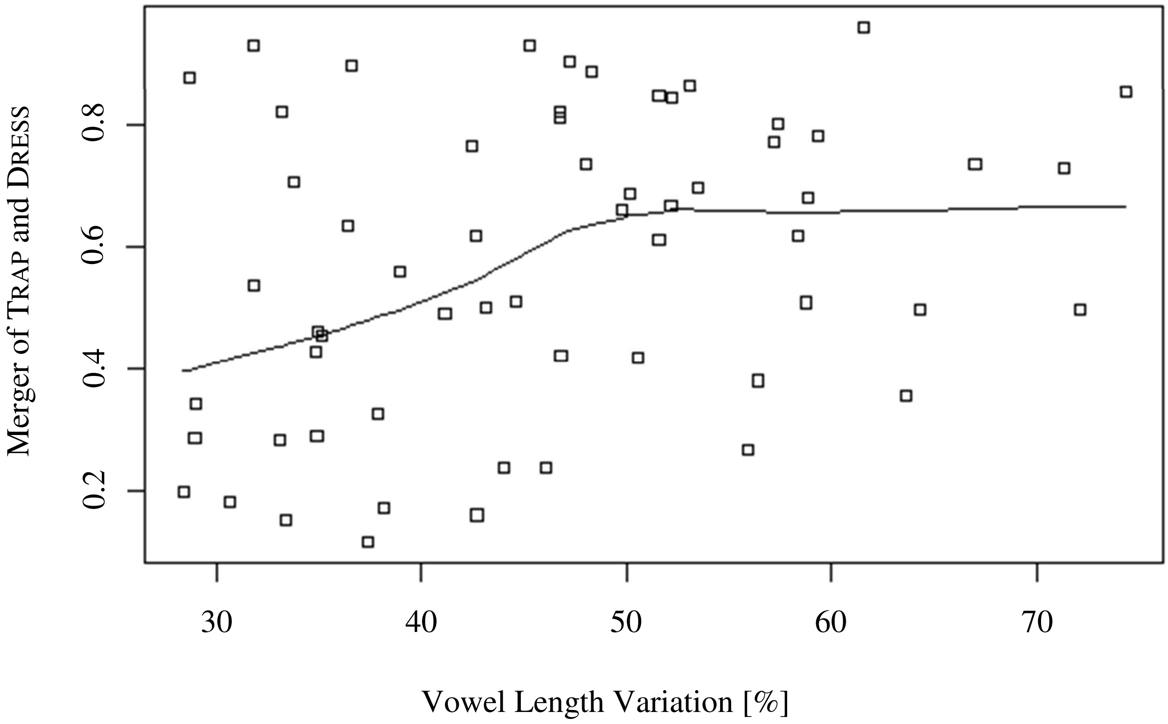


Figure 4-65. Merger of TRAP and DRESS plotted against the mean vowel length variation of each speaker in groups G-UK and G-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)). The merger of TRAP and DRESS was measured using Pillai scores (see section 3.4.3). $p < .00001$, $C = .61$.

It is clearly visible from Figure 4-65 that speakers who have distinct categories for TRAP and DRESS produce clearer vowel length variation. The effect is highly significant ($p < .00001$). However, there are a number of speakers in the upper left corner who pronounce TRAP and DRESS distinctively, but do not produce much vowel length variation. Likewise, there are speakers in the lower right hand corner who do the exact opposite – they produce vowel length variation quite clearly, although they merge TRAP and DRESS. The former cluster of speakers is expected to be quite successful at acquiring vowel length variation if they are made aware of its existence, while speakers in the latter category still need to acquire separate categories for the two vowels, which is relatively difficult.

By now, it should not come as a surprise that the merger of TRAP and DRESS is not statistically significant in either group C-UK ($p = .39$) or C-US ($p = .99$) due to small group size and the similar behavior of the speakers in this group. This is why, as before, the speakers in the two control groups were added to the existing data pool of groups G-UK and G-US. The result is illustrated by Figure 4-66:

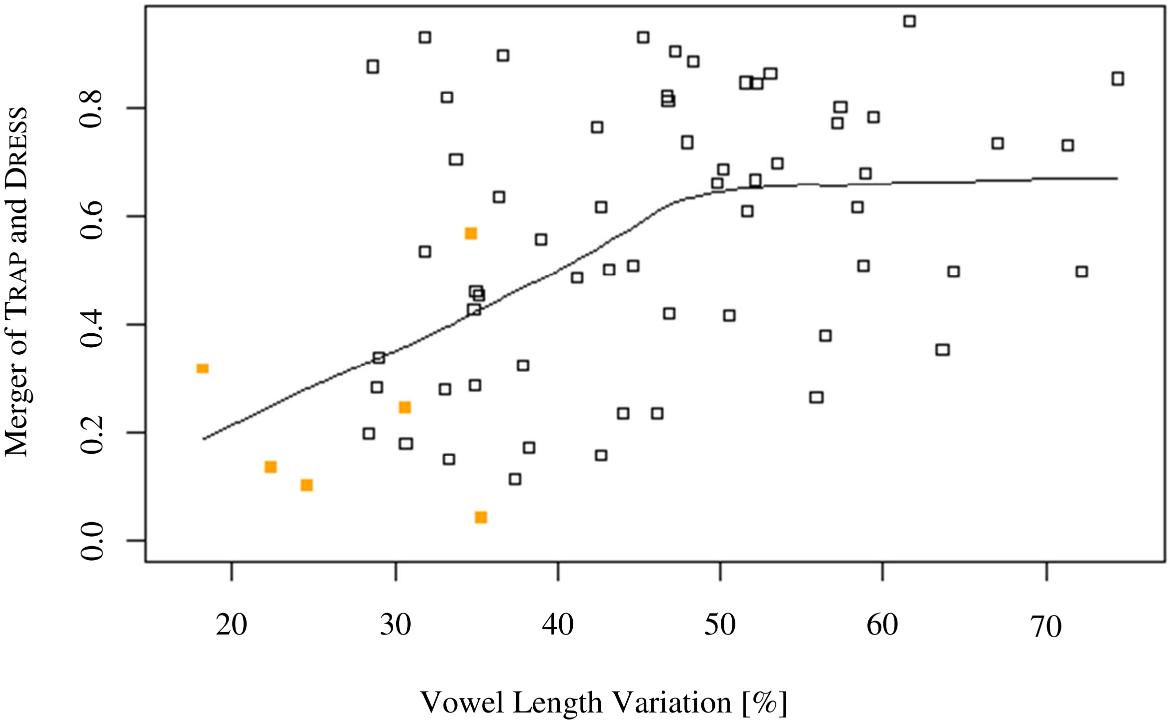


Figure 4-66. Merger of TRAP and DRESS plotted against the number of incorrect answers in the perception experiment by the participants in groups G-US, G-UK, C-UK and C-US. The line represents a non-parametric locally weighted scatterplot smoothing curve (Cleveland (1979)) and takes into account data from all four speaker groups. Merger of TRAP and DRESS was measured using Pillai scores (see section 3.4.3). $p < .0000000001$, $C = .64$. Empty black squares = groups G-UK and G-US; orange squares = groups C-UK and C-US.

Adding the control group to the existing data pool makes a great improvement both to the significance of the merger of TRAP and DRESS as a factor in the production of vowel length variation, and to the fit of the model ($p < .0000000001$, $C = .64$). This is a strong indication that the control group corresponds well with the more advanced groups G-UK and G-US, which is also visible in Figure 4-66 above.

As with devoicing before, we can see the speakers in the control groups C-UK and C-US towards the left-hand side of the graph. However, there seems to be a vertical 3-and-3 divide: Three speakers form a pattern at the lower left end of the scatterplot smoothing curve, while the other three can be found among the less advanced speakers of groups G-UK and G-US. The former three speakers produce hardly any vowel length variation and merge TRAP and DRESS quite strongly. Interestingly, this divide is a perfect match for the two control groups: The least advanced speakers all come from group C-UK, whereas the ones patterning with the speakers from G-UK and G-US are the speakers from group C-US. Since this clear divide has not occurred before, its appearance must not be overstated, particularly because the difference in the production of vowel length variation between groups C-UK and C-US is not statistically significant ($p = .10$). However, it might be an indication that even less experienced learners profit from stronger vowel length variation usually exhibited by American native speakers in relation to British native speakers, and find it easier to mimic it.

In summary, we have seen that the merger of TRAP and DRESS is a factor in both the perception and production of vowel length variation. Adding the control groups to the existing pool of data has again provided us with a clearer picture. As was the case with devoicing, groups C-UK and C-US do not improve the model which links perception and merger of TRAP and DRESS, which was explained by the fact that 5 out of the 6 speakers in these groups performed above average in the perception experiment. Regarding the production of vowel length variation and how this is influenced by a merger of TRAP and DRESS, it has become obvious that the control groups pattern with groups G-UK and G-US. We can conclude that speakers who have separate categories for TRAP and DRESS are significantly better at perceiving and producing vowel length variation.

4.6 Summary

This chapter is designed to give a summary and overview of the results gained from the production and the perception experiment. The results will first be presented in the form of success factors; the tables on pages 194 ff. are designed as a quick overview of the significance (or lack thereof) of the various factors that were tested.

4.6.1 Production

The results gained from the production experiment show that the following factors lead to success in producing vowel length variation in a more native-like manner:

- **Phonological proficiency** strongly influences the production of vowel length variation. Learners who have successfully established two distinct categories for TRAP and DRESS produce significantly greater vowel length variation than those who merge the vowels. Likewise, low levels of devoicing goes hand-in-hand with stronger vowel length variation.
- **Performance in the perception experiment** is a strong predictor for the production of vowel length variation. Perception precedes production; learners who make fewer mistakes in the perception experiment produce significantly greater vowel length variation.

- Although the **length of the stay abroad** significantly influences native-like production of vowel length variation, the **reason for the stay abroad** is an equally important factor, if not more so, since some speakers who stayed a shorter amount of time were shown to be more proficient if they had been exposed to both BICS and CALP (cf. Cummins (1979; 1991)).
- Learners who subjectively **rate themselves as more proficient** produce significantly greater vowel length variation.
- Learners who **converse mostly with native speakers** produce a significantly greater degree of vowel length variation than those who predominantly speak to other non-native speakers.
- There is no effect of gender, the emphasis a learner places on achieving a native-like accent, or active and passive language use. Language use might only be a relevant factor in an immersion setting, but not an instructional one.

Other findings related to the two major groups were the following:

British Group

- British native speakers show most variation in the **vowel** TRAP → LOT → DRESS → KIT. This sequence is matched by group G-UK. The control group produces more variation in DRESS than LOT, which might be an indication that L1 strongly influences this group (DRESS is a similar vowel, LOT is a new one).
- The British native speakers produce most variation before the **consonant** groups of /g, k/ → /b, p/ → /d, t/. Group G-UK inverts the last two consonant groups: /g, k/ → /d, t/ → /b, p/. In the control group C-UK, there is no recognizable pattern: /d, t/ → /b, k/ → /g, p/. The fact that group C-UK produces most variation before /d, t/ might be due to /t/ often being glottalized or unreleased in final position, which renders this opposition and the related vowel length variation particularly salient. This might also be the reason why the two more advanced learner groups show the second highest degree of variation in these two consonants (cf. the results from group G-US below).
- **Phonological proficiency** has no effect on the British native speakers, because they neither merge TRAP and DRESS nor devoice final consonants. In contrast, both factors have a strong effect on group G-UK, with speakers showing no devoicing and two separate vowel categories producing significantly greater vowel length variation. Due

to small group size, there is no significant effect in group C-UK. Therefore, their performance is best viewed in relation to group G-UK's (cf. chapters 4.5.1 and 4.5.2).

- **Word frequency** is a significant influence for the British native speakers and group G-UK, but not for group C-UK. For the British native speakers, this might be due to the fact that less frequent words take longer to recall and are therefore lengthened. Group G-UK might be affected because they are more likely to be confronted with high frequency words, which might then be easier for them to recall and reproduce with accurate vowel length variation.

American Group

- American native speakers show most variation in the **vowel** DRESS → KIT → TRAP → LOT. This sequence is quite unexpected. Group G-US matches the sequence of the British native speakers and group G-UK (TRAP → LOT → DRESS → KIT). The control group C-US matches group C-UK and produces more variation in DRESS than LOT, which is again evidence that L1 influences the control groups quite profoundly.
- The American native speakers produce most variation before the **consonant** groups of /g, k/ → /b, p/ → /d, t/. Group G-US, like group G-UK, inverts the last two consonant groups: /g, k/ → /d, t/ → /b, p/. The control group C-US again shows no recognizable pattern: /d, k/ → /g, p/ → /b, t/.
- **Phonological proficiency** has no effect on the American native speakers. They neither merge TRAP and DRESS nor devoice final consonants. In contrast, both factors are significant in group G-US. Learners who do not exhibit devoicing and have formed two separate vowel categories produce significantly greater vowel length variation. Due to small group size, there is no significant effect in group C-US. Therefore, their performance is best viewed in relation to group G-US's (cf. chapters 4.5.1 and 4.5.2).
- As with the British group before, **word frequency** is a significant influence for the American native speakers and group G-US, but not group C-US. The reasons may be the same as for the British group.

We can conclude from the findings that the production of vowel length variation hinges on a number of different factors. Learners who want to become proficient in producing vowel length variation should seek interaction with native speakers, as several results have illustrated that learners who predominantly speak English with native speakers perform better than those who have more contact with other non-native speakers. Those who stayed abroad longer, particularly because of their family or in order to study at a foreign university, were able to match the native speakers' variation more closely. A stay abroad is particularly helpful if it introduces the learner to different, more formal levels of the language. Learners should also focus on improving their phonological proficiency. If they manage to acquire separate phoneme categories and eliminate final devoicing, they will be more successful at producing vowel length variation in a native-like manner. Last, it might be a valid idea for learners to train their perception, as we have seen that perception precedes production. Therefore, training perception might lead to improved production.

4.6.2 Perception

As with production, the results gained from the perception experiment show that there are a number of factors which have the potential to make a learner successful in perceiving vowel length variation in a native-like manner:

- **Phonological proficiency** strongly influences the perception of vowel length variation. Learners who have two distinct categories for TRAP and DRESS perform significantly better in the perception experiment than those who merge the vowels. Likewise, speakers who do not exhibit final devoicing make fewer mistakes in the perception experiment.
- **Performance in the production experiment** is a significant influence on the performance in the perception experiment. Speakers who produce greater vowel length variation make significantly fewer mistakes in the perception experiment.
- There is no effect of active use of English with native speakers. Learners who converse mostly with native speakers do not perceive vowel length variation better than those who predominantly speak to other non-native speakers.
- There is no effect of gender, the emphasis a learner places on achieving a native-like accent, how proficient a learner rates him-/herself, or active and passive language use.

In the following, the remaining results are listed according to groups:

British Group

- All three groups – British native speakers, G-UK and C-UK – are significantly influenced by the **vowel length variation** exhibited by the two speakers who recorded the tokens for the perception experiment.
- The British native speakers misheard the most tokens which ended in the **consonants** /d, p/ → /b, k/ → /g, t/. Group G-UK shows the same sequence regarding the voiceless consonants /p/ → /k/ → /t/, but inverts the first two voiced consonant groups: /b/ → /d/ → /g/. The control group C-UK matches the findings for group G-UK.
- The British native speakers made most mistakes when the vowel was KIT → DRESS → TRAP → LOT, as did group G-UK. Group C-UK matches this sequence except that LOT is the vowel where most mistakes occur: LOT → KIT → DRESS → TRAP.
- **Phonological proficiency** has no effect on the British native speakers, because they do not merge TRAP and DRESS and do not show devoicing. In contrast, both factors are significant in group G-UK. Learners who do not exhibit devoicing and have established two separate vowel categories perform significantly better in the perception experiment. Due to small group size, there is no significant effect in group C-UK. Therefore, their performance is best viewed in relation to group G-UK's (cf. chapters 4.5.1 and 4.5.2).
- **Word frequency** is not a significant influence for the British native speakers, but it is for group G-UK. It is not significant in group C-UK. This result might be due to the fact that British native speakers are able to use durational cues in the absence of spectral cues for the consonant, no matter what (the frequency of) the token is. In contrast, the learners in group G-UK are more likely to have heard more frequent words before, enabling them to recognize them when presented with them in the perception experiment, even when the final consonant is missing.
- The absolute **length of the vowel** (in ms) was found to be a significant factor in groups G-UK and C-UK, but not for the British native speakers. This indicates that (even advanced) learners need more time to process vowel duration.

American Group

- The **vowel length variation** exhibited by the two speakers who recorded the tokens for the perception experiment has a significant influence on all three groups (American native speakers, G-US and C-US).
- The American native speakers identified the most tokens incorrectly which ended in the **consonants** /b, k/ → /d, t/ → /g, p/. Group G-US shows the same sequence regarding the voiceless consonants /k/ → /t/ → /p/, but inverts the last two voiced consonant groups: /b/ → /g/ → /d/. The control group C-US shows sequences which are quite different: /g, t/ → /b, p/ → /d, k/.
- The American native speakers misidentified most tokens when the **vowel** was TRAP → KIT → DRESS → LOT. Group G-US does not match this sequence: TRAP → LOT → KIT → DRESS. Group C-US is different again: KIT → TRAP → LOT → DRESS.
- **Phonological proficiency** has no effect on the American native speakers who do not merge TRAP and DRESS and do not devoice final consonants. However, both factors have a significant effect on group G-US. Learners who do not display devoicing and who have two separate vowel categories identified significantly more tokens correctly in the perception experiment. Due to small group size, there is no significant effect in group C-US. Therefore, their performance is best viewed in relation to group G-US's (cf. chapters 4.5.1 and 4.5.2).
- **Word frequency** is only a significant factor in group G-US. It has no effect on the American native speakers and group C-US. The reasons for this finding are probably the same as for the British group.
- In results similar to those for the British group, the absolute **length of the vowel** (in ms) is significant in groups G-US and C-US, but not for the American native speakers. This indicates that (even advanced) learners need more time to process vowel duration. This is further evidence that learners need a longer amount of time to process vowel length variation.

The findings of the perception experiment lead to the conclusion that there are fewer factors involved, which makes it quite difficult to give learners an indication of which aspects they should work on to improve their perception. As is the case with production, phonological proficiency is key. However, word frequency, the absolute length of a word or the amount of vowel length variation exhibited by interlocutors cannot be controlled by learners. A second important factor might be perception training. As we have seen, perception precedes production, so that improved perception might have the additional advantage that learners also become more successful in the production of vowel length variation.

The tables on the following pages are a quick overview of the significance (or lack thereof) of the various factors that were tested in the perception and production experiment.

Table 4-17. Overview of results from the British native speakers, groups G-UK and C-UK gained from the production experiment. VLV = vowel length variation.

	British native speakers	**G-UK**	**C-UK**
VLV in vowels ***most → least***	Trap → Lot → Dress → Kit	Trap → Lot → Dress → Kit	Trap → Dress → Lot → Kit
VLV before consonants ***most → least***	/g/ → /b/ → /d/ /k/ → /p/ → /t/	/g/ → /d/ → /b/ /k/ → /t/ → /p/	/d/ → /b/ → /g/ /t/ → /k/ → /p/
Merger of Trap and Dress	No effect (*p* = .73)	Significant influence (*p* = .05)	No effect (*p* = .46)
Devoicing	No effect (*p* = .21)	Significant influence (*p* = .05)	No effect (*p* = .61)
Word frequency	Significant influence (*p* = .02)	Significant influence (*p* = .002)	No effect (*p* = .13)

Table 4-18. Overview of results from the American native speakers, groups G-US and C-US gained from the production experiment. VLV = vowel length variation.

	American native speakers	**G-US**	**C-US**
VLV in vowels ***most → least***	Dress → Kit → Trap → Lot	Trap → Lot → Dress → Kit	Trap→ Dress→ Lot→ Kit
VLV before consonants ***most → least***	/g/ → /b/ → /d/ /k/ → /p/ → /t/	/g/ → /d/ → /b/ /k/ → /t/ → /p/	/d/ → /g/ → /b/ /k/ → /p/ → /t/
Merger of Trap and Dress	No effect (*p* = .51)	Significant influence (*p* = .02)	No effect (*p* = .78)
Devoicing	No effect (*p* = .33)	Significant influence (*p* = .05)	No effect (*p* = .39)
Word frequency	Significant influence (*p* = .004)	Significant influence (*p* = .0005)	No effect (*p* = .26)

Table 4-19. Overview of results gained from the production experiment in analyses where groups G-UK, G-US, C-UK and C-US were pooled together.

	G-UK, G-US, C-UK, C-US
Length of stay abroad ***least similar ← most similar*** **(compared to native speakers)**	*Home ← One year ← Half a year ← One and a half years ← More than two years ← Two years*
Reason for stay abroad ***least similar ← most similar*** **(compared to native speakers)**	*Home ← School ← Travel ← FLA ← University ← Family*
Gender	No effect ($p = .91$)
Language use	No effect (active use of English $p = .37$; passive use of English $p = .56$)
Active use with native speakers	Significant influence ($p = .007$)
Self-rating of phonological proficiency	Significant influence ($p < .0000001$)
Importance of native-like accent	No effect ($p = .19$)
Performance in the perception experiment	Significant influence ($p = .02$)
Merger of Trap and Dress	Significant influence ($p < .0000000001$)
Devoicing	Significant influence ($p = .0005$)

Table 4-20. Overview of results from the British native speakers, groups G-UK and C-UK gained from the perception experiment.

	British native speakers	**G-UK**	**C-UK**
Vowel length variation	Significant influence (p < .00001)	Significant influence (p < .000001)	Significant influence (p < .04)
Mistakes before consonants ***most → least***	/d/ → /b/ → /g/ /p/ → /k/ → /t/	/b/ → /d/ → /g/ /p/ → /k/ → /t/	/b/ → /d/ → /g/ /p/ → /k/ → /t/
Mistakes in vowels ***most → least***	Kit → Dress → Trap → Lot	Kit → Dress → Trap → Lot	Lot → Kit → Dress → Trap
Merger of Trap and Dress	No effect (*p* = .98)	Significant influence (*p* = .05)	No effect (*p* = .16)
Devoicing	No effect (*p* = .10)	Significant influence (*p* = .009)	No effect (*p* = .39)
Word frequency	No effect (*p* = .24)	Significant influence (*p* < .0001)	No effect (*p* = .96)
Length of token	No effect (*p* = .11)	Significant influence (*p* < .0001)	No effect (*p* = .21)

Table 4-21. Overview of results from the American native speakers, groups G-US and C-US gained from the perception experiment.

	American native speakers	**G-US**	**C-US**
Vowel length variation	Significant influence ($p = .05$)	Significant influence ($p < .00000001$)	Significant influence ($p = .007$)
Mistakes before consonants ***most → least***	/b/ → /d/ → /g/ /k/ → /t/ → /p/	/b/ → /g/ → /d/ /k/ → /t/ → /p/	/g/ → /b/ → /d/ /t/ → /p/ → /k/
Mistakes in vowels ***most → least***	Trap → Kit → Dress → Lot	Trap → Lot → Kit → Dress	Kit → Trap → Lot →Dress
Merger of Trap and Dress	No effect ($p = .34$)	Significant influence ($p = .04$)	No effect ($p = .74$)
Devoicing	No effect ($p = .24$)	Significant influence ($p = .04$)	No effect ($p = .61$)
Word frequency	No effect ($p = .21$)	Significant influence ($p < .0000001$)	No effect ($p = .43$)
Length of token	No effect ($p = .64$)	Significant influence ($p < .00001$)	Significant influence ($p = .002$)

Table 4-22. Overview of results gained from the perception experiment in analyses where groups G-UK, G-US, C-UK and C-US were pooled together.

	G-UK, G-US, C-UK, C-US
Length of stay abroad ***least similar ← most similar*** **(compared to native speakers)**	*One year ← One and a half years ← Half a year ← Home ← More than two years ← Two years*
Reason for stay abroad ***least similar ← most similar*** **(compared to native speakers)**	*Travel ←University ← FLA ← Home ← School ← Family*
Gender	No effect (*p* = .26)
Self-rating of phonological proficiency	No effect (*p* < .15)
Importance of native-like accent	No effect (*p* = .09)
Language use	No effect (active use of English *p* = .07; passive use of English *p* = .65)
Active use with native speakers	No effect (*p* = .29)
Performance in the production experiment	Significant influence (*p* = .02)
Merger of TRAP and DRESS	Significant influence (*p* = .03)
Devoicing	Significant influence (*p* < .001)

5 Discussion

The study focus of the present dissertation sought to examine and describe how German L2 learners of English acquire vowel length variation in production and perception. This was done by confronting participants with four distinct production tasks as well as a perception experiment and observing how different speakers compare to native speakers of English. Now that all results have been listed and explained in the previous chapter, it is time to return to the research questions posed in section 2.5.

Which factors make learners successful?

Regarding L2 production, we saw that both attitude and the learning environment of the learners were very important in determining the success at acquiring vowel length variation. Speakers who rated themselves as more proficient, i.e. who believed in themselves, actually produced more vowel length variation than those who doubted their abilities. This suggests that a positive attitude towards one's own abilities, which very probably has a direct influence on motivation, can help learners be successful.

Age, although it was not explored as a variable in this dissertation, did have an effect on language learning, since those learners who went abroad because of their family (i.e., grew up bilingually with a native English-speaking parent and were either born in the UK / the USA, or went their later on in their lives to visit family) performed to the most native-like standard. However, it is not clear whether this is predominantly due to their AOL, increased and more optimal input, their more holistic identification with the language, or a combination of these factors. We must also bear in mind that only 5 out of 60 participants belong in this category, which does not allow any strong statements to be made. Moreover, all of them established German as their dominant native language, which makes them learners of English, even if they were very early and successful learners.

A finding which emphasizes the importance of a stay abroad and the ensuing confrontation with native speakers of the L2 is that word frequency strongly affects production. Since words of higher frequency are more likely to be heard more often by learners, this means that learners have a greater chance of storing the correct pronunciation and reproducing it when necessary. In the study, more frequent words were produced with significantly greater variation than rare words. This warrants the hypothesis that L2 speakers acquire vowel length variation largely by mimicking native speakers. When L2 learners lack input from native speakers, they will face difficulty in production. Interestingly, this showed most clearly in the reading of the minimal pairs. Several participants from the German learner groups hesitated when confronted with certain words (e.g. *cob*) and remarked that they did not know how to pronounce the word, since they had never heard it before. When they were encouraged by the researcher they eventually pronounced the phonemes correctly, but most of them did not use accurate allophonic vowel length.

> *What is the role of the stay abroad of the learners and what is the role of their reason for it?*

A second important factor is the language experience of the learners. Learners who were confronted with native speakers on a more longitudinal level reached a more native-like production of vowel length variation. This was true also for those who conversed with native speakers after they returned to Germany – they were significantly more proficient than those who predominantly spoke to other non-native speakers. However, the stay abroad that the participants completed is a much stronger influence, since within the group of proficient speakers, those who conversed predominantly with non-native speakers after their return did not produce significantly lower vowel length variation than those who continued speaking to native speakers ($p = .19$). This indicates that vowel length variation is strongly implemented into the learners' minds during their time abroad. Therefore, completing a stay abroad, where the learner is confronted with the L2 on a daily basis, and experiences life in all its facets in the L2, is a crucial success factor on a subphonemic level. In addition, we must also bear in mind that the reason for the stay abroad was shown to be of utmost importance: Those learners who also experienced the L2 on a deeper cognitive level (what Cummins (1979; 1991) calls *CALP*) performed to a more native-like degree than those who only experienced the L2 as part of their travel or school exchange (Cummins' *BICS*).

> *Does "attention" make a difference? That is, are learners more successful in formal tasks such as minimal pair and word list reading, where they pay closer attention to their pronunciation?*

This question is not one to be answered easily. As a reminder, the two learner groups G-UK and G-US managed to match their native speaker model in a differing number of tasks, as illustrated by the following table:

Table 5-1. Overview of the tasks in which groups G-UK and G-US match their native speaker model, i.e. when they are *not* significantly different from the British and American native speakers, respectively.

	G-UK				**G-US**			
	KIT	DRESS	TRAP	LOT	KIT	DRESS	TRAP	LOT
Minimal pairs	✓	✓	✓					
Word list	✓	✓	✓					
Sentences	✓		✓				✓	
Text			✓	✓			✓	✓

Group G-UK matches their native speaker model more often in the formal tasks (minimal pairs and word list reading), so that one might assume that attention does make a difference for them. When they pay close attention to their pronunciation, they produce native-like vowel length variation in KIT, DRESS and TRAP. However, they also perform similarly to the British native speakers in both formal tasks (sentence and text reading) in TRAP and in one formal task each in KIT and LOT. Moreover, group G-US only matches their native speaker model in informal tasks. This, in contrast, might suggest that learners are able to match native speakers first in informal tasks, where connected speech phenomena such as liaison or vowel reduction simplify the accurate production of vowel length variation (and where native speakers produce lower variation values due to these phenomena and are therefore easier to match). We must bear in mind, however, that the American native speakers' production of vowel length variation was at odds with what was expected (DRESS → KIT → TRAP → LOT). Thus, it is more accurate to say that in the sentence and text reading tasks in TRAP and LOT, the American native speakers actually approach group G-US's production, rather than the other way around. This is further evidenced when we compare group G-US to the British native speakers, who show much more variation (and the expected sequence). Group G-US cannot match the British native speakers in the tasks that they match the American native speakers in. They are significantly different in the sentence and text reading task in TRAP ($p = .004$ each) and the text reading task in LOT ($p < .0000001$). Likewise, group C-US cannot match the British native speakers in the sentence and text reading task in TRAP ($p = .02$ and $p = .002$, respectively). This suggests group G-US is only able to match the American native speakers in the tasks outlined in Table 5-1 above because the native speakers produce such surprisingly little variation. Therefore, we should concentrate on the learner group G-UK for the analysis. Their behavior supports the hypothesis that formal tasks are easier to match for learners after all. This is possibly due to the fact that in these tasks, learners pay close attention to their pronunciation.

In which vowels (Kit, Dress, Trap, Lot) and coda plosives (/b/, /p/, /d/, /t/, /k/, /g/) can learners most easily match their native speaker target?

In order to discuss this question, we may use Table 5-1 above again, as well as Table 5-2 below, which lists the production of vowel length variation of the individual groups from most to least variation:

Table 5-2. Order of vowel length variation exhibited by the individual groups in the production experiment, from most to least variation.

	Vowel length variation ***most → least***
British native speakers	Trap → Lot → Dress → Kit
G-UK	Trap → Lot → Dress → Kit
C-UK	Trap → Dress → Lot → Kit
American native speakers	Dress → Kit → Trap → Lot
G-US	Trap → Lot → Dress → Kit
C-US	Trap → Dress→ Lot→ Kit

At first sight, Table 5-2 suggests that new sounds might be easiest to match for German learners, since both groups G-UK and G-US produce most variation in Trap and Lot, as do the British native speakers.[75] Moreover, we saw that group G-UK performs similarly to the British native speakers in all tasks in Trap, but only three in Dress, two in Kit and one in Lot. However, when we consider the control groups, which exemplify an earlier learning stage, the tables turn. These groups prefer vowel length variation in the similar vowel Dress rather than Lot. McAllister *et al.* (2002), Pruitt *et al.* (2006) and Kondaurova & Francis (2008) have shown that native phonetic experience with a durational contrast may aid perception of non-native sounds. In the case of this study, native *phonemic* experience with duration may influence *production* of allophonic variation. We have seen that particularly regarding Dress, the phonemic durational differences in German are quite salient, with Dress being realized as long /eː/ and /ɛː/, and short /ɛ/. In contrast, English Lot, realized as /ɒ/ in British and /ɑː/ in American English, is not usually linked to German /ɔ/ and /oː/. Thus, speakers need some experience to be able to produce vowel length variation in Lot in a native-like manner. The development of vowel length variation in Lot is doubtlessly helped

---

[75] Due to their unexpected and inexplicable behavior in the experiment, group Americans will be ignored in this analysis.

by the fact that native speakers produce relatively high levels of variation in this vowel, since it is a low vowel with intrinsically stronger variation. In the end, however, group G-UK only matches the British native speakers in one task in LOT vs. three tasks in KIT and two in DRESS. This suggests that even though German learners produce overall more variation in LOT, the variation is not as native-like as in the two similar vowels KIT and DRESS. This lends credence to the hypothesis that learners find it easier to produce native-like vowel length variation in similar sounds.

On the other hand, we have not examined TRAP yet. This is the vowel in which the British native speakers produce most variation, which makes this variation perceptually highly salient and possibly easier to reproduce. Group G-UK manages to match the British native speakers in all four tasks in this vowel. I would argue, however, that TRAP cannot be fully characterized as a new vowel for this group of learners. It is certainly a new vowel from a systemic point of view, because it does not exist in the German phoneme inventory. However, all learners who participated in this study have completed a language course specifically designed to improve their pronunciation. These language classes focus on teaching German learners to pronounce TRAP and DRESS distinctly. This results in the learners being more aware of the specific characteristics of TRAP than any other of the sounds examined within this dissertation. Therefore, TRAP may be viewed as a “similar” vowel in the sense that it is well known to the German learners. Cebrian (2006) has previously argued in the same direction, namely that his results were influenced by “undergraduate students’ greater metalinguistic knowledge as a result of formal instruction in English linguistics” (Cebrian 2006: 383). In conclusion, group G-UK’s metalinguistic knowledge of the sound TRAP together with the highly salient vowel length variation produced by the native speakers probably led to the result that this is the vowel in which the German learners can match their native speaker model best.

Salience also played a part in the way in which the different coda consonants influence the learners’ production of vowel length variation. Consider Table 5-3 below:

Table 5-3. Order of vowel length variation preceding the six coda consonants exhibited by the individual groups in the production experiment, from most to least.

	Vowel length variation in relation to the preceding coda consonant *most → least*
British native speakers	/g, k/ → /b, p/ → /d, t/
G-UK	/g, k/ → /d, t/ → /b, p/
C-UK	/d, t/ → /b, k/ → /g, p/
American native speakers	/g, k/ → /b, p/ → /d, t/
G-US	/g, k/ → /d, t/ → /b, p/
C-US	/d, k/ → /g, p/ → /b, t/

Both native speaker groups and the two learner groups C-US and C-UK produced most variation in vowels preceding /g, k/. However, while the second largest variation in the native speaker groups was observed before /b, p/, group C-US and C-UK produced least variation before this coda consonant group. Instead, these groups produced more variation before /d, t/. This might be explained by salience: /t/ is often glottalized or unreleased in final position, which renders this opposition and the related vowel length variation particularly salient. The two control group C-US and C-UK do not show a recognizable pattern. Moreover, group C-UK produces more variation before the voiceless consonant /k/ than the voiced coda /g/; group C-US produces greater variation in vowels preceding /p/ than /b/. In conclusion, this means that vowel length variation in vowels preceding the different coda consonants only emerges accurately with experience abroad. Still, even with more experienced speakers, salience of the coda consonant has a strong influence.

Perception

Which factors make learners successful? What is the role of the stay abroad of the learners and what is the role of their reason for it?

The second large part of the present dissertation aimed to study how German learners *perceive* vowel length variation. We saw that the factors which made certain learners successful were not as clear as those observed in the area of production. The length of the stay abroad did not have a clear linear effect on perception, and neither did the reason for the stay abroad. The two most proficient groups were groups *Family* and *School*, i.e. speakers who grew up bilingually and those who went abroad on school exchange during their teenage years. This suggests a link to age. Since age was not a factor studied in this dissertation, however, this is only a first hypothesis and should be researched further. It stands to reason that AOL should have a similarly strong influence on the perception of subphonemic variation as it does on the perception of phonemes.

As was the case with production, word frequency played a significant role in the perception of vowel length variation. This factor was relevant for groups G-UK, G-US and C-US. This again hints at the fact that vowel length variation is a learned feature for L2 speakers, but not for native speakers. Learners have a higher likelihood of hearing frequent words pronounced. They can store these high frequency words more easily and recall them when they are confronted with them in a perception task. In contrast, native speakers are not influenced by word frequency, but are able to use durational cues no matter what the frequency of the word is.

Do learners use the same cues as native speakers in perceiving vowel length variation?

The perception experiment revealed that learners use cues very differently than native speakers. One noteworthy finding was that although all groups were influenced by the vowel length variation exhibited by the speakers of the tokens, the direction of the influence was quite different. The British and American native speakers used vowel shortening as a cue, i.e. they made significantly fewer mistakes when the vowel was relatively short because it appeared before a voiceless consonant. In contrast, none of the learner groups used this cue; instead, all groups focused on vowel lengthening. The longer the vowel, the more successfully learners identified the token as one with a voiced coda consonant. This finding matches the results proposed by van der Feest & Swingly (2011). Furthermore, learners were influenced by the overall length of the token (in milliseconds), which was not the case with the native speakers. This suggests that learners not only use different cues in the perception of vowel length variation, but that they also need exaggerated cues. What is more, learners are not able to fully rely on duration as cues to post-vocalic consonant voicing in the absence of spectral cues of the consonant. While both the British and the American native speakers made significantly fewer mistakes in tokens that they heard later in the trial ($p = .03$ and $p = .01$, respectively), suggesting that they realized at some point that they had to switch to durational cues, the learner groups did not manage this (G-UK $p = .77$, G-US $p = .63$, C-UK $p = .29$, C-US $p = .44$).

Is VLV in similar vowels (Kit, Dress) easier to perceive than in new ones (Trap, Lot)?

In the course of the theoretical analysis of how learners perceive foreign sounds, we examined four different perception models: Flege's SLM, Bohn's PAM, Kuhn's NLM and Escudero's L2LP. We noted that two of the models – the NLM and the PAM – consider L2 similar sounds as no special learning challenge, because the learner can use already established sounds/features. In contrast, the SLM views the acquisition of similar sounds as a great challenge due to equivalence classification, which may block the establishing of categories in the L2. The L2LP assumes an intermediate position and suggests that perceptual learning of similar L2 sounds is a challenge for learners because perceptual mappings of the L1 have to be adjusted. However, new sounds are considered to pose even greater difficulty, because this requires the learner to integrate and use cues which are not part of the L1 system. This view is supported by the NLM and PAM. In contrast, the SLM posits the formation of a category for a new sound to be relatively easy, because the category can be established in the phonological space without influencing or disturbing any pre-existing L1 sound.

The obvious question now is whether participants had more difficulty identifying similar sounds (Kit, Dress) or new ones (Trap, Lot), and whether their behavior supports one of the perception models discussed above. Unfortunately, there is no easy answer to this question because all groups performed quite differently. The individual groups showed the following order of mistakes:

Table 5-4. Order of mistakes regarding vowels made by the individual groups in the perception experiment, from most to least.

	Mistakes in vowels ***most → least***
British native speakers	Kit → Dress → Trap → Lot
G-UK	Kit → Dress → Trap → Lot
C-UK	Lot → Kit → Dress → Trap
American native speakers	Trap → Kit → Dress → Lot
G-US	Trap → Lot → Kit → Dress
C-US	Kit → Trap → Lot →Dress

As is evident from the table, only the British native speakers and the learner group G-UK show the same order. They make most mistakes when the vowels are similar and fewer mistakes when the vowels are new. This would support the hypothesis posited by the SLM. Group C-UK is somewhat similar to the British native speakers and G-UK, except for LOT, where this group misidentified most tokens. Group C-UK also produces less variation in LOT than the British native speakers and group G-UK and prefers the similar vowel DRESS to the new vowel LOT in production. This, in turn, would suggest that inexperienced speakers have more trouble with new sounds. However, this does not fit in with the fact that group C-UK makes fewest mistakes with TRAP, which is also a new sound. Since this is a sound which German learners are trained to perceive and produce in the language classes at university, however, it is quite possible that this specific training had an effect on their perception. LOT, in turn, is not trained as being distinguished from another German sound; therefore, it might be considered as more of a foreign sound, which might lead inexperienced learners to struggle with its perception. We must also bear in mind that the vowel length variation produced by the speakers who recorded the tokens is greater in TRAP and LOT than in KIT and DRESS, which makes it harder to perceive it in higher vowels. It might be interesting to examine learners of a language which has similar low vowels to English, but different high vowels, in order to gauge the effect that the intrinsic vowel length variation has on perception in contrast to the way the two vowel systems relate regarding the “similar” / “new” distinction.

The analysis for the American group yields a vastly dissimilar picture. All groups show different orders, and the learner group G-US makes more mistakes in similar vowels than with new ones. Thus, their performance supports the PAM, the NLM and the L2LP, which argue that the perception of new sounds is a great challenge. The control group C-US makes most mistakes with KIT, which is similar to the British native speakers and group G-UK. In contrast, they make fewest mistakes with DRESS, which is the vowel with the second highest error rate in the British group and G-UK. Thus, the analysis is quite complicated. What is also very interesting is that the American native speakers misidentified most tokens incorrectly when the vowel was TRAP. We should recall that this is the vowel in which this group also produces the lowest variation. This might be an odd coincidence, or it might suggest that there is a larger change going on in this vowel in American English. In any case, the largely opposite performance of the American and the British group both in the production and the perception experiment do not allow a finite statement as to whether similar sounds or new ones pose greater difficulty for learners.

Which tokens including which (missing) coda plosives (/b/, /p/, /d/, /t/, /k/, /g/) are easiest to identify?

In order to answer this question, let us first have a look at the performance of the individual groups in the perception experiment:

Table 5-5. Order of mistakes in the perception experiment in relation to the coda consonant, from most to least.

	Number of mistakes in the perception experiment in relation to the coda consonant ***most → least***
British native speakers	/d/ → /b/ → /g/ /p/ → /k/ → /t/
G-UK	/b/ → /d/ → /g/ /p/ → /k/ → /t/
C-UK	/b/ → /d/ → /g/ /p/ → /k/ → /t/
American native speakers	/b/ → /d/ → /g/ /k/ → /t/ → /p/
G-US	/b/ → /g/ → /d/ /k/ → /t/ → /p/
C-US	/g/ → /b/ → /d/ /t/ → /p/ → /k/

In the British sector, all three groups show the same order of mistakes within the voiceless coda plosives. Interestingly, /t/ shows the lowest error rate, which may again be explained by the salient realization of it. Within the group of voiced consonants, groups G-UK and C-UK listened more to the vowel length variation exhibited by the speakers of the tokens. Since the speakers produced more variation in /g/ than /d/ and /b/, speakers made the according number of mistakes. Within the American group, the findings are slightly different. Although both the American native speakers and group G-US show the same order of mistakes concerning voiceless coda plosives, /t/ does not show the lowest error rate as was the case in the British group. This might be because glottalized /t/ is not a feature of American English, but unreleased /t/ is. Thus, final /t/ might be slightly less salient in American than in British English. The mistakes which occurred when the coda consonant was voiced was not easily explained by the vowel length variation exhibited by the native speakers who recorded the tokens. The high spread within the production of vowel length variation doubtlessly played a role. This means that we can conclude two things regarding the perception of vowel length variation: Firstly, perception is helped by the amount of vowel length variation exhibited by the speaker, but only if this variation is fixed and does not vary too much among different tokens. Secondly, allophonic realizations which make a coda consonant particularly salient help learners tune in to durational differences more easily, thus enabling them to recognize tokens even in the absence of spectral information for the coda consonant.

> *Is success at producing phonemic contrasts an indicator for success at producing and perceiving subphonemic contrasts?*

An issue which can be answered with the data available from this study is that perception precedes production. Many researchers have suggested this, among them Trubetzkoy (1939), Flege (1987; 1991), Neufeld (1988), Llisterri (1995), Rochet (1995) and Flege & MacKay (2004), and the present study has found support for this hypothesis. There are a number of findings which can serve as evidence. First, we found a strong connection between perception and production: Learners who made fewer mistakes in the perception experiment produced significantly higher vowel length variation. At the same time, there were no speakers who perceived vowel length variation poorly but produced it in a native-like manner. However, there are speakers who perceive vowel length variation well but do not produce it. These speakers can be found in the control groups. Five out of the six speakers in groups C-UK and C-US made fewer than average mistakes in the experiment, but do not produce a large amount of vowel length variation. Since the speakers of the control group are more inexperienced than those of the two learner groups G-UK and G-US, they can be said to exemplify an earlier stage of learning where perception is already present, but production is not.

> *Are speakers who merge TRAP and DRESS less able to produce and/or perceive vowel length variation?*

The final part of the analysis focused on the relationship between phonological proficiency and success at producing and perceiving subphonemic vowel length variation. Phonological proficiency was assessed twofold: Firstly, by measuring whether the German learners had managed to establish two separate categories for TRAP and DRESS, and secondly, by examining whether participants exhibited devoicing. These two phonological features are the most difficult for German learners to eradicate, and the easiest one to spot a foreign accent in. In the course of the analysis we found that a merger of TRAP and DRESS has a strong influence both on perception and production: Speakers who merge these two vowels produce significantly lower vowel length variation and perform below average in the perception experiment. In a nutshell, this means that learners who have managed to establish separate spectral categories for TRAP and DRESS also manage to produce and perceive vowel length variation accurately. This finding is directly relevant regarding the representation of sounds in a learner's mind. Sounds are not represented as several distinct cues (i.e. spectral vs. durational cues) but as a bundle of cues which together form the category. Also, these different cues of the segment develop together, at least in production.[76] These bundles, or chunks, are the basis of a mechanism known as *chunking*, which has been observed in many areas of the linguistic system. Chunking has been established as one of the key cognitive mechanisms in humans; "each chunk collects a number of pieces of information from the environment into a single unit" (Gobet *et al.* 2001: 236). This means that learners perceive a chunk consisting of spectral and durational properties of a vowel rather than single cues which they deliberately merge together. Furthermore, learners chunk increased vowel length variation and voiced sounds / shortened vowels followed by voiceless codas, respectively. Evidence of chunking can be found when "primitive stimuli are grouped into larger conceptual groups, such as the manner by which letters are grouped into words, sentences or even paragraphs" (Gobet *et al.* 2001: 237). Similar effects were shown in

- verbal learning (Ebbinghaus (1885), Feigenbaum & Simon (1962; 1984), Simon & Feigenbaum (1964), Brown & McNeill (1966), Simon & Simon (1973)),
- letter perception (Richman & Simon (1989)),
- concept formation (Gobet *et al.* (1997)),
- acquisition of vocabulary (Jones *et al.* (2000)) and grammar (Tomasello (1992), Cook (1994)),
- syntactic categories (Croker *et al.* (2000), Jones *et al.* (2000)),
- lexical recognition and production processes (Kirsner (1994)),
- word associations (Entwisle (1966), Södermann (1993)),
- collocations and idioms (Sinclair (1991), Ellis (1994)), as well as
- phonology (Leather & James (1991), Ellis & Beaton (1993)).

As observed in the present dissertation, the phonological chunking of L2 learners develops with increasing input. In this respect, the two control groups are an important source of evidence because they exemplify the least proficient speakers who have mastered neither the

---

[76] No statement can be made on the *perception* of the difference between TRAP and DRESS and its relation to the perception of vowel length variation, since the perception of spectral differences was not tested in this experiment.

spectral nor the durational cue. The large spread in the data supports the analysis that chunking emerges as learners become more proficient. Even though learners differ in their ease of acquisition, there was an overall linear trend between merger and production/perception of vowel length variation (cf. Figure 4-64 and Figure 4-66).

Is there a link between vowel length variation and devoicing?

The presence of a link between vowel length and devoicing was even clearer, possibly because vowel length variation is a direct cue to post-vocalic consonant voicing. The theoretical chapter outlined two plausible hypotheses: 1) Learners who exhibit strong devoicing might be unable to produce accurate vowel length variation, or 2) learners might use exaggerated vowel length variation in order to compensate for devoicing. We found hypothesis 1 to be correct. Speakers who did not show devoicing produced significantly higher vowel length variation than those who had difficulty producing a voicing contrast in the final plosive. In addition, speakers without devoicing also made significantly fewer mistakes in the perception experiment (cf. Figure 4-60 and Figure 4-62). This finding is rather unfortunate for the less proficient speakers. When learners produce both tokens of a minimal pair with equal vowel length and equal (devoiced) consonants the tokens become indistinguishable, which can seriously impede communication. Vowel length variation would be an ideal cue for German learners in order to handle devoicing, especially because it is easier to acquire than to eradicate devoicing. If German speakers could learn to use vowel length variation accurately, they would be able to communicate successfully even if they had difficulty pronouncing final voiced consonants. After all, devoicing also occurs in native speakers of English, if not as strongly as in German L2 learners. However, with accurate vowel length variation, the intended meaning of the token can still be signaled. A study of whether and how strongly native speakers of English are affected by a foreign accent which shows a combination of inaccurate vowel length variation and devoicing would provide valuable insight here.

Where do we go from here?

It has become clear that vowel length variation is a valuable cue for German learners. On the one hand, it can be used as an additional cue in order to help them establish separate spectral categories for TRAP and DRESS, for instance by making them aware of the fact that TRAP has greater intrinsic duration and therefore also greater vowel length variation than DRESS. We have seen in the course of this study that learners acquire vowel length variation together with spectral cues for each phoneme, so that a focus on vowel length variation might speed up the acquisition of spectral cues. Durational cues, after all, seem to be easier for German learners to employ than spectral ones, since durational cues are already present in the German language as a phonemic cue, but spectral cues for TRAP are absent. Furthermore, since vowel length variation is a direct cue to post-vocalic consonant voicing, acquiring this variation would also help German learners who struggle with devoicing. The latter is one of the most difficult features of a German accent to eradicate, and persists even in many advanced learners. In contrast, we have seen that German learners can learn to use vowel length variation with relative ease. This means that vowel length variation is a valuable cue for learners in order to be able to signal meaning even without the voicing distinction in the coda.

Unfortunately, teaching has for a very long time refrained from focusing on units below the segmental level. Even though instruction in the 1950s and 1960s placed much emphasis on learners attaining a native-like accent, this approach lost more and more prominence between the 1960s and 1980s (cf. Preston (1981)). At that time, many questioned that it was even possible for learners to achieve native-like pronunciation. Therefore, they concentrated their efforts on teaching grammatical constructions and translation instead. In the 1990s, finally, teaching re-discovered pronunciation as a means to communicative fluency. However, many focused on the suprasegmental level and voice quality (cf. Morley (1991), Kruger Bott (2005)) and argued that "[e]mpirical research and pronunciation materials' writers suggest that teaching suprasegmentals before segmentals to intermediate and advanced [non-native speakers] could be more beneficial in a shorter period of time" (Kruger Bott 2005: 5). For instance, Anderson-Hsieh & Koehler (1988), Anderson-Hsieh *et al.* (1992), Derwing *et al.* (1998) and Moyer (1999) all noted that suprasegmental training had a greater positive effect on pronunciation than segmental instruction. This renewed attention to pronunciation training doubtlessly also had to do with the fact that many non-native speakers – refugees and immigrants as well as students and professionals – were interested in improving their accent "because they left their native countries to accommodate in or visit English speaking countries to embrace cultural, economic, and financial opportunities" (Pourhosein Gilakjani 2012: 97, cf. also Derwing & Munro (2009)). This is not so different from the situation today, where globalization has increased international collaboration. People communicate with others from all over the world, not only via e-mail but also through video chats or in meetings. This means that adequate pronunciation is an important means of ensuring successful communication.

Still, the importance of teaching learners aspects of pronunciation below the segmental level has been largely neglected to this day. However, there are a few researchers who advocate that allophones should be included in teaching regimes. For instance, Prator writes that “if ability to speak English is an important objective, we should probably include attention to a few of the most important allophones that are in complementary distribution” (Prator 1971: 71). Similarly, Celce-Murcia *et al.* note that

> [i]n many languages, initial voiceless stops are less strongly aspirated than in English, or are even unaspirated. Speakers of these languages may therefore tend to confuse initial /b, d, g/ in English with their own language’s unaspirated /p, t, k/ in this position. These learners may be misperceived by English native speakers as producing *back* instead of *pack*, or *die* instead of *tie*. In fact what they may be producing is an unaspirated /p/ or /t/ in place of the English aspirated counterparts. They may, of course, also have difficulty in differentiating such minimal word pairs. For these learners, aspiration can provide a valuable clue to perceiving and producing these words. (Celce-Murcia *et al.* 1996: 63)

These researchers have evidently discovered the importance of teaching units below the segmental level. Regarding allophones, Picard thinks that “the highest priority should be given to flaps, especially those that are allophones of /t/” (2007: 333). He justifies the selection of this allophone by two criteria: 1) transferability and 2) differential salience. By transferability, Picard understands the “superior potential pronounceability” of certain allophones, “due to the fact that they are often found to be phonemic in many languages” (2007: 336). In turn, differential salience is defined as “the phonological distance that exists between a particular phone and its corresponding phoneme” (2007: 336). But what about teaching subphonemic units? Picard does not think much of this idea. He writes:

> What should also be disregarded are those subphonemic segments that are the “universal consequences of inherent properties of the human speech-producing mechanism” (Anderson 1976: 340), such as the lengthening of vowels before voiced consonants, e.g., bit vs. bid. (Picard 2007: 336)

The analysis and the findings of the present dissertation have demonstrated quite clearly that Picard’s view cannot be supported. Vowel length variation has been shown to be a very important cue for German learners which can help them signal meaning especially when they have trouble mastering units at the segmental level. Support can be found in Jenkins (2002), who argues in her ELF (*English as a Lingua Franca*) approach that students should use “British English /t/ between vowels in words such as ‘la<u>tt</u>er’, ‘wa<u>t</u>er’ rather than American English flapped [r] [sic!]” (Jenkins 2002: 96). In contrast, she strongly advocates the teaching of vowel length variation. According to her, “[s]hortening of vowel sounds before fortis (voiceless) and maintenance of length before lenis (voiced) consonants, for example the shorter /æ/ in 'sat' as contrasted with the longer /æ/ in 'sad'" is a phonetic requirement of L2 learners of English (Jenkins 2002: 96 f.).

As for Picard's criteria of transferability and differential salience, vowel length variation meets these in part: Vowel length is phonemic in German, so that transferability is given. Concerning differential salience, this is not necessarily the case. Both phones have the same features as the corresponding phoneme; they only differ in length. However, Picard's account is missing a third, very important criterion: an effort vs. use ratio. That is, what is the effort learners have to invest in order to learn a distinction in relation to how useful it is in ensuring ease of communication? This should certainly be considered when gauging which features to teach a foreign learner in an often relatively tight timeframe. Regarding allophonic flaps, the level of effort needed for learners to acquire them might not be high, but what is the use? Speakers might sound more "American" when they use dental flaps rather than [t], but meaning is signaled in exactly the same manner regardless of which of the two allophones is used.[77] In contrast, if we reflect on vowel length variation in this light, it becomes evident that teaching subphonemic features is very worthwhile. As mentioned before, vowel length variation can help German learners establish separate categories for TRAP and DRESS on a segmental level, and help them handle devoicing. This indicates that the level of usefulness of vowel length variation is extremely high. In contrast, the effort it takes learners to acquire this variation is very small. Even the less experienced speakers from the control group can perceive vowel length variation, and a stay abroad of as little as 4 months suffices for learners to start producing it. When we compare this to the difficulty associated with eradicating final devoicing – something many of the advanced learners in groups G-UK and G-US have not managed even after a stay abroad and more than 10 years of formal instruction in English – it is obvious that priority should be given to the subphonemic level.

At this point we should also ask ourselves what the ultimate goal is for L2 learners. Is it native-like competence, or intelligibility? Clearly, vowel length variation brings learners one step closer in both directions, whereas flapping only increases native-like competence, but not intelligibility. While native-like competence might be highly important for some individuals, researchers have shown that this is an ambitious goal which many (adult) learners can never reach, and it might not even be desirable. For instance, Leather (1983) notes that native-like pronunciation of an L2 learner may lead to negative reactions from native speakers (cf. 1983: 199). In addition, some speakers might also want to retain part of their accent "to mark their

---

[77] Cf. Avery & Ehrlich, who state that "[y]ou should not insist on having students pronounce flaps because using a /t/ where native speakers use a flap results in very little loss in comprehensibility" (1992: 42 f.).

L1 identity and to insure that they are not perceived as betraying their loyalty to their L1 community" (Morley 1991: 499). For instance, in a study by Jenkins, participants frequently reacted negatively when asked how they would feel if their accent was deemed native-speaker like: "I don't want to be what I am not. I am Italian, I have my own culture, my roots are Italian", "I feel Polish.... I don't want to sound like an English person, obviously not" (Jenkins 2005: 538). A last point is that when aiming for a native-speaker accent, the two most widely taught national norms are Received Pronunciation and General American English, but some learners might not want to identify with either variety (cf. McArthur (2002)). It thus seems that intelligible pronunciation is a more worthwhile goal for L2 learners. In this regard, Brawn (2010: 115 f.) differentiates learners' pronunciation into three levels:

- Level 1: The speaker is not understood due to errors on the segmental or prosodic level.
- Level 2: The speaker can be understood, but his/her pronunciation is unpleasant to listen to due to heavy accent.
- Level 3: The speaker is understood and his/her pronunciation is pleasant to listen to.

Brawn calls this last level "comfortably intelligible" (cf. also Abercrombie (1963), Scovel (1988)) and names this as the objective L2 learners should aim for. This does not mean, however, that individuals may not want to acquire native-like competence; they should be supported if they so choose (cf. Harmer (2001)). Within the frame of this study, many participants want to become teachers of English and aspire to achieve a native-like accent. This makes sense, too, as it would be counterintuitive to have L2 speakers with heavy accents teach children and expect the children to be able to reach comfortable intelligibility (level 3) with below-par input. Other studies, too, have found that L2 learners place high emphasis on accurate pronunciation and intelligibility. For instance, Derwing & Rossiter (2002) studied 100 adult learners of English with regard to their pronunciation difficulties. They found that "a majority of participants attributed their communication difficulties to pronunciation, at least in part, and over a third felt that their foreign accents were the primary cause of any communication breakdowns" (Derwing & Rossiter 2002: 162, cf. also Derwing (2003)). What is more, foreign accents not only impede communication, but they can also lead to stereotypes, prejudice and xenophobia. Many non-native speakers struggle with stereotypes and negative evaluations due to their accent (cf. Bradac (1990), Lippi-Green (1997), Derwing

(2003), Lindemann (2005), Gluszek & Dovidio (2010), Fuertes *et al.* (2012)). In particular, L2 speakers with accents are perceived as less pleasant to listen to (cf. Bresnahan *et al.* (2002), Lindemann (2003)), less intelligent (cf. Bradac (1990), Lindemann (2003)), less competent (cf. Bresnahan *et al.* (2002), Boyd (2003)), less loyal (cf. Edwards (1982)) and less credible (cf. Lev-Ari & Keysar (2010)). It has been argued that this is the case not because listeners consciously stereotype non-native speakers, but because accented speech is more difficult to process (cf. Munro & Derwing (1995), Lev-Ari & Keysar (2010)). This means that intelligibility is highly important and should be the major goal of pronunciation training.

Last, there is also a political dimension to language learning and L2 speakers becoming proficient and intelligible. After two council resolutions in 1997 and 2002 on the promotion of foreign language learning, the European Commission finally established a new policy at a summit meeting in Barcelona in 2002. According to this “mother tongue + 2” policy, every European citizen is supposed to master his/her native language plus two foreign European languages (Barcelona European Council 2002: 19).[78] This “stresse[s] that the knowledge of languages is one of the basis [sic!] skills each citizen needs in order to take part effectively in the European knowledge society” (Council of the European Union 2008).[79] As the ultimate goal, the EU policy names[80]

> effective communicative ability: active skills rather than passive knowledge. ‘Native speaker’ fluency is not the objective, but appropriate levels of skill in reading, listening, writing and speaking in two foreign languages are required, together with intercultural competencies and the ability to learn languages whether with a teacher or alone. (Action Plan 2004 – 2006: 8)

However, more than 10 years later, many of the *Bundesländer* in Germany have yet to address this in their curricula. It is quite clear that pronunciation training must be integrated from an early age on in order to achieve the goal of multilingual competence and effective communicative ability. Baden-Württemberg has now decided to include pronunciation as a major component of the curriculum in modern foreign languages. From 2014 onwards, foreign language classes will include pronunciation training. In addition, choosing English as a major subject for the *Abitur* will come with a compulsory oral exam in the form of a 10 minute oral presentation and subsequent colloquium.[81] The curriculum emphasizes *die Fähigkeit zur angemessenen Darstellung* (‘the ability to adequately present the topic’) and

---

[78] <http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/71025.pdf>.

[79] <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:320:0001:01:en:HTML>.

[80] cf. Commission of the European Communities 2003 <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:EN:PDF>

[81] Please find example exercises for the colloquium at Cornelsen: <http://www.cornelsen.de/shop/capiadapter/download/get/file/9783069604709_x1PS_KommPruefung_HE_C21.pdf?fileNr=1&sku=3-464-00076588-0&callerId=SBK3>. Note that this file was created for the *Bundesland* of Hessen; unfortunately, no such reference material exists for Baden-Württemberg.

links this to the following conversational abilities (Ringel-Eichinger & Selz 2001/2002: 1, my translation):

- being able to answer questions spontaneously and in an adequate manner
- being able to speak fluently

The students' language skills in the presentation and the colloquium will be graded along the following criteria (Ringel-Eichinger & Selz 2001/2002: 5 ff.):

- well intelligible pronunciation
- fluency
- flexibility in communication
- feedback behavior
- rhetorical strategies
- conversational 'lubricants'

The goals outlined for the new oral exams for the *Abitur* are quite ambitious. Although they are certainly a step in the direction of meeting the aims outlined in the Action Plan of the European Commission, there are several problems which need to be solved. First, time is valuable. Many teachers are already overwhelmed with the teaching load and focus on content rather than pronunciation. The situation in Germany has worsened somewhat since the introduction of G8 – students now complete high school in 8 years rather than 9. With ever less time available, most teachers think that students will be better served knowing literature and cultural studies for their final examinations rather than be good communicators. However, this view is clearly outdated now – students will be required to be able to communicate fluently. It is not clear whether teachers will be able to cope with these completely new requirements. Moreover, students who are now in grade 11 or 12 and will take their *Abitur* within the next few years have not had any specific training and are expected to complete their communication exam successfully without having been prepared for it. This will certainly be frustrating for both students and teachers for a number of years to come.

A second immense problem is that many teachers have no training in teaching pronunciation. When Breitkreutz *et al.* (2001) examined Canadian ESL (English as a Second Language) programs they noted that only about 30% of the teachers had formal training in pronunciation teaching. This number is likely to be similar or even lower in Germany. Moreover, many teachers "lack confidence, skills and knowledge" (MacDonald 2002: 3). This "lack of training [means] that some teachers have serious misgivings about the effectiveness of teaching pronunciation" (Couper 2006: 47). And – as we have seen above – even many researchers do not agree on which levels of pronunciation should be focused on. Quite clearly, the lack of proper instructional material causes insecurity and confusion. It does not help that speakers from different L1 backgrounds will have different needs. A first step might be to consider the following three aspects (Brawn 2010: 117):

- "Understanding how speech sounds are produced and articulated, how English makes systematic contrastive use of sounds, which sequences of sounds occur most frequently together, how sequences of sounds are modified in connected speech, and which patterns of stress, rhythm and intonation occur in English and how do these things differ from the learners [sic!] first language."
- "Predicting and identifying the aspects of pronunciation which are problematic for particular learners, drawing if possible on contrastive awareness of English and the learners' first language."
- "Distinguishing between aspects of pronunciation which are important for learners to acquire in their own speech, and aspects which are perhaps only important for recognition purposes in facilitating listening comprehension."

Unfortunately, the Action Plan of the European Commission is not very explicit as to what should be part of teachers' education. Although it emphasizes the role of the teacher, there is only general information on how teachers need to be trained:

> Language teachers have a crucial role to play in building a multilingual Europe. They, more than teachers of other subjects, are called upon to exemplify the European values of openness to others, tolerance of differences, and willingness to communicate. It is important that they have all had adequate experience of using the target language and understanding its associated culture. All teachers of a foreign language should have spent an extended period in a country where that language is spoken and have regular opportunities to update their training. […] The skills and personal resources required to teach languages well are considerable. Initial training should equip language teachers with a basic 'toolkit' of practical skills and techniques, through training in the classroom; language teachers need the advice of trained mentors as well as regular opportunities to keep their language and teaching skills up to date.
> (Action Plan 2004 – 2006: 10)

It is not quite clear what the "toolkit" the action plan mentions involves, but the ideas outlined above and formulated by Brawn (2010) might be a valid start. It is also positive that the Action Plan mentions that teachers should have spent some time abroad, since this dissertation has shown that a stay abroad as little as 4 months helps speakers acquire subphonetic variation.[82] And evidently, a new curriculum can only be as successful as the

---

[82] The Action Plan outlines the same goal for all students (not just students of foreign languages): "All students should study abroad, preferably in a foreign language, for at least one term, and should gain an accepted language qualification as part of their degree course" (Action Plan 2004 – 2006: 8).

teachers disseminating it. Therefore, if we want students to become fluent communicators, we need to ensure that sufficient time is allocated to this goal and that teachers feel secure in their new role and have the required knowledge in phonetics and phonology. Particularly when we look at the third of Brawn's points – important aspects of pronunciation vs. aspects which are only important for recognition purposes – the lack of consensus as to which units of language are "important" is a major problem and will have to be addressed in order to develop suitable training material. In accordance with the findings of the present study, this dissertation strongly argues that subphonetic units such as vowel length variation should be included. It is important that we make use of the knowledge of how valuable vowel length variation can be to German learners. As Morley puts it,

> it is imperative that students' educational, occupational, and personal/social language needs, *including reasonably intelligible pronunciation*, be served with instruction that will give them communicative empowerment-effective language use that will help them not just to survive, but to succeed.
> (Morley 1991: 489, original emphasis)

Vowel length variation is not currently being taught, and this is a missed opportunity. The usefulness of vowel length variation greatly exceeds the small effort it takes learners to acquire it. A stay abroad of only 4 months suffices for learners to produce and perceive vowel length variation. This finding is not surprising. Flege & Liu (2001) noted that several years spent in a foreign language environment will not lead to a better accent unless the quality of the input is high (similar findings are reported by Rochet (1995), Winitz *et al.* (1995), Bradlow *et al.* (1997), Bradlow *et al.* (1999), Guion & Pederson (2007), Piske (2007) and Sereno & Wang (2007)). A second important point is therefore to ensure the quality of the input we offer to students. At university, language classes are taught by native speakers of English, so this does not pose a problem. However, students at school often do not have teachers who are native speakers. And since research has shown that pronunciation is the first ability to be affected by age, waiting until students enter high school at age 10 or 11 might already be too late in any case. Although this dissertation has demonstrated that even adult learners can acquire vowel length variation with a minimal stay abroad, not all of the participants were successful. Those who were successful were those learners who had managed to establish two separate categories for TRAP and DRESS and had low devoicing. In order to make teaching most effective, it should be done as early as possible when all three features – separate phoneme categories, low levels of devoicing, and accurate production/perception of vowel length variation – can still be easily acquired by learners. Therefore, learners should be confronted with vowel length variation as early as possible.

When is the right time, then? Preschool would be ideal, as the European Commission states:[83]

> It is a priority for Member States to ensure that language learning in kindergarten and primary school is effective, for it is here that key attitudes towards other languages and cultures are formed, and the foundations for later language learning are laid. (Action Plan 2004 – 2006: 7)

Bilingual preschools have become more popular in recent years, due both to the rise of English as an international and global language and the increase of competitiveness on the job market. Many parents feel that their children need to be able to speak English to a native-like standard in order to become successful. According to FMKS (*Verein für Frühe Mehrsprachigkeit an Kitas und Schulen* 'Association for early multilingualism in day nurseries and schools'), there were 680 bilingual preschools in Germany in 2010, 281 of which are German-English bilingual.[84] This number had increased by 25% compared to 2007. Currently, the association is collecting new data and is expecting a similar rise to have occurred within the last three years. This goes to show that there is considerable interest in bilingual preschools. And sensitizing children to vowel length variation during preschool would be ideal for a number of reasons. First, these preschools offer immersion programs in which contact to a native speaker (or someone with native-like competence) of English is possible at least half of the time. In addition, the preschools listed by FMKS work according to the "one person – one language" principle, which means that native German speakers will only speak German to the children, whereas native English speakers will only speak English. In this way, children learn the foreign language not in a decontextualized instructional setting, but in their natural everyday environment. Through this contact with native speakers, children will be confronted with vowel length variation in an implicit manner similar to the way the adult learners of the present study experienced this during their stay abroad. In addition, vowel length variation could be taught in a more explicit manner – for instance through songs or rhymes which specifically focus on this. There are already many phonological awareness songs which concentrate on different vowel sounds, so vowel length variation would only be an extension of this. One could invent new lyrics and sing them to songs the children already know. For instance, there could be a song called "Doug has a duck" sung to the music of "Old MacDonald had a farm". Doug's animals could then include a sick pig, a sad cat, etc.

Rhymes might also be a good way of teaching vowel length variation because they emphasize the fact that words only rhyme when they have equal codas, which lead to equal

---

[83] <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:EN:PDF>.

[84] FMKS 2010: <www.fmks-online.de/_wd_showdoc.php?pic=776>.

vowel length variation. This might be especially helpful in the beginning, because "[c]hildren appear to be better able to capture and gain control over larger units of sound before smaller units of sound" (Yopp & Yopp 2000: 132, cf. also Treiman & Zukowski (1991), Stahl & Murray (1994)). Rhymes could teach children that *had* rhymes with *sad* or *bad*, tokens which all have longer allophonic duration. In contrast, *hat* rhymes with *sat* and *bat*, which have shorter duration. Many of the traditional rhymes already include words with vowel length variation. As an example, consider the lap rhyme *Leg over, leg over* and the finger play *Five Fat Peas*:[85]

Leg over, **leg** over,
As the **dog** went to Dover;
When he came to a stile
Hop! He went over.

Five **fat** peas in a pea **pod pressed**
One grew, two grew, and so **did** all the rest
They grew and grew and they never **stopped**
They grew so **big** that the pea **pod popped**.

Last, preschool teachers could read out stories which draw attention to vowel length variation. One example is the story about Little Bear and Little Tiger which was part of this dissertation. This way of teaching vowel length variation might be especially helpful if kinesthetic cues are incorporated. For instance, the story can be read with different voices for the characters and played out using hand puppets.

Even if L2 learning starts at a young age, however, it will not be successful unless it continues over the span of a few years. This is another problem in Germany. Unfortunately, there are only few bilingual primary schools – 108, according to FMKS (data from 2007/2008).[86] Of these 108 schools, 44 are English-German bilingual. Baden-Württemberg has only 5 bilingual primary schools (of which 2 are English-German bilingual). This shows that Germany has much catching up to do on this level. Many primary schools do not introduce foreign languages at all; some introduce languages in grade 3 (usually on a voluntary basis with one or two hours of instruction per week). Considering that bilingual preschools have become so popular, it is a shame that primary schools lack behind. It would be ideal if children could build on the knowledge they have acquire in preschool institutions in primary school.

The European Commission advertises CLIL – Content and Language Integrated Learning, "in which pupils learn a subject through the medium of a foreign language" (Action

---

[85] Note that the tokens in bold print include vowel + consonant cluster sequences which were not part of the present study but which are also subject to VLV.
[86] FMKS 2008: <http://www.fmks-online.de/_wd_showdoc.php?pic=544>.

Plan 2004 – 2006: 8). CLIL has been shown to be highly successful and has received widespread recognition as a result (cf. Wode (1995; 2009b), Zydatiß (2000)). In Germany, this type of immersion learning has been successfully tested at two schools, namely the Claus Rixen School in Altenholz near Kiel and at the Trilingual International Primary School in Magdeburg (cf. Wode (2002; 2004; 2009b), Kersten (2005), Bongartz (2007)). Results showed that children acquired very good comprehension skills, although active language use only developed during primary school. Children were able to communicate in English on school topics by grade 4 and had no inhibitions to use their L2. When vocabulary was missing, they naturally reformulated or described what they meant. They even read texts age-appropriate for L1 English-speaking children, even though reading skills had not been specifically taught (cf. FMKS 2011: 9).[87] This goes to show how successful immersion programs in preschool and primary school can be.

CLIL is thought to be so successful because it represents a type of learning similar to L1 acquisition. Moreover, since children learn the L2 "on the go", there is no pressure. Rather, children make the experience that foreign languages are useful and that they can learn new concepts in a foreign language. Some more advantages of this approach are listed by the European Commission:

> It can provide effective opportunities for pupils to use their new language skills now, rather than learn them now for use later. It opens doors on languages for a broader range of learners, nurturing self-confidence in young learners and those who have not responded well to formal language instruction in general education. It provides exposure to the language without requiring extra time in the curriculum. (Action Plan 2004 – 2006: 8)

Naturally, in order to guarantee that CLIL works, native speaker teachers (or teachers with native-speaker skills) are required. The European Commission advises the "exchange [of] teachers between Member States; such teachers may work as teachers of their mother tongue, teachers of another language or as teachers of another subject through their mother tongue" (Action Plan 2004 – 2006: 10 f.). Unfortunately, there are still many administrative and legal obstacles preventing the easy putting into practice of this idea.

In any case, SLA at primary school can only be successful if it is natural and if there is sufficient and adequate input. Teachers need to make sure that activities in the classroom are age appropriate. Wesche (2002: 358 f.) identifies the following vital conditions:

---

[87] FMKS 2011: <www.fmks-online.de/_wd_showdoc.php?pic=817>.

- Language contact starts as early as possible.
- Language contact is intense and contextualized through motivating activities.
- Language contact takes place over an extended period (initially 100% immersion with a native speaker teacher, followed by bilingual teaching with at least 50% instruction in the L2).
- L1 has ongoing support.
- Teachers make pedagogical and linguistic adjustments to provide comprehensible input.

In the bilingual 108 primary schools that FMKS lists, at least half of the subjects are taught in the children's L2. This might initially be an ambitious aim for several reasons. Firstly, as mentioned above, schools need to make sure that they have enough competent teachers. This should be possible if we follow the suggestion that the teachers at preschool should be native speakers, "because they need to be able to react fluently, in an idiomatic and age appropriate manner" (Wode 2009a: 32, my translation)[88], but that this is not necessary in primary school (cf. also Fischer 2007: 33). Instead, teachers in primary school should have "communicative competence in the L2 as well as profound knowledge of the subject and knowledge of teaching methodologies. Moreover, students who want to become teachers must be trained in psycholinguistics, especially in the area of language acquisition" (Wode 2009a: 33, my translation).[89] Of course, this means that universities need to make sure that they offer relevant courses and degrees. Thus, providers of all levels of education, from preschool to university, need to work together. In reality, this is not the case yet and will have to change in order for the aim of multilingualism to be an achievable one. First important steps have been taken, however. For instance, many of the *Pädagogische Hochschulen* (PH, 'schools of education') in Baden-Württemberg offer a degree in European teaching, which includes "European oriented social and cultural competences" (PH Freiburg 2013: 1, my translation).[90] Specifically, students are trained in bilingual learning and teaching. In addition, the degree includes a compulsory semester abroad as advised by the European Commission, and students are obliged to teach at least 8 hours bilingually during their 4 week practical course at a school.[91] Unfortunately, degrees from PH are only valid for primary schools (*Grundschule*), *Hauptschule* and (*Werk-*)*Realschule*. Students who want to become teachers at *Gymnasium*

---

[88] <http://cms.awo-sh.de/cms/fileadmin/awo-sh/Dokumente/PDF/Alten1.pdf>.

[89] <http://cms.awo-sh.de/cms/fileadmin/awo-sh/Dokumente/PDF/Alten1.pdf>.

[90] <https://www.ph-freiburg.de/fileadmin/dateien/studium/europalehramt/Fuer_Bewerber_und_Interessierte/PH_Freiburg_Europalehramt_G.pdf>.

[91] <https://www.ph-freiburg.de/fileadmin/dateien/studium/europalehramt/Fuer_Bewerber_und_Interessierte/PH_Freiburg_Europalehramt_G.pdf>.

have to study at university, where no specialized degree in bilingual teaching and learning is available as yet.

A second problem with the introduction of immersion programs at primary schools is that parents might be concerned that their children might be overwhelmed by being expected to grasp the concepts of a subject in a foreign language. Even though more is better,[92] schools might have to start small in order to gain acceptance and receive support from outside. Regarding possible worries expressed by parents, it is important to note that children are not required or even expected to be fully competent in their L2 from the start. Instead, they are encouraged to answer questions and give input in their L1 within the first year (cf. FMKS 2011: 4). Moreover, it is essential to integrate parents into the project. Not only do they need to be informed about what happens in school, but they also need to take time to speak and read German at home so that they support their children's L1. In the case of French immersion at a Canadian kindergarten in St. Lambert, Wesche notes that a large part of the success came from parents' support:

> The program was optional, and learners were thus 'volunteers', whose parents had made an effort to get them into the program and were positive about it and about their learning French. This support and encouragement probably led to children's enhanced motivation for immersion learning. […] Both languages were valued by parents, the immediate community, and the larger society. […] School funding and decision making was under local political control, which meant that parent activism could lead to innovation in local school programs. A well organized and informed group of parents, supported by experts in the field, could convince a local school board to experiment with a new program.
> (Wesche 2002: 359 f., emphasis omitted)

In case of strong hesitation on the parents' side, it might also be possible to begin immersion with subjects which include many pictures, numbers and symbols, such as *Heimat- und Sachkunde* (a subject focusing on local history, knowledge of the area, general science and general knowledge) and mathematics, before expanding the scope to other areas such as biology, geography, history and arts (cf. Burmeister & Daniel (2002), Rymarczyk (2003), Fischer (2007), Ministerium für Kultus, Jugend und Sport Baden-Württemberg (2013)). However, it is doubtful if a program can be fully effective when parents have too many worries. In the end, we must start somewhere if we want to further multilingualism in Germany. Acceptance will come when this method gains ground in society, and parents will see that immersion does not mean fostering an elite (cf. Fischer 2007: 34), but strengthening the confidence and linguistic skills of all children.

---

[92] The children at the Claus Rixen school had 70% of all classes in their L2 (English), while the children at the school in Magdeburg only had 50% of all lessons in English. In language tests, the children from the Claus Rixen school scored better than those from Magdeburg (FMKS 2011: 5).

It is unfortunate that there are still so few bilingual primary schools, and it will certainly take time to change this. This means that for some years to come, children will be confined to starting a second language in grade 3 (with only one or two hours of instruction per week), or even only starting it in secondary school in grade 5. Although this is not ideal, it might not be an immense problem as far as pronunciation is concerned. Research has shown that the childhood language memory remembers segments it was confronted with for years even in the absence of exposure. This means that children from a bilingual preschool might still be able to remember units on the segmental and even subsegmental level when they enter secondary school, even if they did not have any exposure to their L2 in primary school. For instance, Oh *et al.* compared the perception and production of Korean sounds of "childhood speakers who had spoken Korean regularly for a few years during childhood to those of two other groups: 1) childhood hearers who had heard Korean regularly during childhood but had spoken Korean minimally, if at all, and 2) novice learners" (Oh *et al.* 2003: Abstract). At the time of testing, all subjects were first-year college students. The researchers summarize the findings as follows:

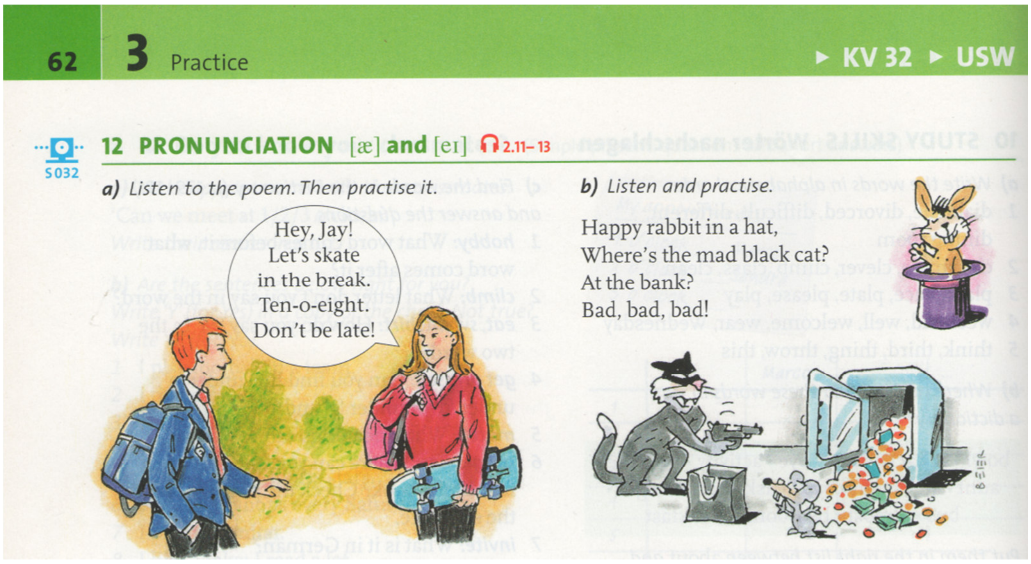
> Childhood speakers were as good as native speakers at hearing the phonemic contrasts of their childhood language, outperforming the novice learners. Their phoneme production was quite native-like and outperformed novice learners and childhood hearers, highlighting the benefits of childhood speaking experience. Furthermore, their accent rating scores suggest that this advantage in phonology goes beyond just individual phonemes. Together, these findings not only underscore the importance of early language experience, but also suggest that the benefits of early language experience. (Oh *et al.* 2003: 11)

These findings suggest that when it comes to pronunciation, an early start is crucial, and that such an early start can make the difference between accentuated speech / imperfect perception and more native-like production and perception. Other studies have supported the results (cf. Au & Romo (1997), Knightley *et al.* (2003), Werker & Tees (2005), Footnick (2007), Au *et al.* (2008), Bowers *et al.* (2009), Oh *et al.* (2010)).[93] In turn, this supports the arguments made in this dissertation, namely that preschool is ideal for children to start learning their L2 (particularly the phonology of it), including vowel length variation. In the absence of foreign language classes in primary schools, vowel length variation could then again be incorporated in pronunciation training at secondary schools.

This leads us to the question of what pronunciation training at secondary school currently looks like, and how it could be optimized. As mentioned before, Baden-Württemberg has changed its curriculum in order to incorporate pronunciation as a major component from grade 5 onwards. In order to meet these new requirements, school books now feature

---

[93] For counterevidence, cf. Pallier *et al.* (2003), Ventureyra *et al.* (2004), Au & Oh (2009).

pronunciation exercises. One of the most widely used books in English classes is *English G 21* by Cornelsen. After each unit, the book contains a green practice section which includes exercises on phonemes and allophones. Unfortunately, subphonemic units, such as vowel length variation, are not yet part of pronunciation training, although this dissertation has shown what valuable cues for German learners they are. However, it would be very easy to incorporate them into the existing exercises. Consider Figure 5-1 below:

62 **3** Practice ▸ **KV 32** ▸ **USW**

S032 **12 PRONUNCIATION** [æ] **and** [eɪ] 2.11–13

a) *Listen to the poem. Then practise it.*

Hey, Jay!
Let's skate
in the break.
Ten-o-eight.
Don't be late!

b) *Listen and practise.*

Happy rabbit in a hat,
Where's the mad black cat?
At the bank?
Bad, bad, bad!

Figure 5-1. Pronunciation exercise focusing on phonemes. Source: Cornelsen English G 21 A1 2006: 62.

This exercise focuses on training the distinction between [æ] and [eɪ]. [æ] is of interest regarding vowel length variation. Rhyme b) is ideal to teach vowel length variation, since it includes a number of monosyllabic words subject to it:

> Happy rabbit in a **hat**,
> Where's the **mad black cat**?
> **At** the bank?
> **Bad**, **bad**, **bad**!

The book could quite easily add an exercise c) which focuses on the length of [æ] in the individual words, so that children train to perceive vowel length variation. This could be supported by listening to the rhyme – the red headphone symbol next to the heading indicates that the rhyme can be listened to on the accompanying CD. It is recorded by a young British girl who pronounces vowel length variation clearly:

Table 5-6. Length of the tokens and their vowel length variation (VLV) in the *Happy rabbit in a hat* rhyme as recorded by the British girl on the CD accompanying the Cornelsen English G 21 A1 course book. Note that the coda consonants of *black* and *bad* do not match; since there are no other tokens available, however, they were paired with each other nonetheless.

Token	Length (ms)	VLV (%)
hat	143.27	43.19
mad	217.92	
cat	130.24	116.23
bad	351.08	
at	110.89	101.24
bad	270.06	
black	114.86	111.41
bad	299.91	

The auditive support from the recording can strengthen the children's impression of vowel length variation. Furthermore, the book could add visual support of vowel length variation in the vocabulary section. Currently, the book works with IPA transcription, but only employs lengthening signs for the long vowels. Incorporating single lengthening signs (e.g. [bæˑd]) in order to signal that a vowel is lengthened before a voiced coda would be an additional visual cue for children. This visual support is crucial for children, as is evidenced by a different pronunciation exercise:[94] In unit two, children are made aware of the existence of the three allomorphs of the plural {-s} – [s], [z] and [ɪz] (cf. Cornelsen English G 21 A1 2006: 42). They are given a number of words, for instance *snakes*, *boxes* and *boys*, and are asked to group them together according to their allomorph. One teacher who tried out this exercise with her students reported that the children were unable to hear a distinction between the three allomorphs, although they produced them distinctly. By the end of the exercise, the children were frustrated and confused because there was no visual cue to indicate which allomorph was the correct one.[95] Since the children automatically produced the correct allomorph due to assimilation to the preceding consonant, however, it is questionable whether one should include such an exercise in the first place. It might be more useful to point children to subphonemic units such as vowel length variation which they do not automatically produce

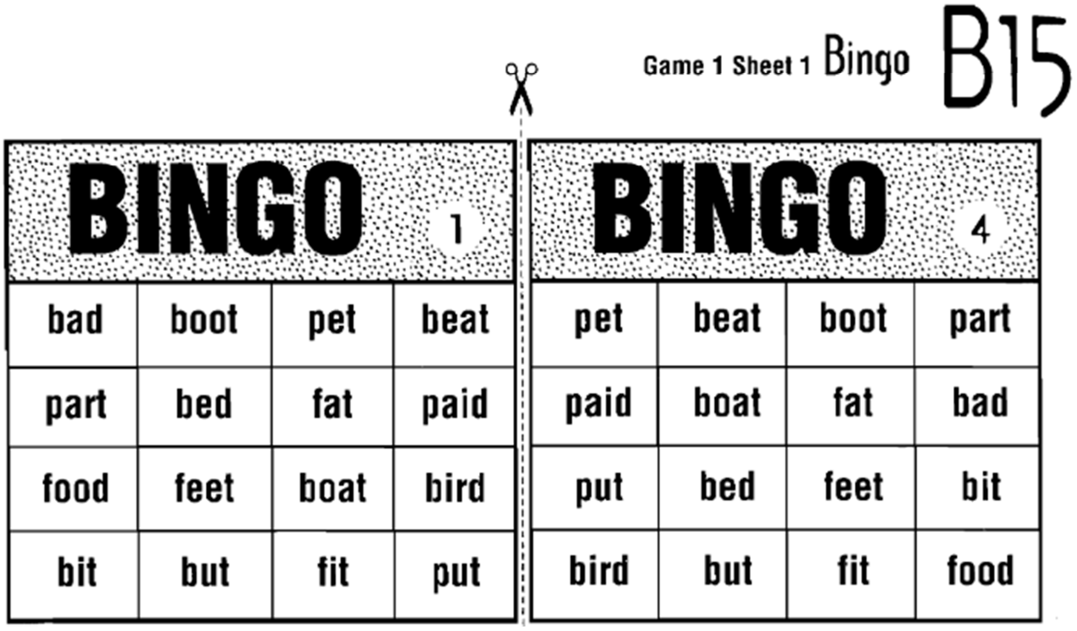
---

[94] Cf. also Gass & Selinker (2008: 335), who argue that selective attention is a major factor in learning. Similar findings can be found in Schmidt (1990; 1993; 2001), Kellogg & Dare (1989). Unfortunately, the findings relate to morphology and syntax; phonology has been mostly ignored (but cf. Couper (2006), Counselman (2010)).

[95] Consider Couper (2006: 52) here, who says: "Explanations can be helpful, but they are not always effective, as it is often very difficult to get learners to focus on the right thing. In other words, teachers must recognise that learners' phonological concepts will be different from those of the native speaker, and that it is therefore critical that teachers make sure their students understand the metalanguage they are using."

correctly, and emphasize this with the addition of a visual cue in the transcription. In the example above, this would be particularly helpful for the acquisition of TRAP, as this dissertation has shown that vowel length variation can help Germans acquire separate spectral categories for this new vowel.

Secondary school also poses the advantage that children are slightly older and can cope with the L2 alphabet, which offers the possibility of new approaches to fostering phonemic awareness. This is important "because it contributes to receptive and productive fluency" (Moyer 1999: 100). Unfortunately, the course book series English G 21 only features pronunciation exercises up to book A3 (grade 8). This is clearly insufficient, since children cannot be expected to have fully mastered pronunciation after only 3 years.[96] Therefore, it is important that teachers continue to design pronunciation games and incorporate them into their schedules. For instance, children could play a version of bingo with words including vowel length variation.[97] Consider Figure 5-2 below:



Game 1 Sheet 1 Bingo B15

BINGO			1
bad	boot	pet	beat
part	bed	fat	paid
food	feet	boat	bird
bit	but	fit	put

BINGO			4
pet	beat	boot	part
paid	boat	fat	bad
put	bed	feet	bit
bird	but	fit	food

Figure 5-2. Example bingo sheet. Source: Hancock 1996: 71.

These two bingo sheets already contain a number of words which include vowel length variation as studied in this dissertation: *bad*, *pet*, *bed*, *fat*, *bit*, *fit*. It would be quite simple to change the other words and design the bingo sheet in a way so that each vowel + consonant combination appears once. In addition, this game is well suited for children to train not only

---

[96] Some teachers have reported that pronunciation training becomes very difficult at around grade 8, because students enter puberty and do not want to communicate in front of their classmates anymore due to fear of being bullied or mocked.

[97] More games, for instance jigsaws, mazes, board and card games, guessing activities, etc., can be found in Hancock (1996).

their perception, but also production. Child A could read a word from a list and child B crosses off the word he or she has heard. If the final consonant is devoiced or the vowel is not lengthened before a voiced coda, child B can give feedback that the word in question was not clearly pronounced. Children need input that they have not been understood in order to learn that intelligible speech is key for communication.[98] Another useful feedback exercise is dictation. Children dictate each other sentences which do not make sense when the words are not pronounced carefully. Walker (2010: 89) lists a number of reasons why dictation improves pronunciation:

- "[I]t leaves a written record of any 'breakdowns' in communication. Differences in the learners' respective texts will immediately reveal where they failed to understand each other."
- "[T]he written records of the activity allow learners to accurately identify the pronunciation items at fault. This will raise their awareness of problem areas for them at an individual level."
- "Dictation also allows listeners time to think about the context and make adjustments to what they thought they heard on the basis of what seems reasonable […]. In this respect, it is a valuable tool for encouraging listeners to accommodate receptively to the speaker."

One slight drawback of dictation might be that it should only be used when learners have mastered orthography on an intermediate level. Otherwise, it is quite possible that the focus switches to mistakes in orthography, because this is what students are used to being corrected on. Therefore, the teacher needs to ensure that students understand that their partners misspelled a word because they misperceived it, not because they *do not know* how to spell the intended word.

A similar exercise to dictation which exemplifies how to use minimal pairs with contrasting vowel length is discussed by Walker (2010: 83). She suggests that child A asks a question or reads a statement and child B responds with the appropriate option depending on which word was heard:

---

[98] Cf. Moyer (1999: 92), who found that the presence of feedback lead to a closer to native rating in a pronunciation task. Cf. also Couper (2006: 52).

1. Can you get him a cap / cab ?	a) I'll ring for one now.
	b) He can have mine.
2. We'll have to wade / wait.	a) Yes, the water's rising fast.
	b) That's OK. I'm not in a hurry.

This exercise again requires children to give each other feedback on their pronunciation. It is conceivable that children will more easily grasp the importance of vowel length variation when they experience that it can cause communication breakdowns. At the same time, students can work on repairing communication problems together by trying to pronounce the intended word more carefully a second time. This would then also train their self-confidence as second language speakers. After all, "[o]ne of the keys to successful communication between competent language users is their ability to work together to construct understanding" (Walker 2010: 89).

In conclusion, we have seen that there is an abundance of exercises and games teachers can rely on to train the perception and production of vowel length variation. Some may need slight adjustment, but it is important that teachers continue to offer students pronunciation training, particularly because regular course books do not include sufficient practice material. Although efforts are being made to revise curricula, we are still miles away from meeting the aim of multilingualism in Europe. It is crucial that we promote bilingual preschools and primary schools in order to exposure children to an L2 as early as possible. An early start is not only important in the light of the CP discussion, but also because it is the perfect age for children to be open about the L2 culture. Regarding new aims for instruction, it has been demonstrated here that vowel length variation is a useful cue for German learners. Therefore, this dissertation strongly argues that it should be incorporated into pronunciation instruction. We can use a passage from Derwing & Munro (2005) summarize the challenges that lie ahead:[99]

> One of the most important challenges in the coming years is an emphasis on greater collaboration between researchers and practitioners to encourage more classroom-relevant research. […] Ideally, teacher preparation programs should provide ESL teachers with sufficient background to enable them to assess their students' pronunciation problems and to critically evaluate research findings, materials, and techniques to determine their applicability for their students. At the same time, researchers need to understand classroom dynamics and students so that they can work in concert with teachers to ensure appropriate research methodology and meaningful findings. In the meantime, applied linguists with an interest in pronunciation should ensure that ESL teacher preparation programs offer courses in pronunciation pedagogy firmly rooted in existing research. Researchers and teachers owe this to ESL students, many of whom view pronunciation instruction as a priority. (Derwing & Munro 2005: 392)

---

[99] ESL = *English as a Second Language*.

We also owe it to students to find out more about what the relevant features for a second language learner are, and how to teach these features effectively. This dissertation has broken new ground in showing that subphonemic features are highly valuable for German learners, and further studies should build on this. We are clearly beyond the point where we believe that focusing on the segmental level alone is sufficient for L2 learners to become proficient speakers. However, there are still many questions to be answered. Since this dissertation is the first to focus on vowel length variation and how its perception and production is acquired by L2 speakers, it investigated a number of variables simultaneously. It is now time to examine each factor separately in order to gain a deeper understanding of how it operates. This would be particularly interesting with regard to whether learners can more easily match native speakers in similar vowels or new ones. Due to the largely opposite performance of the British and the American native speakers in this study, this dissertation was only able to cautiously argue that German learners find it easier to produce vowel length variation in similar vowels. For the same reason, it was not possible to find clear support for one of the four perception models which were discussed. It would be of great value to examine learners of a language which has similar low vowels to English, but different high vowels, in order to gauge the effect that intrinsic vowel length variation has on perception in contrast to the way the two vowel systems relate regarding the “similar” / ”new” distinction. This would further our understanding of the interaction between languages in the phonological space of learners.

Furthermore, although age was not explored as a variable, results indicated that this, too, had an effect on language learning. Therefore, subsequent studies should take this factor into account. More longitudinal studies would be helpful in order to receive a clearer picture of how the perception and production of vowel length variation emerges. This would shed some light on why some speakers who perceive vowel length variation do not produce it. Since the perception experiment was somewhat decontextualized, research would certainly profit from experiments focusing on real-life utterances. Perception training should also be evaluated.

Since this dissertation has shown that vowel length variation is especially helpful for learners who merge TRAP and DRESS and display devoicing, this leads us to a very important question which should certainly be the focus of further research: How strongly are native speakers of English affected by a foreign accent which shows a combination of inaccurate vowel length variation and devoicing? And, out of the two cues, which one is more helpful in signaling meaning?

In the end, achieving a native-like accent is a difficult endeavor. Especially when the feature in question is subphonemic in the target language, i.e. below the consciousness of even native speakers, it may seem impossible to master. This dissertation has shown that it is not – and I do sincerely hope that it inspires many more studies on the acquisition of subphonemic units.

6 References

Aaltonen, O., O. Eerola, A. Hellström, E. Uusipaikka & A. H. Lang. 1997. "Perceptual Magnet Effect in the Light of Behavioral and Psychophysiological Data." *J. Acoust. Soc. Am.* 101 (2): 1090–1105.

Abercrombie, D. 1963. *Problems and Principles in Language Study*. London: Longman.

Aitken, A. J. 1981. "The Scottish Vowel Length Rule." In: M. Benskin & M. L. Samuels (eds.). *So Meny People Longages and Tonges: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*. Edinburgh: Edinburgh University Press. 131–157.

Aitken, A. J. 1984. "Scottish Accents and Dialects." In: P. Trudgill (ed.). *Language in the British Isles*. Cambridge: CUP. 94–114.

Alonso-Nanclares, L., J. Gonzalez-Soriano, J. R. Rodriguez & J. DeFelipe. 2008. "Gender Differences in Human Cortical Synaptic Density." *Proc. Natl. Acad. Sci. USA* 105: 14615–14619.

Altenberg, E. & R. Vago. 2006. "Theoretical Implications of an Error Analysis of Second Language Phonology Production." *Language Learning* 33 (4): 427–447.

Anderson, S. 1976. "Nasal Consonants and the Internal Structure of Segments." *Language* 52: 326–344.

Anderson-Hsieh, J. & K. Koehler. 1988. "The Effect of Foreign Accent and Speaking Rate on Native Speaker Comprehension." *Language Learning* 38: 561–593.

Anderson-Hsieh, J., R. Johnson & K. Koehler. 1992. "The Relationship between Native Speaker Judgements of Nonnative Pronunciation and Deviance in Segmentals, Prosody and Syllable Structure." *Language Learning* 42: 529–555.

Asher, J. J. & R. Garcia. 1969. "The Optimal Age to Learn a Foreign Language." *Modern Language Journal* 53: 334–341.

Atlas zur Deutschen Alltagssprache. 2011. *Aussprache Tag und Zeug*. Available online at <http://www.philhist.uni-augsburg.de/lehrstuehle/germanistik/sprachwissenschaft/ada/runde_1/f15a-b/> [Last accessed January 04, 2014]

Atlas zur Deutschen Alltagssprache. 2011. *Länge des betonten Vokals*. Available online at <http://www.philhist.uni-augsburg.de/lehrstuehle/germanistik/sprachwissenschaft/ada/runde_2/f22a-c/> [Last accessed January 04, 2014]

Au, T. K. & J. S. Oh. 2009. "Korean as a Heritage Language." In: P. Li, C. Lee, G. B. Simpson & Y. Kim (eds.). *Handbook of East Asian Psycholinguistics, Part III: Korean Psycholinguistics*. Cambridge: CUP. 268–275.

Au, T. K. & L. F. Romo. 1997. "Does Childhood Language Experience Help Adult Learners?" In: H. C. Chen (ed.). *The Cognitive Processing of Chinese and Related Asian Languages*. Hong Kong: Chinese University Press. 417–441.

Au, T. K., J. S. Oh, L. M. Knightley, S. A. Jun & L. F. Romo. 2008. “Salvaging a Childhood Language.” *Journal of Memory and Language* 58: 998–1011.

Auer, P., P. Gilles & H. Spiekermann. 2002. “Introduction: Syllable Cut and Tonal Accents. Two ‘Exceptional Prosodies’ of Germanic and Some Thoughts on their Mutual Relationship.” In: P. Auer, P. Gilles & H. Spiekermann. *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer. 1–10.

Avery, P. & S. Ehrlich. 1992. *Teaching American English Pronunciation*. Oxford: OUP.

Aylett, M. & A. Turk. 2004. “The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech.” *Language and Speech* 47 (1): 31–56.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: CUP.

Baker, W., P. Trofimovich, M. Mack & J. E. Flege. 2002. “The Effect of Perceived Phonetic Similarity on Non-Native Sound Learning by Children and Adults.” In: B. Skarabela, S. Fish & A. Do (eds.). *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville: Cascadilla Press. 36–47.

Barcelona European Council. 2002. “Presidency Conclusions – Barcelona, 15 and 16 March 2002”. Available online at <http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/71025.pdf> [Last accessed January 08, 2014]

Bard, E. & M. Aylett. 1999. “The Dissociation of Deaccenting, Giveness and Syntactic Role in Spontaneous Speech.” *ICPhs 99*: 1753–1756.

Barry, W. 1979. “Complex Encoding in Word-Final Voiced and Voiceless Stops.” *Phonetica* 36: 361–372.

Becker, T. 1996. “Zur Repräsentation der Vokallänge in der deutschen Standardsprache.” *Zeitschrift für Sprachwissenschaft* 15 (1): 3–21.

Becker, T. 2002. “Silbenschnitt und Silbenstruktur in der deutschen Standardsprache der Gegenwart.” In: P. Auer, P. Gilles & H. Spiekermann (eds.) *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer. 87–101.

Beckmann, M. E. & J. Edwards. 1990. “Lengthening and Shortenings and the Nature of Prosodic Constituency.” In: J. Kingston & M. E. Beckmann (eds.). *Papers in Laboratory Phonology I*. Cambridge: CUP. 152–178.

Belasco, S. 1953. “The Influence of Articulation of Consonants on Vowel Duration.” *J. Acoust. Soc. Am.* 25: 1015–1016.

Bell, A., J. M. Brenier, M. Gregory, C. Girarnd & D. Jurafsky. 2009. “Predictability Effects on Durations of Content and Function Words in Conversational English.” *Journal of Memory and Language* 60 (1): 92–111.

Benders, T. & P. Escudero. (in preparation). “Vowel Duration in the Perception of Monolingual Peruvian Spanish Listeners.”

Bent, T., A. R. Bradlow & B. L. Smith. 2007. “Segmental Errors in Different Word Positions and their Effects on Intelligibility of Non-Native Speech.” In: O.-S. Bohn & M. J. Munro (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: Benjamins. 331–347.

Best, C. 1994. “The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model.” In: J. C. Goodman & H. C. Nusbaum (eds.). *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge: MIT Press. 167–224.

Best, C. 1995. “A Direct Realist View of Cross-Language Speech Perception.” In: W. Strange (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press. 171–204.

Best, C. & W. Strange. 1992. “Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants.” *Journal of Phonetics* 20: 305–331.

Best, C. & M. Tyler. 2007. “Nonnative and Second-Language Speech Perception. Commonalities and Complementarities.” In: O.-S. Bohn & M. J. Munro (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: Benjamins. 13–34.

Bialystok, E. & K. Hakuta. 1999. “Confounded Age: Linguistic and Cognitive Factors in Age Differences for Second Language Acquisition.” In: D. Birdsong (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah: Lawrence Erlbaum. 161–181.

Bloomfield, L. 1933. *Language*. New York: Henry Holt.

Boersma, P. 1998. *Functional Phonology*. Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P. & B. Hayes. 2001. “Empirical Tests of the Gradual Learning Algorithm.” *Linguistic Inquiry* 32: 45–86.

Boersma, P. & D. Weenink. 2012. Praat: Doing Phonetics by Computer. [Computer Program]. Version 5.1.32. <www.praat.org>.

Boersma, P., P. Escudero & R. Hayes. 2003. “Learning Abstract Phonological from Auditory Phonetic Categories: An Integrated Model for the Acquisition of Language-Specific Sound Categories.” In: M. J. Sole, D. Recansens & J. Romero (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions. 1013–1016.

Bohn, O.-S. 1995. “Cross-Language Perception in Adults: First Language Transfer Doesn’t Tell it All.” In: W. Strange (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press. 379–410.

Bohn, O.-S. & J. E. Flege. 1990. “Interlingual Identification and the Role of Foreign Experience in L2 Vowel Perception.” *Applied Psycholinguistics* 11: 303–328.

Bohn, O.-S. & J. E. Flege. 1992. “The Production of New and Similar Vowels by Adult German Learners of English.” *Studies in Second Language Acquisition* 14: 131–158.

Bongaerts, T. 1999. “Ultimate Attainment in L2 Pronunciation: The Case of Very Advanced Late L2 Learners.” In: D. Birdsong (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. London: Erlbaum. 133–159.

Bongaerts, T., B. Planken & E. Schils. 1995. “Can Late Starters Attain a Native Accent in a Foreign Language? A Test of the Critical Period Hypothesis.” In: D. Singleton & Z. Lengyel (eds.). *The Age Factor in Second Language Acquisition.* Clevedon: Multilingual Matters. 30–50.

Bongaerts, T., C. Summeren, B. Planken & E. Schils. 1997. “Age and Ultimate Attainment in the Pronunciation of a Foreign Language.” *Studies in Second Language Acquisition* 19: 447–465.

Bongartz, C. M. 2007. “Studying Classroom Language: Curricular Planning and Proficiency Assessment in Young Foreign Language Learners.” In: S. Volk-Birke & J. Lippert (ed.). *Proceedings of the Anglistentag 2006 in Halle*. Trier: WVT. 387–396.

Bosch, L., A. Costa & N. Sebastián-Gallés. 2000. “First and Second Language Vowel Perception in Early Bilinguals.” *European Journal of Cognitive Psychology* 12: 189–221.

Bowers, J. S., S. L. Mattys & S. H. Gage. 2009. “Preserved Implicit Knowlegde of a Forgotten Childhood Language.” *Psychological Science* 20: 1064–1969.

Boyd, S. 2003. “Foreign-Born Teachers in the Multilingual Classroom in Sweden: The Role of Attitudes to Foreign Accent.” *International Journal of Bilingual Education and Bilingualism* 6: 283–295.

Boyle, J. 1987. “Perspectives on Stress and Intonation in Language Learning.” *System 15*: 189–195.

Bradac, J. J. 1990. “Language Attitudes and Impression Formation.” In: H. Giles & W. P. Robinson (eds.). *Handbook of Language and Social Psychology*. London: John Wiley. 387–412.

Bradlow, A. R. & D. B. Pisoni. 1999. “Recognition of Spoken Words by Native and Non-Native Listeners: Talker-, Listener- and Item-Related Factors.” *J. Acoust. Soc. Am.* 106 (4): 2074–2085.

Bradlow, A. R., D. B. Pisoni, R. Akahane-Yamada & Y. Tohkura. 1997. “Training Japanese Listeners to Identify English /r/ and /l/: Some Effects of Perceptual Learning on Speech Production.” *J. Acoust. Soc. Am.* 101 (4): 2299–2310.

Bradlow, A. R., R. Akahane-Yamada, D. B. Pisoni, Y. Tohkura. 1999. “Training Japanese Listeners to Identify English /r/and /l/: Long-Term Retention of Learning in Perception and Production.” *Perception & Psychophysics* 61 (5): 977–985.

Braunschweiler, N. 1997. “Integrated Cues of Voicing and Vowel Length in German: A Production Study.” *Language and Speech* 40: 353–376.

Brawn, J. R. 2010. “Teaching Pronunciation Gets a Bad R.A.P.: A Framework for Teaching Pronunciation.” Hankuk University of Foreign Studies. Available online at <http://www.tesolreview.org/down/2-6.%20James%20R.%20Brawn.pdf> [Last accessed January 03, 2014]

Breitkreutz, J., T. M. Derwing & M. J. Rossiter. 2001. “Pronunciation Teaching Practices in Canada.” *TESL Canada Journal* 19: 51–61.

Bresnahan, M. J., R. Ohashi, R. Nebashi, W. Y. Liu & S. M. Shearman. 2002. “Attitudinal and Affective Response Toward Accented English.” *Language and Communication* 22: 171–185.

Browman, C. P. & L. Goldstein. 1989. “Articulatory Gestures as Phonological Units.” *Phonology* 6: 201–251.

Brown, R. & D. McNeill. 1966. “The ‘Tip of the Tongue’ Phenomenon.” *Journal of Verbal Learning and Verbal Behavior* 5: 325–337.

Buder, E. & C. Stoel-Gammon. 2002. “American and Swedish Children’s Acquisition of Vowel Duration: Effects of Vowel Identity and Final Stop Voicing.” *J. Acoust. Soc. Am.* 111: 1854–1864.

Burmeister, P. & A. Daniel. 2002. “How Effective Is Late Partial Immersion? Some Findings of a Secondary School Program in Germany.” In: P. Burmeister, T. Piske & A. Rohde (eds.). *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag Trier. 499–515.

Burnham D. K. 1986. “Developmental Loss of Speech Perception: Exposure to and Experience with a First Language.” *Applied Psycholinguistics* 7: 207–240.

Bush, N. 2001. “Frequency Effects and Word-Boundary Palatalization in English.” In: J. Bybee & P. Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins. 255–280.

Bybee, J. & P. Hopper. 2001. “Introduction to Frequency and Emergence of Linguistic Structure.” In: J. Bybee & P. Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins. 1–24.

Campbell, W. N. & S. D. Isard. 1991. “Segment Durations in a Syllable Frame.” *Journal of Phonetics* 19: 37–47.

Canepàri, L. 2010. *The Pronunciation of English around the World: Geo-Social Applications of the Natural Phonetics & Tonetics Method*. Munich: Lincom Europa.

Catford, J. C. 1965. *A Linguistic Theory of Translation*. Oxford: OUP.

Catford, J. C. & D. B. Pisoni. 1970. “Auditory vs. Articulatory Training in Exotic Sounds.” *Modern Language Journal* 54: 477–481.

Cebrian, J. 2006. “Experience and the Use of Non-Native Duration in L2 Vowel Categorization.” *Journal of Phonetics* 34: 372–387.

Celce-Murcia, M., D. M. Brinton & J. M. Goodwin. 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: CUP.

Chen, M. 1970. “Vowel Length as a Function of the Voicing of the Consonant Environment.” *Phonetica* 22: 129–159.

Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Chun, D. 1988. “The Neglected Role of Intonation in Communicative Competence and Proficiency.” *Modern Language Journal* 72: 295–303.

Cleveland, W. S. 1979. “Robust Locally Weighted Regression and Smoothing Scatterplots.” *J. Amer. Statist. Assoc*. 74: 828–830.

Cohen, J. 1992. “A Power Primer.” *Psychological Bulletin* 112 (1): 155–159.

Commission of the European Communities. 2003. “Promoting Language Learning and Linguistic Diversity: An Action Plan 2004 – 2006.” Available online at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:EN:PDF> [Last accessed January 08, 20414]

Conboy, B. T. & D. L. Mills. 2006. “Two Languages, One Developing Brain: Event-Related Potentials to Words in Bilingual Toddlers.” *Developmental Science* 9: F1–F2.

Cook, V. J. 1994. “The Metaphor of Access to Universal Grammar in L2 Learning.” In: N. Ellis (ed.). *Implicit and Explicit Learning of Languages*. London: Academic Press. 477–502.

Cook, V. J. 2002. “Background to the L2 User.” In: V. J. Cook (ed.). *Portraits of the L2 User*. Clevedon: Multilingual Matters. 1–28.

Cornelsen / Context 21 / Hessen. 2012. “Kommunikationsprüfung in Hessen – Abitur Englisch.” Available online at <http://www.cornelsen.de/shop/capiadapter/download/get/file/9783069604709_x1PS_KommPruefung_HE_C21.pdf?fileNr=1&sku=3-464-00076588-0&callerId=SBK3> [Last accessed January 08, 2014]

Cornelsen / H. Schwarz (ed.). 2006. *English G 21 A1*. Berlin: Cornelsen.

Council of the European Union. 2008. “Council Resolution of 21 November 2008 on a European Strategy for Multilingualism.” Available online at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:320:0001:01:en:HTML> [Last accessed January 08, 2014]

Counselman, D. 2010. *Improving Pronunciation Instruction in the Second Language Classroom*. Doctoral Dissertation, Pennsylvania State University.

Couper, G. 2003. “The Value of an Explicit Pronunciation Syllabus in ESOL Teaching.” *Prospect* 18 (3): 53–70.

Couper, G. 2006. “The Short and Long-Term Effects of Pronunciation Instruction.” *Prospect* 21: 46–66.

Croker, S., J. M. Pine & F. Gobet. 2000. “Modelling the Optional Infinitive Phenomena: A Computational Account of Tense Optionality in Children’s Speech.” *Proceedings of the 3rd International Conference on Cognitive Modelling*. Veenendaal: Universal Press. 78–85.

Crowther, C. S. & V. Mann. 1992. “Native Language Factors Affecting Use of Vocalic Cues to Final Consonant Voicing in English.” *J. Acoust. Soc. Am.* 92 (2): 711–722.

Crystal, T. H. &. A. S. House. 1982. “Segmental Durations in Connected Speech Signals: Preliminary Results.” *J. Acoust. Soc. Am.* 72: 705–716.

Crystal, T. H. & A. S. House. 1988a. “Segmental Durations in Connected-Speech Signals: Current Results.” *J. Acoust. Soc. Am.* 83 (4): 1553–1573.

Crystal, T. H. & A. S. House. 1988b. “The Duration of American English Vowels: An Overview.” *Journal of Phonetics* 16: 263–284.

Crystal, T. H. & A. S. House. 1988c. “Segmental Durations in Connected-Speech Signals: Syllabic Stress.” *J. Acoust. Soc. Am.* 83 (4): 1574–1585.

Cummins, J. 1979. “Cognitive/Academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some Other Matters.” *Working Papers on Bilingualism* 19: 121–129.

Cummins, J. 1991. “Conversational and Academic Language Proficiency in Bilingual Contexts.” *AILA Review* 8: 75–89.

Denes, P. 1955. “Effect of Duration on the Perception of Voicing.” *J. Acoust. Soc. Am.* 27: 761–764.

Dehaene-Lambertz, G. & T. Gliga. 2004. “Common Neural Basis for Phoneme Processing in Infants and Adults.” *Journal of Cognitive Neuroscience* 16: 1375–1387.

DeKeyser, R. 2000. “The Robustness of Critical Period Effects in Second Language Acquisiton.” *Studies in Second Language Acquisition* 22: 449–553.

De Lacy, P. 1998. "The Effect of Consonant Clusters on Vowel Duration in English." 1–20. Available online at <http://www.pauldelacy.net/webpage/docs/delacy-1998-effect% 20of% 20consonant%20clusters%20on%20vowel%20duration.pdf> [Last accessed May 23, 2013]

Derwing, T. M. 2003. "What do ESL Students Say about their Accents?" *Canadian Modern Language Review* 59: 547–566.

Derwing, T. M. & M. J. Munro. 2005. "Second Language Accent and Pronunciation Teaching: A Research-Based Approach." *TESOL Quarterly* 39 (3): 379–397.

Derwing, T. M. & M. J. Munro. 2009. "Putting Accent in its Place: Rethinking Obstacles to Communication." *Language Teaching* 42: 476–490.

Derwing, T. M. & M. J. Rossiter. 2002. "ESL Learners' Perceptions of their Pronunciation Needs and Strategies." *System* 30: 155–166.

Derwing, T. M., M. J. Munro & G. Wiebe. 1997. "Pronunciation Instruction for 'Fossilized' Learners: Can it Help?" *Applied Language Learning* 8: 217–235.

Derwing, T. M., M. J. Munro & G. Wiebe. 1998. "Evidence in Favour of a Broad Framework for Pronunciation Instruction." *Language Learning* 48: 393–410.

Dickerson, L. 1975. "The Learner's Interlanguage as a System of Variable Rules." *TESOL Quarterly* 9: 401–407.

Diehl, R. L. & M. A. Walsh. 1986. "An Auditory Basis for Rate Normalization in Stops and Glides." *J. Acoust. Soc. Am.* 80: 125. Abstract.

Dietrich, C., D. Swingley & J. F. Werker. 2007. "Native Language Governs Interpretation of Salient Speech Sound Differences at 18 Months." *Proceedings of the National Academy of Sciences of the United States of America* 104 (41): 16027–16031.

DiSimoni, F. G. 1974. "Influence of Consonant Environment on Duration of Vowels in the Speech of Three-, Six-, and Nine-Year-Old Children." *J. Acoust. Soc. Am.* 55 (2): 362–363.

Dörnyei, Z. & P. Skehan. 2003. "Individual Differences in L2 Learning." In: C. Doughty & M. Long (eds.). *Handbook of Second Language Acquisition*. Malden: Blackwell. 589–630.

Ebbinghaus, H. 1885. *Memory: A Contribution to Experimental Psychology*. New York: Teachers College. Available online at <http://psychclassics.yorku.ca/Ebbinghaus/> [Last accessed December 03, 2013]

Eckert, H. & W. Barry. 2002. *The Phonetics and Phonology of English Pronunciation*. Trier: WVT.

Edwards, J. 1982. "Language Attitudes and their Implications among English Speakers." In: E. B. Ryan & H. Giles (eds.). *Attitudes Toward Language Variation: Social and Applied Contexts*. London: Edward Arnold. 20–33.

Eefting, W. 1988. "Temporal Variation in Natural Speech: Some Explorations." *Proc. Speech 1988: 7th FASE Symposium*. 503–507.

Eilers, R. E. 1977. "Context Sensitive Perception by Infants." *J. Acoust. Soc. Am.* 61: 1321–1336.

Eilers, R. E., D. H. Bull, K. Oller, & D. L. Lewis. 1984. "The Discrimination of Vowel Duration by Infants." *J. Acoust. Soc. Am* 75: 1213–1218.

Eimas, P. D. 1975. "Auditory and Phonetic Coding of the Cues for Speech: Discrimination of the [r-l] Distinction by Young Infants." *Perception & Psychophysics* 18 (5): 341–347.

Elert, C.-C. 1964. *Phonological Studies of Quantity in Swedish*. Stockholm: Almquist & Wiksell.

Elliott, R. E. 1995. "Field Independence/Dependence, Hemispheric Specialization, and Attitude in Relation to Pronunciation Accuracy in Spanish as a Foreign Language." *The Modern Language Journal* 79: 356–371.

Ellis, N. C. & A. Beaton. 1993. "Psycholinguists Determinants of Foreign Language Vocabulary Learning." *Language Learning* 43: 559–617.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: OUP.

Entwisle, D. R. 1966. *Word Associations of Young Children*. Baltimore: Johns Hopkins Press.

Escudero, P. 2000. *Developmental Patterns in the Adult L2 Acquisition of New Contrasts: The Acoustic Cue Weighting in the Perception of Scottish Tense/Lax Vowels in Spanish Speakers*. Unpublished M.Sc. Thesis, University of Edinburgh. Available online at <http://uvafon.hum.uva.nl/paul/p2/papers/escuderoMScThesis.pdf> [Last accessed August 15, 2013]

Escudero, P. 2001. "The Role of the Input in the Development of L1 and L2 Sound Contrasts: Language-Specific Cue Weighting for Vowels." In: A. H-J. Do, L. Domínguez & A. Johansen (eds.). *Proceedings of the 25th Annual Boston University Conference on Language Development Vol. 1*. Somerville: Cascadilla Press. 250–261.

Escudero, P. 2002. "The Perception of English Vowel Contrasts: Acoustic Cue Reliance in the Development of New Contrasts." In: J. Leather & A. James (eds.). *New Sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt. 122–131.

Escudero, P. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Doctoral Dissertation, University of Utrecht. Available online at <http://igitur-archive.library.uu.nl/dissertations/2005-1109-200037/index.htm> [Last accessed August 15, 2013]

Escudero, P. 2006. “The Phonological and Phonetic Development of New Vowel Contrasts in Spanish Learners of English.” In: B. O. Baptista & M. A. Watkins (eds.). *English With a Latin Beat: Studies in Portugues/Spanish-English Interphonology*. Amsterdam: Benjamins. 149–161.

Escudero, P. 2009. “The Linguistic Perception of ‘Similar’ L2 Sounds.” In: P. Boersma & S. Hamann (eds.). *Phonology in Perception*. Berlin: Mouton the Gruyter. 151–190.

Escudero, P. & P. Boersma. 2003. “Modelling the Perceptual Development of Phonological Contrasts with Optimality Theory and the Gradual Learning Algorithm.” In S. Arunachalam, E. Kaiser & A. Williams (eds.). *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8: 71–85.

Escudero, P. & P. Boersma. 2004. “Bridging the Gap between L2 Speech Perception Research and Phonological Theory.” *Stud. Second Lang. Acquis.* 26: 551–585.

Escudero, P. & L. Polka. 2003. “A Cross-Language Study of Vowel Categorization and Vowel Acoustics.” *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions. 861–864.

Escudero, P., T. Benders & S. C. Lipski. 2009. “Native, Non-Native and L2 Perceptual Cue Weighting for Dutch Vowels: The Case of Dutch, German, and Spanish Listeners.” *Journal of Phonetics* 37: 452–465.

Feigenbaum, E. A. & H. A. Simon. 1962. “A Theory of the Serial Position Effect.” *British Journal of Psychology* 53: 307–320.

Feigenbaum, E. A. & H. A. Simon. 1984. “EPAM-Like Models of Recognition and Learning.” *Cognitive Science* 8: 305–336.

Fidelholtz, J. 1975. “Word Frequency and Vowel Reduction in English.” *CLS* 75: 200–213.

Fintoft, K. 1961. “The Duration of Some Norwegian Speech Sounds.” *Phonetica* 7: 19–39.

Fischer, R. M. & R. N. Ohde. 1990. “Spectral and Duration Properties of Front Vowels as Cues to Final Stop-Consonant Voicing.” *J. Acoust. Soc. Am.* 88 (3): 1250–1259.

Fischer, U. 2007. “Von Herausforderungen und Chancen.” *Grundschule* 4: 32–34.

Fischer-Jørgensen, E. 1964. “Sound Duration and Place of Articulation.” *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 17: 175–207.

Flege, J. E. 1987. “A Critical Period for Learning to Pronounce Foreign Languages?” *Applied Linguistics* 8: 162–177.

Flege, J. E. 1988. “Factors Affecting Degree of Perceived Foreign Accent in English Sentences.” *J. Acoust. Soc. Am.* 84: 70–79.

Flege, J. E. 1991. “Perception and Production: The Relevance of Phonetic Input to L2 Phonological Learning.” In T. Heubner & C. Ferguson (eds.). *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Amsterdam: Benjamins. 249–290.

Flege, J. E. 1992. “Speech Learning in a Second Language.” In: C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.). *Phonological Development: Models, Research, Implications*. Timonium: York Press. 565–604.

Flege, J. E. 1995. “Second Language Speech Learning: Theory, Findings and Problems.” In: W. Strange (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press: 233–277.

Flege, J. E. 1998. Second-Language Learning: The Role of Subject and Phonetic Variables.” *STiLL-1998*: 1–8.

Flege, J. E. & O.-S. Bohn. 1989. “The Perception of English Vowels by Native Spanish Speakers.” *J. Acoust. Soc. Am.* 85 (Suppl. 1): 85 A.

Flege, J. E. & K. L. Fletcher. 1992. “Talker and Listener Effects on Degree of Perceived Foreign Accent.” *J. Acoust. Soc. Am.* 91: 370–389.

Flege, J. E. & J. Hillenbrand. 1984. “Limits on Phonetic Accuracy in Foreign Language Speech Production.” *J. Acoust. Soc. Am.* 76 (3): 708–721.

Flege, J. E. & S. Liu. 2001. “The Effect of Experience on Adults’ Acquisition of a Second Language.” *Stud. Second Lang. Acquis.* 23: 527–552.

Flege, J. E. & I. R. A. MacKay. 2004. “Perceiving Vowels in a Second Language.” *Stud. Second Lang. Acquis.* 26: 1–34.

Flege, J. E., O.-S. Bohn & S. Jang. 1997. “Effects of Experience on Non-Native Speakers’ Production and Perception of English Vowels.” *Journal of Phonetics* 25: 437–470.

Flege, J. E., M. Munro & I. R. A. MacKay. 1995. “The Effect of Age of Second Language Learning on the Production of English Vowels.” *Speech Communication* 16: 1–26.

Flege, J. E., M. J. Munro & L. Skelton. 1992. “Production of the Word-Final /t/ – /d/ Contrast by Native Speakers of English, Mandarin, and Spanish.” *J. Acoust. Soc. Am.* 92: 128–143.

Flege, J. E., C. Schirru & I. R. A. MacKay. 2003. “Interaction between the Native and Second Language Phonetic Subsystems.” *Speech Communication* 40: 467–491.

Flege, J. E., G. Yeni-Komshian & H. Liu. 1999. “Age Constraints on Second Language Acquisition.” *Journal of Memory & Language* 41: 78–104.

FMKS – Verein für frühe Mehrsprachigkeit an Kindertageseinrichtungen und Schulen. 2008. “Ranking: Anzahl bilingualer Grundschulen in den Bundesländern und in großen Städten.” Available online at <http://www.fmks-online.de/_wd_showdoc.php?pic=544> [Last accessed January 08, 2014]

FMKS – Verein für frühe Mehrsprachigkeit an Kindertageseinrichtungen und Schulen. 2010. “Bei bilingualen Kitas in Deutschland bleibt das Saarland Spitzenreiter.” Available online at <http://www.fmks-online.de/_wd_showdoc.php?pic=776> [Last accessed January 08, 2014]

FMKS – Verein für frühe Mehrsprachigkeit an Kindertageseinrichtungen und Schulen. 2011. “Immersion in Primary School – A Guide.” Available online at <www.fmks-online.de/_wd_showdoc.php?pic=817> [Last accessed January 08, 2014]

Footnick, R. 2007. “A Hidden Language: Recovery of a ‘Lost’ Language is Triggered by Hypnosis.” In: B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (eds.). *Languag Attrition: Theoretical Perspectives*. Amsterdam: Benjamins. 169–187.

Fourakis, M. 1986. “An Acoustic Study of the Effects of Tempo and Stress on Segmental Intervals.” *Phonetica* 43: 172–188.

Fourakis, M. 1991. “Tempo, Stress and Vowel Reduction.” *J. Acoust. Soc. Am*. 90 (4): 1816–1827.

Fox, R. A. & D. Terbeek. 1977. “Dental Flaps, Vowel Duration and Rule Ordering in American English.” *Journal of Phonetics* 5: 27–34.

Francis, A. L. & H. C. Nusbaum. 2002. “Selective Attention and the Acquisition of New Phonetic Categories.” *Journal of Experimental Psychology: Human Perception and Performance* 28 (2): 349–366.

Francis, A. L., K. Baldwin & H. C. Nusbaum. 2000. “Effects of Training on Attention to Acoustic Cues.” *Perception and Psychophysics* 62 (8): 1668–1680.

Frisch, S. A., N. R. Large & D. B. Pisoni. 2000. “Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Words.” *Journal of Memory and Language* 42: 481–496.

Fuertes, J. N., W. H. Gottdiener, H. Martin, T. C. Gilbert & H. Giles. 2012. “A Meta-Analysis of the Effects of Speakers’ Accents on Interpersonal Evaluations.” *European Journal of Social Psychology* 42: 120–133.

Furhop, N. & J. Peters. 2013. *Einführung in die Phonologie und Graphematik*. Stuttgart: Metzler.

Gahl, S. & S. M. Garnsey. 2004. “Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation.” *Language* 80 (4): 748–775.

Garnes, S. 1974. “Suprasegmental Aspects of Icelandic Vowel Quality.” *Ohio State University, Working Papers in Linguistics* 17: 144–159.

Gass, S. 1984. “Development of Speech Perception and Speech Production in Adult Second Language Learners.” *Applied Psycholinguistics* 5: 51–74.

Gass, S. & L. Selinker. 2008. *Second Language Acquisition: An Introductory Course*. New York: Routledge.

Gay, T. 1977. "Effects of Speaking Rate on Vowel Formant Movements." *STL-QPSR* 18 (2-3): 8–30.

Gay, T. 1978. "Effects of Speaking Rate on Vowel Formant Movements." *J. Acoust. Soc. Am.* 63: 223–230.

Gerosa, M., S. Lee, D. Guiliani & S. S. Narayanan. 2006. "Analyzing Children's Speech: An Acoustic Study of Consonants and Consonant-Vowel Transition." *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* 393–396. Available online at <http://sail.usc.edu/aigaion2/index.php/attachments/single/193> [Last accessed August 01, 2013]

Gluszek, A. & J. F. Dovidio. 2010. "The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication." *Personality and Social Psychology Review* 14 (2): 214–237.

Gobet, F., P. C. R. Lane, S. Croker, P. C-H. Cheng, G. Jones, I. Oliver & J. M. Pine. 2001. "Chunking Mechanisms in Human Learning." *TRENDS in Cognitive Sciences* 5 (6): 236–243.

Gobet, F., H. B. Richman, J. J. Staszewski & H. A. Simon. 1997. "Goals, Representations, and Strategies in a Concept Attainment Task: The EPAM Model." *The Psychology of Learning and Motivation* 37: 265–290.

Gopal, H. S. 1990. "Effects of Speaking Rate on the Behavior of Tense and Lax Vowel Durations." *Journal of Phonetics* 18: 497–518.

Gopal, H. S. 1996. "Generalizability of Current Models of Vowel Duration." *Phonetica* 53: 1–32.

Griffin, Z. M. & K. Bock. 1998. "Constraint, Word Frequency, and Levels of Processing in Spoken Word Production." *Journal of Memory and Language* 38: 313–338.

Guion, S. G. & E. Pederson. 2007. "Investigating the Role of Attention in Phonetic Learning." In: O.-S. Bohn & M. J. Munro (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: Benjamins. 57–77.

Hacquard, V., M. A. Walter & A. Marantz. 2006. "The Effect of Inventory on Vowel Perception in French and Spanish: An MEG Study." *Brain & Language* 100: 295–300.

Halle, M. & K. Stevens. 1967. "Mechanisms of Glottal Vibration for Vowels and Consonants." *J. Acoust. Soc. Am.* 41 (6): 1613–1613.

Halle, M., G. Hughes & J.-P. Radley. 1957. "Acoustic Properties of Stop Consonants." *J. Acoust. Soc. Am.* 29: 107–116.

Hancock, M. 1996. *Pronunciation Games*. Cambridge: CUP. Available online at: <http://www.anthonyteacher.com/teachingbooks/Pronunciation_Games.pdf> [Last accessed December 02, 2013]

Harmer, J. 2001. *The Practice of English Language Teaching*. Harlow: Pearson Education.

Hay, J., P. Warren & K. Drager. 2006. "Factors Influencing Speech Perception in the Context of a Merger-in-Progress." *Journal of Phonetics* 34 (4): 458–484.

Heffner, R.-M. S. 1937. "Notes on the Length of Vowels." *American Speech* 12: 128–134.

Hillenbrand, J. M., M. J. Clark & R. A. Houde. 2000. "Some Effects of Duration on Vowel Recognition." *J. Acoust. Soc. Am.* 108 (6): 3013–3022.

Hillenbrand, J. M., L. A. Getty, M. J. Clark & K. Wheeler. 1995. "Acoustic Characteristics of American English Vowels." *J. Acoust. Soc. Am.* 97: 3099–3111.

Hillenbrand, J., D. R. Ingrisano, B. L. Smith & J. E. Flege. 1984. "Perception of the Voiced-Voiceless Contrast in Syllable-Final Stops." *J. Acoust. Soc. Am.* 76: 18–26.

Hoard, J. E. 1966. "Juncture and Syllable Structure in English." *Phonetica* 15 (2): 96–109.

Hogan, J. T. & A. J. Roszypal. 1980. "Evaluation of Vowel Duration as a Cue for the Voicing Distinction in the Following Word-Final Consonant." *J. Acoust. Soc. Am.* 67: 1764–1771.

Holt, L. L. & A. J. Lotto. 2006. "Cue Weighting in Auditory Categorization: Implications for First and Second Language Acquisition." *J. Acoust. Soc. Am.* 119: 3059–3071.

Hoole, P., C. Mooshammer & H. G. Tillmann. 1994. "Kinematic Analysis of Vowel Production in German." *Proceedings of ICSLP 94*: 53–55.

House, A. S. & G. Fairbanks. 1953. "The Influence of Consonants Environment upon the Secondary Acoustical Characteristics of Vowels." *J. Acoust. Soc. Am.* 25: 105–113.

Huang, C. B. 1986. "The Effect of Formant Trajectory and Spectral Shape on the Tense/Lax Distinction in American Vowels." *IEEE ICASSP*: 893–896.

Hurford, J. 1991. "The Evolution of the Critical Period for Language Acquisition." *Cognition* 40 (3): 159–201.

Imada, T., Y. Zhang, M. Cheour, S. Taulu, A. Ahonen & P. Kuhl. 2006. "Infant Speech Perception Activates Broca's Area: A Developmental Magnetoencephalography Study." *NeuroReport* 17: 957–962.

Ingram, D. 1989. *Phonological Disability in Children*. London: Whurr.

Iverson, P. & P. Kuhl. 1996. "Influences of Phonetic Identification and Category Goodness on American Listeners' Perception of /r/ and /l/." *J. Acoust. Soc. Am.* 92: 1130–1140.

Iverson, P., P. K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Kettermann & C. Siebert. 2003. “A Perceptual Interference Account of Acquisition Difficulties for Non-Native Phonemes.” *Cognition* 87: B47–B57.

Jacewicz, E., R. A. Fox & J. Salmons. 2007. “Vowel Duration in Three American English Dialects.” *Am. Speech* 82: 367–385.

James, A. R. 1985. “Phonetic Transfer and Phonological Explanation: Some Theoretical and Methodological Issues.” In: E. Kellerman & M. Sharwood (eds.). *Cross-Language Influence in Second Language Acquisition*. Oxford: Pergamon Press.

Javkin, H. R. 1975. “Perceptual Basis of Vowel-Duration Differences.” *J. Acoust. Soc. Am.* 58 (S1): 58-58.

Jeng, J.-Y. & G. Weismer. 2004. “The Influence of Speaking Rates and Utterance Units on Segmental Duration of Mandarin.” 1–11. Available online at <http://scitation.aip.org/confst/ASA/data/2/4pSC3.pdf> [Last accessed October 02, 2013]

Jenkins, J. 2002. “A Sociolinguistically Based, Empirically Researched Pronunciation Syllabus for English as an International Language.” *Applied Linguistics* 23 (1): 83–103.

Jenkins, J. 2005. “Implementing an International Approach to English Pronunciation: The Role of Teacher Attitudes and Identity.” *TESOL Quarterly* 39 (3): 535–543.

Jespersen, O. 1904. *Lehrbuch der Phonetik*. Leipzig: Teubner.

Jessen, M. 2002. “Spectral Balance in German and its Relevance for Syllable Cut Theory.” In: P. Auer, P. Gilles & H. Spiekermann (eds.) *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer. 153–180.

Jones, D. 1950. *The Phoneme: Its Nature and Use.* Cambridge: W. Heffer & Sons.

Jones, G, F. Gobet & J. M. Pine. 2000. “Learning Novel Sound Patterns.” *Proceedings of the 3rd International Conference on Cognitive Modelling*. Veenendaal: Universal Press. 169–176.

Jurafsky, D. 1996. “A Probabilistic Model of Lexical and Syntactic Access and Disambiguation.” *Cognitive Science* 20: 137–194.

Jurafsky, D., A. Bell, M. Gregory & W. D. Raymond. 2000. “Probabilistic Relations between Words: Evidence from Reduction in Lexical Production.” In: J. Bybee & P. Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. 229–254.

Kang, S. 2013. “Neural Basis of the Word Frequency Effect and its Relation to Lexical Processing.” *UC Berkeley Phonology Lab Annual Report 2012*. Available online at <http://linguistics.berkeley.edu/phonlab/annual_report/documents/2012/Kang_word_freq.pdf> [Last accessed September 30, 2013]

Kehoe, M. & C. Stoel-Gammon. 2001. “Development of Syllable Structure in English-Speaking Children with Particular Reference to Rhymes.” *Journal of Child Language* 28 (2): 393–432.

Kellogg, R. T. & R. S. Dare. 1989. “Explicit Memory for Unattended Information.” *Bulletin of the Psychonomic Society* 27: 409–412.

Kersten, K. 2005. “Bilinguale Kindergärten und Grundschulen: Wissenschaft und Praxis im Kieler Immersionsprojekt.” In: P. Baron (ed.). *Bilingualität im Kindergarten und in der Primarstufe. Bessere Zukunftschancen für unsere Kinder*. Opole: Niemieckie Towarzystwo Oswiatowe. 22–33.

Kessinger, R. H. & S. E. Blumstein. 1998. “Effects of Speaking Rate on Voice-Onset Time and Vowel Production: Some Implications for Perception Studies.” *Journal of Phonetics* 26: 117–128.

Kewley-Port, D., R. Akahane-Yamada & K. Aikawa. 1996. “Intelligibility and Acoustic Correlates of Japanese Accented English Vowels.” *Proceedings of ICSLP 96*: 450–453.

Kirsner, K. 1994. “Implicit Processes in Second Language Learning. In: N. Ellis (ed.). *Implicit and Explicit Learning of Languages*. London: Academic Press. 283–312.

Klatt, D. H. 1973. “Interaction between two Factors that Influence Vowel Duration.” *J. Acoust. Soc. Am.* 54 (4): 1102–1104.

Klatt, D. H. 1975. “Vowel Lengthening is Syntactically Determined in a Connected Discourse.” *J. Acoust. Soc. Am.* 59: 1208–1221.

Kluender, K. R., R. L. Diehl & B. A. Wright. 1988. “Vowel-Length Differences before Voiced and Voiceless Consonants: An Auditory Explanation.” *Journal of Phonetics* 16: 153–169.

Kluge, D. C., A. S. Rauber, M. S. Reis & R. A. H. Bion. 2007. “The Relationship between the Perception and Production of English Nasal Codas by Brazilian Learners of English.” *Proceedings of Interspeech 2007*: 2297–2300.

Knightley, L. M., S. A. Jun, J. S. Oh & T. K. Au. 2003. “Production Benefits of Childhood Overhearing.” *J. Acoust. Soc. Am.* 114: 465–474.

Ko, E. 2007. “Acquisition of Vowel Duration in Children Speaking American English.” *Proceedings of Interspeech 2007*: 1881–1884.

Ko, E., M. Soderstrom & J. Morgan. 2009. “Development of Perceptual Sensitivity to Extrinsic Vowel Duration in Infants Learning American English.” *J. Acoust. Soc. Am.* 126 (5): 134–139.

Koenig, L. L., J. C. Lucero & E. Perlman. 2008. “Speech Production Variability in Fricatives of Children and Adults: Results of Functional Data Analysis.” *J. Acoust. Soc. Am.* 124 (5): 3158–3170.

Kohler, K. 1979. “Dimensions in the Perception of Fortis and Lenis Plosives.” *Phonetica* 36 (4–5): 332–343.

Kondaurova, M. V. & A. L. Francis. 2008. “The Relationship between Native Allophonic Experience with Vowel Duration and Perception of the English Tense/Lax Vowel Contrast by Spanish and Russian Listeners.” *J. Acoust. Soc. Am.* 124 (6): 3959–3971.

Krashen, S. D. & T. D. Terrell. 1988. *The Natural Approach: Language Acquisition in the Classroom*. Oxford: Pergamon Press.

Krashen, S. D., M. H. Long & C. Scarcella Robin. 1979. “Age, Rate, and Eventual Attainment in Second Language Acquisition.” *TESOL Quarterly* 13 (4): 573–582.

Krause, S. E. 1982. “Vowel Duration as a Perceptual Cue to Postvocalic Consonant Voicing in Young Children and Adults.” *J. Acoust. Soc. Am.* 71: 990–995.

Kroos, C. 1996. *Eingipflige und zweigipflige Vokale des Deutschen? Kinematische Analyse der Gespanntheitsopposition im Standarddeutschen*. M. A. Thesis, University of Munich.

Kruger Bott, A.-M. 2005. *Computer-Aided Self-Access Pronunciation Materials Designed to Teach Stress in American English*. Brigham Young University.

Kuhl, P. 1979. “Speech Perception in Early Infancy: Perceptual Constancy for Spectrally Dissimilar Vowel Categories.” *J. Acoust. Soc. Am.* 66: 1668–1679.

Kuhl, P. 1983. “Perception of Auditory Equivalence Classes for Speech in Early Infancy.” *Infant Behavior and Development* 6: 263–285.

Kuhl, P. 2000. “A New View of Language Acquisition.” *Proceedings of the National Academy of Sciences 97*: 11850–11857.

Kuhl, P. 2004. “Early Language Acquisition: Cracking the Speech Code.” *Nature Reviews Neuroscience* 5 (11): 831–843.

Kuhl, P. & P. Iverson. 1995. “Linguistic Experience and the ‘Perceptual Magnet Effect.’” In: W. Strange (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press. 121–154.

Kuhl, P., K. A. Williams, F. Lacerda, K. N. Stevens & B. Lindblom. 1992. “Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age.” *Science* 255: 606–608.

Kuhl, P., B. T. Conboy, S. Coffrey-Corina, D. Padden, M. Rivera-Gaxiola & T. Nelson. 2008. “Phonetic Learning as a Pathway to Language: New Data and Native Language Magnet Theory Expanded (NLM-e).” *Philosophical Transactions of the Royal Society B* 363: 979–1000.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Philadelphia Press.

Labov, W. 1994. *Principles of Linguistic Change Vol. 1: Internal Factors*. Oxford: Blackwell.

Labov, W. & M. Baranowski. 2006. “50 msec.” *Language Variation and Change* 18: 1–18.

Laefur, C. 1992. “Patterns of Voicing-Conditioned Vowel Duration in French and English.” *Journal of Phonetics* 20: 411–440.

Landauer, T. K. & L. A. Streeter. 1973. “Structural Differences between Common and Rare Words: Failure of Equivalence Assumptions for Theories of Word Recognition.” *Journal of Verbal Learning and Verbal Behavior* 12: 119–131.

Langstrof, C. 2006a. “Acoustic Evidence for a Push-Chain Shift in the Intermediate Period of New Zealand English.” *Language Variation and Change* 18 (2): 141–164.

Langstrof. C. 2006b. *Vowel Change in New Zealand English – Patterns and Implications*. Doctoral Dissertation, University of Canterbury.

Leather, J. 1983. “Second-Language Pronunciation Learning and Teaching.” *Language Teaching* 16 (3): 198–219.

Leather, J. & A. James. 1991. “The Acquisition of Second Language Speech.” *Studies in Second Language Acquisition* 13: 305–314.

Lee, S., A. Potamianos & S. Narayanan. 1999. “Acoustics of Children’s Speech: Developmental Changes of Temporal and Spectral Parameters.” *J. Acoust. Soc. Am.* 105 (3): 1455–1468.

Lehiste, I. 1960. “An Acoustic-Phonetic Study of Internal Open Juncture.” *Phonetica* 5 (1): 5–54.

Lehiste, I. 1970. *Suprasegmentals*. Cambridge: MIT Press.

Lehiste, I. 1975. “Some Factors Affecting the Duration of Syllabic Nuclei in English.” In: G. Drachmann (ed.). *Proceedings of the First Salzburg Conference on Linguistics*. Tübingen: Gunter Narr. 81–104.

Lenneberg, E. 1967. *Biological Foundations of Language*. New York: Wiley.

Lev-Ari, S. & B. Keysar. 2010. “Why Don’t we Believe Non-Native Speakers? The Influence of Accent on Credibility.” *Journal of Experimental Social Psychology* 46: 1093–1096.

Lieberman, P. 1963. “Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech.” *Language and Speech* 6: 172–187.

Lightner, T. M. 1972. *Problems in the Theory of Phonology Vol 1: Russian Phonology and Turkish Phonology*. Edmonton: Linguistic Research.

Lindblom, B. 1968. “Temporal Organization of Speech Production.” *Speech Transmission Laboratories, Quarterly Progress Status Report*. No. 2–3: 1–5.

Lindblom, B. 1983. “Economy of Speech Gestures.” In: P. F. MacNeilage (ed.). *The Production of Speech*. New York: Springer. 217–246.

Lindblom, B. & K. Rapp. 1973. *Some Temporal Regularities of Spoken Swedish.* University of Stockholm.

Lindemann, S. 2003. “Koreans, Chinese or Indians? Attitudes and Ideologies about Nonnative English Speakers in the United States.” *Journal of Sociolinguistics* 7: 348–364.

Lindemann, S. 2005. “Who speaks ‘Broken English’? US Undergraduates’ Perception of Nonnative English.” *International Journal of Applied Linguistics* 15: 187–212.

Lippi-Green, R. 1997. *English with Accents: Language, Ideology, and Discrimination in the United States*. New York: Routledge.

Lisker, L. 1957. “Closure Duration and the Intervocalic Voiced-Voiceless Distinctions in English.” *Language* 33: 42–49.

Lisker, L. 1974. “On ‘Explaining’ Vowel Duration Variation.” *Glossa* 8: 223–246.

Lisker, L. 1978. “Rapid vs. Rabid: A Catalogue of Acoustic Features that May Cue the Distinction.” *Haskins Laboratories Status Report on Speech Research* 54: 127–132.

Llisterri, J. 1995. “Relationships between Speech Production and Speech Perception in a Second Language.” *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 92–99.

Logan, J. S., S. E. Lively & D. B. Pisoni. 1991. “Training Japanese Listeners to Identify English /r/ and /l/: A First Report.” *J. Acoust. Soc. Am.* 89: 874–886.

Long, M. H. 1990. “Maturational Constraints on Language Development.” *Studies in Second Language Acquisition* 12 (3): 251–285.

Long, M. H. 2005. “Problems with Supposed Counter-Evidence to the Critical Period Hypothesis.” *IRAL* 43: 287–317.

Luce, P. A. & J. C. Luce. 1985. “Contextual Effects on Vowel Duration, Closure Duration, and the Consonant/Vowel Ratio in Speech Production.” *J. Acoust. Soc. Am.* 78: 1949–1957.

Maack, A. 1953. “Die Beeinflussung der Sonantendauer durch die Nachbarkonsonanten.” *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 7: 104–128.

Maas, U. 2002. “Die Anschlusskorrelation des Deutschen im Horizont einer Typologie der Silbenstruktur.” In: P. Auer, P. Gilles & H. Spiekermann. *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer. 11–34.

Maas, U. & D. Tophinke. 1993. “Loser und fester Anschluß. Versuch der Neubewertung einer abgelegten Kategorie.” In: J. Schmidt-Radefeldt & A. Harder (eds.). *Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdtke*. Tübingen: Narr. 133–151.

Macdonald, D., G. Yule & M. Powers. 1994. “Attempts to Improve English L2 Pronunciation: The Variable Effects of Different Types of Instruction.” *Language Learning* 44: 75–100.

MacDonald, S. 2002. “Pronunciation – Views and Practices of Reluctant Teachers.” *Prospect* 17 (3): 3–18.

Mack, M. 1982. “Voicing-Dependent Vowel Duration in English and French: Monolingual and Bilingual Production.” *J. Acoust. Soc. Am.* 71 (1): 173–178.

Maclagan, M. & J. Hay. 2007. “Getting Fed Up with our Feet: Contrast Maintenance and the New Zealand English ‘Short’ Front Vowel Shift.” *Language Variation and Change* 19 (1): 1–25.

Major, R. 1993. “Sociolinguistic Factors in Loss and Acquisition of Phonology.” In K. Hyltenstam & A. Viberg (eds.). *Progession and Regression in Language: Sociocultural, Neuropsychological and Linguistic Perspectives*. Cambridge: CUP. 463–478.

Mattingly, I. 1968. “Synthesis by Rule of General American English.” *Supplement to Status Report on Speech Research*. Haskin Laboratories.

Mattys, S. L., P. W. Jusczyk, P. A. Luce & J. L. Morgan. 1999. “Phonotactic and Prosodic Effects on Word Segmentation in Infants.” *Cognitive Psychology* 38 (4): 465–494.

Maye, J. C., J. F. Werker & L. A. Gerken. 2002. “Infant Sensitivity to Distributional Information Can Affect Phonetic Discrimination.” *Cognition* 82: B 101–B 111.

Mayo, L., M. Florentine & S. Buus. 1997. “Age of Second-Language Acquisition and Perception of Speech in Noise.” *Journal of Speech, Language, and Hearing Research* 40: 686–693.

McAllister, R., J. E. Flege & T. Piske. 2002. “The Influence of L1 on the Acquisition of Swedish Quantity by Native Speakers of Spanish, English and Estonian.” *Journal of Phonetics* 30: 229–258.

McArthur, T. 2002. *The Oxford Guide to World English*. Oxford: OUP.

McCandliss, B. D., J. A. Fiez, A. Protopapas, M. Conway & J. L. McClelland. 2002. “Success and Failure in Teaching the [r]–[l] Contrast to Japanese Adults: Tests of a Hebbian Model of Plasticity and Stabilization in Spoken Language Perception.” *Cognitive, Affective, & Behavioral Neuroscience* 2 (2): 89–108.

McClelland, J. L., A. Thomas, B. D. McCandliss & J. A. Fiez. 1999. “Understanding Failures of Learning: Hebbian Learning, Competition for Representational Space, and Some Preliminary Experimental Data.” In: J. Reggia, E. Ruppin & D. Glanzman (eds.). *Brain, Behavioral, and Cognitive Disorders: The Neurocomputational Perspective*. Oxford: Elsevier. 75–80.

Meyer, E. A. 1903. *Englische Lautdauer: Eine experimentalphonetische Untersuchung*. Leipzig: Harrassowitz.

Meyer, E. A. 1904. “Zur Vokaldauer im Deutschen.” In: *Nordiska studier. Tillegnade Adolf Noreen på hans 50-årsdag den 13 Mars 1904 af Studiekamrater och lärjungar*. Uppsala. 347–356.

Miller, J. L., F. Grosjean & C. Lomanto. 1984. “Articulation Rate and its Variability in Spontaneous Speech: A Reanalysis and Some Implications.” *Phonetica* 41: 215–225.

Mills, D. L., K. Plunkett, C. Prat & G. Schafer. 2005. “Watching the Infant Brain Learn Words: Effects of Vocabulary Size and Experience.” *Cogn. Dev*. 20: 19–31.

Ministerium für Kultus, Jugend und Sport Baden-Württemberg. 2013. *Bilingualer Unterricht: deutsch-englisch an allgemein bildenden Gymnasien*. Available online at <http://www.kultusportal-bw.de/site/pbs-bw/get/documents/KULTUS.Dachmandant/KULTUS/kultusportal-bw/pdf/Bilingualer-Unterricht.pdf> [Last accessed January 04, 2014]

Minnick-Fox, M. & K. Maeda. 1999. “Categorization of American English Vowels by Japanese Speakers.” In: J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.). *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: University of California. 1437–1440.

Moon, S.-J. & B. Lindblom. 1994. “Interaction between Duration, Context, and Speaking Style in English Stressed Vowels.” *J. Acoust. Soc. Am*. 96 (1): 40–55.

Morley, J. 1991. “The Pronunciation Component in Teaching English to Speakers of Other Languages.” *TESOL Quarterly* 25 (1): 51–74.

Morrison, G. S. 2002. “Perception of English /□/ and /□/ by Japanese and Spanish Listeners: Longitudinal Results.” In: G. S. Morrison & L. Zsoldos (eds.). *Proceedings of the North West Linguistics Conference 2002*. Burnaby: Simon Fraser University Linguistics Graduate Student Association. 29–48.

Morrison, G. S. 2008a. “L1-Spanish Speakers’ Acquisition of the English /□/–/□/ Contrast: Duration-Based Perception Is Not the Initial Developmental Stage.” *Language & Speech* 51: 285–315.

Morrison, G. S. 2008b. “Perception of Synthetic Vowels by Monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish Listeners.” *Canadian Acoustics* 36 (4): 17–23.

Moyer, A. 1999. “Ultimate Attainment in L2 Phonology.” *Studies in Second Language Acquisition* 21: 81–108.

Moyer, A. 2004. *Age, Accent and Experience in Second Language Acquisition. An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters.

Moyer, A. 2007. “Empirical Considerations on the Age Factor in L2 Phonology.” *Issues in Applied Linguistics* 15 (2): 109–127.

Mugitani, R., F. Pons, L. Fais, C. Dietrich, J. F. Werker & S. Amano. 2009. “Perception of Vowel Length by Japanese- and English-Learning Infants.” *Dev. Psychol.* 45: 236–247.

Muñoz, C. 2004. “The Interplay of Age and Exposure in the Development of Second Language Skills in an Instructed Foreign Language Context.” *Paper presented at the Annual Meetings of the American Association of Applied Linguistics*, Portland, OR, May 1–4.

Munro, M. J. 1993. “Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings.” *Language & Speech* 36: 39–66.

Munro, M. J. & T. M. Derwing. 1995. “Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners.” *Language Learning* 45 (1): 73–97.

Munro, M. J., J. E. Flege & I. R. A. MacKay. 1996. “The Effects of Age of Second-Language Learning on the Production of English Vowels.” *Applied Psycholinguistics* 17: 313–334.

Munson, B. 2007. “Lexical Access, Lexical Representation, and Vowel Production.” In: J. S. Cole & J. I. Hualde (eds.). *Laboratory Phonology 9*. Berlin: Mouton. 201–228.

Murray, R. W. 2000. “Syllable Cut Prosody in Early Middle English.” *Language* 76: 617–654.

Murray, R. W. 2002. “Accents and a Medieval English Phonologist.” In: D. Restle & D. Zaefferer (eds). *Sounds and Systems. Studies in Structure and Change. A Festschrift for Theo Vennemann*. Berlin/New York: Mouton de Gruyter. 91–120.

Myers, S. 2012. “Final Devoicing: Production and Perception Studies.” In: T. Borowsky, S. Kawahara, T. Shinya & M. Sugahara (eds.). *Prosody Matters: Essays in Honor of Elisabeth Selkirk*. London: Equinox Press. 148–180.

Naeser, M. A. 1970. *The American Child’s Acquisition of Differential Vowel Duration*. Doctoral Dissertation, University of Wisconsin.

Narayanan, S. & D. Jurafsky. 1998. “Bayesian Models of Human Sentence Processing.” *COGSCI-98*: 752–757.

Nearey, T. M. 1989. “Static, Dynamic and Relational Properties in Vowel Perception.” *J. Acoust. Soc. Am.* 85 (5): 2088–2113.

Nittrouer, S. 2000. “Learning to Apprehend Phonetic Structure from the Speech Signal: The Hows and Whys.” *Paper presented at the 8th Meeting of the International Clinical Phonetics and Linguistics Association*, Queen Margaret University College, Edinburgh.

Neufeld, G. 1988. “Phonological Asymmetry in Second-Language Learning and Performance.” *Language Learning* 38 (4): 531–559.

Nooteboom, S. G. 1972. “Some Timing Factors in the Production and Perception of Vowels.” *Occasional Papers, University of Essex* 13: 49–76.

Oh, J. S., T. K. Au & S. A. Jun. 2010. “Early Childhood Language Memory in the Speech of Perception of International Adoptees.” *Journal of Child Language* 37: 1123–1132.

Oh, J. S., S. A. Jun, L. M. Knightly & T. K. Au. 2003. “Holding on to Childhood Language Memory.” *Cognition* 86 (3): B53–64. Available online at <http://linguistics.ucla.edu/people/jun/Cognition-Oh%20et%20al.pdf> [Last accessed November 25, 2013]

Oldfield, R. C. & A. Wingfield. 1965. “Response Latencies in Naming Objects.” *Quarterly Journal of Experimental Psychology* 4: 272–281.

Oller, K. 1973. “The Effect of Position in Utterance on Speech Segment Duration in English.” *J. Acoust. Soc. Am.* 54 (5): 1235–1247.

Ortega, L. 2009. *Understanding Second Language Acquisition*. London: Hodder Arnold.

Ostry, D. J. & K. G. Munhall. 1985. “Control of Rate and Duration of Speech Sounds.” *J. Acoust. Soc. Am.* 77 (2): 640–648.

Oyama, S. 1976. “A Sensitive Period for the Acquisition of a Non-Native Phonological System.” *Journal of Psycholinguistic Research* 5: 261–283.

Pallier, C., S. Dehaene, J.-B. Poline, D. LeBihan, A.-M. Argenti, E. Dupoux & J. Mehler. 2003. “Brain Imaging of Language Plasticity in Adopted Adults: Can a Second Language Replace the First?” *Cerebral Cortex* 13 (2): 155–161.

Patkowski, M. 1990. “Age and Accent in a Second Language: A Reply to James Emil Flege.” *Applied Linguistics* 11: 73–78.

Penfield, W. 1965. “Conditioning the Uncommitted Cortex for Language Learning”. *Brain* 88: 787–798.

Pennington, M. 1989. “Teaching Pronunciation from the Top Down.” *RELC Journal* 20: 20–38.

Pennington, M. & J. Richards. 1986. “Pronunciation Revisited.” *TESOL Quarterly* 20: 207–225.

Perales, S., M. del Pilar Garcia Mayo & J. M. Liceras. 2004. “The Acquisition of English Sentential Negation by Bilingual (Basque/Spanish) Children: The Age Factor in an Institutional Setting.” *Paper presented at the Annual Meetings of the American Association of Applied Linguistics*, Portland, OR, May 1–4.

Perlmutter, M. 1989. “Intelligibility Rating of L2 Speech Pre- and Post-Intervention.” *Perceptual and Motor Skills* 68: 515–521.

Peterson, G. E. 1955. “An Oral Communication Model.” *Language* 31 (3): 414–427.

Peterson, G. E. & I. Lehiste. 1960. “Duration of Syllable Nuclei in English.” *J. Acoust. Soc. Am.* 32 (6): 693–703.

Peterson, G. E. & I. Lehiste. 1961. "Transitions, Glides and Diphthongs." *J. Acoust. Soc. Am.* 33 (3): 268–277.

PH Freiburg. 2013. "Europalehramt an der Pädagogischen Hochschule Freiburg – Grundschule." Available online at <https://www.ph-freiburg.de/fileadmin/dateien/studium/europalehramt/Fuer_Bewerber_und_Interessierte/PH_Freiburg_Europalehramt_G.pdf> [Last accessed January 08, 2014]

Picard, M. 2007. "On Teaching the Pronunciation of Subphonemic Segments in English." *SKY Journal of Linguistics* 20: 331–344.

Pierrehumbert, J. 2002. "Word-Specific Phonetics." In: C. Gussenhoven & N. Varner (eds.). *Laboratory Phonology 7*. Berlin: Mouton de Gruyter. 101–139.

Piller, I. 2002. "Passing for a Native Speaker: Identity and Success in Second Language Learning." *Journal of Sociolinguistics* 6: 179–206.

Pinker, S. 1994. *The Language Instinct*. New York: Morrow.

Piroth, H. G. & P. M. Janker. 2004. "Speaker-Dependent Differences in Voicing and Devoicing of German Obstruents." *Journal of Phonetics* 32: 81–109.

Piske, T. 2007. "Implications of James E. Flege's Research for the Foreign Language Classroom." In: O.-S. Bohn & M. J. Munro (ed.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: Benjamins. 301–314.

Piske, T., I. R. A. MacKay & J. E. Flege. 2001. "Factors Affecting Degree of Foreign Accent in an L2: A Review." *Journal of Phonetics* 29: 191–215.

Pluymaekers, M., M. Ernestus & R. H. Baayen. 2005. "Articulatory Planning is Continuous and Sensitive to Informational Redundancy." *Phonetica* 62: 146–159.

Port, R. F. 1981. "Linguistic Timing Factors in Combination." *J. Acoust. Soc. Am.* 69 (1): 262–274.

Port, R. & M. O'Dell. 1985. "Neutralization of Syllable-Final Voicing in German." *Journal of Phonetics* 13: 455–471.

Pourhosein Gilakjani, A. 2012. "The Significance of Pronunciation in English Language Teaching." *Canadian Center of Science and Education* 5 (4): 96–107.

Prator, C. H. 1971. "Phonetics vs. Phonemics in the ESL Classroom: When is Allophonic Accuracy Important?" *TESOL Quarterly* 5: 61–72.

Preston, D. 1981. "The Ethnography of TESOL." *TESOL Quarterly* 15: 105–116.

Pruitt, J. S., J. J. Jenkins & W. Strange. 2006. "Training the Perception of Hindi Dental and Retroflex Stops by Native Speakers of English and Japanese." *J. Acoust. Soc. Am.* 119: 1684–1696.

Purcell, E. T. & R. W. Suter. 1980. “Predictors of Pronunciation Accuracy: A Reexamination.” *Language Learning* 30: 271–287.

Raphael, L. J. 1972. “Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristics of Word-Final Consonants in American English.” *J. Acoust. Soc. Am.* 51: 1296–1303.

Raphael, L. J., M. F. Dorman & D. Geffner. 1980. “Voicing-Conditioned Durational Differences in Vowels and Consonants in the Speech of Three- and Four-Year Old Children.” *Journal of Phonetics* 8: 335–341.

Rauber, A. S., P. Escudero, R. Bion & B. O. Baptista. 2005. “The Interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese.” *Proceedings of Interspeech 2005*: 2913–2916.

Richman, H. B. & H. A. Simon. 1989. “Context Effects in Letter Perception: Comparison of Two Theories.” *Psychological Review* 3: 417–432.

Ringel-Eichinger, A. & U. Selz. 2001/2002. “Vorbereitung und Durchführung der mündlichen Abiturprüfung im Fach Englisch.“ *Landesbildungsserver Baden-Württemberg*. <http://www.schule-bw.de/schularten/gymnasium/zentralepruefungen/abitur/handreichungenmuendlichepruefung/handreichungen_englisch.pdf> [Last accessed January 03, 2014]

Rochet, B. 1995. “Perception and Production of Second-Language Speech Sounds by Adults.” In: W. Strange (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press. 379–410.

Rymarczyk, J. 2003. “Fine Arts in English? A Plea for the Extension of the Canon of Bilingual Subjects.” In: H. Weinstock (ed.). *English and American Studies in German 2002. Summaries of Theses and Monographs. A Supplement to Anglia*. Tübingen: Niemeyer. 156–158.

Saffran, J. R., R. N. Aslin & E. L. Newport. 1996. “Statistical Cues in Language Acquisition: Word Segmentation by Infants.” *COGSCI-96*: 376–380.

Sapon, S. 1952. “An Application of Psychological Theory to Pronunciation Problems in Second Language Learning.” *Mod. Lang. J.* 36: 111–114.

Sato, C. 1985. “Task Variation in Interlanguage Phonology.” In: S. Gass & C. Madden (eds.). *Input in Second Language Acquisition*. Cambridge: Newbury House. 181–196.

Savin, H. B. 1963. “Word-Frequency Effect and Errors in the Perception of Speech.” *J. Acoust. Soc. Am.* 35: 200–206.

Scharf, D. J. 1964. “Vowel Duration in Whispered and in Normal Speech.” *Language Speech* 7: 89–97.

Schmid, P. & G. Yeni-Komshian. 1999. "The Effects of Speaker Accent and Target Predictability on Perception of Mispronunciation." *Journal of Speech, Language, and Hearing Research* 42: 56–64.

Schmidt, R. 1990. "The Role of Consciousness in Second Language Learning." *Applied Linguistics* 11: 129–158.

Schmidt, R. 1993. "Awareness and Second Language Acquisition." *Annual Review of Applied Linguistics* 13: 206–226.

Schmidt, R. 2001. "Attention." In: P. Robinson (ed.). *Cognition and Second Language Instruction*. Cambridge: CUP. 3–32.

Schumann, J. H. 1975. "Affective Factors and the Problem of Age in Second Language Acquisition." *Language Learning* 25: 205–235.

Schumann, J. H. 1978. "Social and Psychological Factors in Second Language Acquisition." In: J. C. Richards (ed.). *Understanding Second & Foreign Language Learning*. Rowley: Newbury House Publishers. 163–178.

Scovel, T. 1969. "Foreign Accents, Language Acquisition, and Cerebral Dominance." *Language Learning* 19: 245–253.

Scovel, T. 1988. *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. New York: Harper & Row.

Scovel, T. 2000. "A Critical Review of the Critical Period Research." *Annual Review of Applied Linguistics* 20: 213–223.

Sebastian-Galles, N. & S. Soto-Faraco. 1999. "Online Processing of Native and Non-Native Phonemic Contrasts in Early Bilinguals." *Cognition* 72: 111–123.

Seidenberg, M. S. & M. C. MacDonald. 1999. "A Probabilistic Constraints Approach to Language Acquisition and Processing." *Cognitive Science* 23: 569–588.

Sereno, J. A. & Y. Wang. 2007. "Behavioral and Cortical Effects of Learning a Second Language." In: O.-S. Bohn & M. J. Munro (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: Benjamins. 239–258.

Sharma, A. & M. F. Dorman. 1998. "Exploration of the Perceptual Magnet Effect Using the Mismatch Negativity Auditory Evoked Potential." *J. Acoust. Soc. Am* 104 (1): 511–517.

Sheldon, A. & W. Strange. 1982. "The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence that Speech Production Can Precede Speech Perception." *Applied Psycholinguistics* 3 (3): 243–261.

Shrosbree, M., N. Tetsu, I. Grenon & M. Kubota. 2011. “Effect of Speech Rate on a Vowel Contrast and Implications for Second Language Training.” *Proceedings of ICPhS XVII*: 8421–1845. Available online at <http://www.icphs2011. hk/resources/OnlineProceedings/RegularSession/Shrosbree/Shrosbree.pdf> [Last accessed September 30, 2013]

Sievers, E. 1901. *Grundzüge der Phonetik* [5th edition]. Leipzig: Breitkopf and Hartel.

Sievers, E. 1980. *Grundzüge der Lautphysiologie: zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen* [reprint of the original 1876 (Leipzig) and 1872 (Weimar) edition]. Hildesheim: Olms.

Simon, D. P. & H. A. Simon. 1973. “Alternative Uses of Phonemic Information in Spelling.” *Review of Educational Research* 43: 115–137.

Simon, E. & T. D’Hulster. 2012. “The Effect of Experience on the Acquisition of a Non-Native Vowel Contrast.” *Language Sciences* 34: 269–283.

Simon, H. A. & E. A. Feigenbaum. 1964. “An Information Processing Theory of Some Effects of Similarity, Familiarity, and Meaningfulness in Verbal Learning.” *Journal of Verbal Learning and Verbal Behavior* 3: 385–396.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Singleton, D. 2000. “Age and Second Language Acquisition.” *Annual Review of Applied Linguistics* 21: 77–89.

Slis, I. H. & A. Cohen. 1969. “On the Complex Regulating the Voiced-Voiceless Distinction II.” *Language and Speech* 7: 89–97.

Slowiaczek, L. M. & D. Dinnsen. 1985. “On the Neutralizing Status of Polish Word-Final Devoicing.” *Journal of Phonetics* 13: 352–341.

Smith, B. L. 1978. “Temporal Aspects of English Speech Production: A Developmental Perspective.” *Journal of Phonetics* 6: 37–67.

Smith, B. L. 1992. “Relationships between Duration and Temporal Variability in Children’s Speech.” *J. Acoust. Soc. Am.* 91: 2165–2174.

Smith, B. L., R. Hayes-Harb, M. Bruss & A. Harker. 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.

Smith, B. L., M. K. Kenney & S. Hussain. 1996. “A Longitudinal Investigation of Duration and Temporal Variability in Children’s Speech Production.” *J. Acoust. Soc. Am.* 99 (4): 2344–2349.

Söderman, T. 1993. “Word Associations of Foreign Language Learners and Native Speakers – Different Response Types and their Relevance to Lexical Development.” In: B. Hammarberg (ed.). *Problems, Process, and Product in Language Learning*. Abo: Abo Academi. 157–169.

Song, J. Y. & K. Demuth. 2008. “Compensatory Vowel Lengthening for Omitted Coda Consonants: A Phonetic Investigation of Children’s Early Representations of Prosodic Words.” *Language and Speech* 51 (4): 385–402.

Song, J. Y., K. Demuth & S. Shattuck-Hufnagel. 2012. “The Development of Acoustic Cues to Coda Contrasts in Young Children Learning American English.” *J. Acoust. Soc. Am.* 131 (4): 3036–3050.

Spiekermann, H. 2000. *Silbenschnitt in deutschen Dialekten*. Tübingen: Niemeyer.

Stack, J. W. 1993. “Effects of Speaking Rate and Stress on Vowel Durations and Formant Structures.” *J. Acoust. Soc. Am.* 93 (4): 2296–2296.

Stahl, S. A. & B. A. Murray. 1994. “Defining Phonological Awareness and its Relationship to Early Reading.” *Journal of Educational Psychology* 86: 221–234.

Stevens, K. N. 1959. “Effect of Duration upon Vowel Identification.” *J. Acoust. Soc. Am.* 31: 109(A).

Stoel-Gammon, C. & E. Buder. 1998. “Effects of Postvocalic Voicing on the Duration of High Front Vowels in Swedish and American English: Developmental Data.” *Proceedings of the Acoustical Society of America* 4: 2989–90.

Stoel-Gammon, C., E. Buder & M. Kehoe. 1995. “Acquisition of Vowel Duration: A Comparison of English and Swedish.” *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 30–34.

Stokes, J. 2001. “Factors in the Acquisition of Spanish Pronunciation.” *ITL Review of Applied Linguistics* 131/132: 63–84.

Strange, W. 1995. “Phonetics of Second-Language Acquisition: Past, Present, Future.” In: K. Elenius & P. Branderud (eds.). *Proceedings of the 1995 International Congress of Phonetic Sciences*. Stockholm: Arne Stomberg. 76–83.

Strange, W., J. J. Jenkins & T. L. Johnson. 1983. “Dynamic Specification of Coarticulated Vowels.” *J. Acoust. Soc. Am.* 74: 695–705.

Suter, R. W. 1976. “Predictors of Pronunciation Accuracy in Second Language Learning.” *Language Learning* 26: 233–253.

Swingley, D. 2009. “Onsets and Codas in 1.5-Year-Olds’ Word Recognition.” *Journal of Memory and Language* 60: 252–269.

Tahta, S., M. Wood & K. Loewenthal. 1981. “Foreign Accents: Factors Relating to Transfer of Accent from the First Language to a Second Language.” *Language & Speech* 24: 265–272.

Tarone, E. 1982. “Systematicity and Attention in Interlanguage.” *Language Learning* 32: 69–82.

Tauberer, J. 2010. *Learning [Voice]*. Doctoral Dissertation, University of Pennsylvania. Available online at <http://repository.upenn.edu/edissertations/288/> [Last accessed July 31, 2013]

Tauberer, J. & K. Evanini. 2009. “Intrinsic Vowel Duration and the Post-Vocalic Voicing Effect: Some Evidence from Dialects of North American English.” *Proceedings of Interspeech 2009*: 2211–2214.

Thompson, I. 1991. “Foreign Accents Revisited: The English Pronunciation of Russian Immigrants.” *Language* Learning 41: 177–204.

Tily, H., S. Gahl, I. Arnon, N. Snider, A. Kothari & J. Bresnan. 2009. “Syntactic Probabilities Affect Pronunciation Variation in Spontaneous Speech.” *Language and Cognition* 1/2: 147–165.

Tomasello, M. 1992. *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: CUP.

Treiman, R. & A. Zukowski. 1991. “Levels of Phonological Awareness.” In: S. A. Brady & D. P. Shankweiler (eds.). *Phonological Processes in Literacy*. Hillsdale: Erlbaum. 67–83.

Tröster-Mutz, S. 2004. “Die Realisierung von Vokallängen: erlaubt ist, was Sp[a(:)]ß macht?” *SKY Journal of Linguistics* 17: 249–265.

Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie*. Prague.

Umeda, N. 1975. “Vowel Duration in American English.” *J. Acoust. Soc. Am.* 58: 434–445.

Valdman, A. 1976. *Introduction to French Phonology and Morphology*. Rowley: Newbury House.

van der Feest, S. V. H. 2007. *Building a Phonological Lexicon: The Acquisition of the Dutch Voicing Contrast in Perception and Production*. Doctoral Dissertation, Radboud Universiteit Nijmegen. Available online at <http://suzanne.vanderfeest.nl/OutlineDiss.pdf> [Last accessed July 31, 2013]

van der Feest, S. V. H & D. Swingley. 2011. “Dutch and English Listeners’ Interpretation of Vowel Duration.” *J. Acoust. Soc. Am.* 129 (3): 57–63.

van Santen, J. P. H. 1992. “Contextual Effects on Vowel Duration.” *Speech Communication* 11: 513–546.

van Santen, J. P. H. 1994. “Assignment of Segmental Duration in Text-to-Speech Synthesis.” *Computer Speech and Language* 8: 95–128.

van Santen, J. P. H. & J. P. Olive. 1990. “The Analysis of Contextual Effects in Segmental Duration.” *Computer Speech Language* 4: 359–390.

Vennemann, T. 1991. “Syllable Structure and Syllable Cut Prosodies in Modern Standard German.” In: P. M. Bertinetto, M. Kenstowicz & M. Loporcaro (eds.). *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*. Torino: Rosenberg & Sellier. 211–243.

Vennemann, T. 1994. “Universelle Nuklearphonologie mit epiphänomenaler Silbenstruktur.” In: K. H. Ramers, H. Vater & H. Wode (eds.). *Universale phonologische Prozesse und Strukturen.* Tübingen: Niemeyer. 7–54.

Vennemann, T. 2000. “From Quantity to Syllable Cuts: On So-Called Lengthening in the Germanic Languages.” *Italian Journal of Linguistics / Rivista di Linguistica* 12: 251–282.

Ventureyra, V. A. G., C. Pallier & H. Y. Yoo. 2004. “The Loss of First Language Phonetic Perception in Adopted Koreans.” *Journal of Neurolinguistics* 17: 79–91.

Walley, A. C. & T. D. Carrell. 1983. “Onset Spectra and Formant Transitions in the Adult’s and Child’s Perception of Place of Articulation in Stop Consonants.” *J. Acoust. Soc. Am.* 73: 1011–1022.

Walker, R. 2010. *Teaching the Pronunciation of English as a Lingua Franca*. Oxford: OUP.

Walsh, T. & K. Diller. 1981. “Neurolinguistic Considerations on the Optimum Age for Second Language Learning.” In: K. Diller (ed.). *Individual Differences and Universals in Language Learning Aptitude*. Rowley: Newbury House. 3–21.

Walsh, T. & F. Parker. 1983. “Vowel Length and Vowel Transition Cues to [±Voice] in Post-Vocalic Stops.” *Journal of Phonetics* 11: 407–412.

Wang, X. & M. J. Munro. 1999. “The Perception of English Tense-Lax Vowel Pairs by Native Mandarin Speakers: The Effect of Training on Attention to Temporal and Spectral Cues.” *Proceedings of the 14th International Congress of Phonetic Sciences*. 125–128.

Wängler, H. H. & R. Weiss. 1975. “Experimental Approach to the Study of Vowel Perception in German.” *Phonetica* 32: 180–199.

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.” *J. Acoust. Soc. Am.* 71: 187–195.

Wardrip-Fruin, C. & S. Peach. 1984. “Developmental Aspects of the Perception of Acoustic Cues in Determining the Voicing Feature of Final Stop Consonants.” *Language and Speech* 27 (4): 367–379.

Warner, N., E. Good, A. Jongman & J. Sereno. 2006. “Orthographic vs. Morphological Incomplete Neutralization Effects.” *Journal of Phonetics* 34: 285–293.

Warner, N., A. Jongman, J. Sereno & R. Kemps. 2004. “Incomplete Neutralization and Other Sub-Phonemic Durational Differences in Production and Perception: Evidence from Dutch.” *Journal of Phonetics* 32: 251–276.

Weinreich, U. 1968. *Languages in Contact. Findings and Problems*. The Hague: Mouton.

Weismer, G., D. A. Dinnsen & M. A. Elbert. 1981. “A Study of the Voicing Distinction Associated with Omitted, Word-Final Stops.” *J. Speech Hear. Disord.* 46: 320–327.

Weiss, R. 1976. *The Perception of Vowel Length and Quality in German: An Experimental-Phonetic Investigation*. Hamburg: Helmut Buske.

Weiss, R. 1977. “The Phonemic Significance of the Phonetic Factors of Vowel Length and Quality in German.” *Phonologica 1976*: 271–276.

Weiss, R. 1978. “The German Vowel: Phonetic and Phonemic Considerations.” *Hamburger Phonetische Beiträge* 25: 461–475.

Weiss, W. 1992. “Perception and Production in Accent Training.” *Revue de Phonétique Appliquée* 102: 69–81.

Wells, J. C. 1962. *A Study of the Formants of the Pure Vowels of British English*. M. A. Thesis, University of London.

Wells, J. C. 1982. *Accents of English*. Cambridge: CUP.

Werker, J. F. & R. C. Tees. 2005. “Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain.” *Developmental Psychobiology* 46: 233–251.

Wesche, M. B. 2002. “Early French Immersion: How has the Original Canadian Model Stood the Test of Time?” In: P. Burmeister, T. Piske & A. Rohde (eds.). *An Integrated View of Language Development. Papers in Honor of Henning Wode.* Trier: Wissenschaftlicher Verlag Trier. 357–379.

Winitz, H., B. Gillespie & J. Starcev. 1995. “The Development of English Speech Patterns of a 7-Year-Old Polish-Speaking Child.” *Journal of Psycholinguistic Research* 24: 117–143.

Wode, H. 1995. *Lernen in der Fremdsprache: Grundzüge von Immersion und bilingualem Unterricht.* Ismanung: Hueber.

Wode, H. 2002. “Fremdsprachenvermittlung in Kita, Grundschule und Sekundarbereich: Ein integrierter Ansatz.” In: C. Finkbeiner (ed.). *Perspektiven Englisch 3*. Braunschweig: Schroedel. 33–42. Available online at <www.fmks-online.de/_wd_showdoc.php?pic=561> [Last accessed January 03, 2014].

Wode, H. 2004. “Frühes Fremdsprachenlernen. Englisch ab Kita und Grundschule: Warum? Wie? Was bringt es?” Kiel: Verein für frühe Mehrsprachigkeit an Kitas und Schulen FMKS e.V. Available online at <www.fmks-online.de/_wd_showdoc.php?pic=172> [Last accessed January 03, 2014]

Wode, H. 2009a. “Die erste Sprache als Wegbereiter der Zweiten.” Available online at <http://cms.awo-sh.de/cms/fileadmin/awo-sh/Dokumente/PDF/Alten1.pdf> [Last accessed January 08, 2014]

Wode, H. 2009b. *Frühes Fremdsprachenlernen in bilingualen Kindergärten und Grundschulen*. Braunschweig: Westermann.

Wolf, C. 1978. “Voicing Cues in English Final Stops.” *Journal of Phonetics* 6: 299–309.

Wright, C. E. 1979. “Duration Differences between Rare and Common Words and their Implications for the Interpretation of Word Frequency Effects.” *Memory & Cognition* 7 (6): 411–419.

Xu, F. & J. B. Tenenbaum. 2007. “Word Learning as Bayesian Inference.” *Psychological Review* 114: 245–272.

Yeni-Komshian, G., J. E. Flege & S. Liu. 2000. “Pronunciation Proficiency in the First and Second Languages of Korean-English Bilinguals.” *Bilingualism: Language and Cognition* 3: 131–149.

Yeni-Komshian, G., M. Robbins & J. E. Flege. 2001. “Effects of Word Class Differences on L2 Pronunciation Accuracy.” *Applied Psycholinguistics* 22: 283–299.

Yopp, H. K. & R. H. Yopp. 2000. “Supporting Phonemic Awareness Development in the Classroom.” *The Reading Teacher* 54 (2): 130–143.

Yule, G. & D. Macdonald. 1995. “The Different Effects of Pronunciation Teaching.” *IRAL* 34 (4): 345–350.

Zaidi, Z. F. 2010. “Gender Differences in Human Brain: A Review.” *The Open Anatomy Journal* 2: 37–55.

Zimmerman, S. A. & S. M. Sapon. 1958. “Note on Vowel Duration Seen Crosslinguistically.” *J. Acoust. Soc. Am.* 30: 152–153.

Zydatiß, W. 2000. *Bilingualer Unterricht in der Grundschule: Entwurf eines Spracherwerbskonzepts für zweisprachige Immersionsprogramme*. Ismaning: Hueber.

Appendix

Appendix A – Participants

A.1 – German Learners (Groups G-UK and G-US)

A.2 – Control Group (Groups C-UK and C-US)

A.3 – British Native Speakers

A.4 – American Native Speakers

A.1 – German Learners (Groups G-UK and G-US)

Subject ID	Gender	Age	Birthplace	Group	Place of stay	Length of stay (months)	Mother's nationality	Father's nationality
001	female	25	Berlin	G-UK	Birmingham	10	German	English
002	male	20	Walsrode	G-US	Massachusetts	10	German	German
003	female	20	Freiburg	G-US	Washington	7	German	German
004	female	24	Blaubeuren	G-US	North Dakota	10	German	German
005	female	29	Emmendingen	G-US	Connecticut	12	German	German
006	female	25	Saarbrücken	G-UK	Edinburgh	16	German	German
007	female	21	Lörrach	G-US	Connecticut	7	German	German
008	female	25	Villingen	G-UK	Stratford-upon-Avon	8	German	German
009	female	27	Kaltenkirchen	G-US	Illinois	11	German	German
010	male	26	Ettenheim	G-US	Wisconsin	10	German	German
011	female	24	Freiburg	G-US	Minnesota	13	German	German
012	male	27	Freiburg	G-US	Massachusetts	13	German	German
013	female	23	Landstuhl	G-US	Wisconsin	24	German	American
014	female	26	Öttingen	G-UK	Southampton	22	German	German

(cont.)

Subject ID	Perception score	Accent importance	Accent self-rating	Pillai score devoicing	Pillai score merger	Comment
001	12	10	8	0.65864	0.23644	
002	13	8	8	0.70146	0.41621	
003	12	5	3	0.12946	0.48878	
004	10	9	7	0.69685	0.81401	
005	19	8	6	0.40327	0.42057	
006	12	8	7	0.51594	0.68138	
007	17	9	8	0.60731	0.66092	
008	14	9	7	0.72018	0.50950	
009	16	8	6	0.54683	0.50090	
010	18	6	6	0.59231	0.17188	
011	13	8	5	0.80580	0.28863	
012	19	10	7	0.17976	0.77327	
013	11	10	9	0.64051	0.89861	
014	8	7	5	0.54742	0.49725	

Subject ID	Gender	Age	Birthplace	Group	Place of stay	Length of stay (months)	Mother's nationality	Father's nationality
015	female	26	Düren	G-UK	Edinburgh	36	German	German
016	male	30	Königs-Wusterhausen	G-UK	Brighton	10	German	German
017	female	25	Ulm	G-UK	Lancaster	9	German	German
018	female	24	Freiburg	G-UK	St. Albans	9	German	German
019	female	26	Freudenstadt	G-UK	Birmingham	10	German	German
020	male	24	Bad Mergentheim	G-UK	Guildford	6	German	German
021	male	24	Tübingen	G-UK	Birmingham	10	German	German
022	female	22	Freiburg	G-UK	Isle of Wight	6	Czech	German
023	male	26	Freiburg	G-UK	London	9	German	German
024	male	29	Regensburg	G-US	Tennessee	11	German	German
025	female	25	Ludwigsburg	G-US	Kentucky	11	German	German
026	female	23	Müllheim	G-US	California	6	German	German
027	female	25	Freiburg	G-US	Michigan	10	German	German
028	male	26	Werl	G-US	New Jersey	6	German	German
029	male	27	Paderborn	G-UK	Aberdeen	12	German	German

(cont.)

Subject ID	**Perception score**	**Accent importance**	**Accent self-rating**	**Pillai score devoicing**	**Pillai score merger**	**Comment**
015	6	9	8	0.62051	0.86586	
016	18	4	4	0.15516	0.50827	
017	13	6	6	0.70569	0.63517	
018	20	10	8	0.46304	0.61129	
019	8	10	6	0.81843	0.73727	
020	14	2	8	0.70027	0.61835	
021	12	7	6	0.81427	0.35552	
022	4	10	9	0.59689	0.84559	
023	12	7	5	0.63708	0.82149	
024	13	10	8	0.83324	0.69695	
025	16	8	6	0.00679	0.23498	
026	5	7	6	0.32882	0.26536	
027	17	8	8	0.84070	0.55827	
028	7	9	7	0.34413	0.17996	
029	12	10	9	0.18202	0.96123	

Subject ID	Gender	Age	Birthplace	Group	Place of stay	Length of stay (months)	Mother's nationality	Father's nationality
030	female	20	Frankfurt	G-US	Arkansas	11	German	German
031	male	20	Viersen	G-US	Kentucky	6	German	German
032	male	26	Pfortzheim	G-US	Iowa	10	German	German
033	female	27	Haltingen	G-UK	Reading	6	German	German
034	male	30	Offenburg	G-UK	Dundee	5	German	German
035	female	23	Fürth	G-UK	Wellington	10	German	German
036	female	29	Ludwigsburg	G-US	Illinois	96	American	German
037	female	20	Jülich	G-UK	Reading	6	German	German
038	female	24	Schleswig	G-UK	Durham	10	German	German
039	female	24	Bodensee	G-UK	Guildford	5	German	German
040	female	24	Tübingen	G-UK	Birmingham	7	German	German
041	female	25	Heilbronn	G-UK	Glasgow	10	German	German
042	female	25	Offenburg	G-US	Minnesota	10	German	German
043	male	21	Rheinfelden	G-UK	Luton	6	English	German
044	male	28	Horb	G-US	Virginia	6	German	German

(cont.)

Subject ID	Perception score	Accent importance	Accent self-rating	Pillai score devoicing	Pillai score merger	Comment
030	13	9	4	0.26506	0.19800	
031	16	8	7	0.02277	0.34091	
032	23	7	3	0.18202	0.28153	
033	9	10	10	0.64178	0.85566	
034	18	3	3	0.29775	0.45452	
035	13	8	2	0.29998	0.28485	
036	10	10	10	0.76687	0.78258	
037	10	9	6	0.12795	0.68822	
038	14	9	7	0.25272	0.53505	
039	13	10	7	0.77682	0.76611	
040	10	9	8	0.70285	0.90640	
041	12	7	8	0.44425	0.80357	
042	9	8	3	0.54520	0.42721	
043	14	6	8	0.55528	0.84843	
044	5	6	3	0.67030	0.82241	

Subject ID	Gender	Age	Birthplace	Group	Place of stay	Length of stay (months)	Mother's nationality	Father's nationality
045	male	26	Neu-Ulm	G-US	Missouri	15	Canadian	American
046	female	26	Kirchheim-Teck	G-UK	Wellington	9	German	German
047	female	27	Freudenstadt	G-UK	Leeds	5	German	German
048	female	25	Heilbronn	G-UK	Suffolk	6	German	German
049	male	24	Freudenstadt	G-US	Maine	9	German	German
050	male	21	Schondort	G-US	Indiana	12	German	German
051	male	29	Siegen	G-US	Pennsylvania	48	French	German
052	female	21	Ostfildern	G-US	Massachussets	5	German	German
053	female	24	Würzburg	G-UK	Nottingham	10	English	German
054	female	26	Berlin	G-US	Massachussets	9	German	German
055	female	24	Ravensburg	G-UK	Jersey Channel Islands	5	German	German
056	female	21	Freiburg	G-US	Pennsylvania	18	German	German
057	male	20	München	G-US	New Mexico	18	German	German
058	female	23	Sindelfingen	G-US	Michigan	10	German	German
059	female	22	Bad Säckingen	G-UK	Manchester	48	German	English
060	male	30	Freiburg	G-UK	Brighton	6	German	Scottish

(cont.)

Subject ID	Perception score	Accent importance	Accent self-rating	Pillai score devoicing	Pillai score merger	Comment
045	15	9	9	0.15325	0.87795	
046	18	10	5	0.03735	0.32589	
047	22	8	7	0.30586	0.70724	
048	15	9	7	0.15671	0.66707	
049	10	10	8	0.51153	0.61720	
050	14	9	7	0.33674	0.15890	
051	12	7	5	0.71028	0.88797	
052	19	9	4	0.29220	0.46161	
053	10	10	9	0.65042	0.93262	
054	15	8	6	0.19621	0.37972	
055	12	10	6	0.70621	0.14972	
056	7	9	7	0.55282	0.73662	
057	16	9	6	0.39118	0.11514	
058	10	9	9	0.14801	0.49796	
059	14	7	5	0.67073	0.93288	
060	13	8	6	0.73981	0.73127	

A.2 – Control Group (Groups C-UK and C-US)

Subject ID	Gender	Age	Birthplace	Group	Mother's nationality	Father's nationality	Perception score	Accent importance	Accent self-rating	Comment
301	male	21	Rheinau	C-UK	German	German	10	8	2	
302	female	20	Rottweil	C-UK	German	German	12	9	4	
303	male	20	Spaichingen	C-UK	German	German	17	3	3	
304	male	21	Bietigheim	C-US	German	German	8	4	5	
305	female	20	Todtmoos	C-US	German	German	11	9	4	
306	male	21	Alpirsbach	C-US	German	German	21	8	7	

A.3 – British Native Speakers

Subject ID	Gender	Age	City	Mother's nationality	Father's nationality	Perception score	Comment
201	female	21	London	German	Malaysian	9	
202	female	23	Aberdeen	German	English	13	English accent
203	male	25	Luton	English	English	19	
204	male	21	Gloucester	English	English	6	
205	male	29	London	English	English	16	
206	female	21	Cardiff	Welsh	Iraqi	14	
207	female	23	Colchester	English	Irish	6	
208	female	22	Stuttgart	German	English	16	English-only household
209	female	21	Great Yarmouth	French	English	19	
210	male	27	Crewe	English	English	13	

A.4 – American Native Speakers

Subject ID	Gender	Age	State	Mother's nationality	Father's nationality	Perception score	Comment
101	female	26	Nebraska	American	American	10	
102	male	27	Virginia	American	American	10	
103	male	24	Pennsylvania	German	American	6	
104	male	25	Pennsylvania	American	American	13	
105	female	23	Georgia	American	American	11	
106	male	24	New Mexico	American	American	4	
107	male	21	Connecticut	Indian	American	16	
108	female	21	Indiana	American	American	6	
109	male	23	California	American	American	4	
110	male	22	Oregon	American	American	10	

Appendix B – Questionnaire and Consent Form

B.1 – Questionnaire

B.2 – Consent Form

B.1 – Questionnaire

Background Information

The information you provide here is confidential. It will only be viewed by the researcher.

Name:_____________________ Age:_______________

Nationality:_________________

Birth place:__________________ Mother tongue:__________________

Parents' birth countries / nationalities:____

1) In which English-speaking country/countries have you stayed for at least 1 month? Please give state/county and city and length of stay (in months):

____

____

1a) How old were you when you stayed there?______________________________________

1b) What was the reason you went there?____

2) How old were you when you started learning English?_____________________________

3) Have you ever lived together with an English-speaking person (here in Germany?)

☐ yes ☐ no

3a) If yes, for how long?____

3b) What nationality was this native speaker?_______________________________________

4) What nationality was/were your English teacher(s) at school? (Please include any private tutors you may have had)____

4b) Which variety/ies of English did your teacher(s) at school speak?____________________

5) Which variety of English do you speak?____

6) Towards which variety of English do you orientate yourself?________________________

7) How often do you use English ACTIVELY, e.g. in conversations, speaking on the phone, etc.:

☐ every day ☐ every other day ☐ several times per week ☐ once per week

☐ other:____________________(please specify)

8) Do you predominantly speak English with ☐ native speakers or ☐ non-native speakers?

8a) If native speakers, which variety do they speak?__________________________________

9) How often do you use English PASSIVELY, e.g. watching movies, listening to music, reading books, etc.:

☐ every day ☐ every other day ☐ several times per week ☐ once per week

☐ other:____________________(please specify)

10) Which variety of English do you have contact with in passive interactions (watching movies, listening to music, etc.)

☐ only American English ☐ only British English ☐ other:_________(please specify)

☐ a combination of varieties, namely:____
(please specify in percentages if you can, e.g. 60% American English, 40% British English)

11) How important is a native-like accent to you? 1 = not at all important, 10 = very important:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

12) How native-like do you judge your accent? 1 = not at all native-like, 10 = completely native-like:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

13) Which languages apart from German and English do you speak?

____

14) Other language-related information which might be relevant (impairment, bilingualism, etc.)

____

B.2 – Consent Form

Participant Information Sheet and Consent Form

Why do I do this study? – I am interested in your perception and production of English sounds. I need to collect data from English-speaking students to allow me to compare the results with native speakers.

What will participation involve? – This research involves filling in a short questionnaire about yourself, listening to tokens and stating which word you heard as well as completing a short experiment where you will read words and sentences and speak about your stay abroad.

How long will participation take? The entire procedure will last about 30 minutes.

As an informed participant of this experiment, I understand that:

1. My participation is voluntary and I may cease to take part in this experiment at any time, without penalty.
2. I can ask questions at any time before or during the experiment.
3. I am aware that I am being recorded and that the recording will be analysed by the researcher.
4. My name and information I give in the questionnaire will only be viewed by the researcher. In publications, first letters of my first name and surname will be used (e.g. A. M.).
5. All my questions about the study have been satisfactorily answered.

I have read and understood the above, and give consent to participate:

Participant's Signature:__________________________________ Date:__________

I have explained the above and answered all questions asked by the participant:

Researcher's Signature:__________________________________ Date:__________

Appendix C – Tasks

C.1 – Minimal Pairs

C.2 –Word List

C.3 – Sentences

C.4 – Text

C.1 – Minimal Pairs

trick	trig
slop	slob
bit	bid
peck	peg
nag	knack
lock	log
wed	wet
dead	debt
pick	pig
set	said
bet	bed
beck	beg
peck	peg
cob	cop
ebb	ep
lit	lid
bad	bat
had	hat

deb	dep
ad	at
tab	tap
wick	wig
cap	cab
back	bag
lip	lib
lack	lag
not	nod
rot	rod
hit	hid
got	God
dock	dog
rip	rib
clock	clog
bed	bet
mop	mob
crip	crib

C.2 – Word List

Test tokens in bold print

pin	pain	pen	pun	pan
beat	**bit**	**bid**	bitter	bidder
pot	**log**	**dog**	sleep	**slob**
had	hood	would	**bat**	**mat**
rip	**rib**	read	real	rude
not	note	**nod**	**lab**	**nag**
step	**back**	**bag**	**beg**	**bad**
clock	**dock**	short	**God**	for
pig	**peg**	**peck**	**pick**	**pack**
heat	**hit**	**hid**	**web**	**ebb**
leap	**lip**	lead	**lab**	**lib**
treat	**step**	weak	**wig**	**brick**
bed	**bet**	**set**	**said**	**speck**
tap	**cap**	**top**	**hop**	**cob**

C.3 – Sentences

Test tokens in bold print

Is she wearing a **wig**? Maybe someone should **knit** her a **hat**.

The **trick** is to take one **step** at a time. Come on, give it a **shot**.

The **pig** is actually a very intelligent animal, and no other mammal but the **bat** can fly.

Linda, can you move your chair a **bit**? My **back** is already against the **brick** wall.

The highest **bid** is only at 2 dollars, so don’t **brag**.

You’re such a **slob**. At least **mop** the floor!

This **ad** really **hit** home.

She will be a **deb** this year.

It’s not so easy to find the perfect **red lipstick** - you need to know your **lip** shape.

Don’t shave over the sink or the drain will **clog**.

Peter says I **hog** the sheets at night. But he hogs the whole **bed**!

God formed Eve from Adam’s **rib**.

I’ve **got** a **bet** going with Jim about who can teach our **dog** some tricks.

Come away from the fountain – you’ll get **wet**!

You could be **dead** 10 years from now.

Don’t **beg**, I’m not giving you any candy.

Matt really likes Anna, and all she **did** was give him a **peck** on the cheek.

Oh! Where did you **get** this nice **bag**?

Tap on wood three times for good luck.

Does your son go to **prep** school? – **Yep**, he does.

Could you call me a **cab**, please? My **clock** says it’s time to leave.

There is a **lack** of women in politics.

Don’t just **nod**. Say “Yes, Sir!”

Always **lock** the doors. This **block** isn’t safe at night.

You need a good fishing **rod** to catch **big** fish.

I got pulled over by a **cop** last week for running the light.

They announced a new flash**mob** on facebook today!

Mind the **gap**. It can make you **trip** quite easily.

Can I get the **tab**, please?

For all I care you can **rot** in hell. I could **stab** you for what you did to me.

I almost broke my **leg** last week when the streets suddenly iced up.

I’ve got this croak in my **neck** that I just can’t seem to get **rid** of.

Hannah bought a light blue **crib** and now she found out she’s having a girl!

Ebb and flow are influenced by the moon.

Nothing can be deleted once it’s been on the world wide **web**.

It filled me with **dread** when she started to **sob**.

You’re such a **squib**. **Snap** out of it!

Can I have my sandwich with an **egg** on **top?**

Can I write you a **cheque**?

The **flag** of the anarchist movement is **black**.

I’m **glad** you’re not **sad** anymore.

C.4 – Text

Test tokens in bold print

This story is about a little bear and a little tiger, who live in a nice house. The little bear goes fishing and the little tiger does the cooking. One day, Little Tiger said: “I have a **job** for you. Go down to the river and fish us a trout for supper!”

So Little Bear took his fishing **rod** and set out. He **sat** down on the last **step** of the stairs which led to the river and threw out his fishing rod. Not even a minute passed before his fishing rod twitched.

“That was **quick,**” thought Little Bear. He pulled the fishing rod out of the water. On it hung a small green **frog**.

“Oh, that’s too **bad**. I don’t like frogs,” **said** Little Bear and threw the frog back into the water. After some time the fishing rod twitched again. This time, there was a **crab** hanging on it.

“I can’t believe it!” said Little Bear. “I have to catch a trout! They are Little Tiger’s favorites. If I come home without one, he’ll think I spent the day lazying around in the sun. He says I’m **glib** at finding excuses anyway.”

Suddenly, he felt a **tap** on his shoulder.

“Who is it? I’m in the **mid** of catching my supper!” cried Little Bear and turned around. A **black** spider sat in a **cobweb** which hung down from a small **twig** of a nearby tree. She smiled.

"You have to try just a little **bit** harder," said the spider. "You have to believe in yourself. Trust me, I know all about fishing."

"I think you just like to **brag,**" said Little Bear. "Spiders don't even like fish."

"Oh don't **fret,**" said the spider, annoyed that Little Bear didn't believe her. "Just try again." And gone she was.

"Thanks for the great **tip,**" said Little Bear sarcastically when she was out of ear**shot**. But he did try again. He had to sit by the river for a long time. He was tired and hungry and slowly lost interest. But then, when he had almost given up hope, his fishing rod twitched again.

"A trout!" cried Little Bear, almost in **shock**. It was really a trout, Little Tiger's favorite! Little Bear was happy he was finally able to go home to Little Tiger and show him his catch. It was about time, too. The sun was already setting and his left **leg** was sore and his **neck** hurt.

Down by the river the spider smiled as she watched Little Bear **hop** all the way home.

Text: frei nach Janosch "Oh, wie schön ist Panama".

© Janosch-Motive mit freundlicher Genehmigung der Janosch film & medien AG.

Appendix D – Vowel Plots

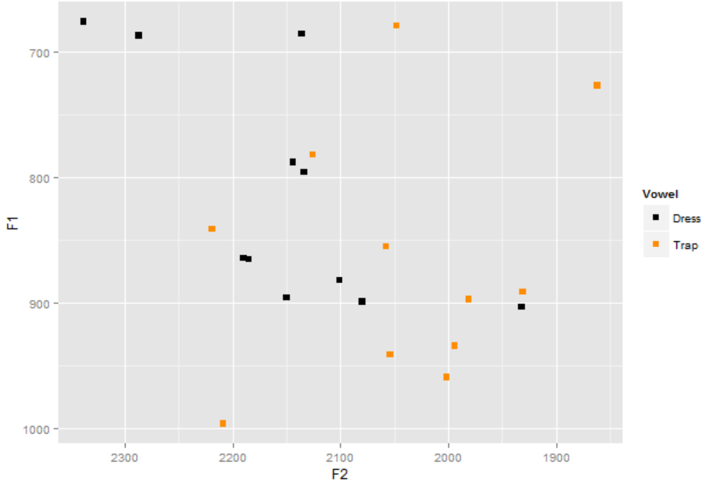
D.1 – German Learners (Groups G-UK and G-US)

D.2 – Control Group (Groups C-UK and C-US)

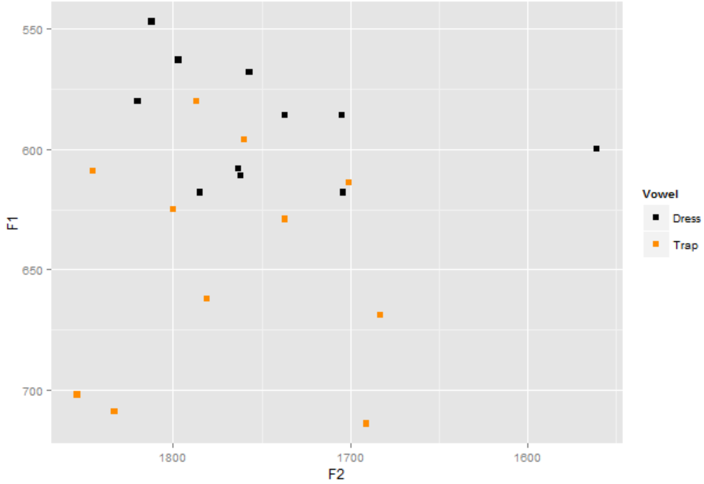
D.3 – British Native Speakers

D.4 – American Native Speakers

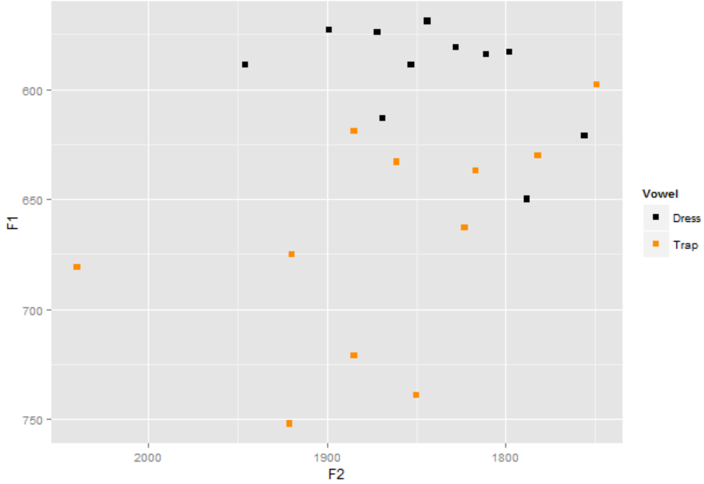
D.1 – German Learners (Groups G-UK and G-US)



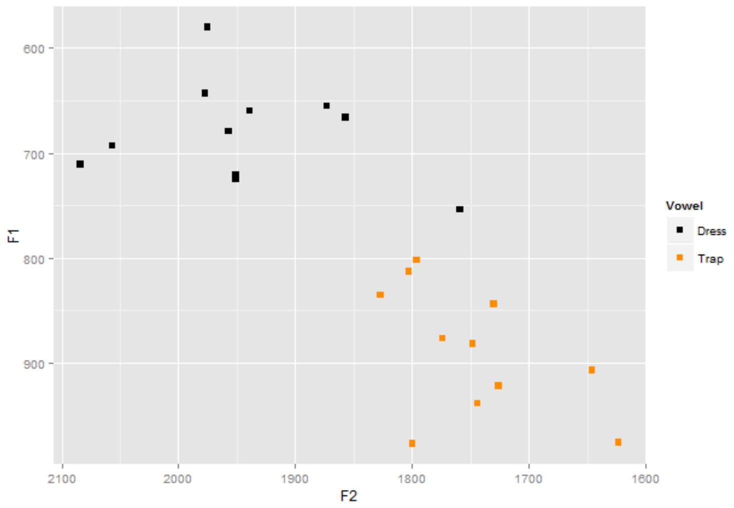
D.1–1. Vowel plot of speaker #001, corresponding to a Pillai score of .24.



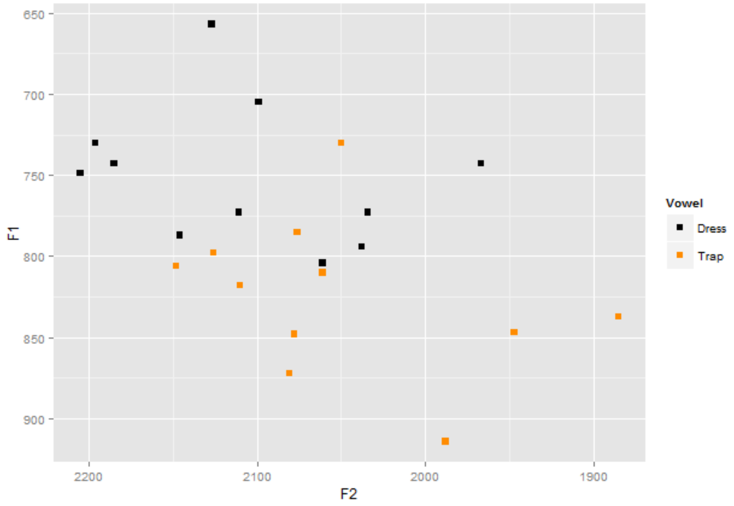
D.1–2. Vowel plot of speaker #002, corresponding to a Pillai score of .42.



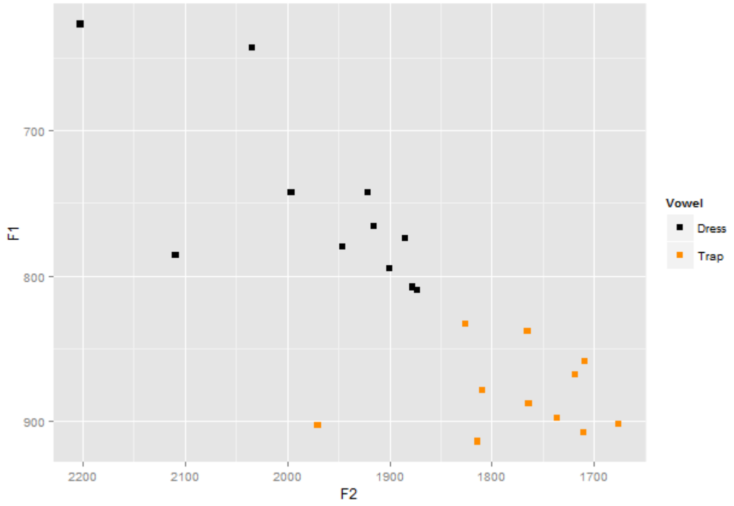
D.1–2. Vowel plot of speaker #003, corresponding to a Pillai score of .49.



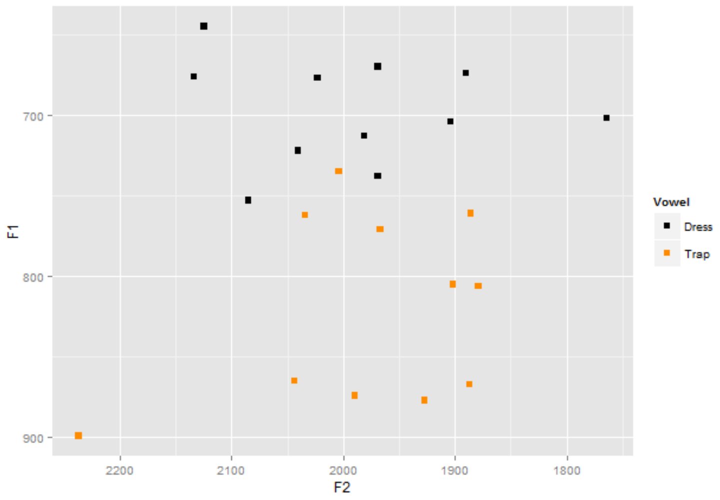
D.1–2. Vowel plot of speaker #004, corresponding to a Pillai score of .81.



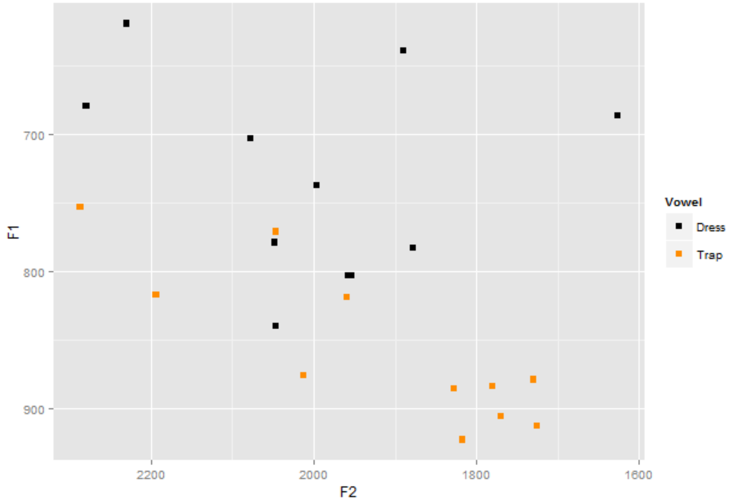
D.1–5. Vowel plot of speaker #005, corresponding to a Pillai score of .42.



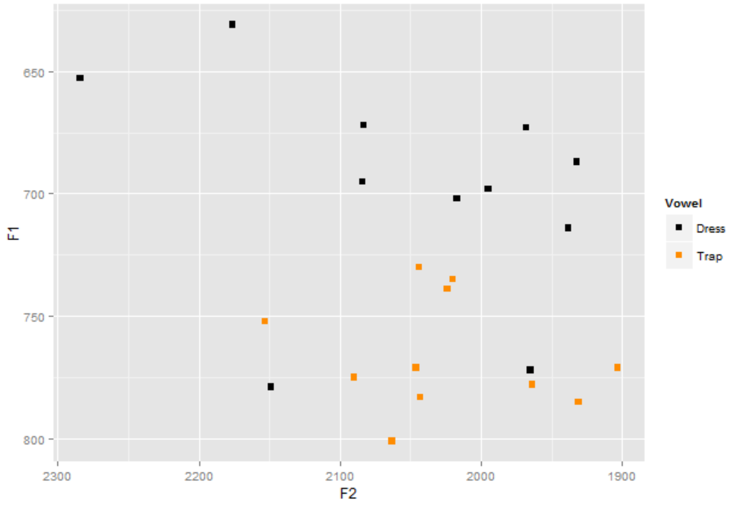
D.1–6. Vowel plot of speaker #006, corresponding to a Pillai score of .68.



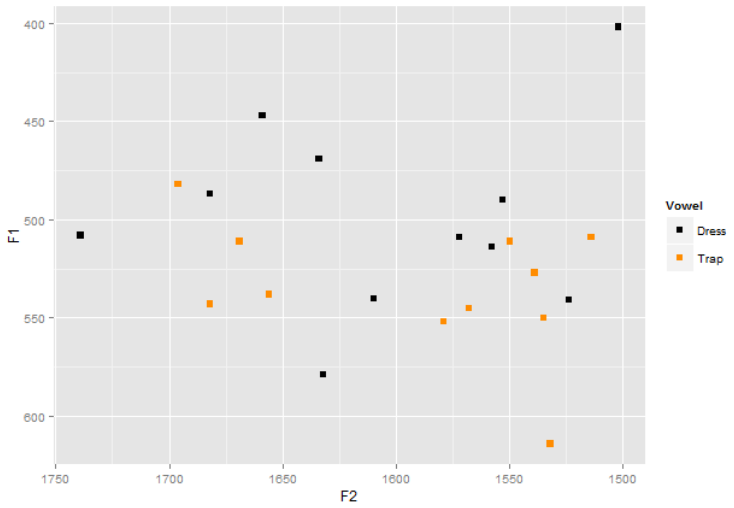
D.1–7. Vowel plot of speaker #007, corresponding to a Pillai score of .66.



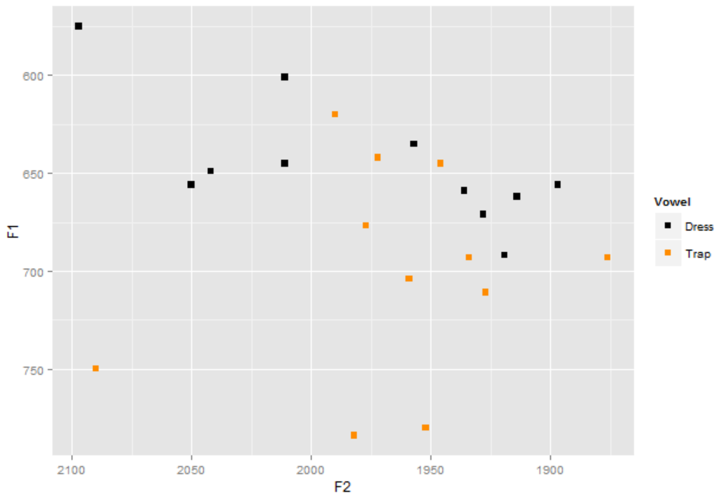
D.1–8. Vowel plot of speaker #008, corresponding to a Pillai score of .51.



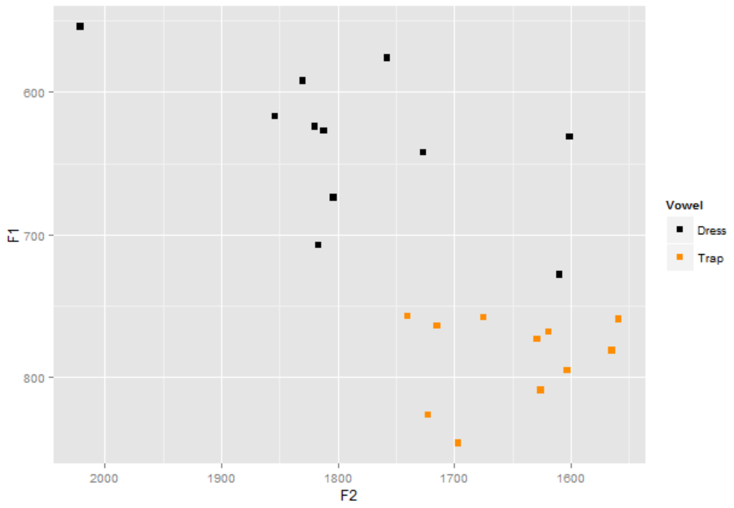
D.1–9. Vowel plot of speaker #009, corresponding to a Pillai score of .51.



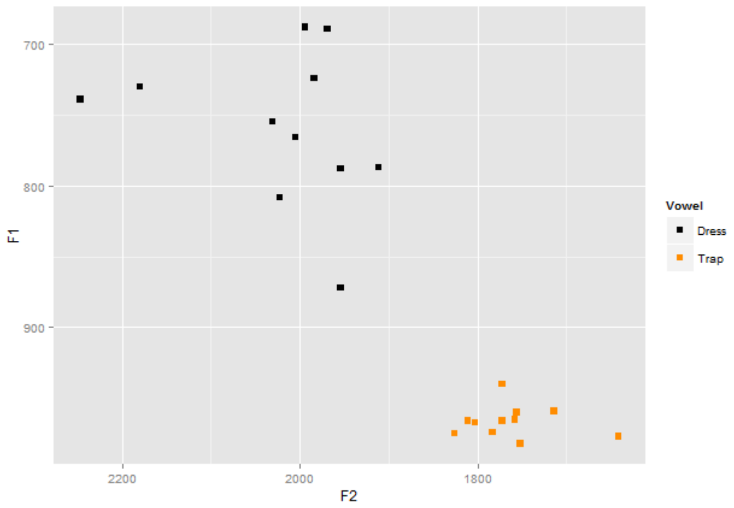
D.1–10. Vowel plot of speaker #010, corresponding to a Pillai score of .17.



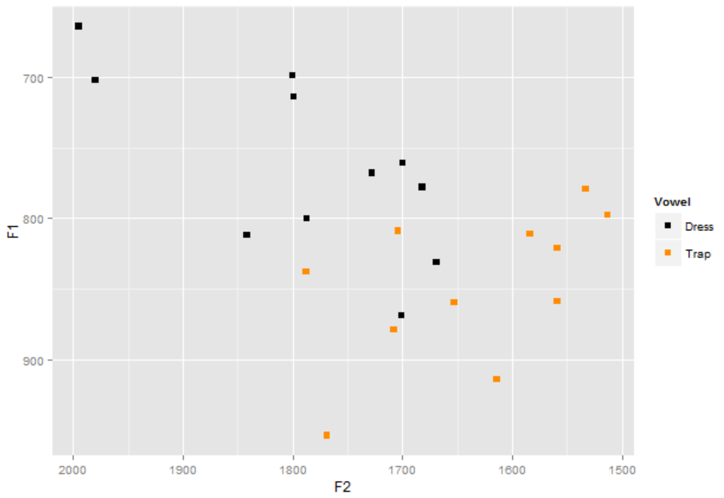
D.1–11. Vowel plot of speaker #011, corresponding to a Pillai score of .29.



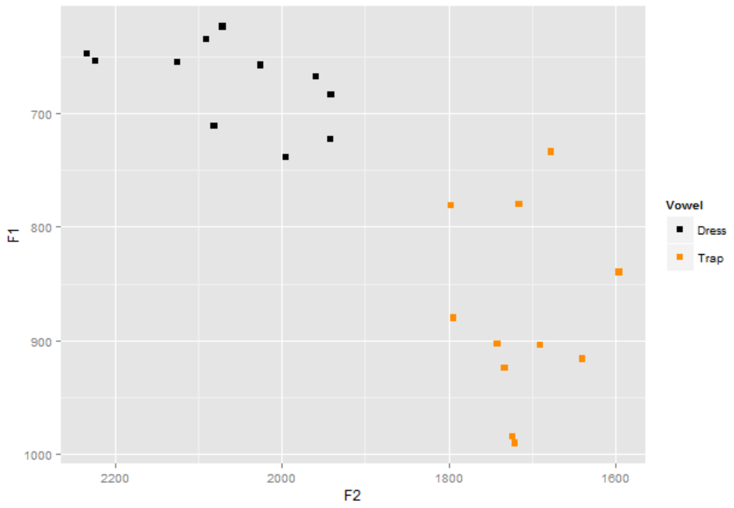
D.1–12. Vowel plot of speaker #012, corresponding to a Pillai score of .77.



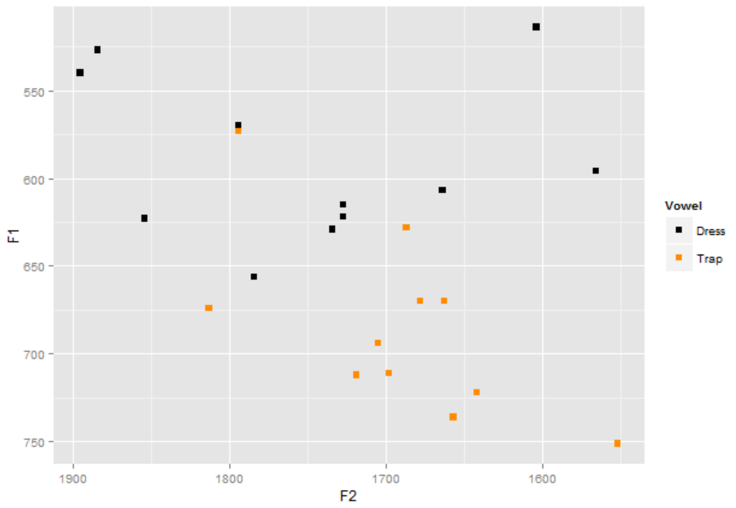
D.1–13. Vowel plot of speaker #013, corresponding to a Pillai score of .90.



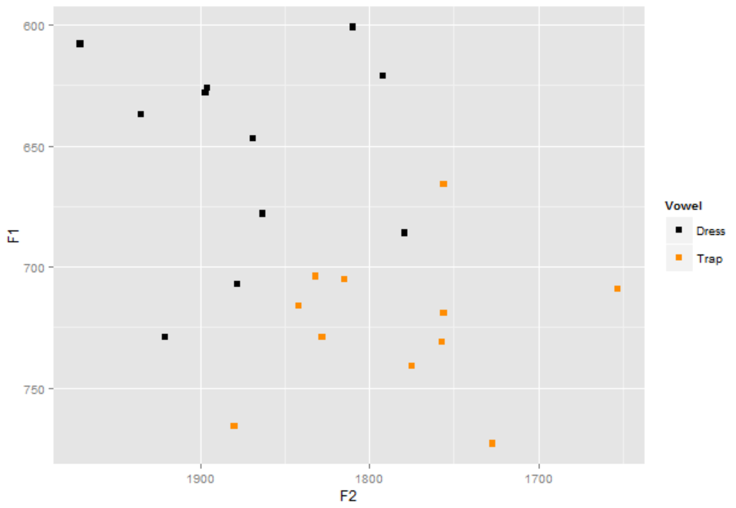
D.1–14. Vowel plot of speaker #014, corresponding to a Pillai score of .50.



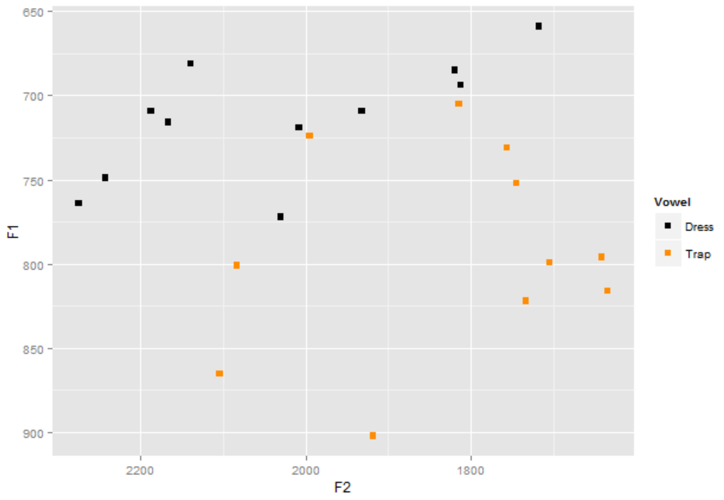
D.1–15. Vowel plot of speaker #015, corresponding to a Pillai score of .87.



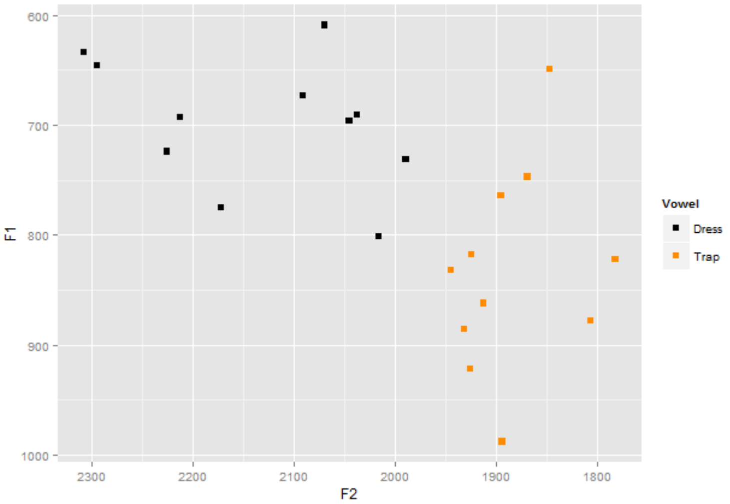
D.1–16. Vowel plot of speaker #016, corresponding to a Pillai score of .51.



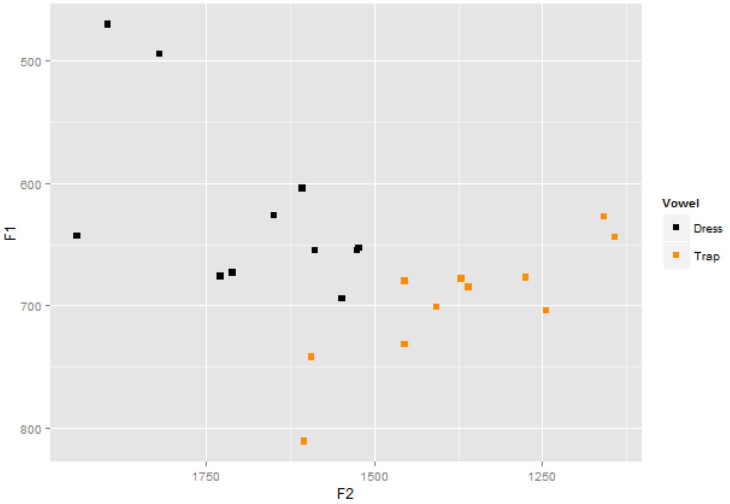
D.1–17. Vowel plot of speaker #017, corresponding to a Pillai score of .64.



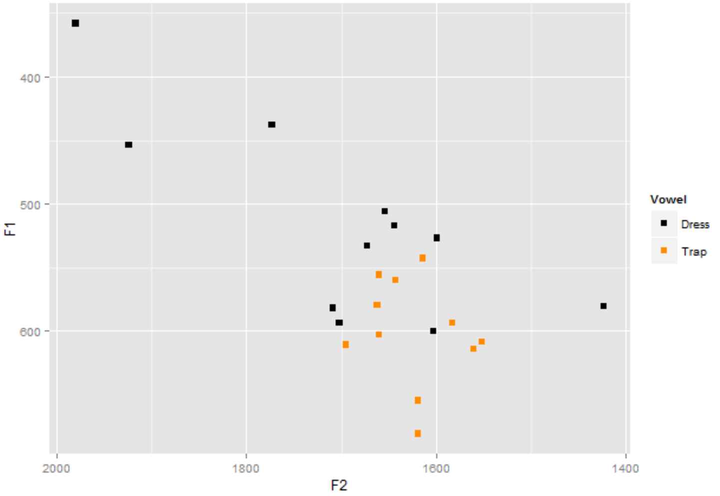
D.1–18. Vowel plot of speaker #002, corresponding to a Pillai score of .61.



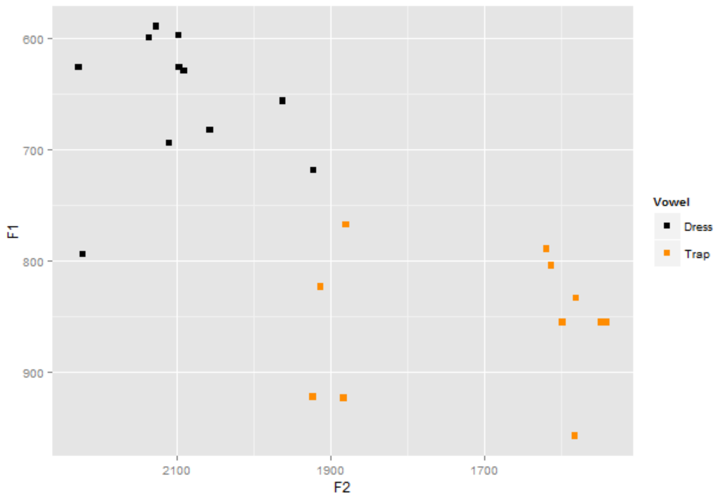
D.1–19. Vowel plot of speaker #019, corresponding to a Pillai score of .74.



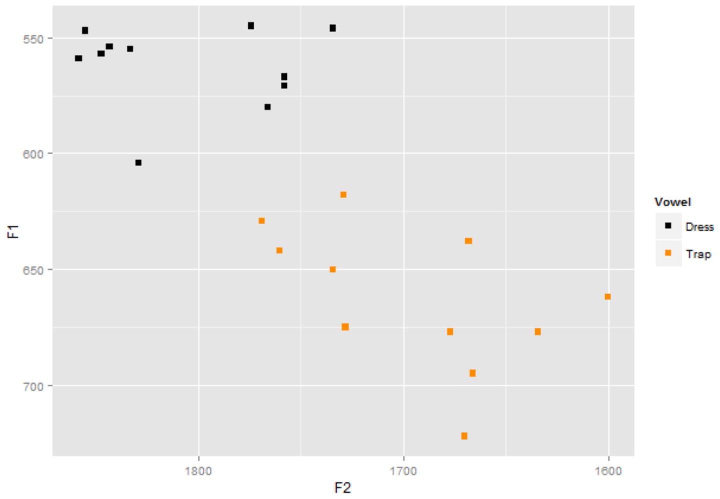
D.1–20. Vowel plot of speaker #020, corresponding to a Pillai score of .62.



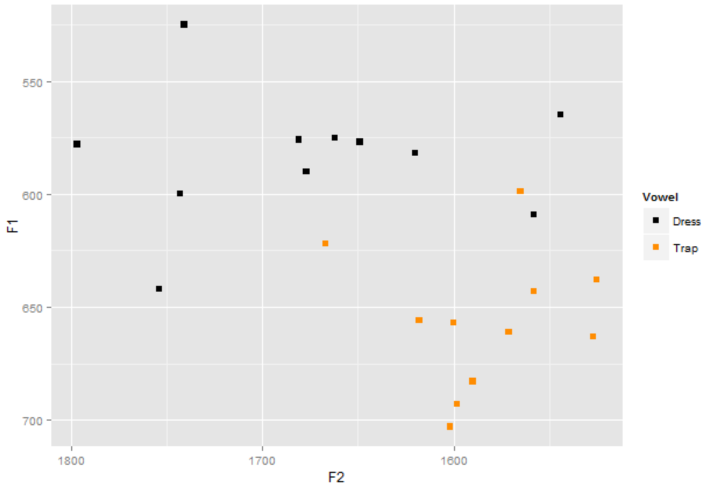
D.1–21. Vowel plot of speaker #002, corresponding to a Pillai score of .36.



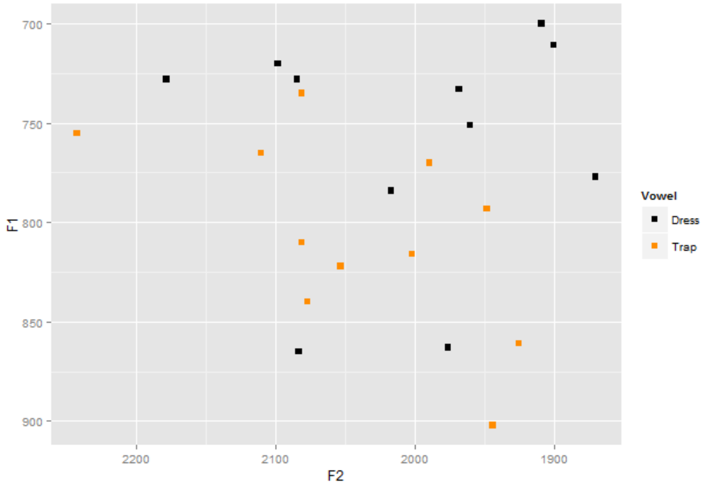
D.1–22. Vowel plot of speaker #002, corresponding to a Pillai score of .85.



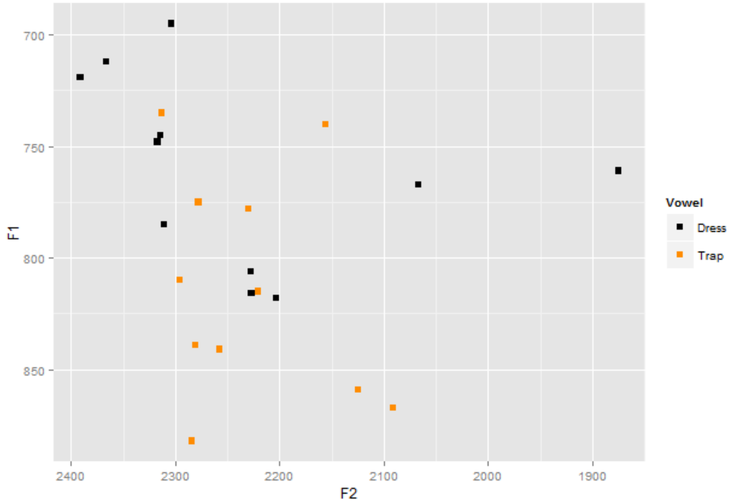
D.1–23. Vowel plot of speaker #023, corresponding to a Pillai score of .82.



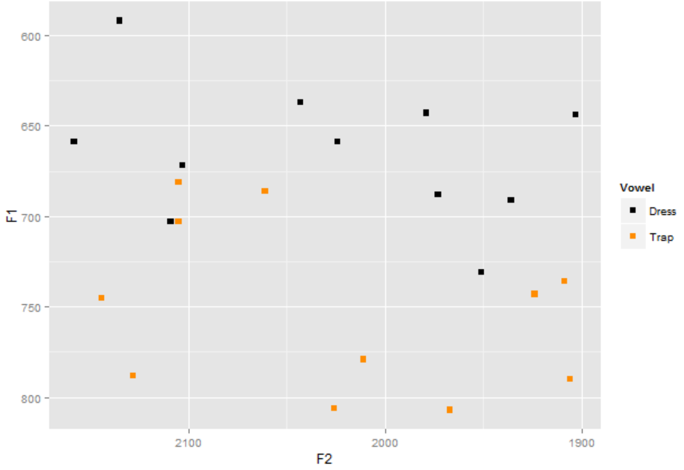
D.1–24. Vowel plot of speaker #024, corresponding to a Pillai score of .70.



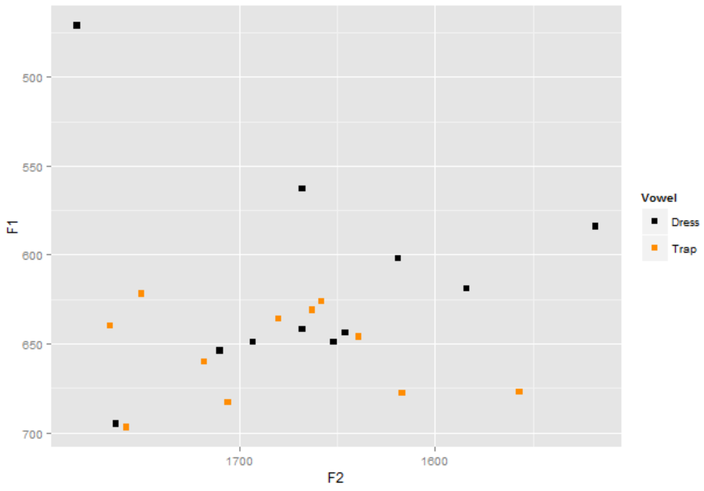
D.1–25. Vowel plot of speaker #025, corresponding to a Pillai score of .23.



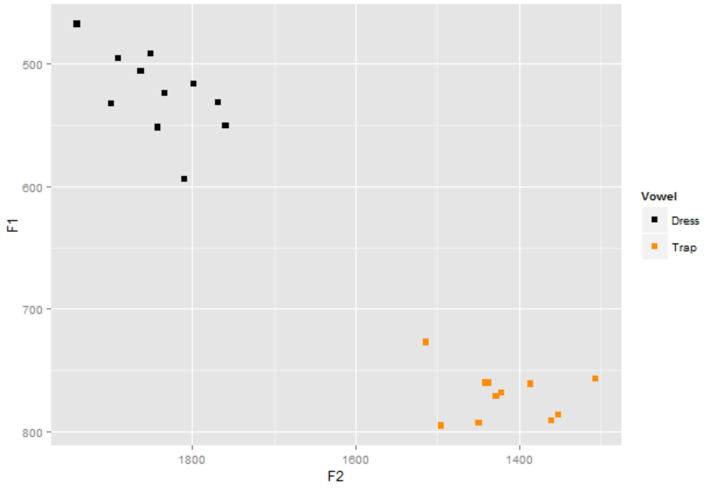
D.1–26. Vowel plot of speaker #026, corresponding to a Pillai score of .27.



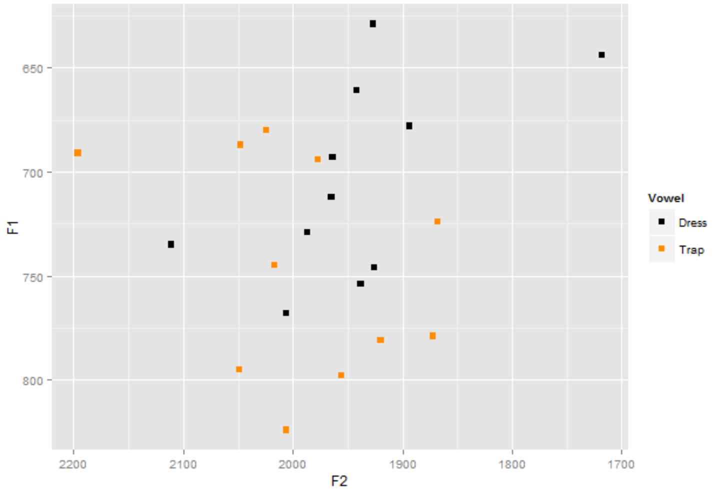
D.1–27. Vowel plot of speaker #027, corresponding to a Pillai score of .56.



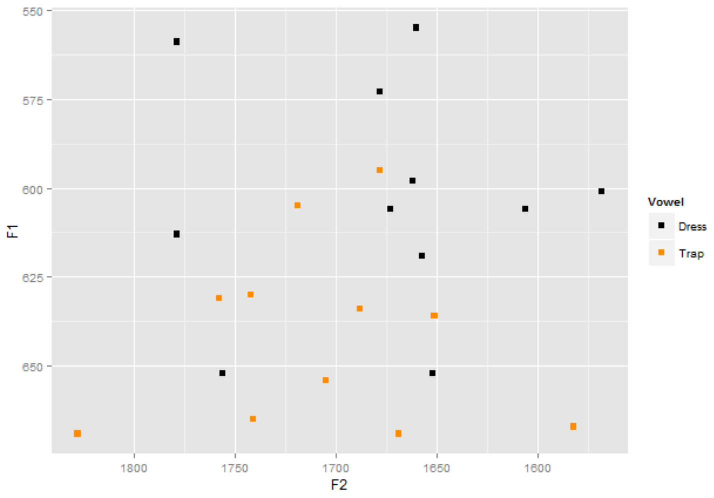
D.1–28. Vowel plot of speaker #028, corresponding to a Pillai score of .18.



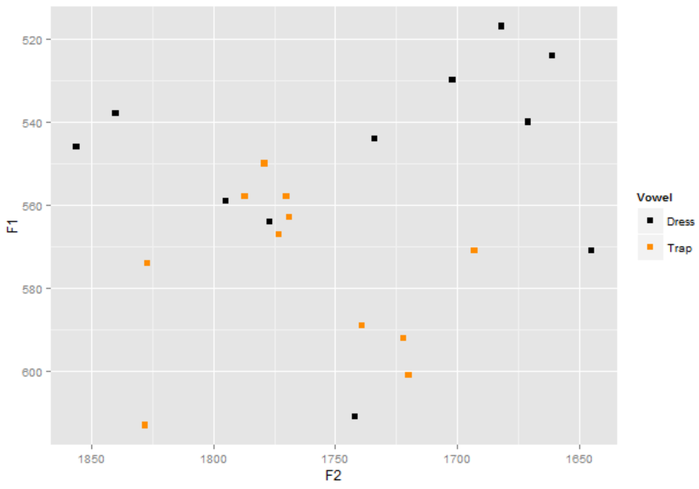
D.1–29. Vowel plot of speaker #029, corresponding to a Pillai score of .96.



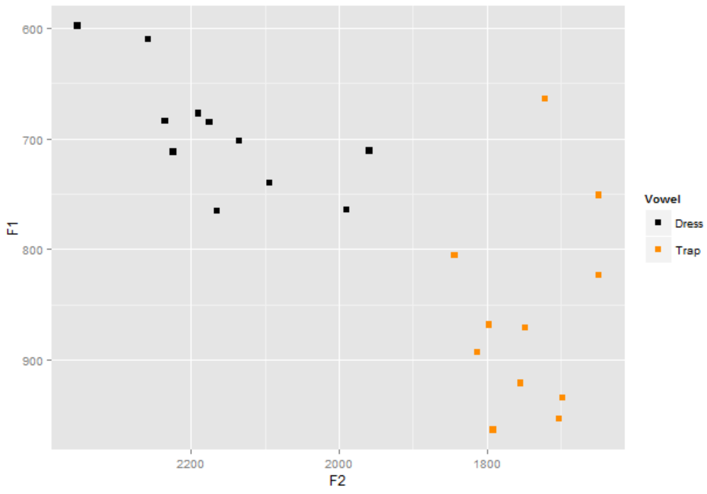
D.1–30. Vowel plot of speaker #030, corresponding to a Pillai score of .20.



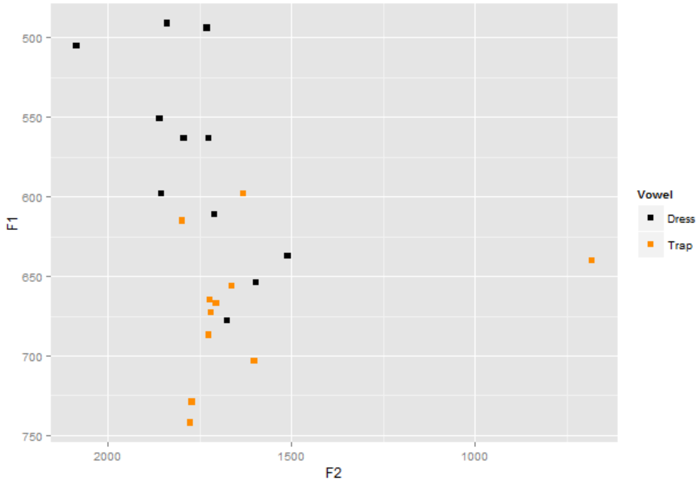
D.1–31. Vowel plot of speaker #031, corresponding to a Pillai score of .34.



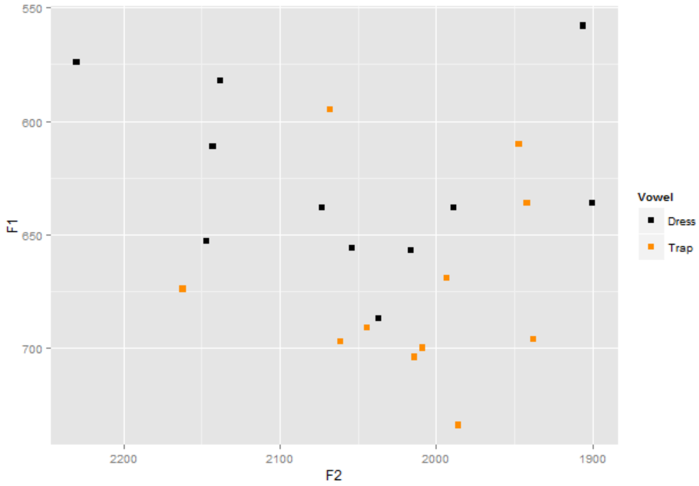
D.1–32. Vowel plot of speaker #0320, corresponding to a Pillai score of .28.



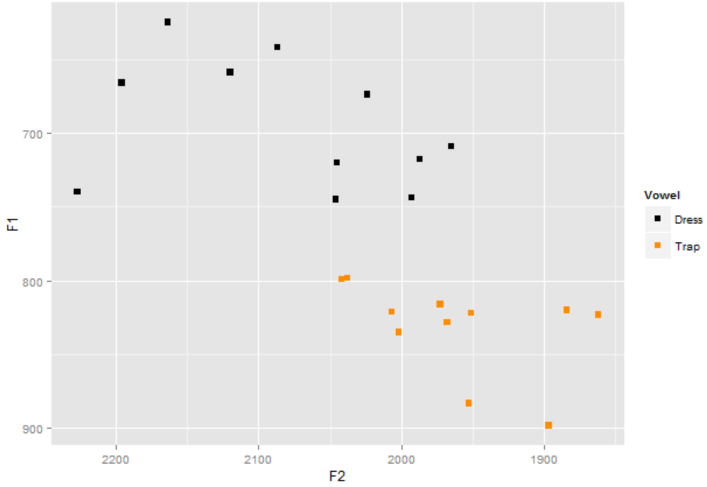
D.1–33. Vowel plot of speaker #033, corresponding to a Pillai score of .86.



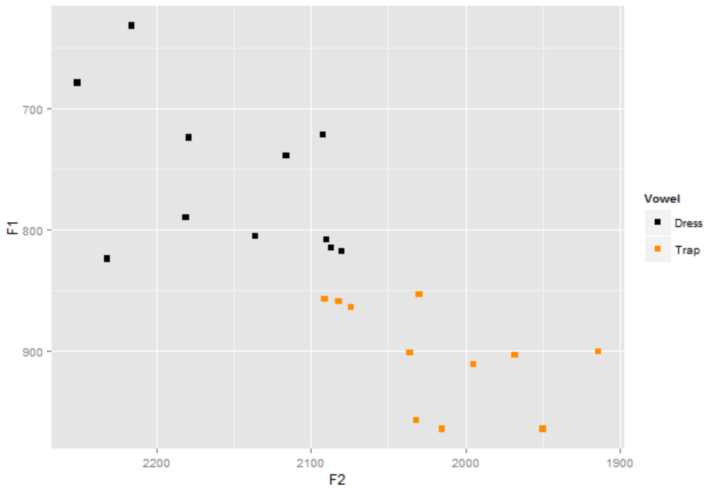
D.1–34. Vowel plot of speaker #034, corresponding to a Pillai score of .45.



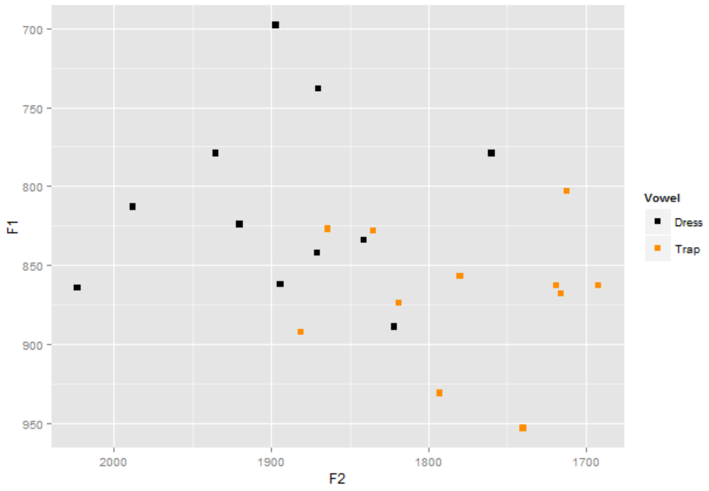
D.1–35. Vowel plot of speaker #035, corresponding to a Pillai score of .28.



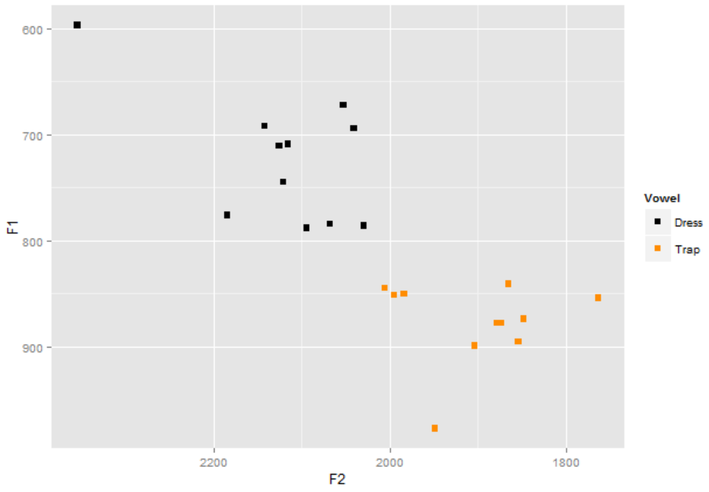
D.1–36. Vowel plot of speaker #036, corresponding to a Pillai score of .78.



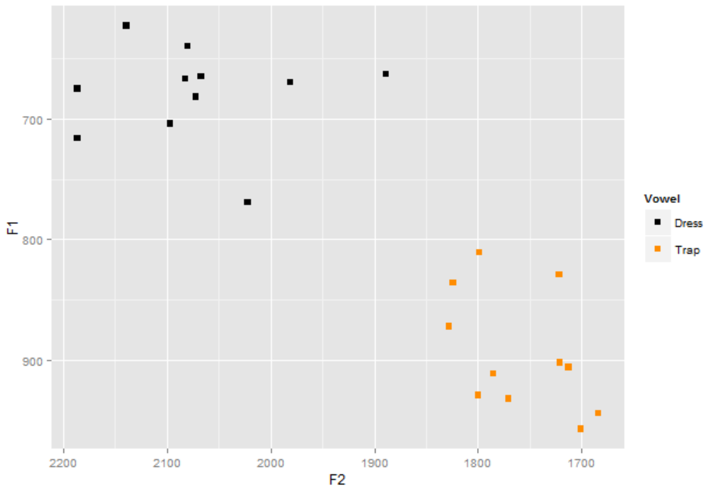
D.1–37. Vowel plot of speaker #037, corresponding to a Pillai score of .69.



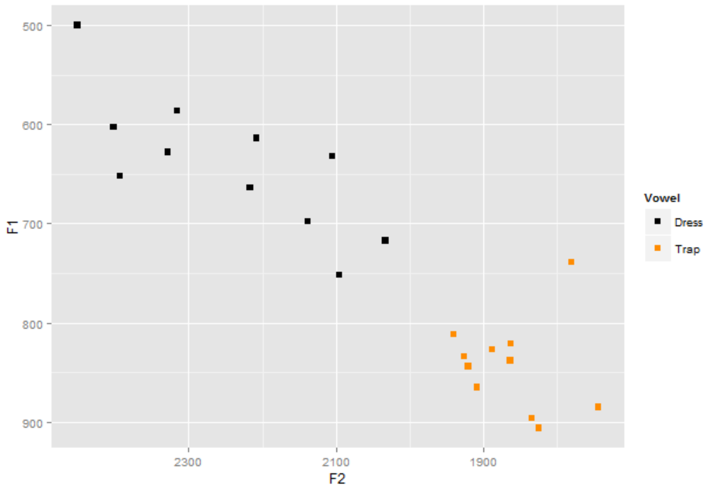
D.1–38. Vowel plot of speaker #038, corresponding to a Pillai score of .53.



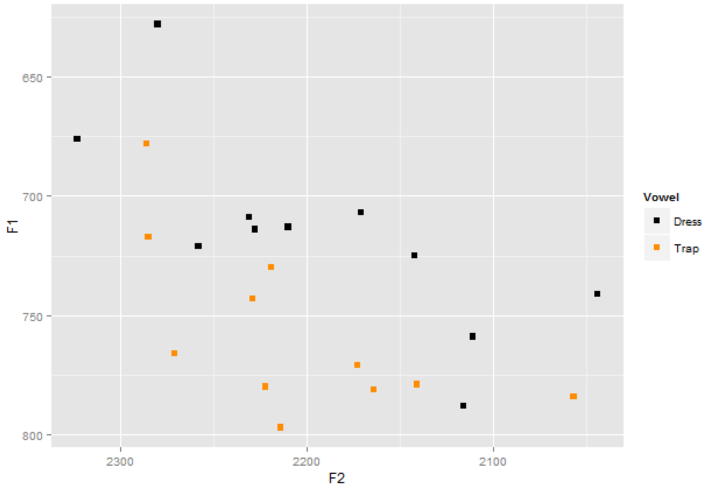
D.1–39. Vowel plot of speaker #039, corresponding to a Pillai score of .77.



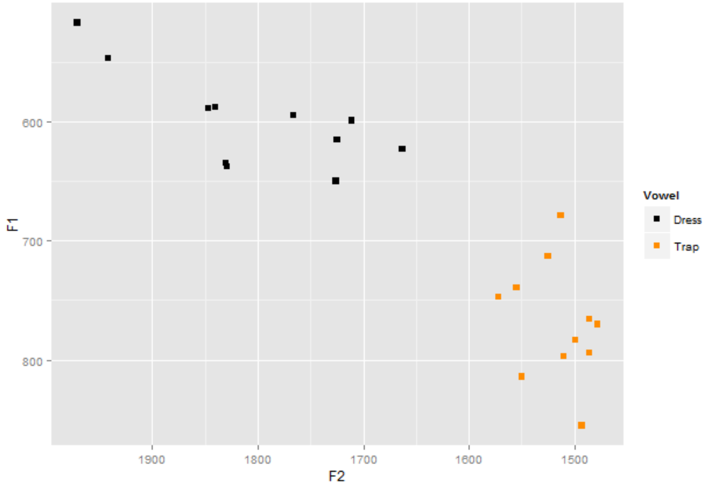
D.1–40. Vowel plot of speaker #040, corresponding to a Pillai score of .91.



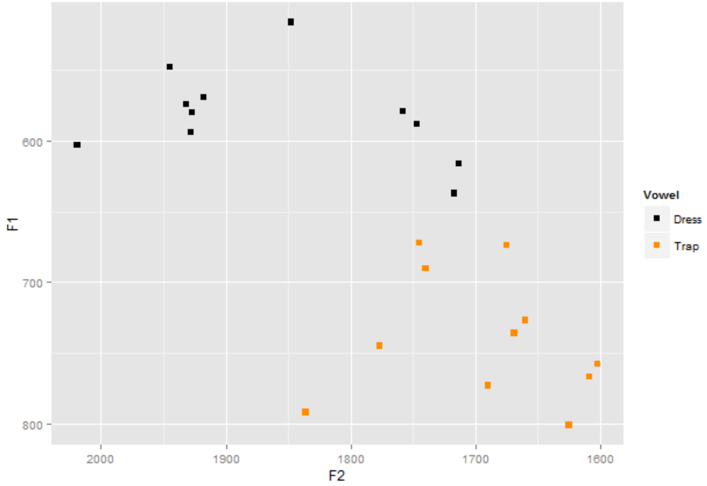
D.1–41. Vowel plot of speaker #041, corresponding to a Pillai score of .80.



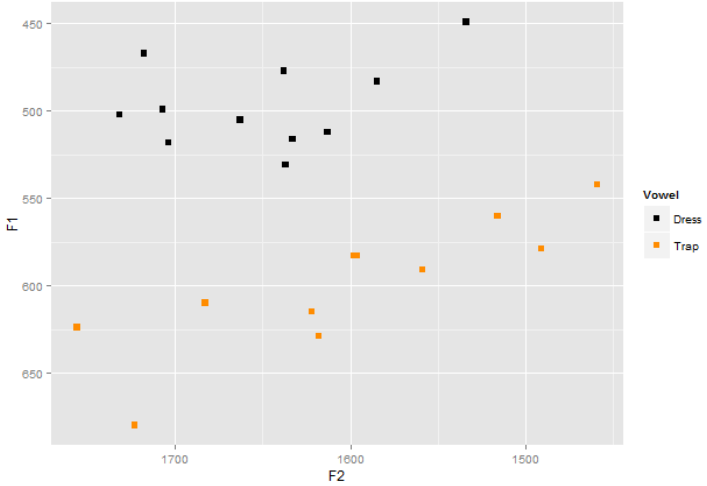
D.1–42. Vowel plot of speaker #042, corresponding to a Pillai score of .43.



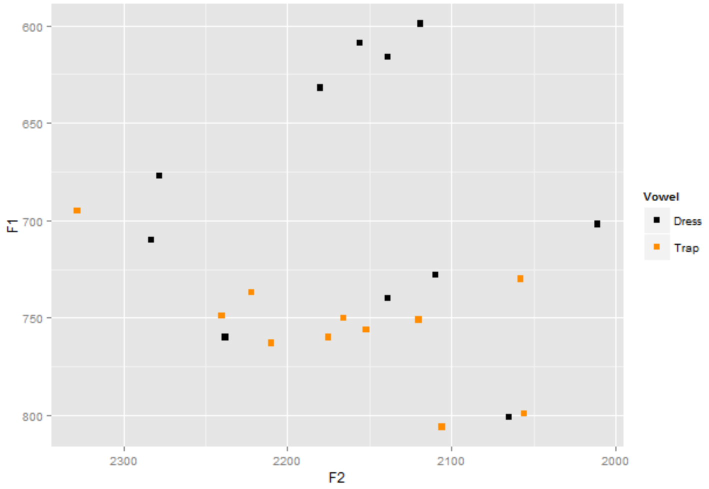
D.1–43. Vowel plot of speaker #043, corresponding to a Pillai score of .85.



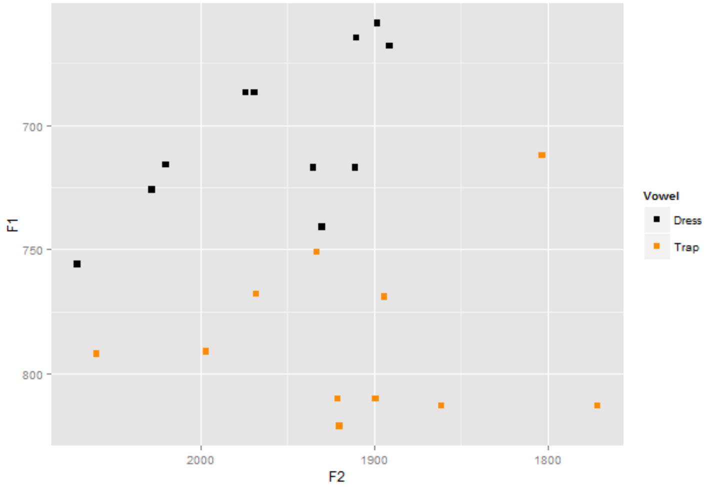
D.1–44. Vowel plot of speaker #044, corresponding to a Pillai score of .82.



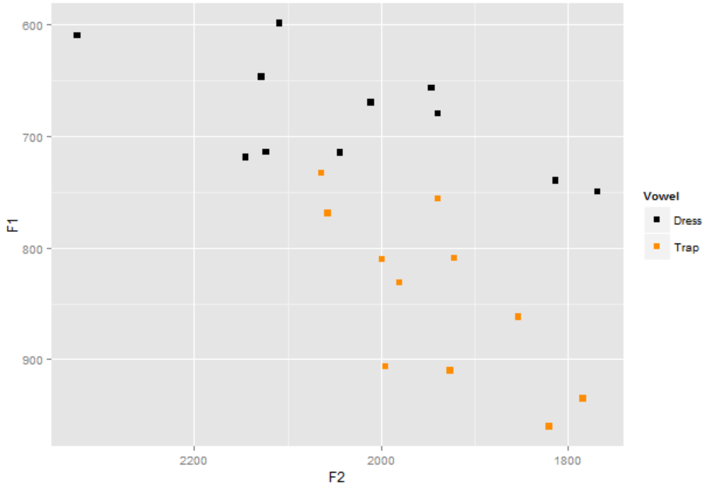
D.1–45. Vowel plot of speaker #045, corresponding to a Pillai score of .88.



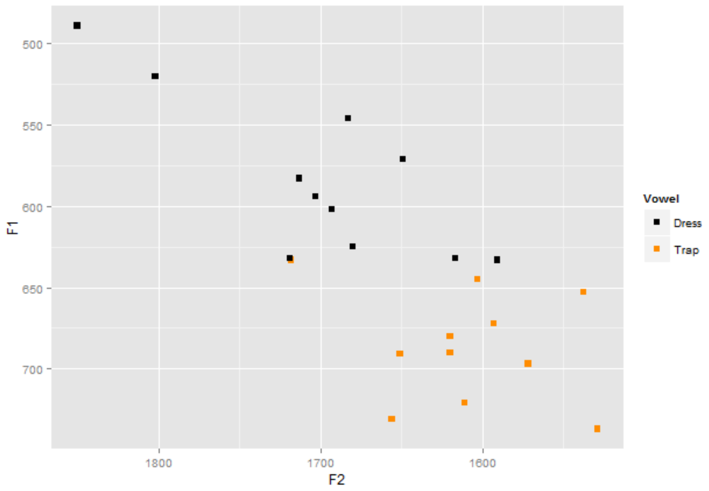
D.1–46. Vowel plot of speaker #046, corresponding to a Pillai score of .33.



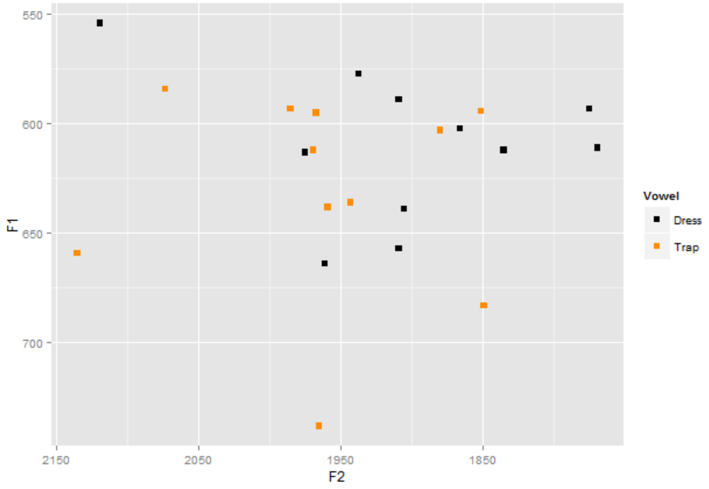
D.1–47. Vowel plot of speaker #047, corresponding to a Pillai score of .71.



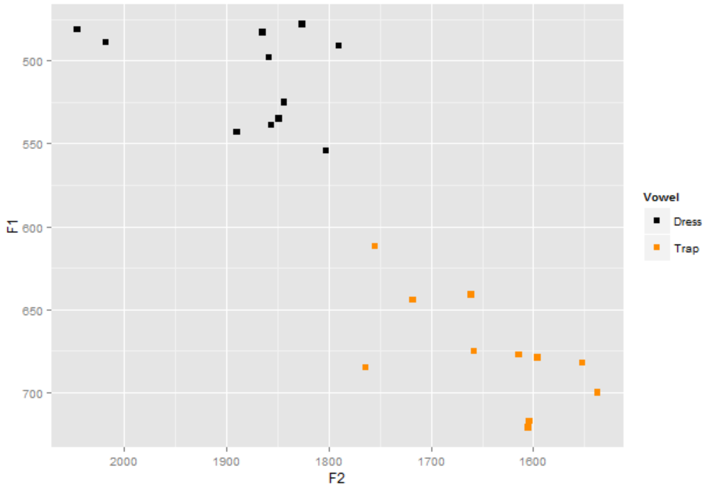
D.1–48. Vowel plot of speaker #048, corresponding to a Pillai score of .67.



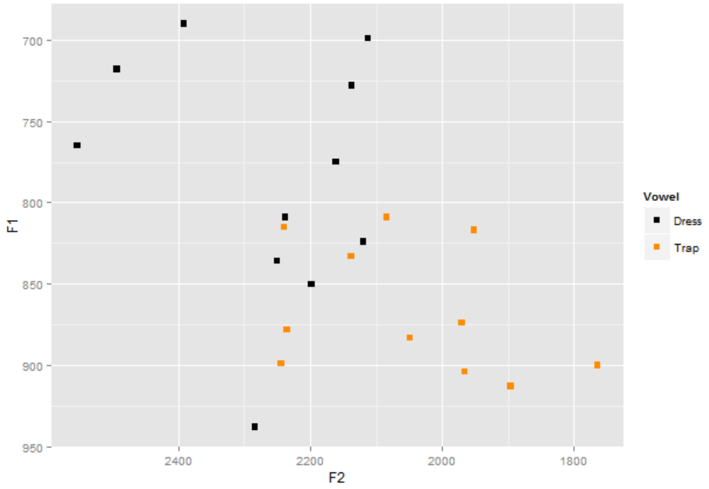
D.1–49. Vowel plot of speaker #049, corresponding to a Pillai score of .61.



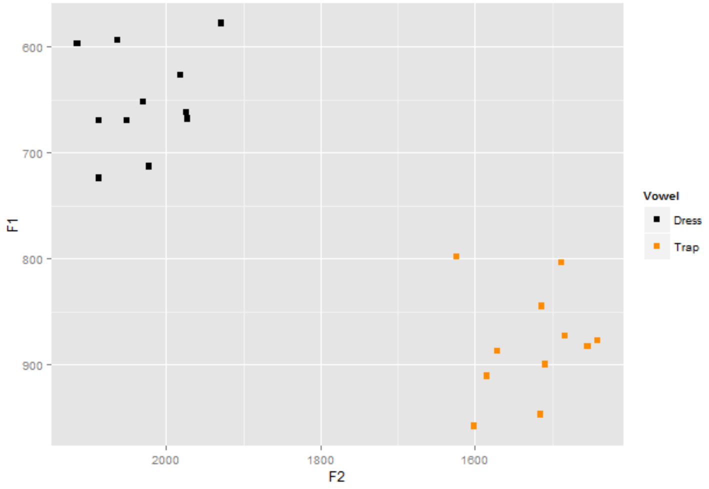
D.1–50. Vowel plot of speaker #050, corresponding to a Pillai score of .16.



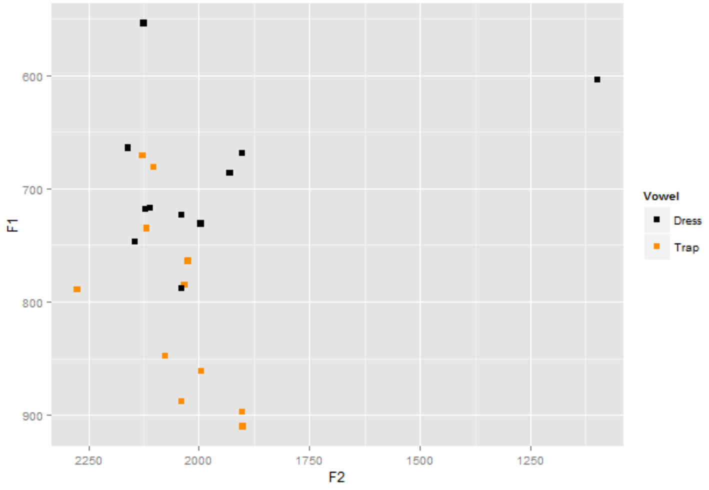
D.1–51. Vowel plot of speaker #051, corresponding to a Pillai score of .89.



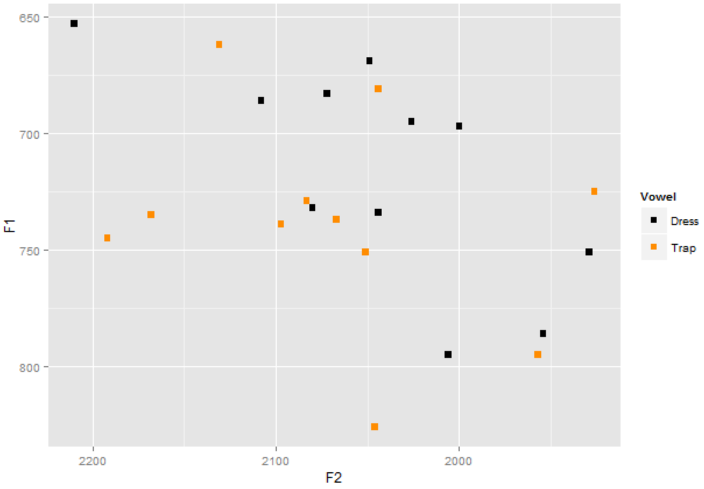
D.1–52. Vowel plot of speaker #052, corresponding to a Pillai score of .46.



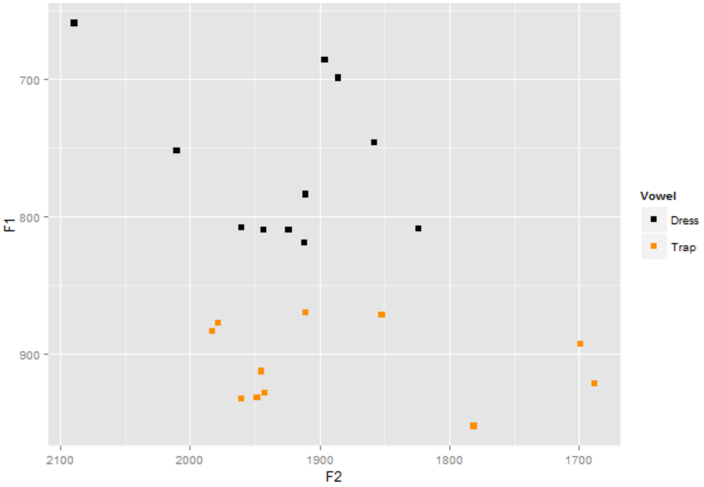
D.1–53. Vowel plot of speaker #053, corresponding to a Pillai score of .93.



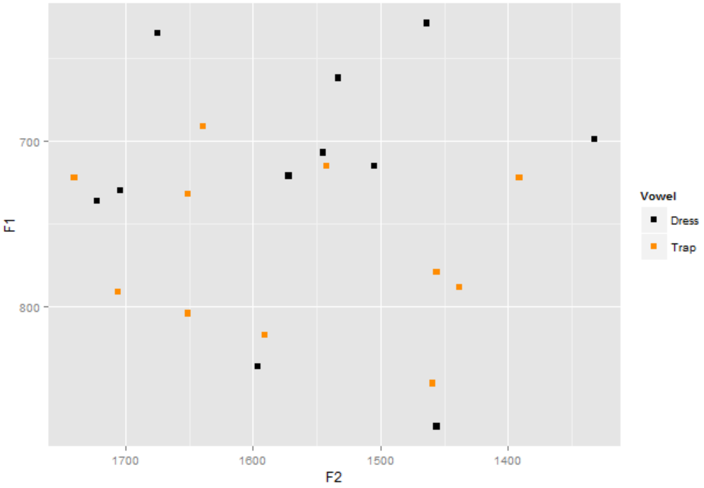
D.1–54. Vowel plot of speaker #054, corresponding to a Pillai score of .38.



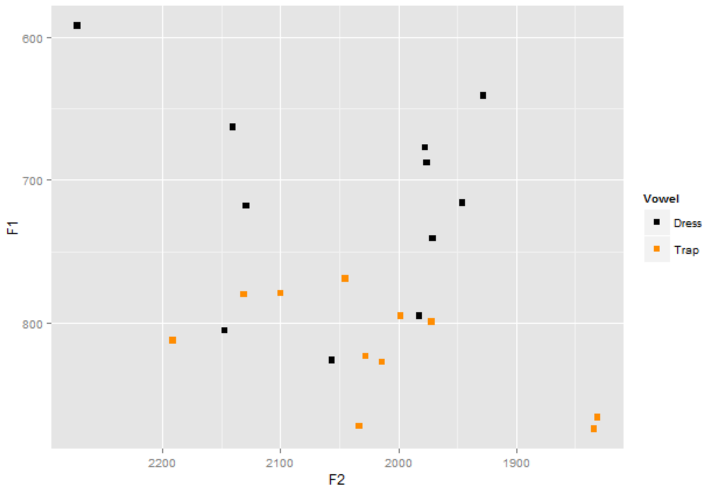
D.1–55. Vowel plot of speaker #055, corresponding to a Pillai score of .15.



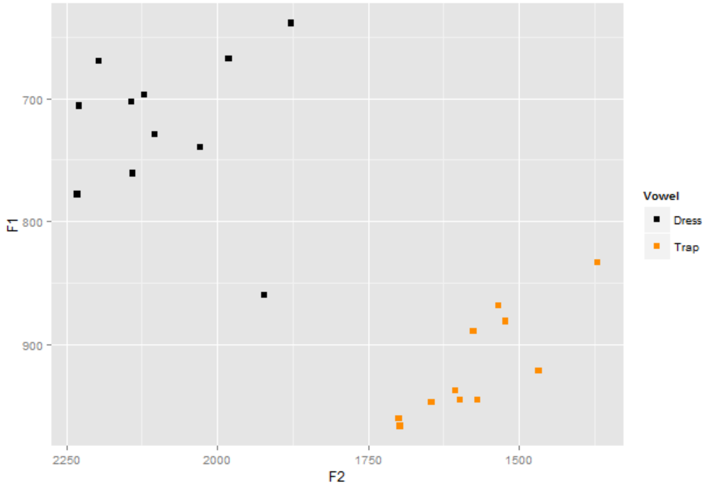
D.1–56. Vowel plot of speaker #056, corresponding to a Pillai score of .74.



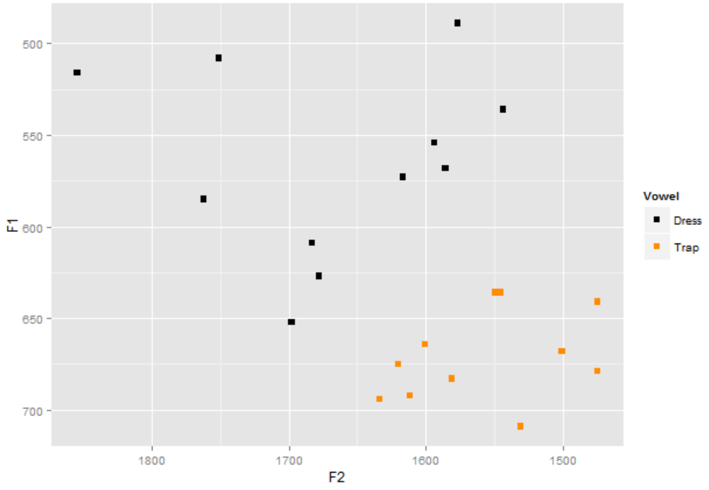
D.1–57. Vowel plot of speaker #057, corresponding to a Pillai score of .12.



D.1–58. Vowel plot of speaker #058, corresponding to a Pillai score of .50.

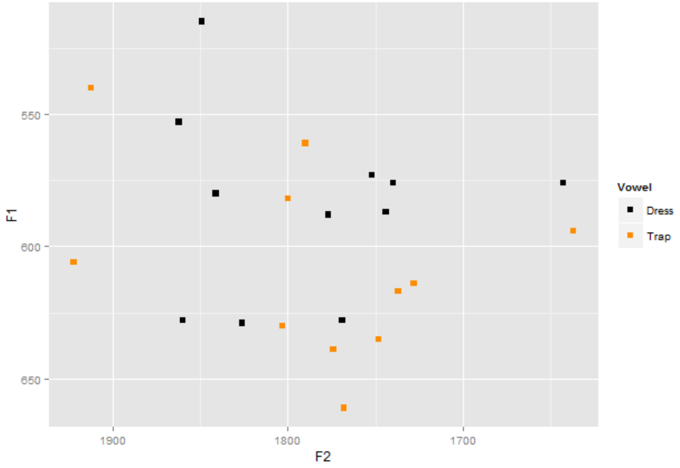


D.1–59. Vowel plot of speaker #059, corresponding to a Pillai score of .93.

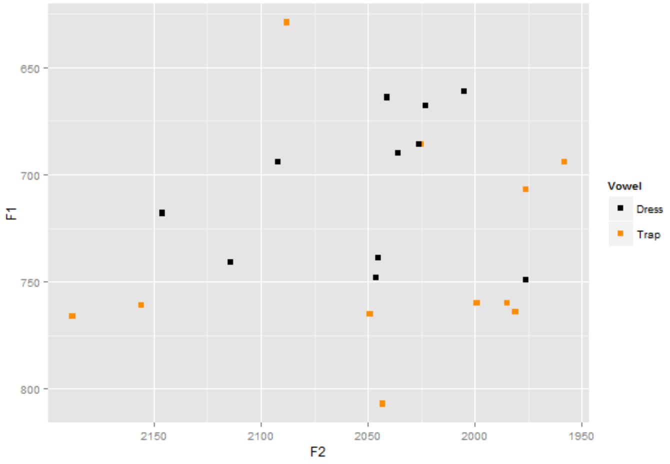


D.1–60. Vowel plot of speaker #060, corresponding to a Pillai score of .73.

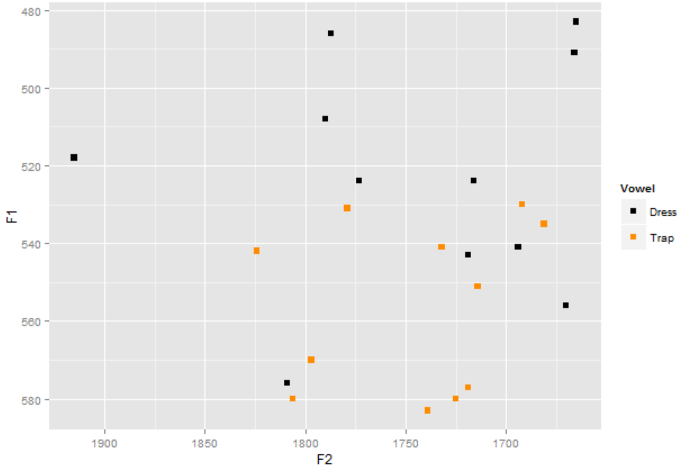
D.2 – Control Group (Groups C-UK and C-US)



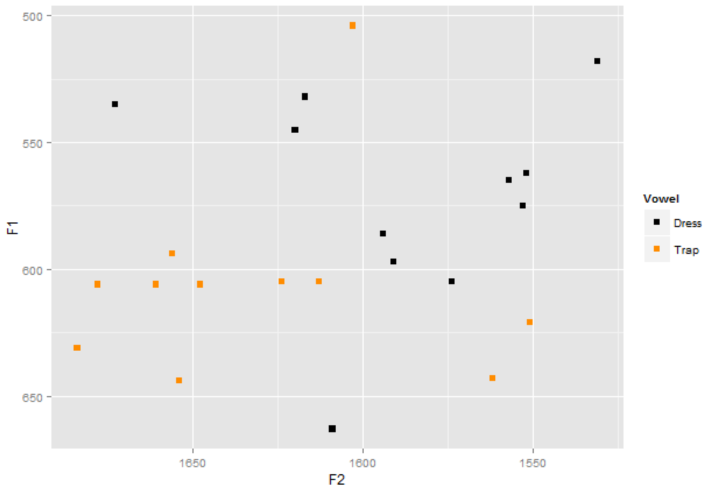
D.2–1. Vowel plot of speaker #301, corresponding to a Pillai score of .10.



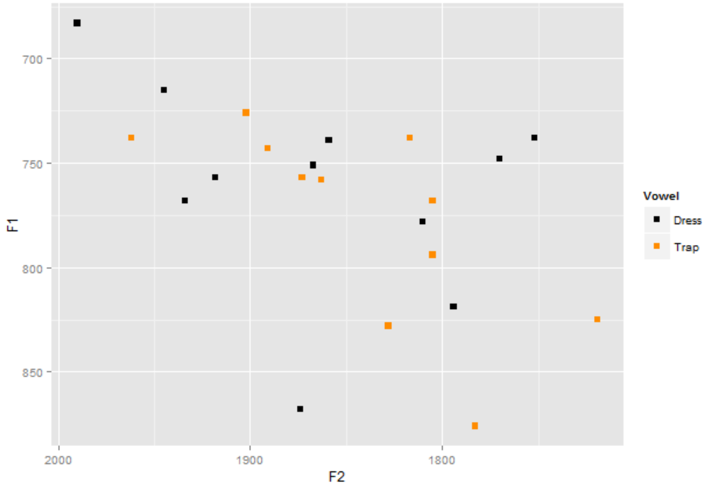
D.2–2. Vowel plot of speaker #302, corresponding to a Pillai score of .14.



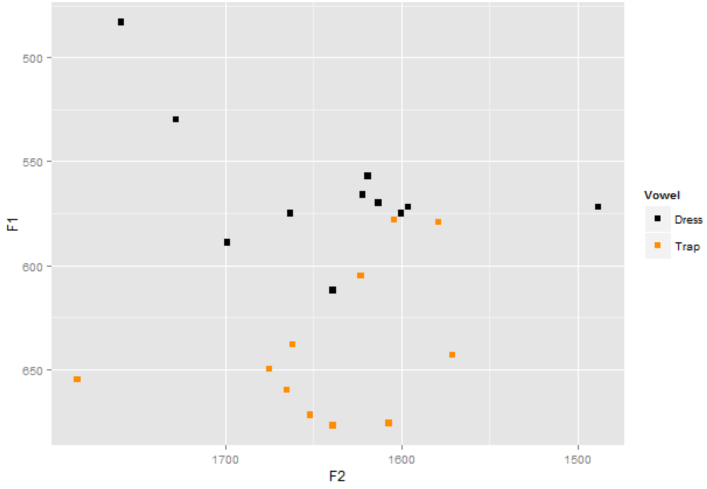
D.2–3. Vowel plot of speaker #303, corresponding to a Pillai score of .32.



D.2–4. Vowel plot of speaker #304, corresponding to a Pillai score of .25.

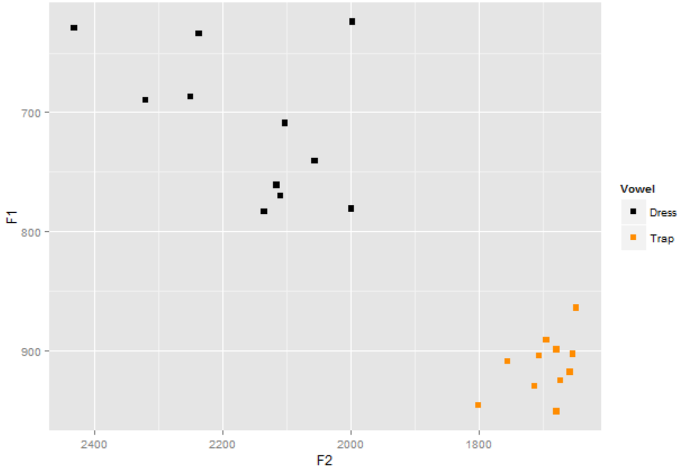


D.2–5. Vowel plot of speaker #305, corresponding to a Pillai score of .04.

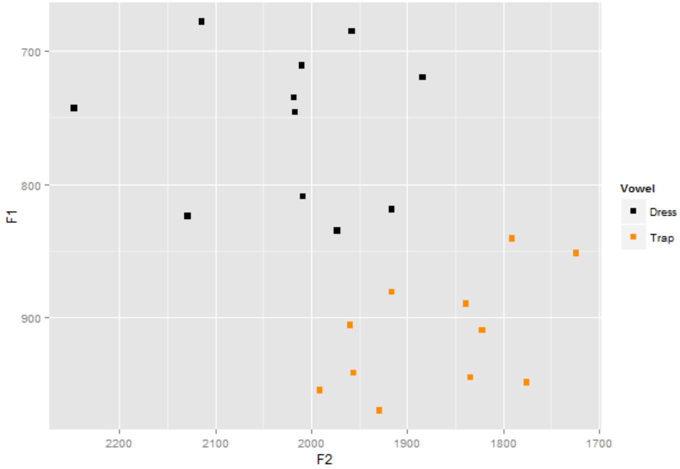


D.2–6. Vowel plot of speaker #306, corresponding to a Pillai score of .57.

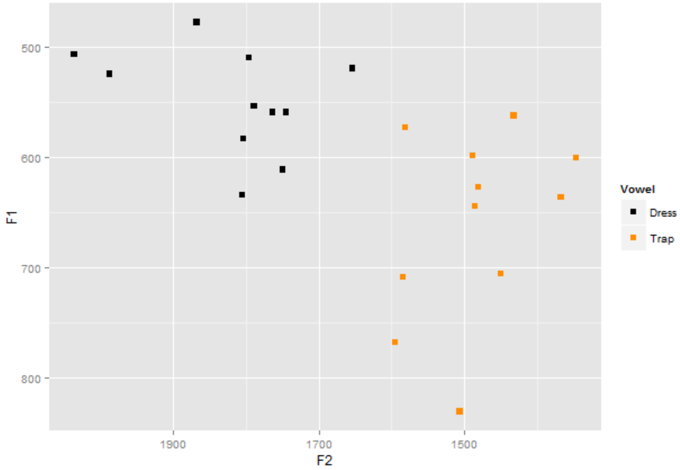
D.3 – British Native Speakers



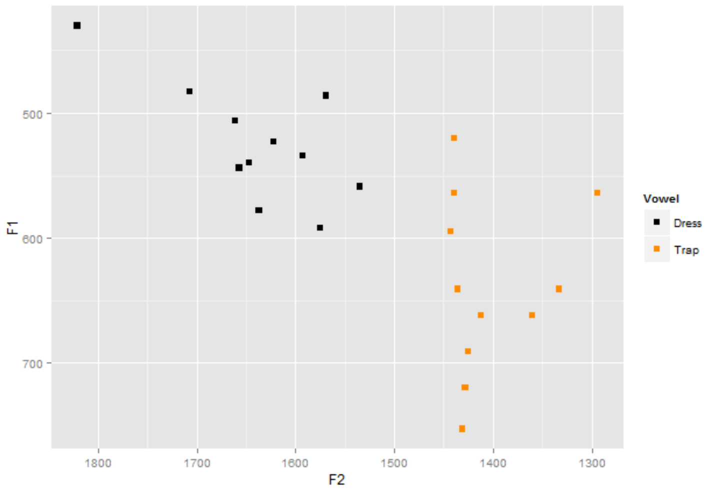
D.3–1. Vowel plot of speaker #201, corresponding to a Pillai score of .89.



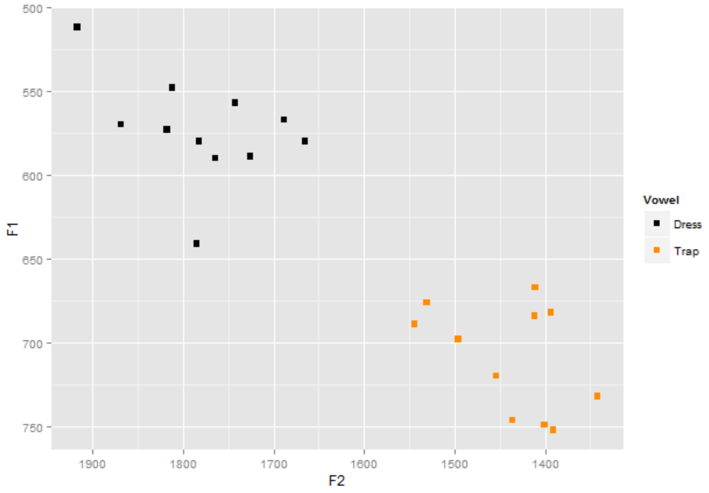
D.3–2. Vowel plot of speaker #202, corresponding to a Pillai score of .80.



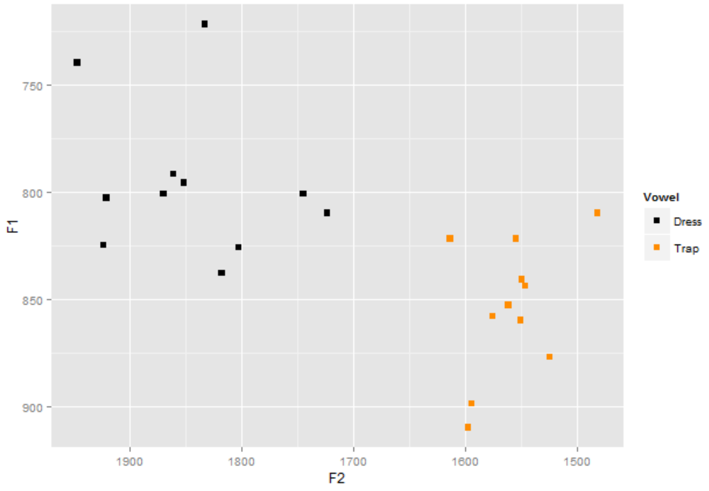
D.3–3. Vowel plot of speaker #203, corresponding to a Pillai score of .81.



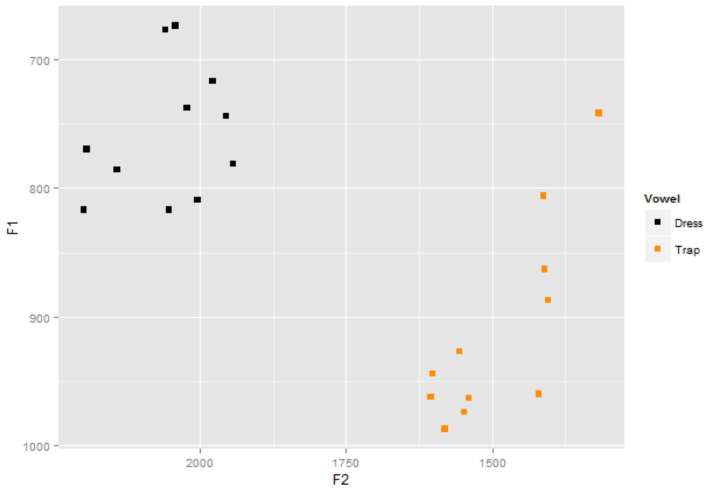
D.3–4. Vowel plot of speaker #204, corresponding to a Pillai score of .79.



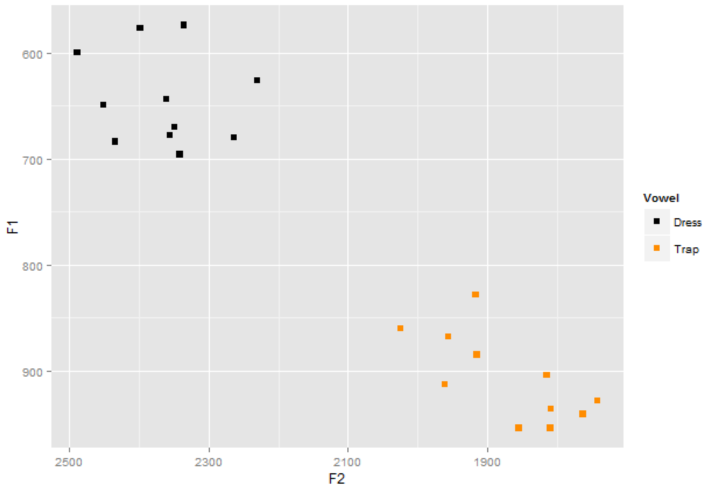
D.3–5. Vowel plot of speaker #205, corresponding to a Pillai score of .89.



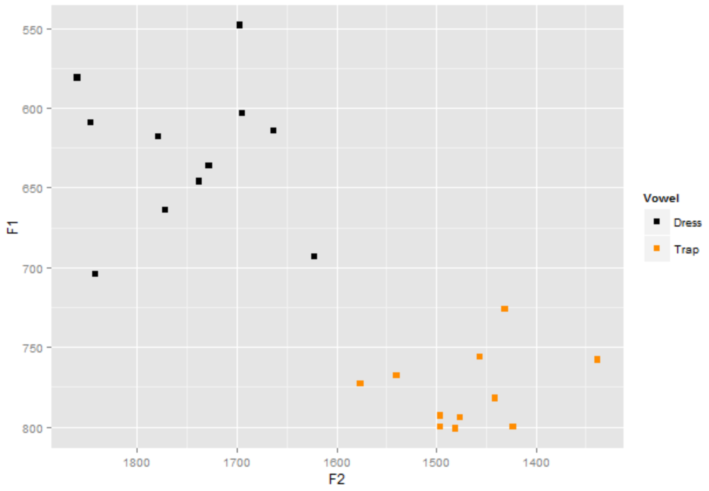
D.3–6. Vowel plot of speaker #206, corresponding to a Pillai score of .88.



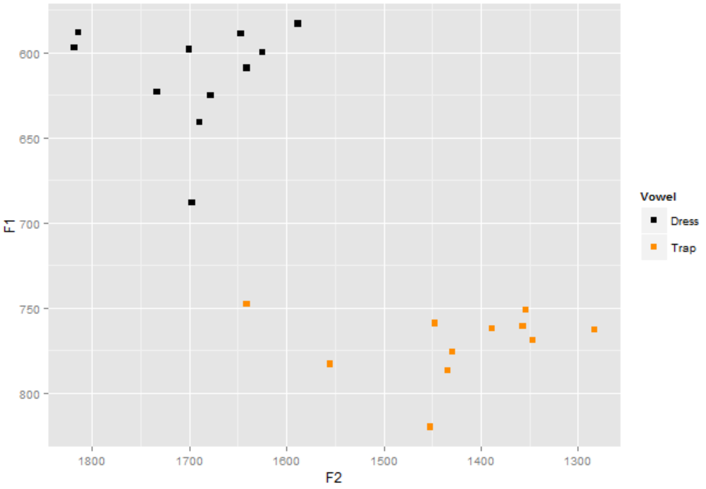
D.3–7. Vowel plot of speaker #207, corresponding to a Pillai score of .96.



D.3–8. Vowel plot of speaker #208, corresponding to a Pillai score of .97.

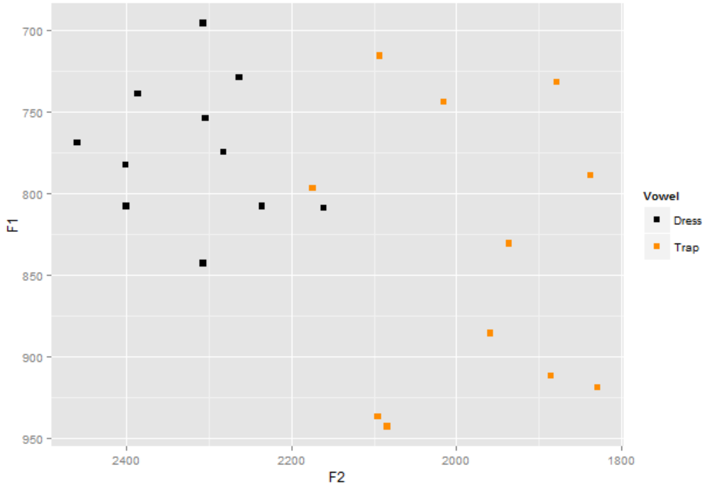


D.3–9. Vowel plot of speaker #209, corresponding to a Pillai score of .90.

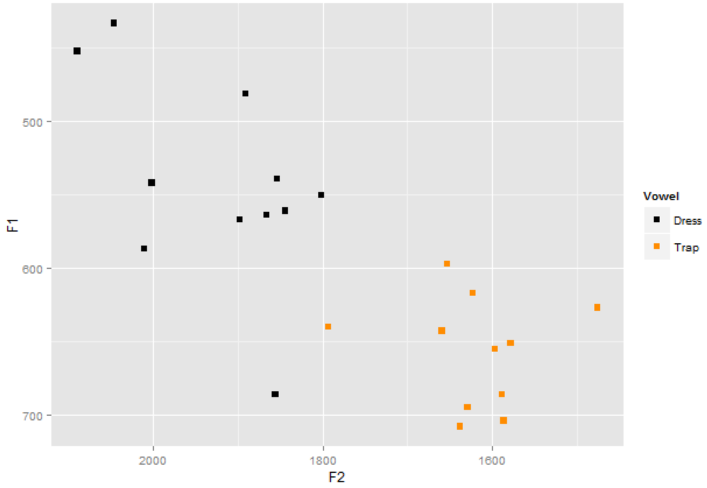


D.3–10. Vowel plot of speaker #210, corresponding to a Pillai score of .93.

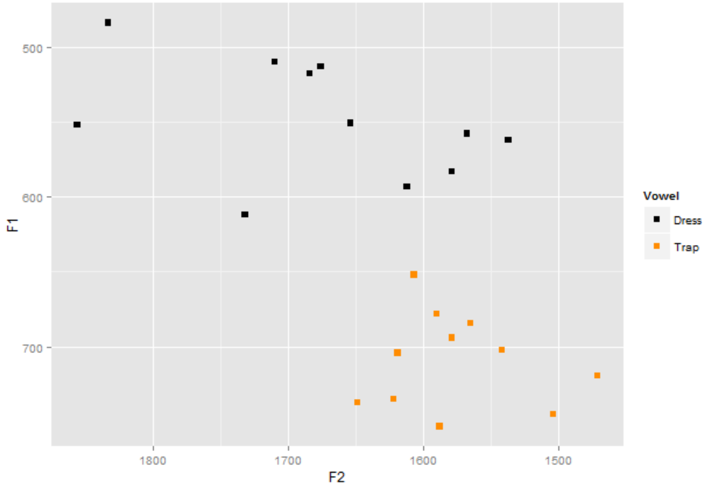
D.4 – American Native Speakers



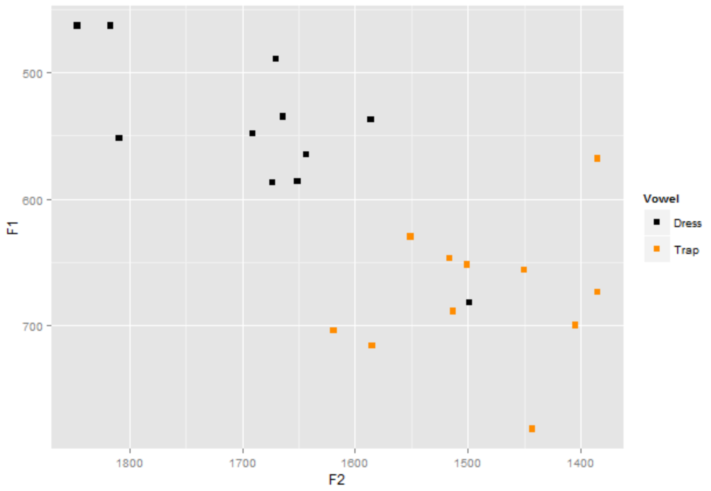
D.4–1. Vowel plot of speaker #101, corresponding to a Pillai score of .75.



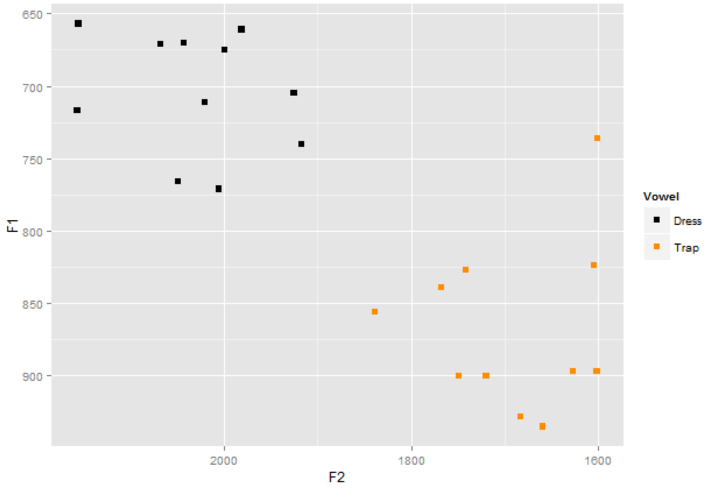
D.4–2. Vowel plot of speaker #102, corresponding to a Pillai score of .78.



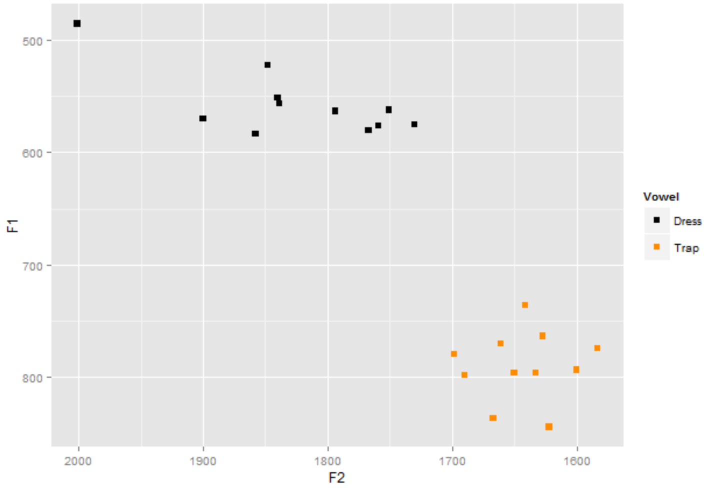
D.4–3. Vowel plot of speaker #103, corresponding to a Pillai score of .85.



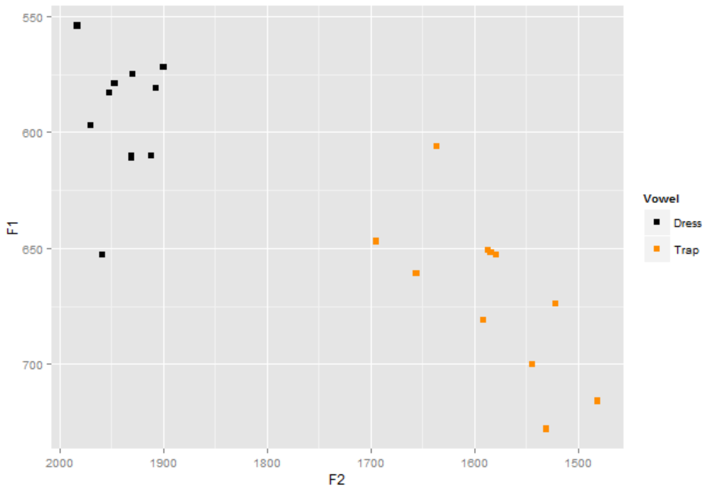
D.4–4. Vowel plot of speaker #104, corresponding to a Pillai score of .65.



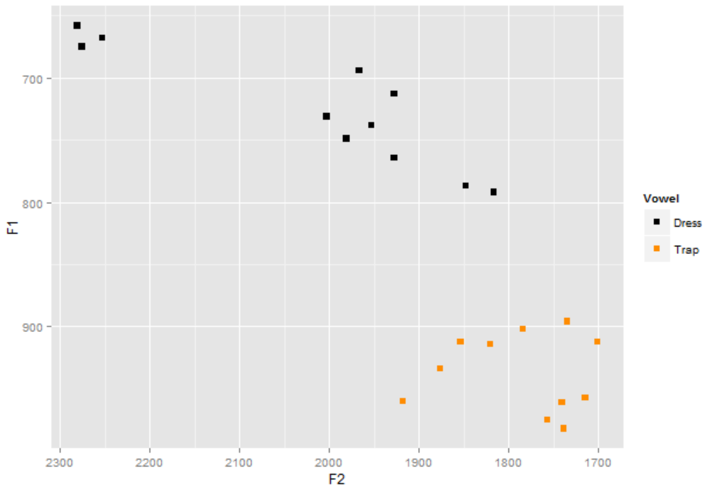
D.4–5. Vowel plot of speaker #105, corresponding to a Pillai score of .89.



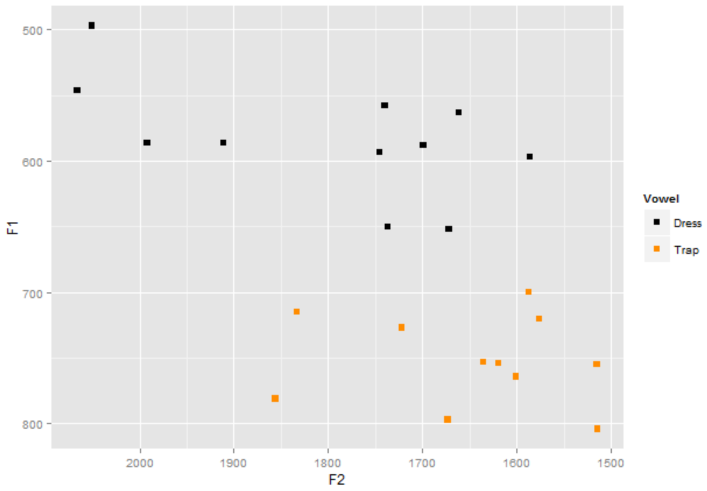
D.4–6. Vowel plot of speaker #106, corresponding to a Pillai score of .94.



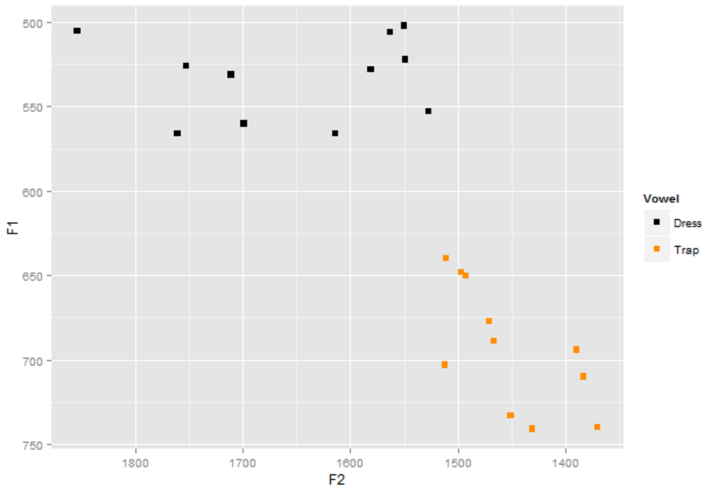
D.4–7. Vowel plot of speaker #107, corresponding to a Pillai score of .94.



D.4–8. Vowel plot of speaker #108, corresponding to a Pillai score of .91.



D.4–9. Vowel plot of speaker #109, corresponding to a Pillai score of .84.



D.4–10. Vowel plot of speaker #110, corresponding to a Pillai score of .88.

Appendix E – Devoicing Scatterplots

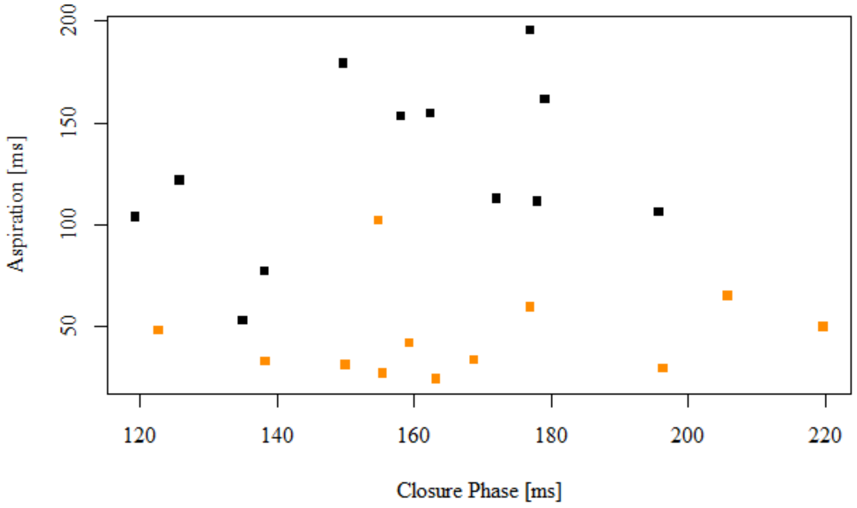
E.1 – German Learners (Groups G-UK and G-US)

E.2 – Control Group (Groups C-UK and C-US)

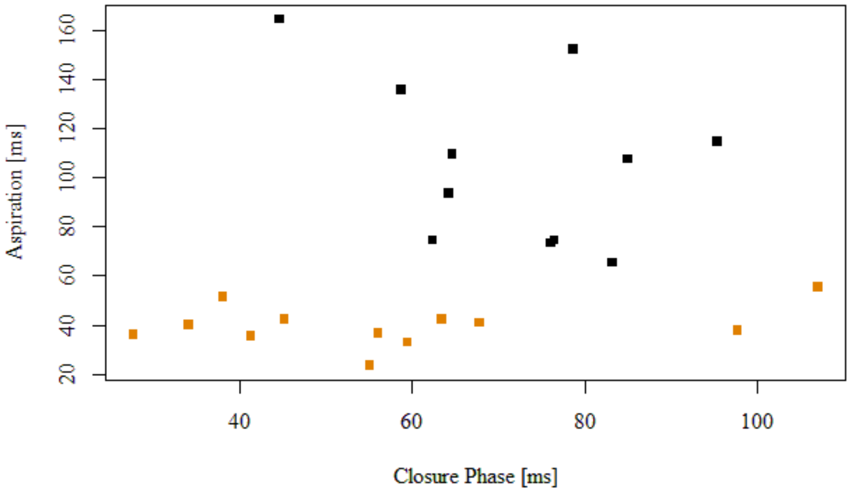
E.3 – British Native Speakers

E.4 – American Native Speakers

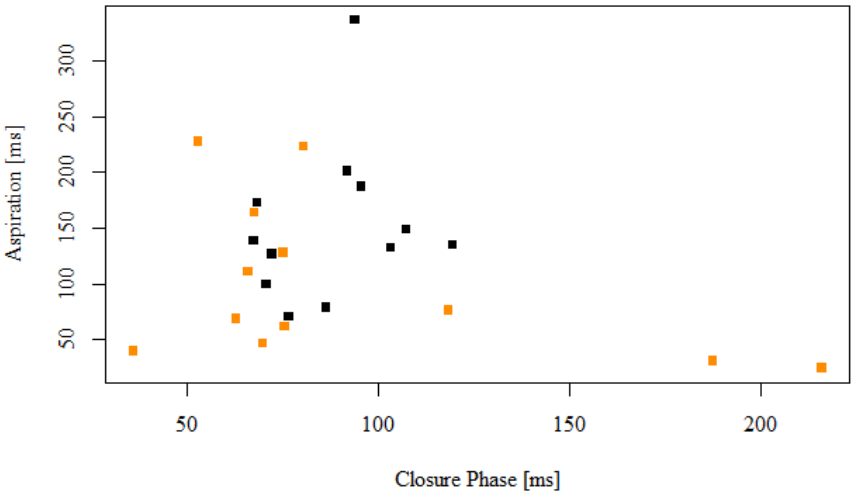
E.1 – German Learners (Groups G-UK and G-US)



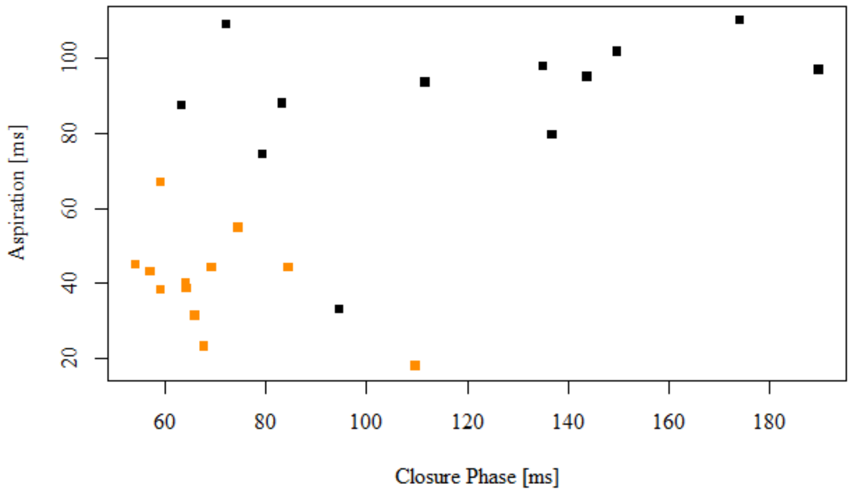
E.1–1. Devoicing scatterplot of speaker #001, corresponding to a Pillai score of .66.



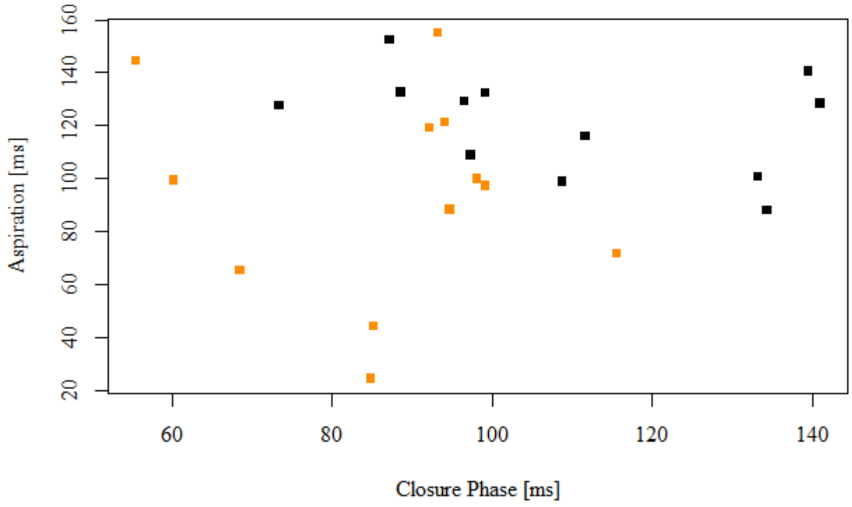
E.1–2. Devoicing scatterplot of speaker #002, corresponding to a Pillai score of .70.



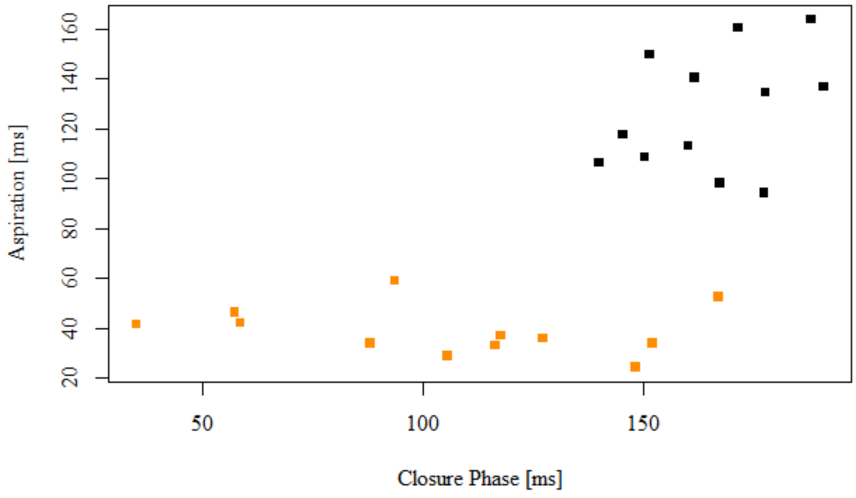
E.1–3. Devoicing scatterplot of speaker #003, corresponding to a Pillai score of .13.



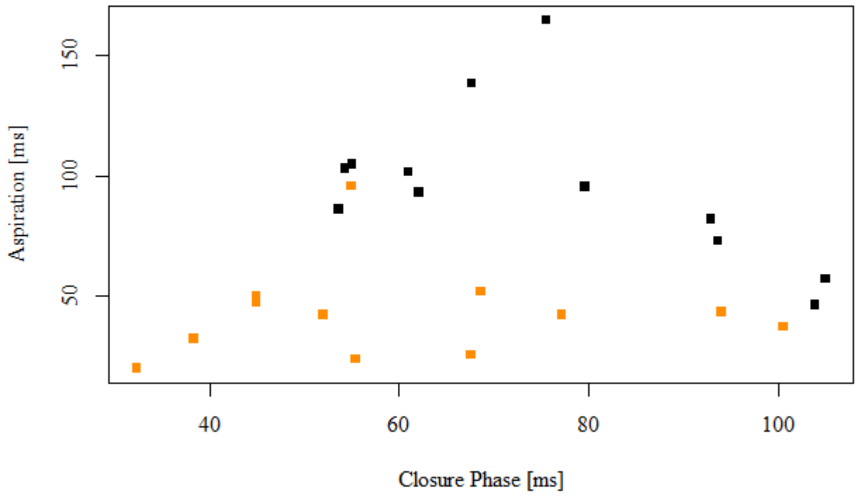
E.1–4. Devoicing scatterplot of speaker #004, corresponding to a Pillai score of .70.



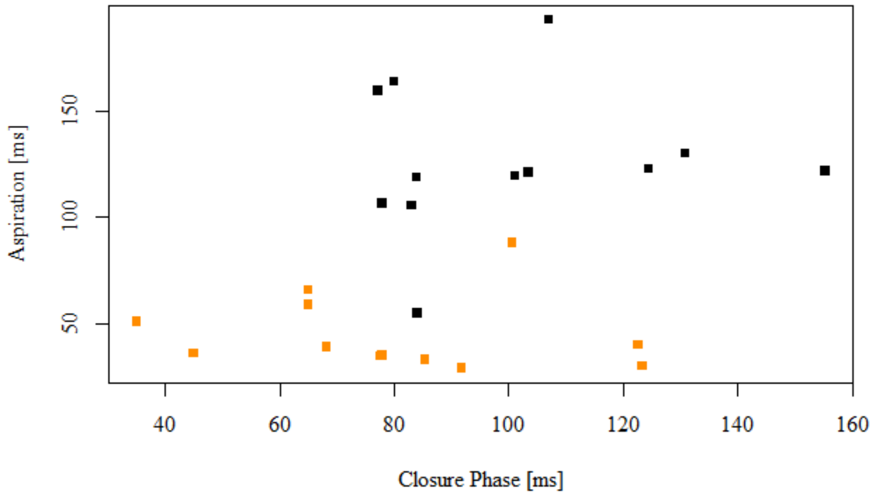
E.1–5. Devoicing scatterplot of speaker #005, corresponding to a Pillai score of .40.



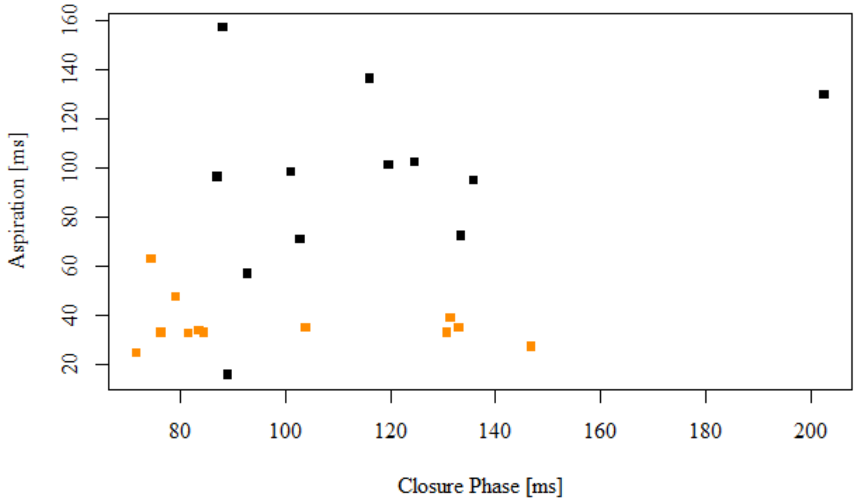
E.1–6. Devoicing scatterplot of speaker #006, corresponding to a Pillai score of .52.



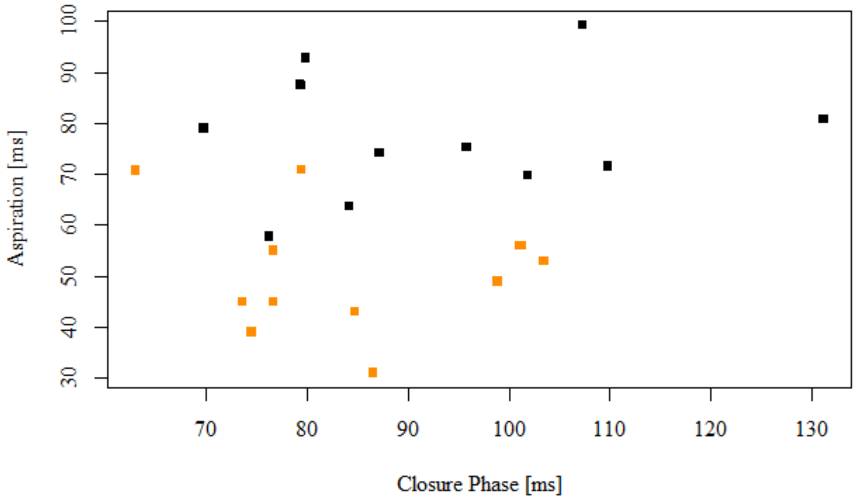
E.1–7. Devoicing scatterplot of speaker #007, corresponding to a Pillai score of .61.



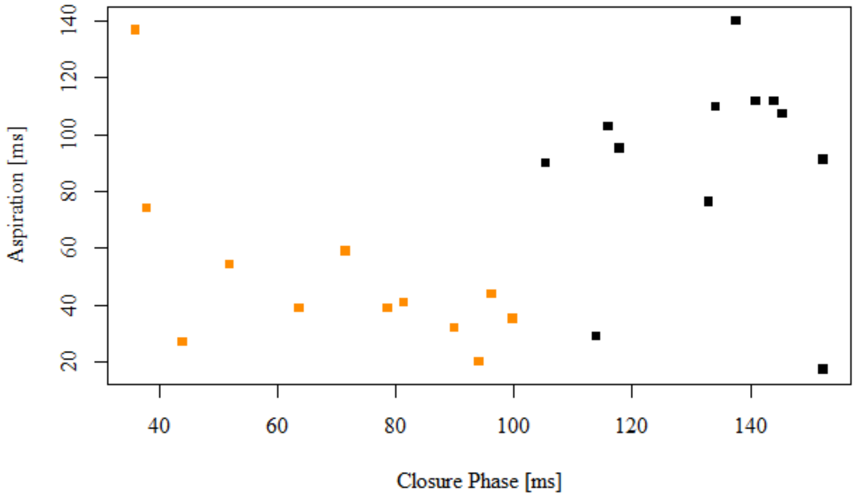
E.1–8. Devoicing scatterplot of speaker #008, corresponding to a Pillai score of .72.



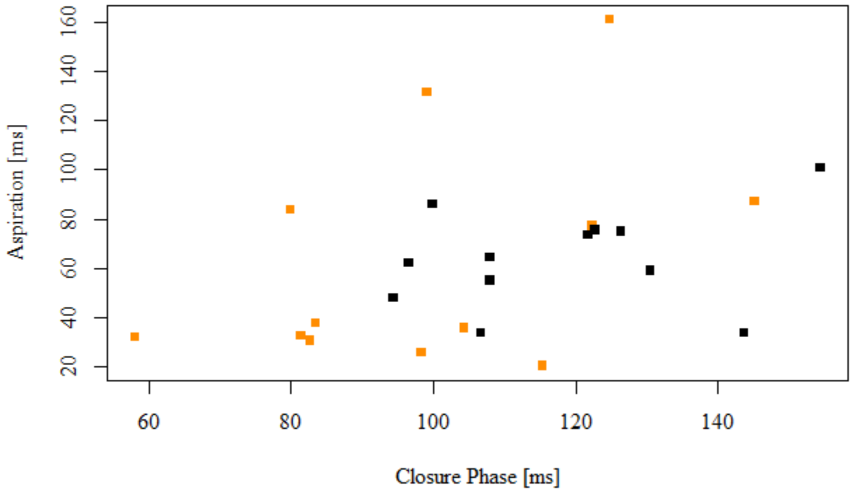
E.1–9. Devoicing scatterplot of speaker #009, corresponding to a Pillai score of .55.



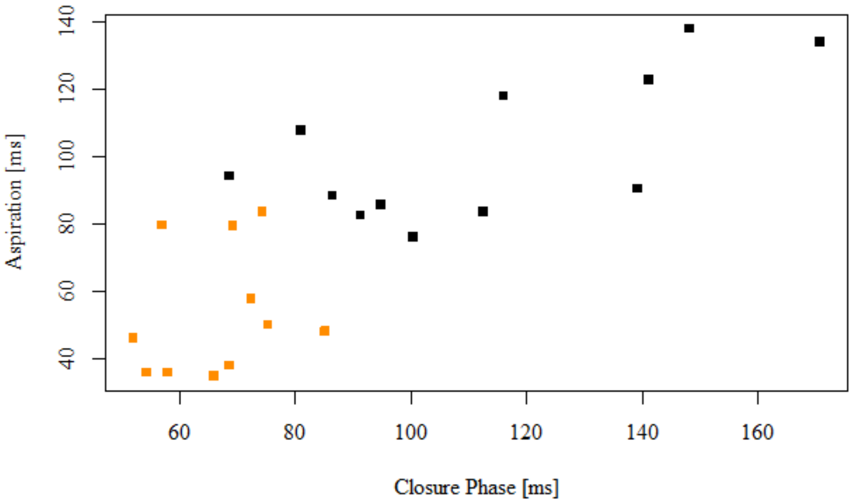
E.1–10. Devoicing scatterplot of speaker #010, corresponding to a Pillai score of .59.



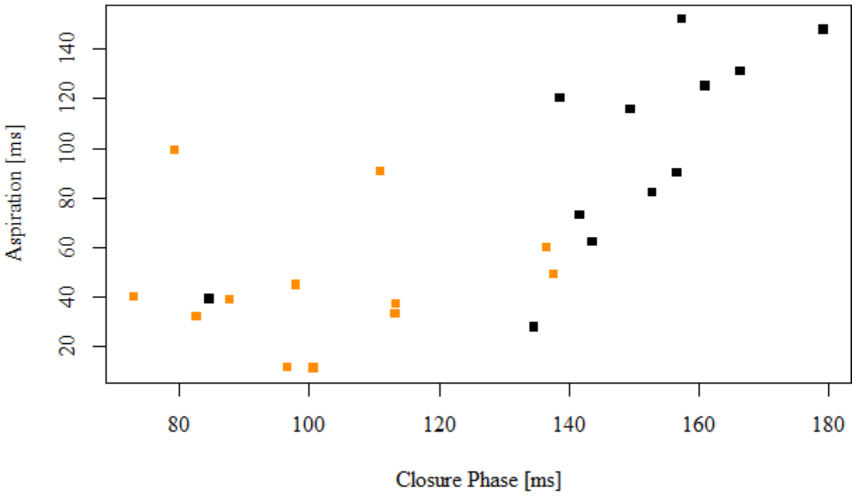
E.1–11. Devoicing scatterplot of speaker #011, corresponding to a Pillai score of .81.



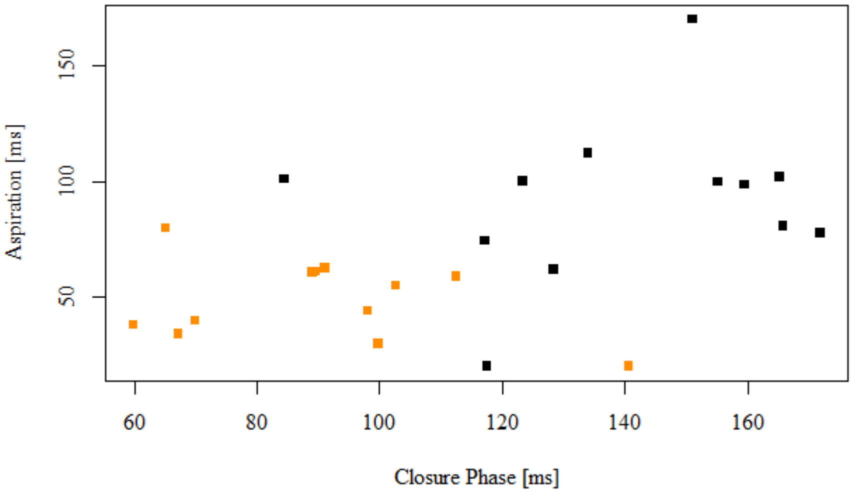
E.1–12. Devoicing scatterplot of speaker #012, corresponding to a Pillai score of .18.



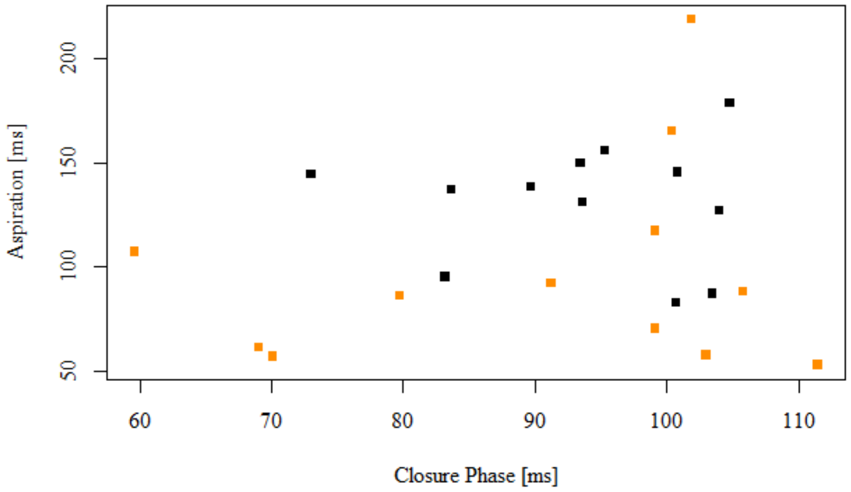
E.1–13. Devoicing scatterplot of speaker #013 corresponding to a Pillai score of .64.



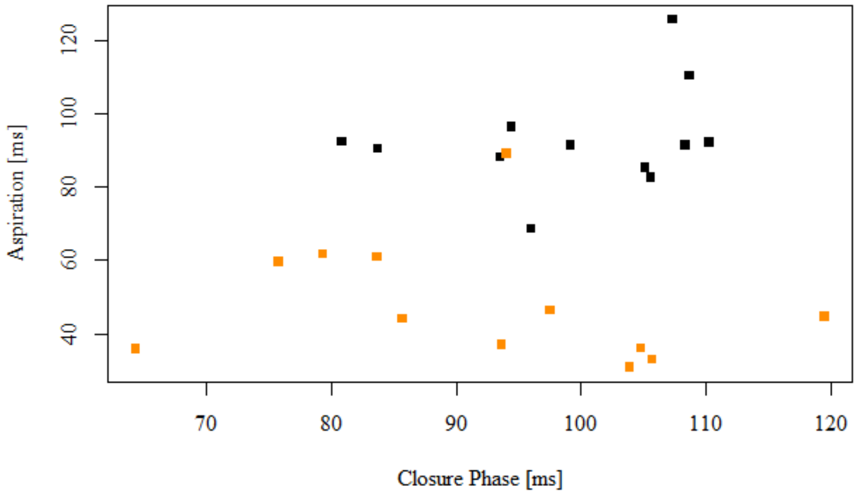
E.1–14. Devoicing scatterplot of speaker #014, corresponding to a Pillai score of .55.



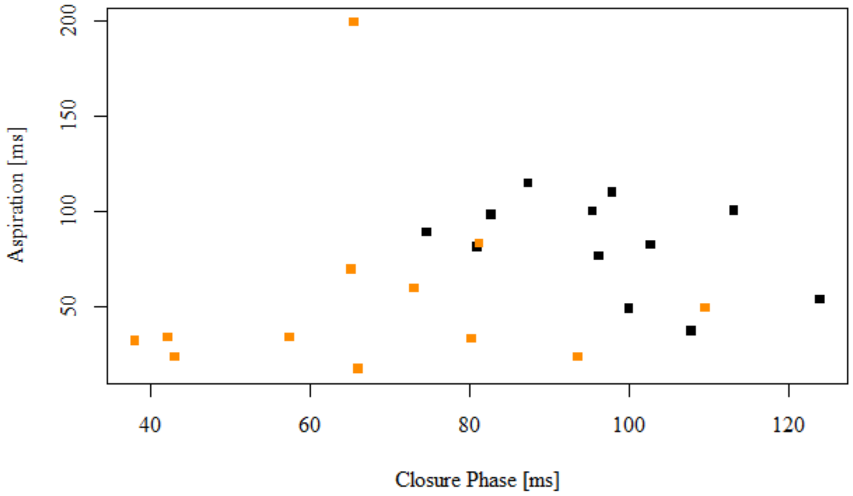
E.1–15. Devoicing scatterplot of speaker #015, corresponding to a Pillai score of .62.



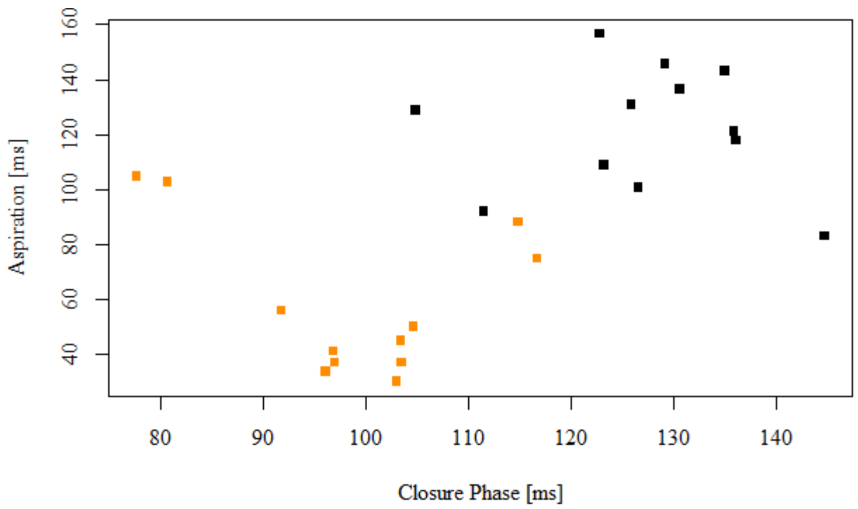
E.1–16. Devoicing scatterplot of speaker #016, corresponding to a Pillai score of .16.



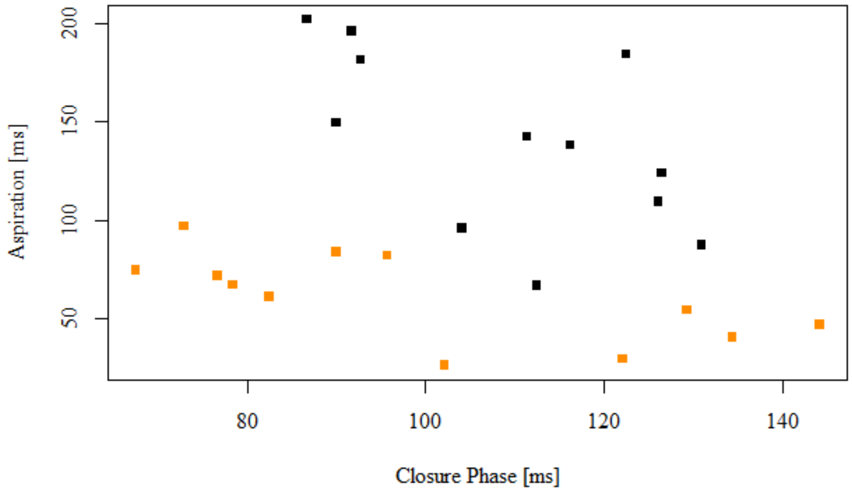
E.1–17. Devoicing scatterplot of speaker #017, corresponding to a Pillai score of .71.



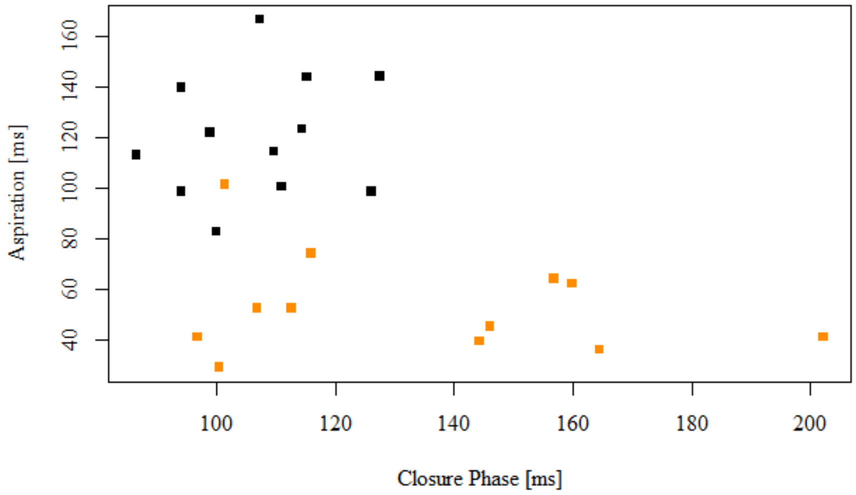
E.1–18. Devoicing scatterplot of speaker #018, corresponding to a Pillai score of .46.



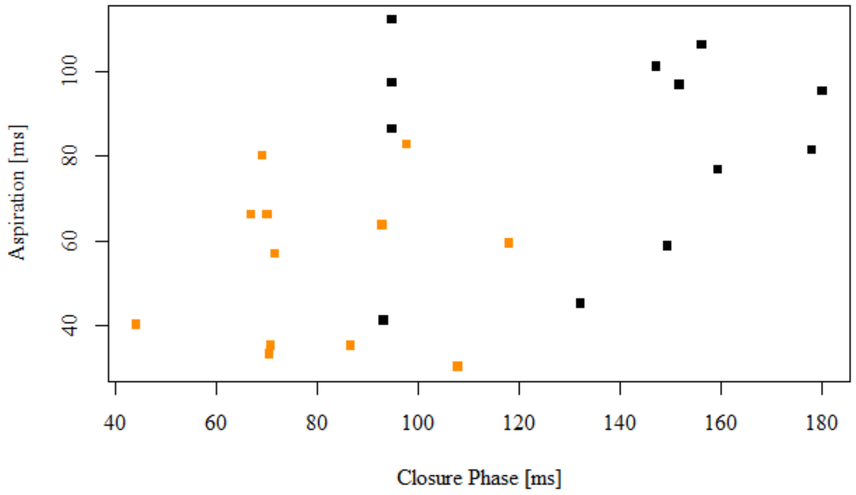
E.1–19. Devoicing scatterplot of speaker #019, corresponding to a Pillai score of .82.



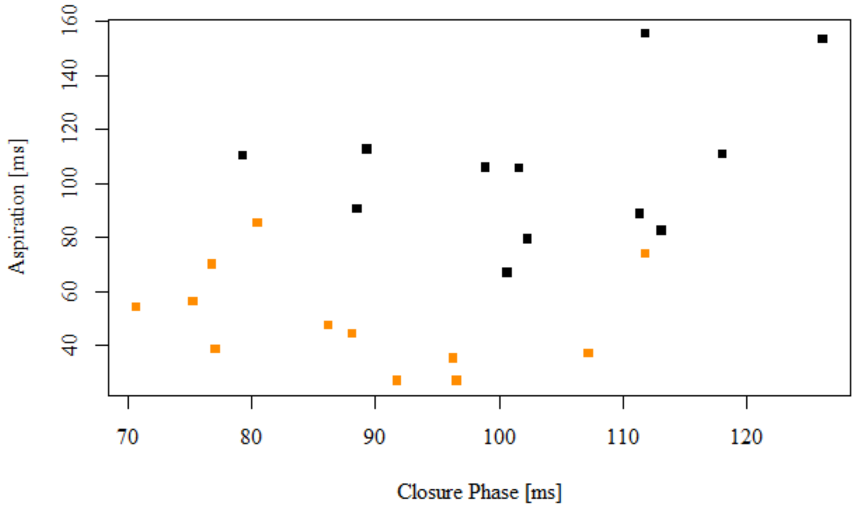
E.1–20. Devoicing scatterplot of speaker #020, corresponding to a Pillai score of .70.



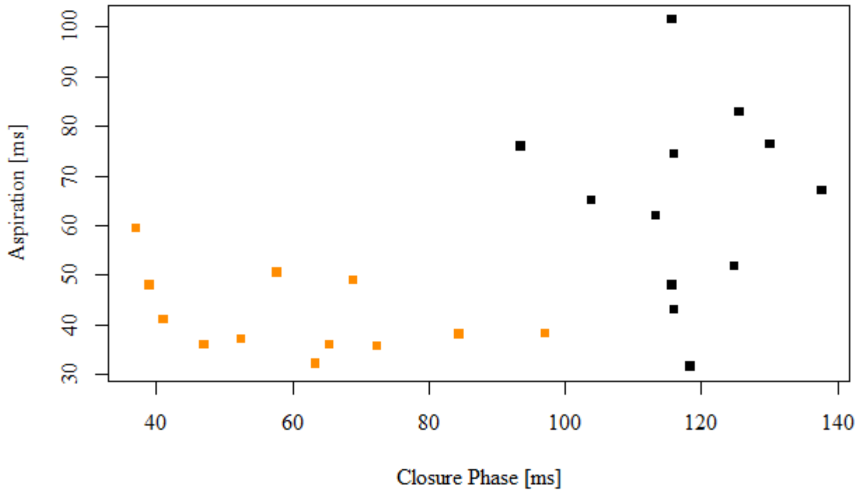
E.1–21. Devoicing scatterplot of speaker #021, corresponding to a Pillai score of .810.



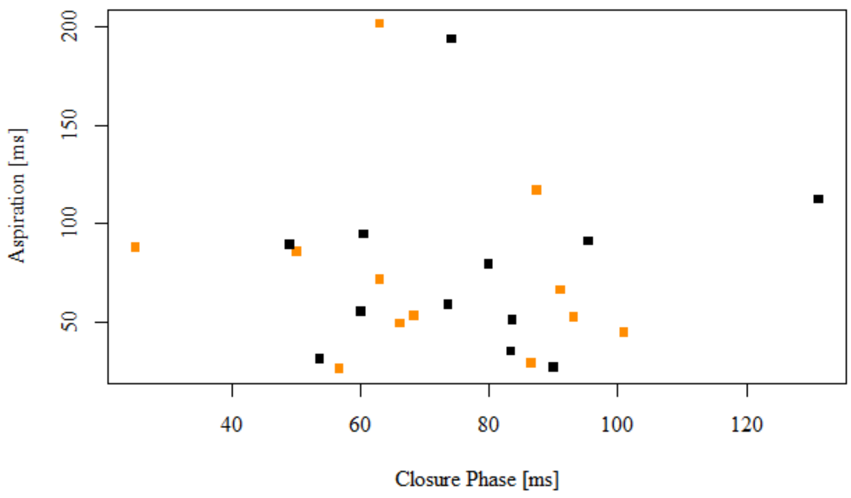
E.1–22. Devoicing scatterplot of speaker #022, corresponding to a Pillai score of .60.



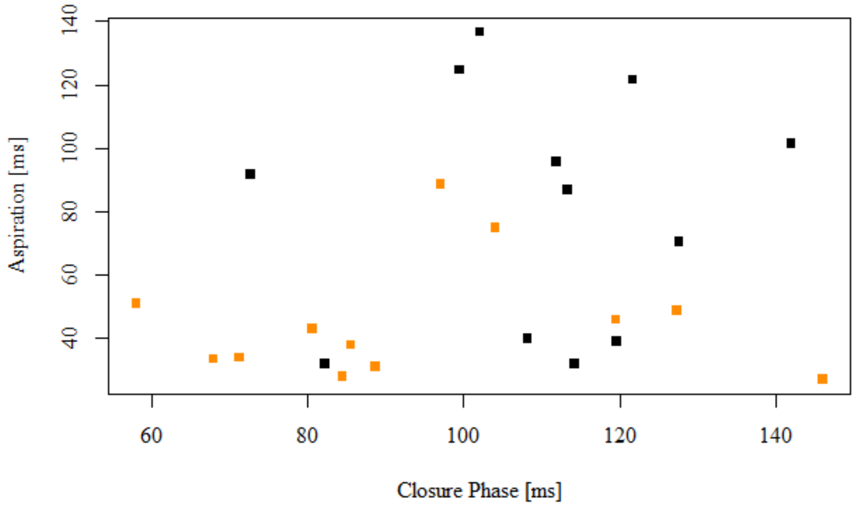
E.1–23. Devoicing scatterplot of speaker #023, corresponding to a Pillai score of .64.



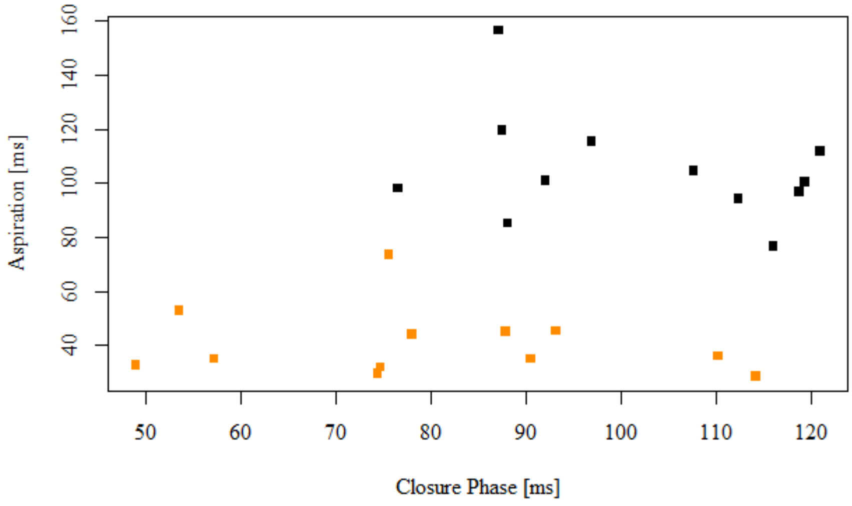
E.1–24. Devoicing scatterplot of speaker #024, corresponding to a Pillai score of .83.



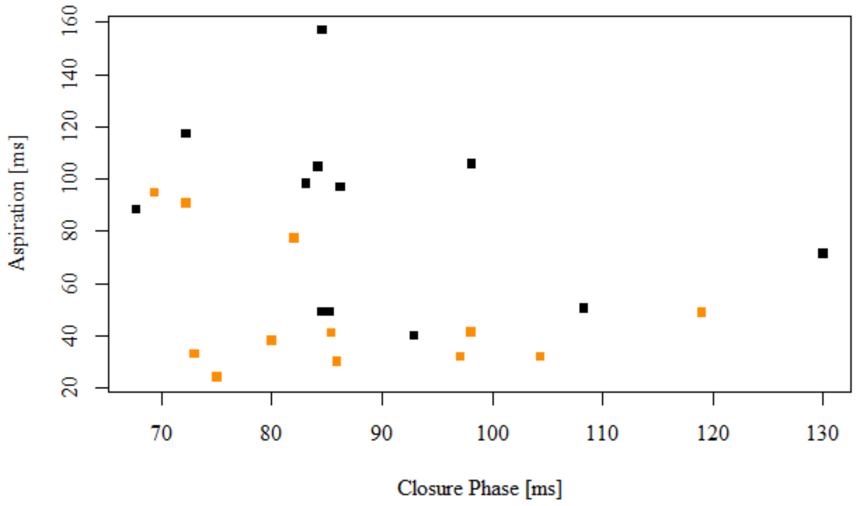
E.1–25. Devoicing scatterplot of speaker #025, corresponding to a Pillai score of .01.



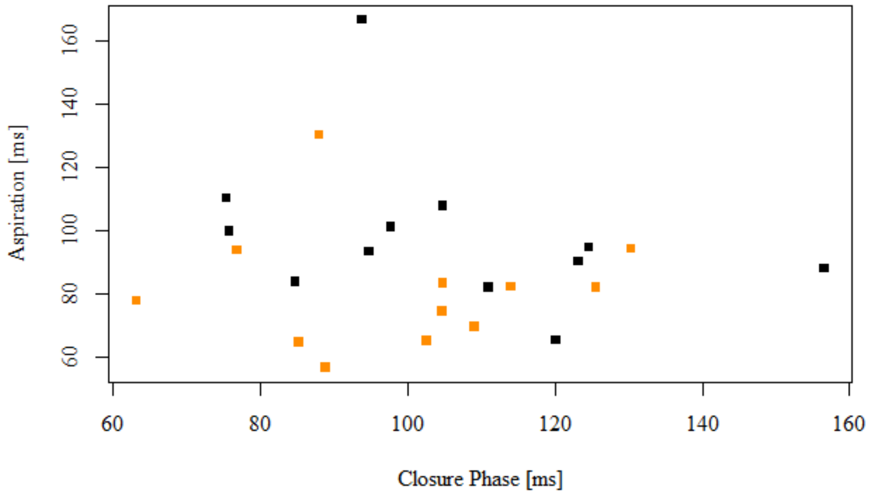
E.1–26. Devoicing scatterplot of speaker #026, corresponding to a Pillai score of .33.



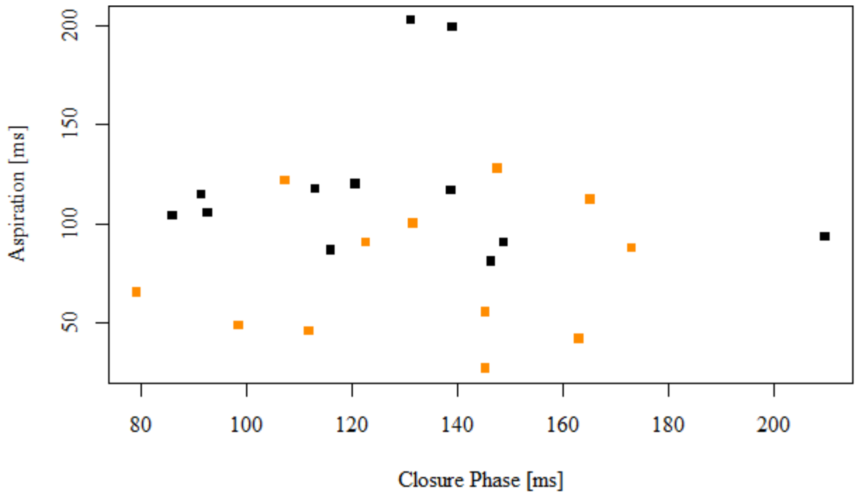
E.1–27. Devoicing scatterplot of speaker #027, corresponding to a Pillai score of .84.



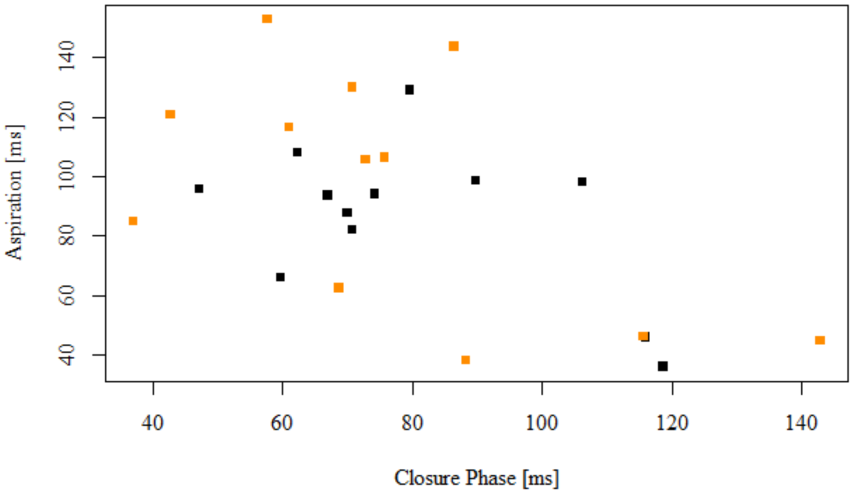
E.1–28. Devoicing scatterplot of speaker #028, corresponding to a Pillai score of .34.



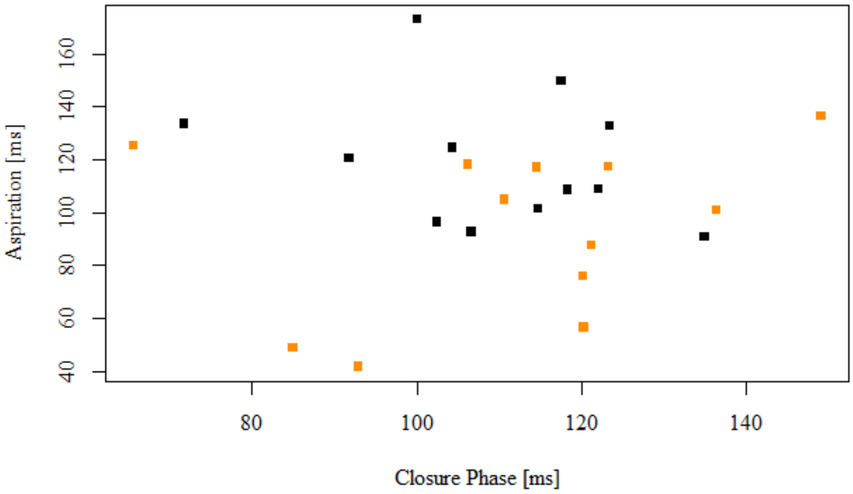
E.1–29. Devoicing scatterplot of speaker #029, corresponding to a Pillai score of .18.



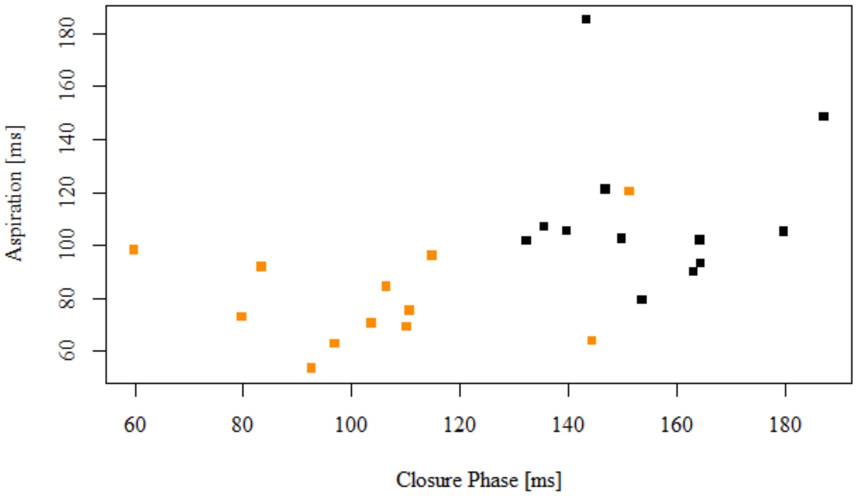
E.1–30. Devoicing scatterplot of speaker #030, corresponding to a Pillai score of .27.



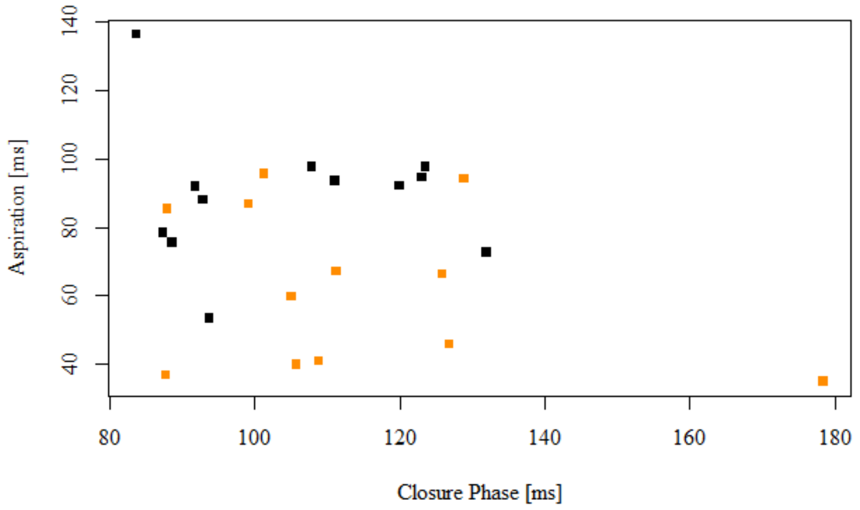
E.1–31. Devoicing scatterplot of speaker #031, corresponding to a Pillai score of .02.



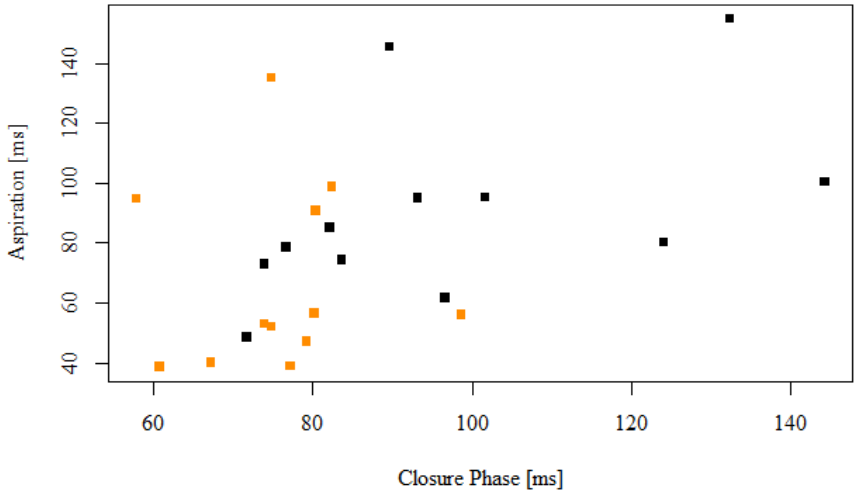
E.1–32. Devoicing scatterplot of speaker #032, corresponding to a Pillai score of .18.



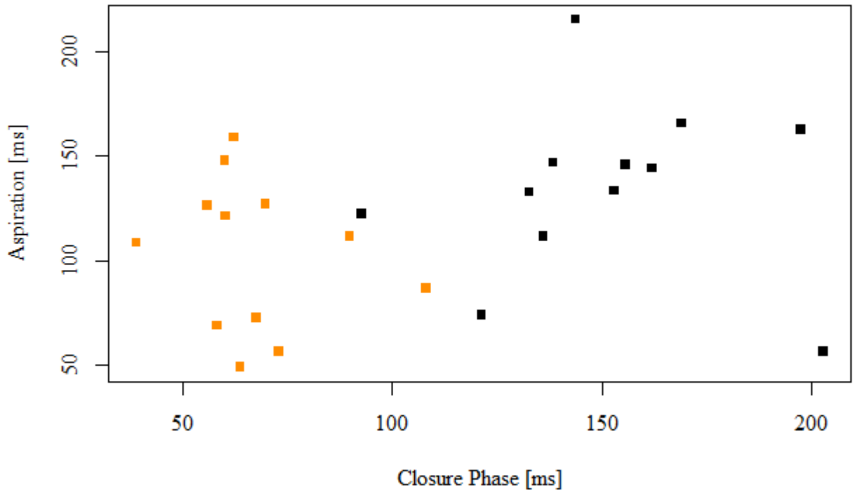
E.1–33. Devoicing scatterplot of speaker #033, corresponding to a Pillai score of .64.



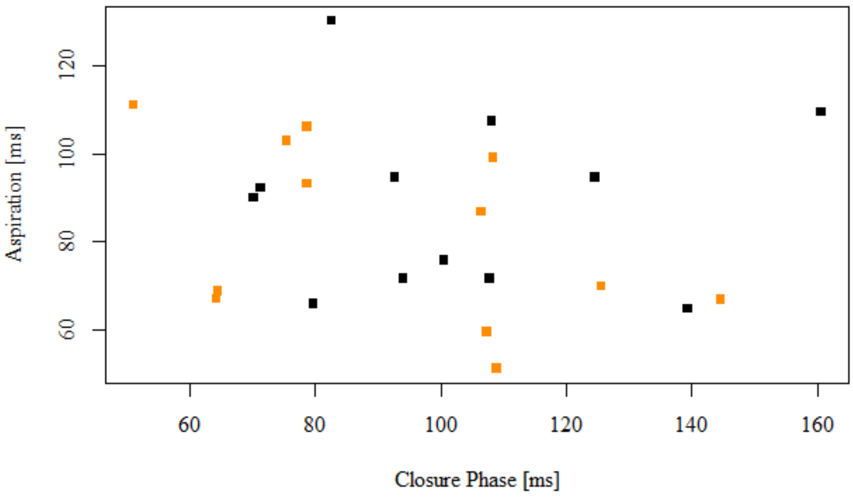
E.1–34. Devoicing scatterplot of speaker #034 corresponding to a Pillai score of .30.



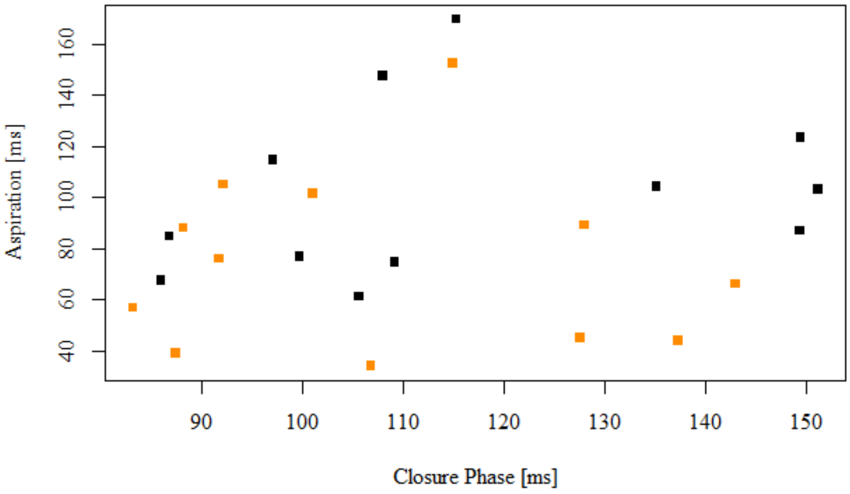
E.1–35. Devoicing scatterplot of speaker #035, corresponding to a Pillai score of .30.



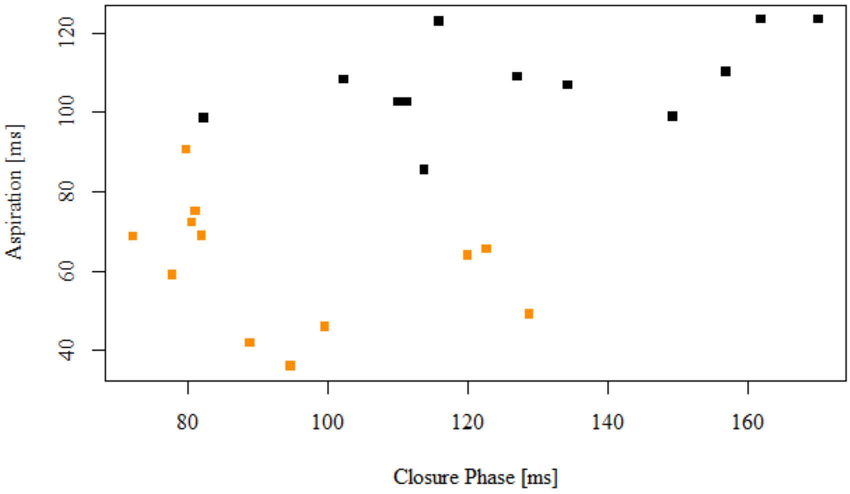
E.1–36. Devoicing scatterplot of speaker #036, corresponding to a Pillai score of .77.



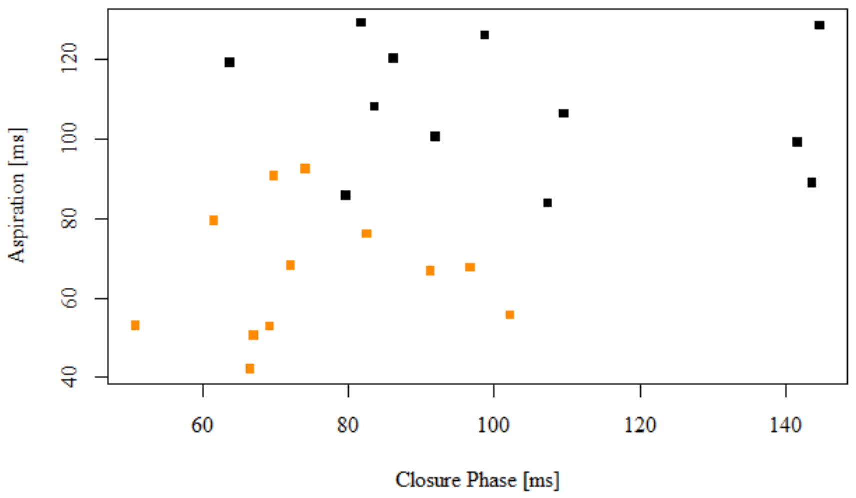
E.1–37. Devoicing scatterplot of speaker #037, corresponding to a Pillai score of .13.



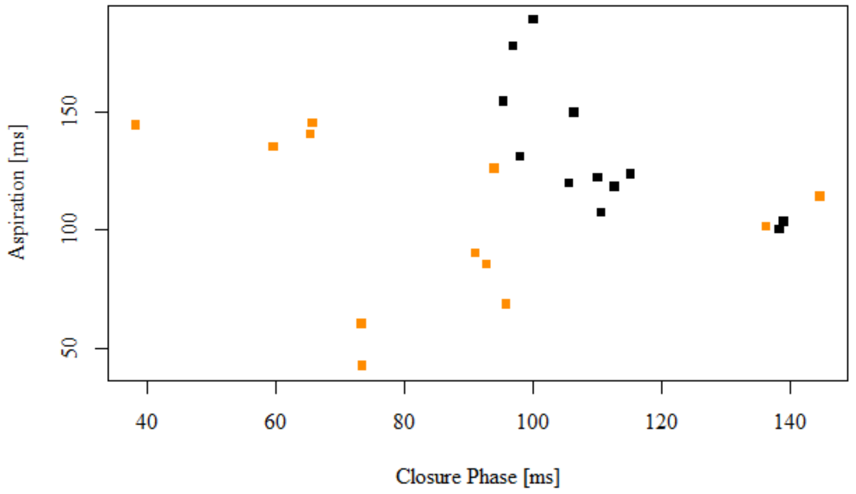
E.1–38. Devoicing scatterplot of speaker #038, corresponding to a Pillai score of .25.



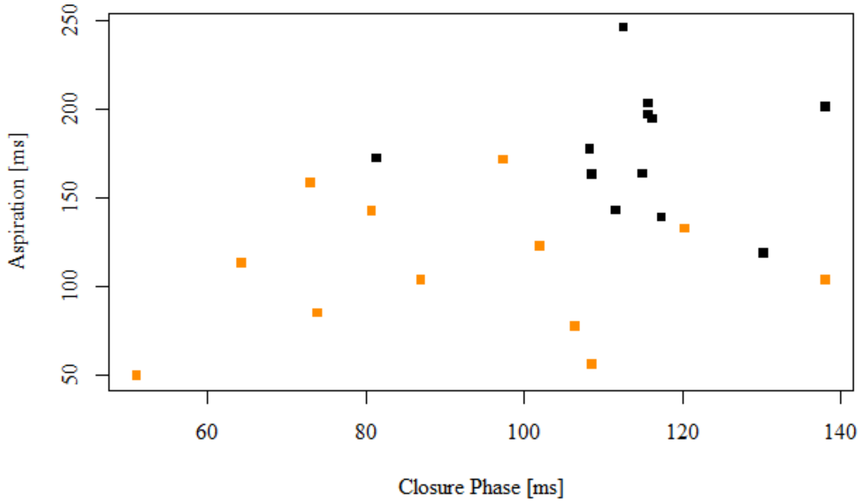
E.1–39. Devoicing scatterplot of speaker #039, corresponding to a Pillai score of .78.



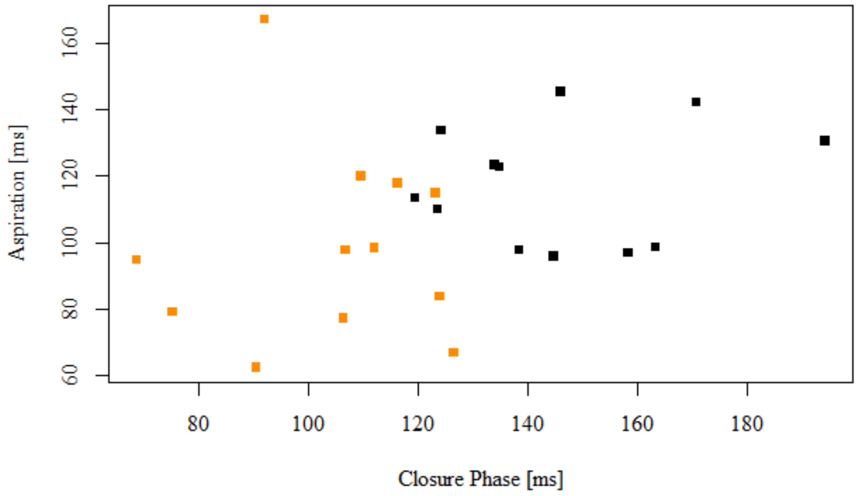
E.1–40. Devoicing scatterplot of speaker #040, corresponding to a Pillai score of .70.



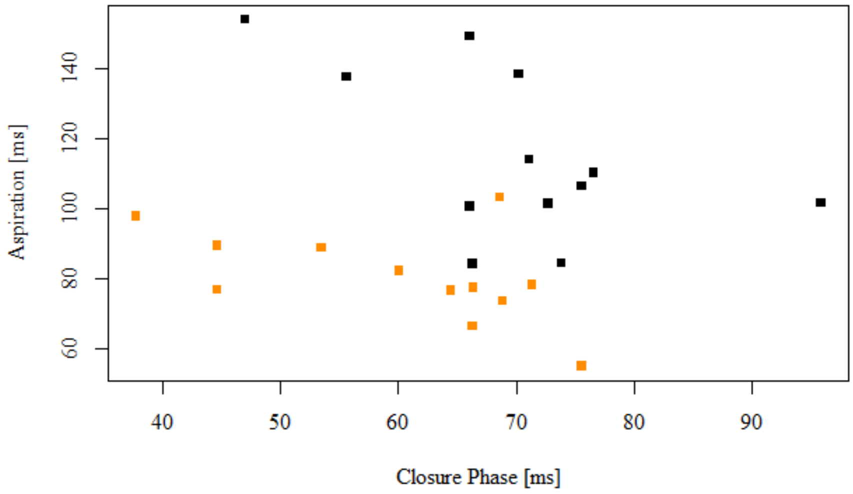
E.1–41. Devoicing scatterplot of speaker #041, corresponding to a Pillai score of .44.



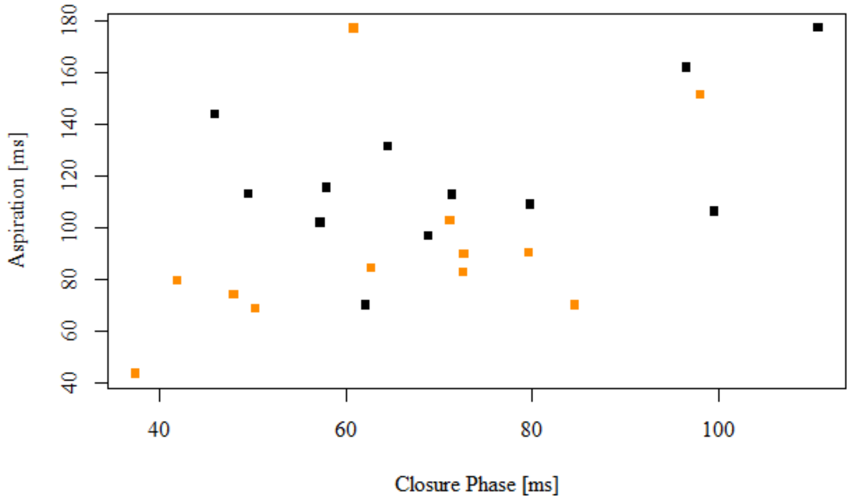
E.1–42. Devoicing scatterplot of speaker #042, corresponding to a Pillai score of .55.



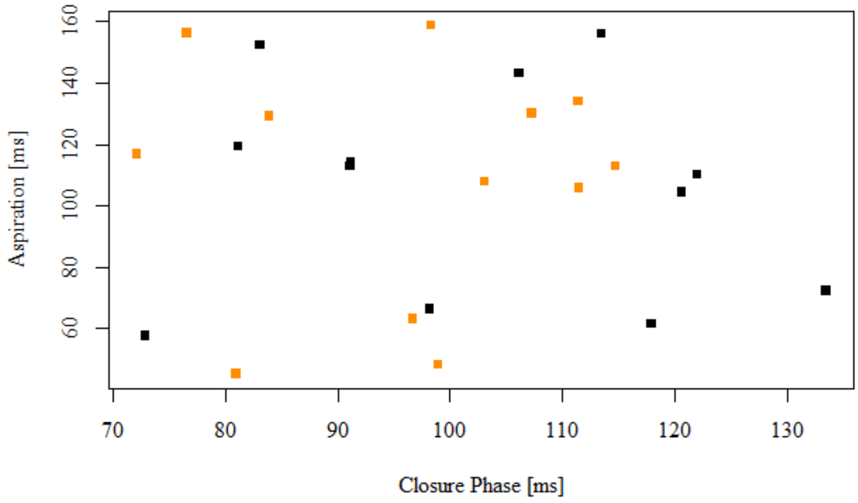
E.1–43. Devoicing scatterplot of speaker #043, corresponding to a Pillai score of .560.



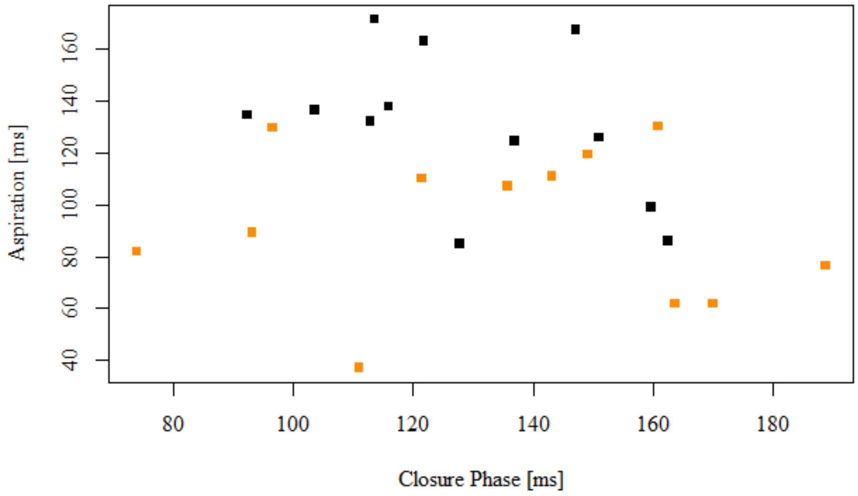
E.1–44. Devoicing scatterplot of speaker #044, corresponding to a Pillai score of .67.



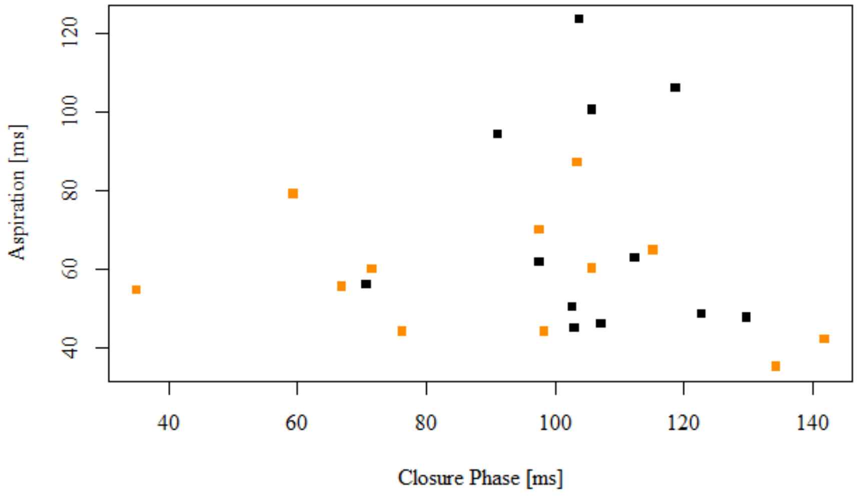
E.1–45. Devoicing scatterplot of speaker #045, corresponding to a Pillai score of .15.



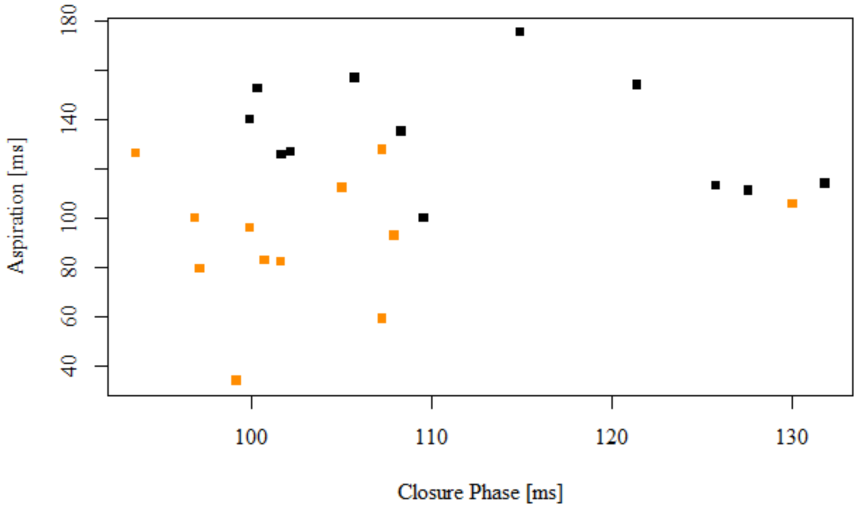
E.1–46. Devoicing scatterplot of speaker #046, corresponding to a Pillai score of .04.



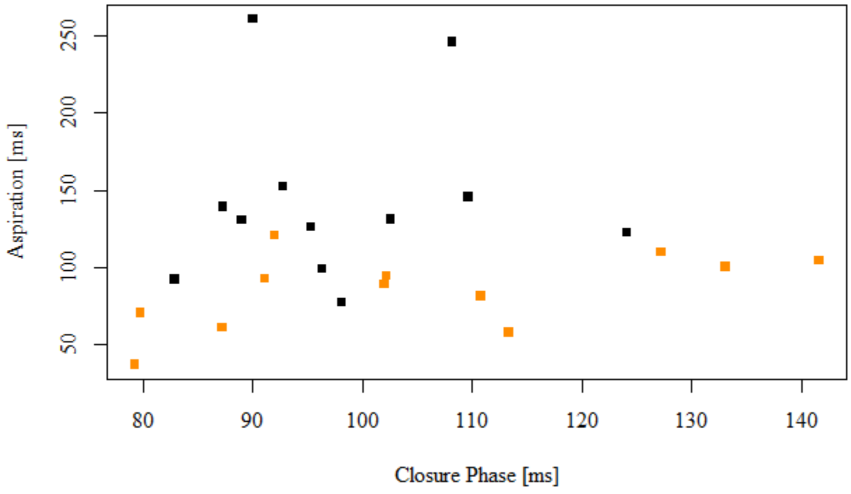
E.1–47. Devoicing scatterplot of speaker #047, corresponding to a Pillai score of .31.



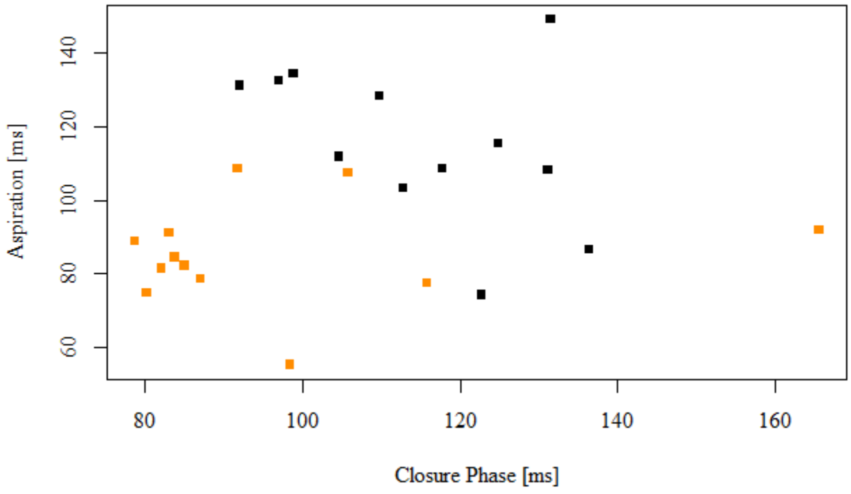
E.1–48. Devoicing scatterplot of speaker #048, corresponding to a Pillai score of .16.



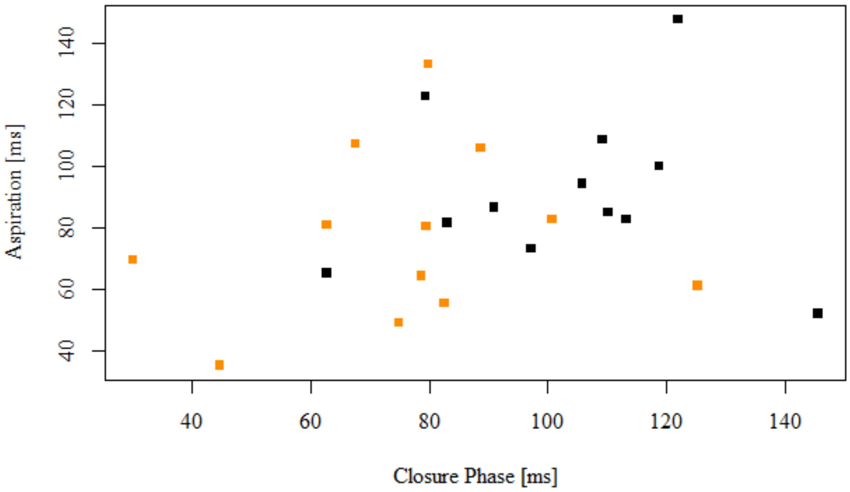
E.1–49. Devoicing scatterplot of speaker #049, corresponding to a Pillai score of .51.



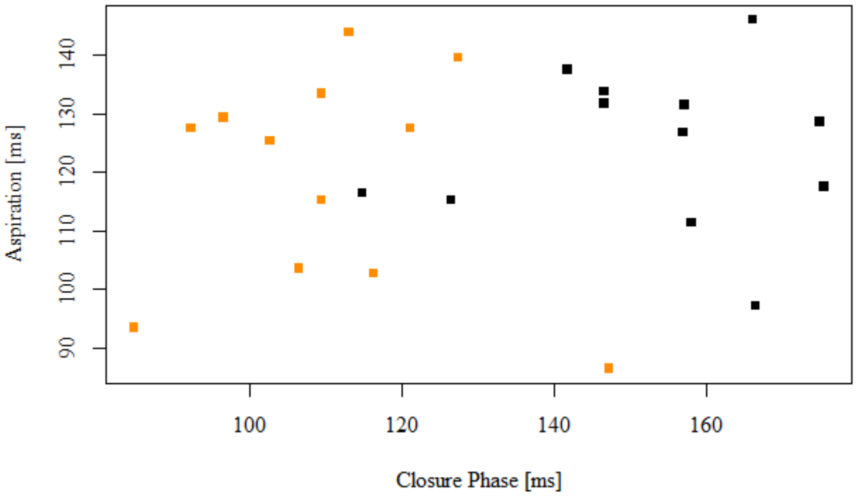
E.1–50. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .34.



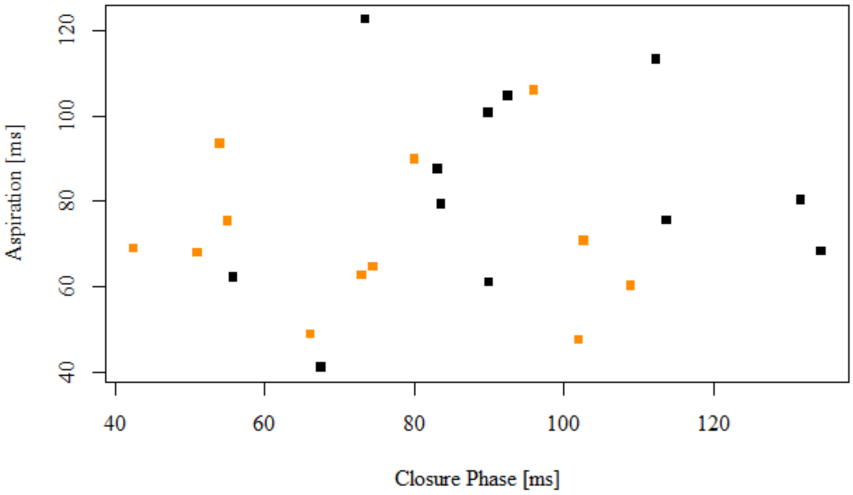
E.1–51. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .71.



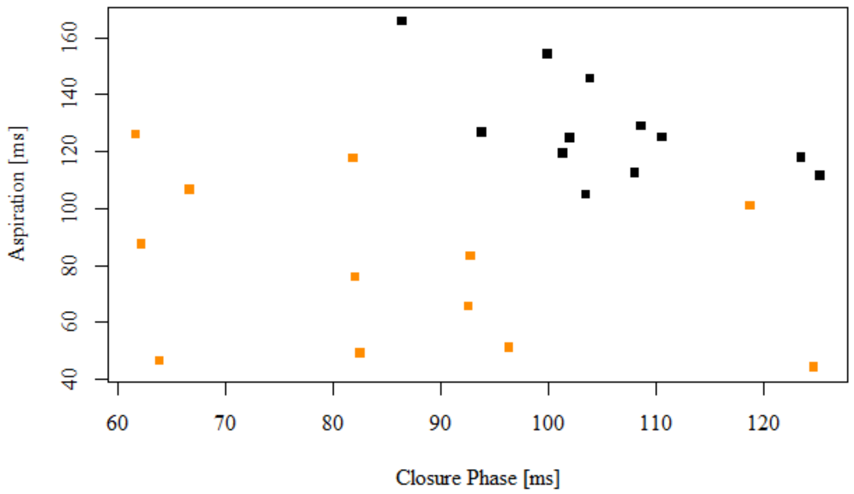
E.1–52. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .29.



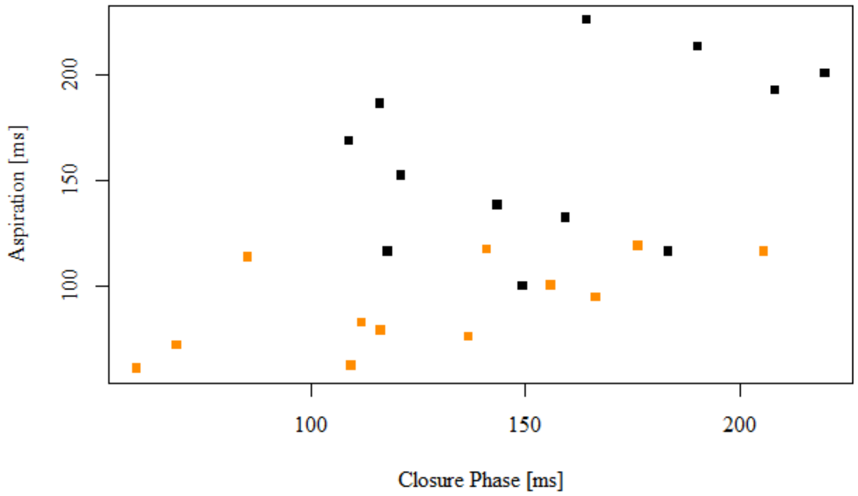
E.1–53. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .65.



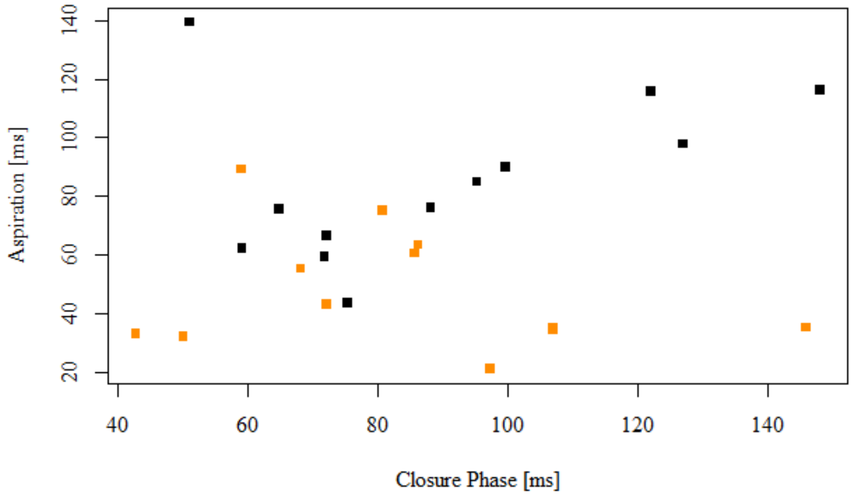
E.1–54. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .20.



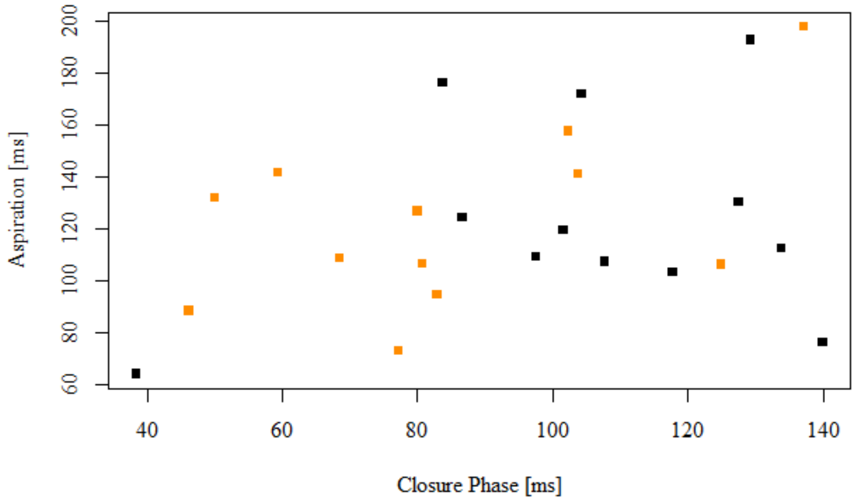
E.1–55. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .71.



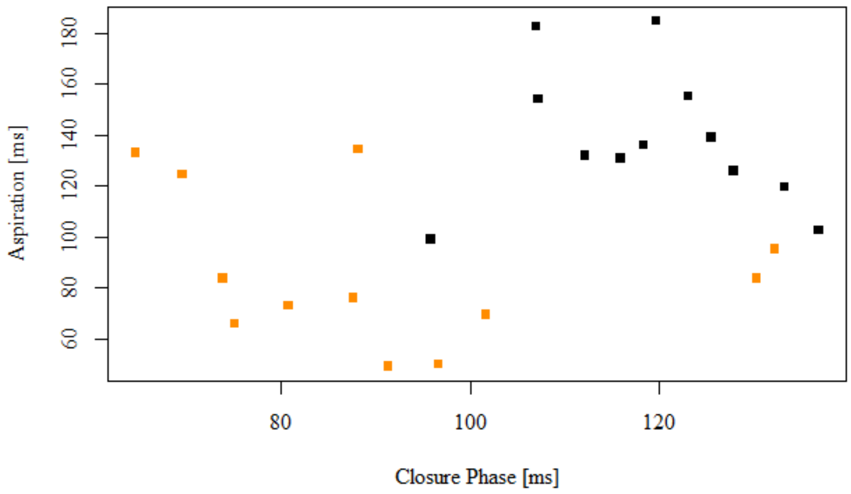
E.1–56. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .55.



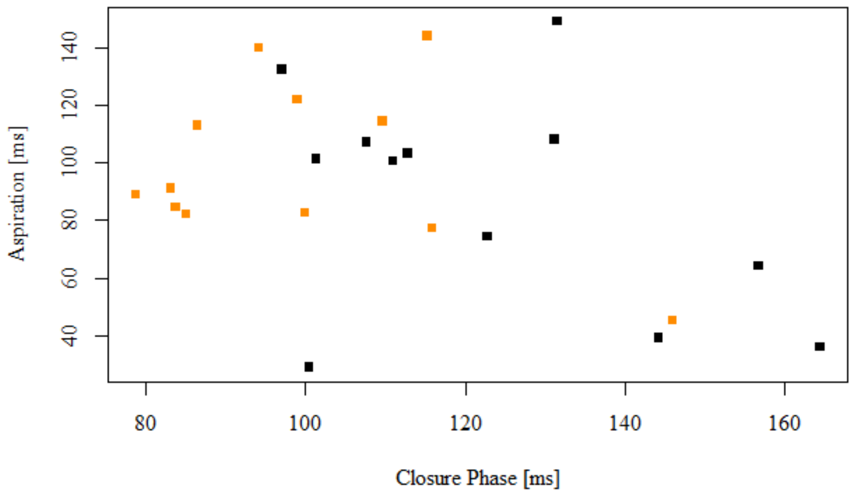
E.1–57. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .39.



E.1–58. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .15.

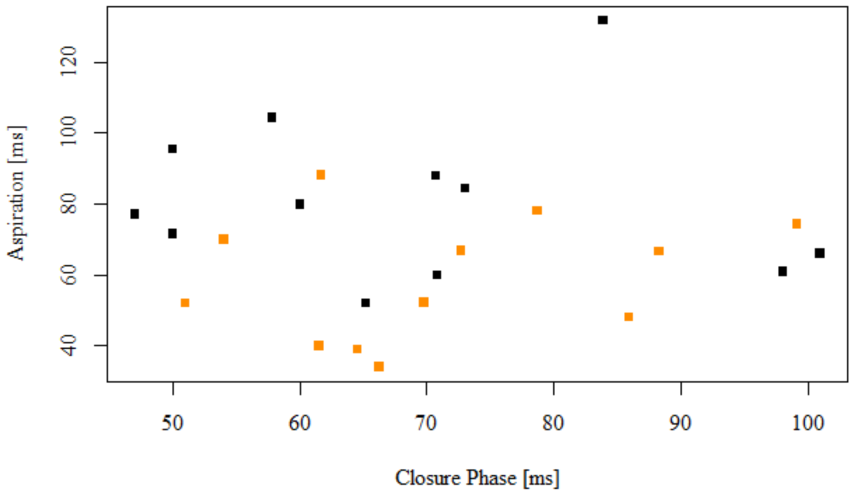


E.1–59. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .67.

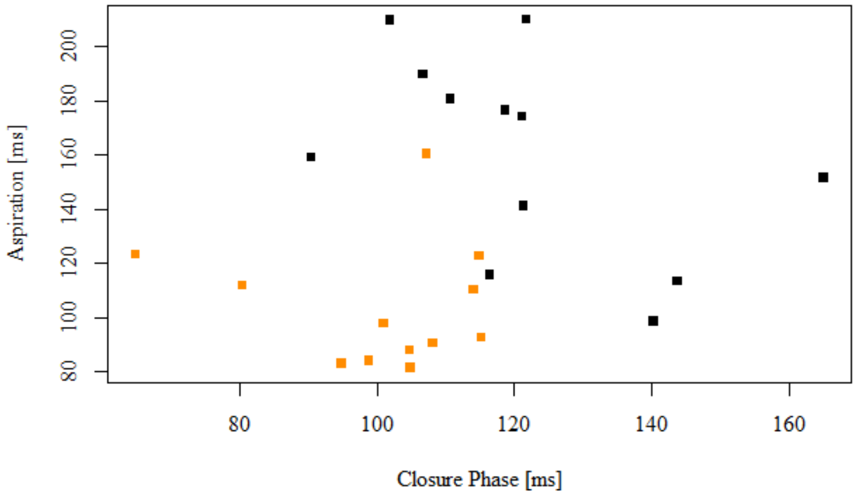


E.1–60. Devoicing scatterplot of speaker #050, corresponding to a Pillai score of .74.

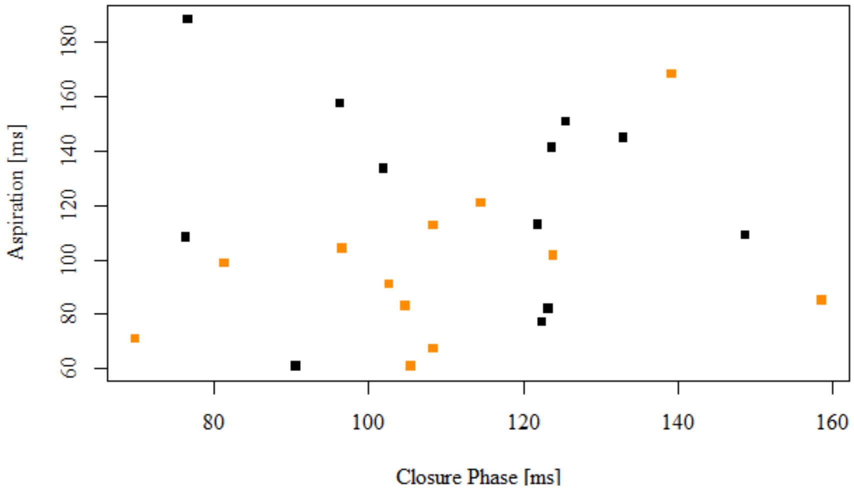
E.2 – Control Group (Groups C-UK and C-US)



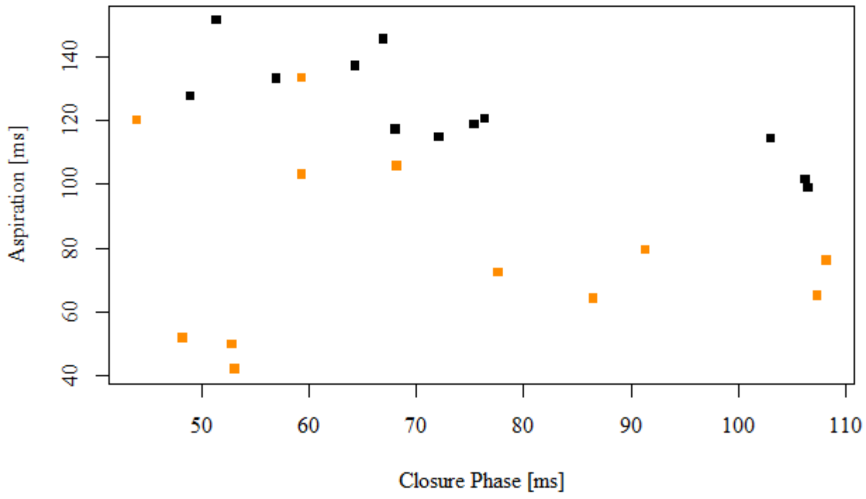
E.2–1. Devoicing scatterplot of speaker #301, corresponding to a Pillai score of .30.



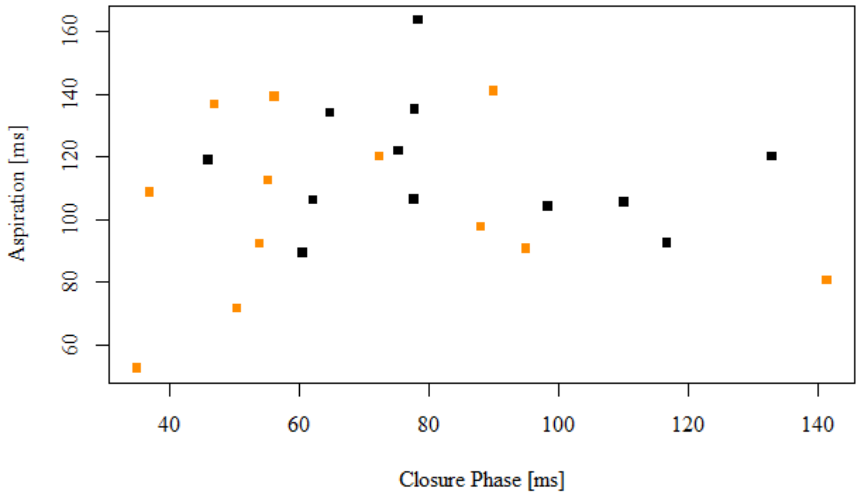
E.2–2. Devoicing scatterplot of speaker #302, corresponding to a Pillai score of .55.



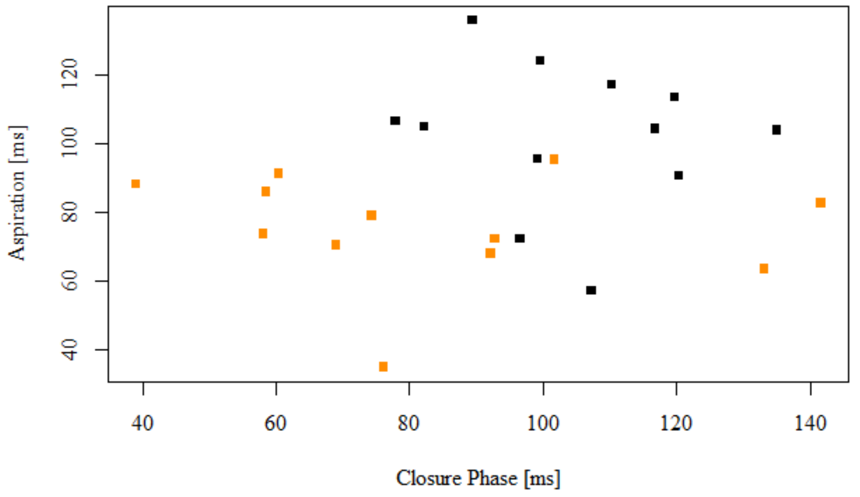
E.2–3. Devoicing scatterplot of speaker #303, corresponding to a Pillai score of .13.



E.2–4. Devoicing scatterplot of speaker #304, corresponding to a Pillai score of .53.

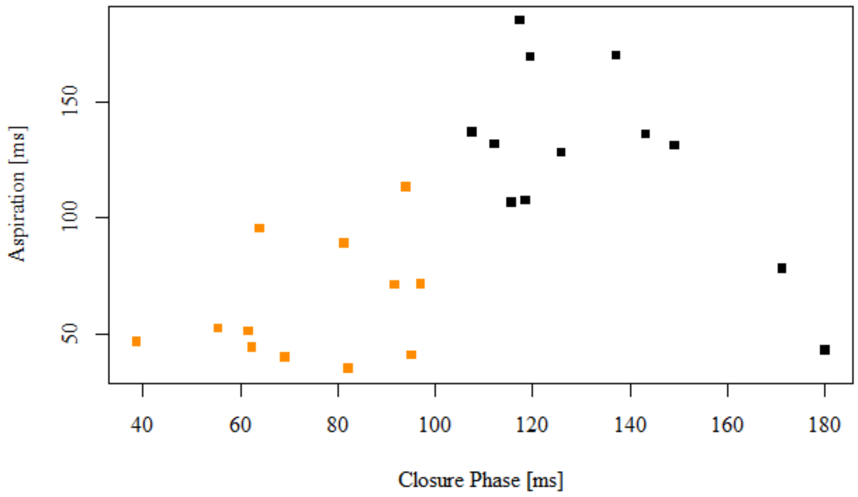


E.2–5. Devoicing scatterplot of speaker #305, corresponding to a Pillai score of .14.

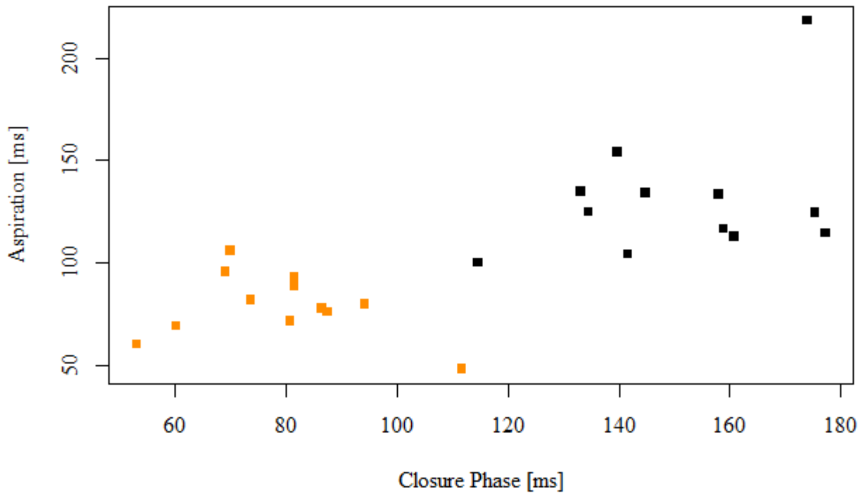


E.2–6. Devoicing scatterplot of speaker #306, corresponding to a Pillai score of .45.

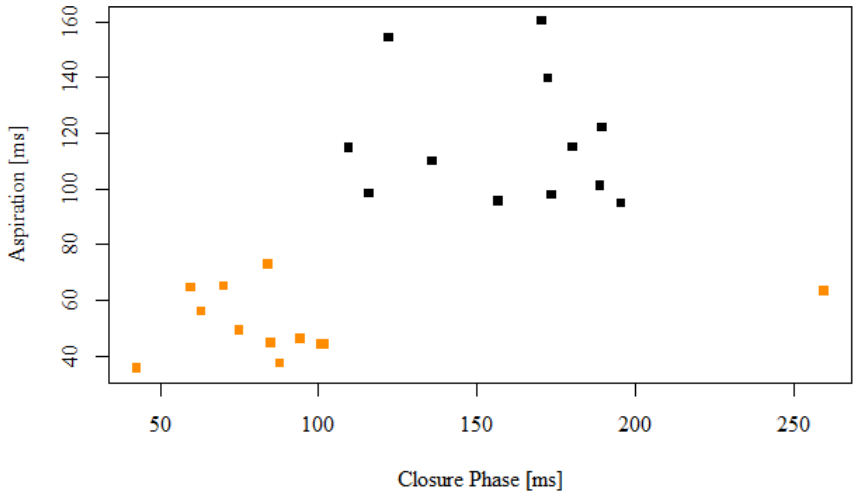
E.3 – British Native Speakers



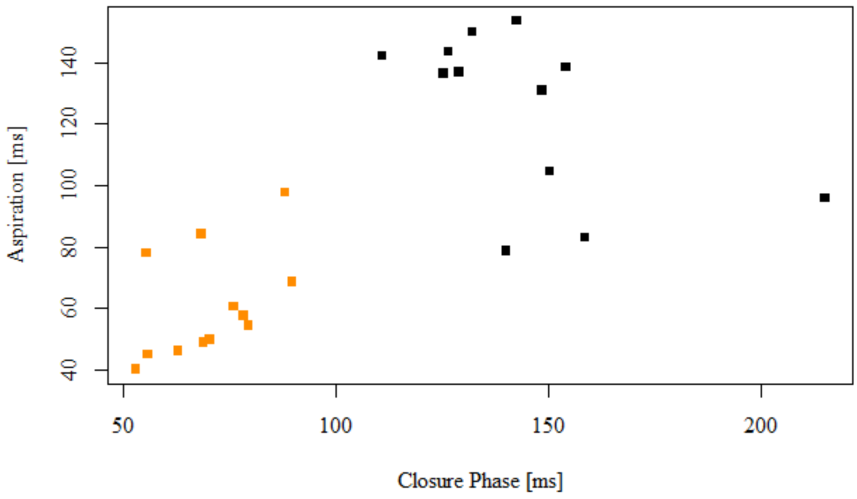
E.3–1. Devoicing scatterplot of speaker #201, corresponding to a Pillai score of .82.



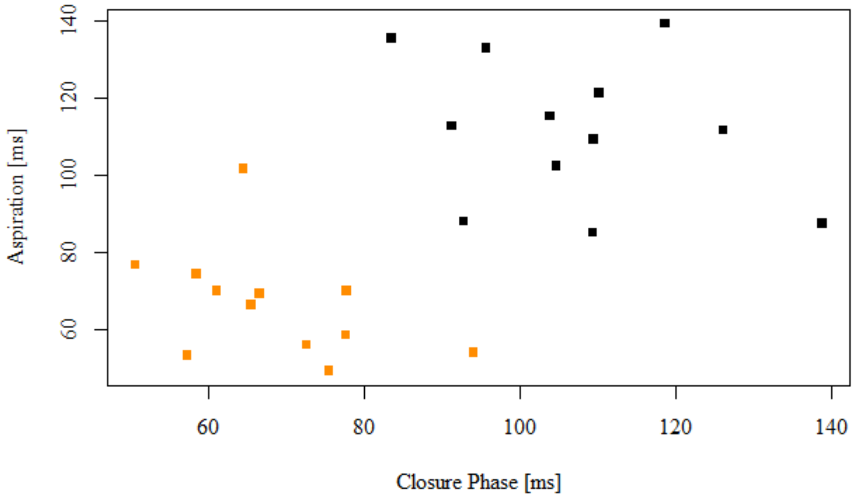
E.3–2. Devoicing scatterplot of speaker #202, corresponding to a Pillai score of .84.



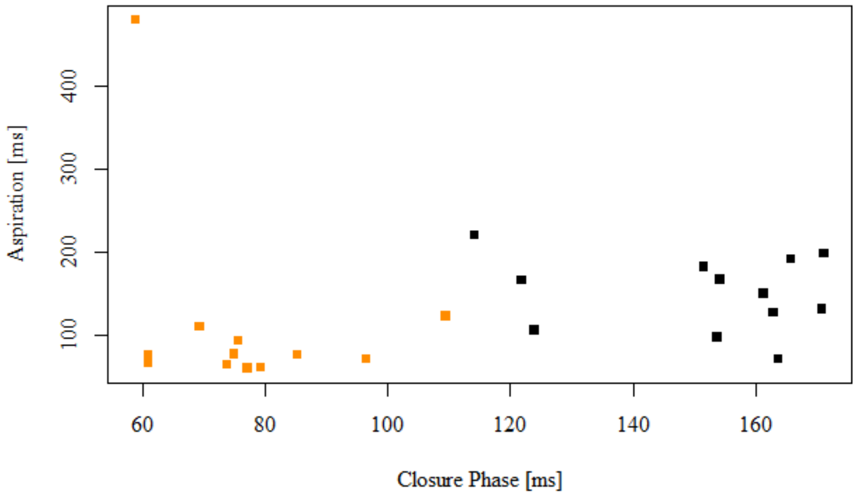
E.3–3. Devoicing scatterplot of speaker #203, corresponding to a Pillai score of .80.



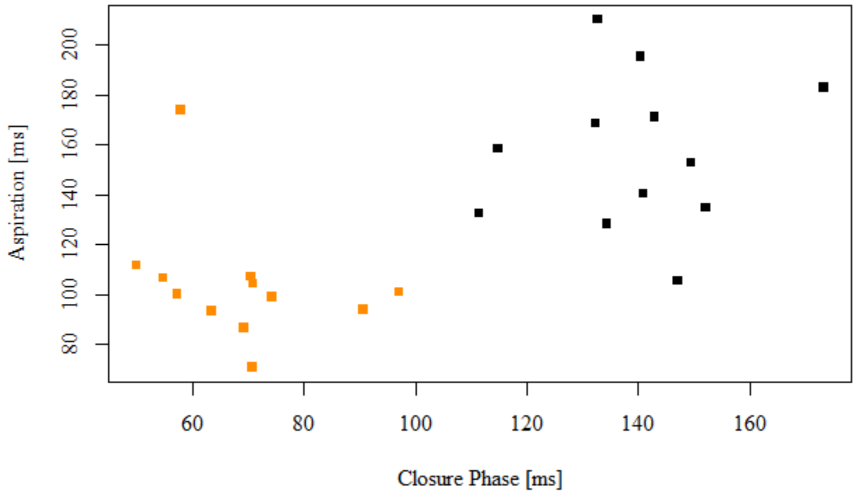
E.3–4. Devoicing scatterplot of speaker #204, corresponding to a Pillai score of .89.



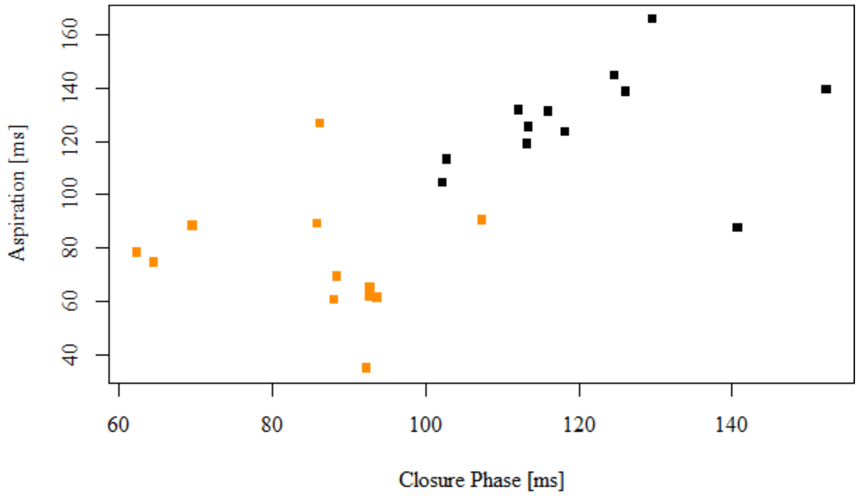
E.3–5. Devoicing scatterplot of speaker #205, corresponding to a Pillai score of .86.



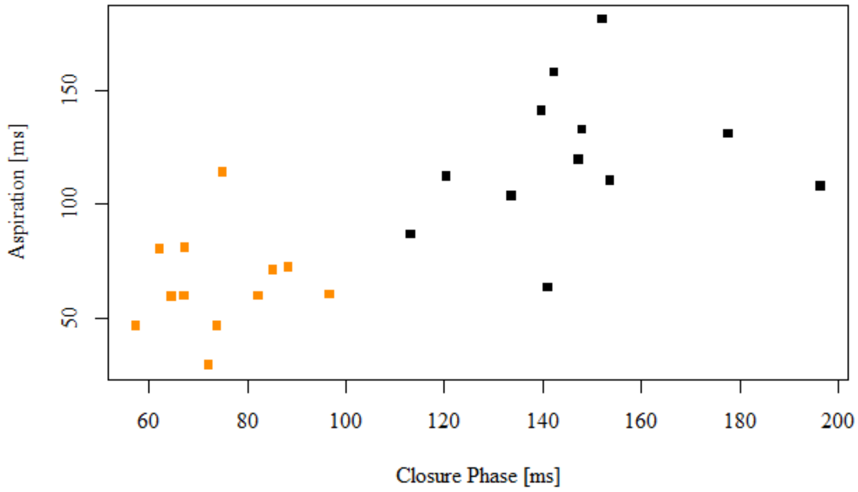
E.3–6. Devoicing scatterplot of speaker #206, corresponding to a Pillai score of .84.



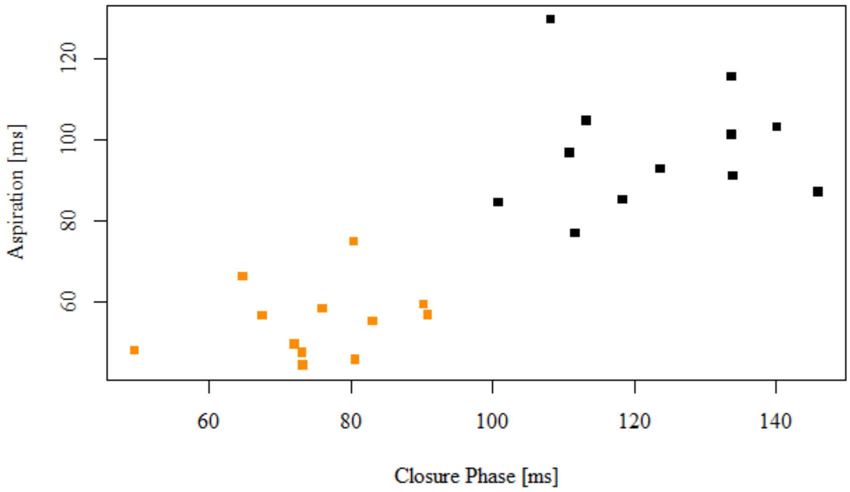
E.3–7. Devoicing scatterplot of speaker #207, corresponding to a Pillai score of .88.



E.3–8. Devoicing scatterplot of speaker #208, corresponding to a Pillai score of .62.

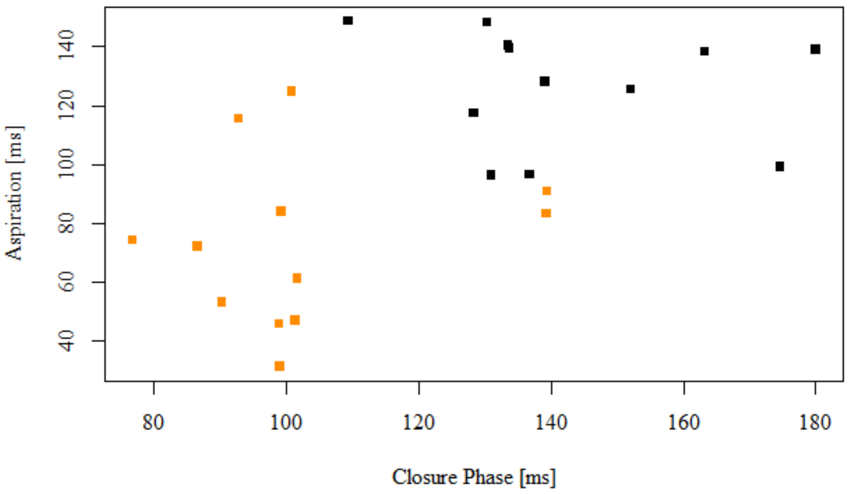


E.3–9. Devoicing scatterplot of speaker #209, corresponding to a Pillai score of .83.

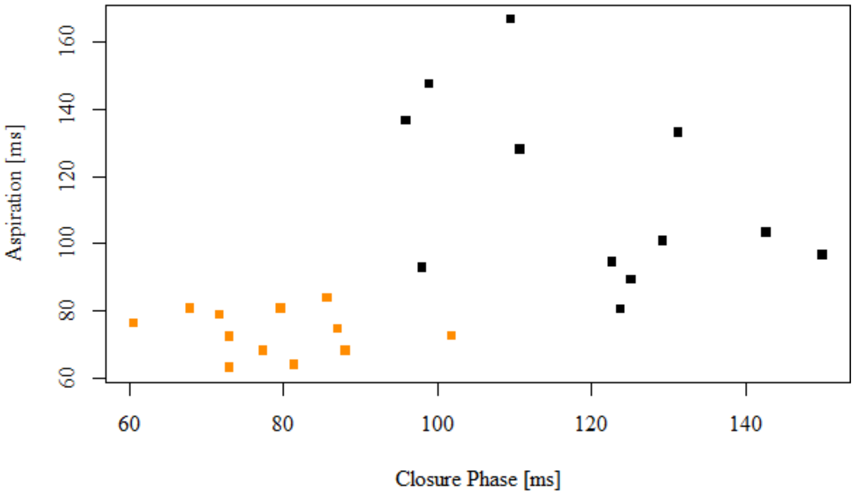


E.3–10. Devoicing scatterplot of speaker #210, corresponding to a Pillai score of .86.

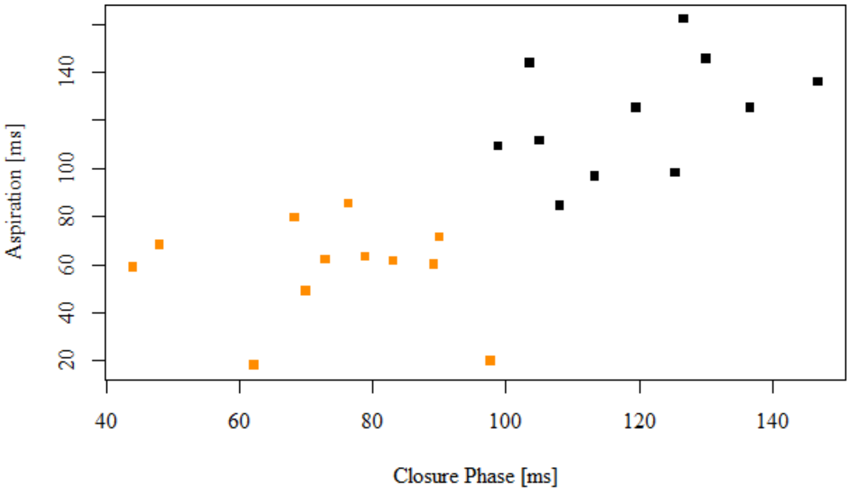
E.4 – American Native Speakers



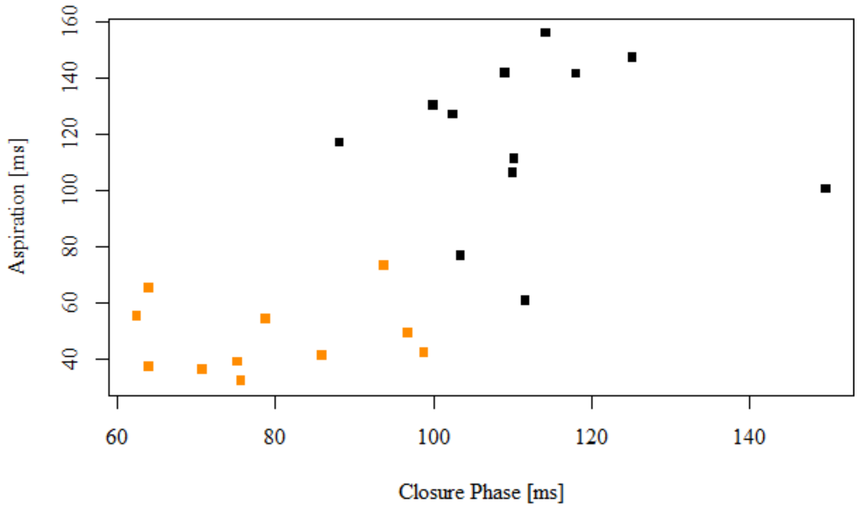
E.4–1. Devoicing scatterplot of speaker #101, corresponding to a Pillai score of .70.



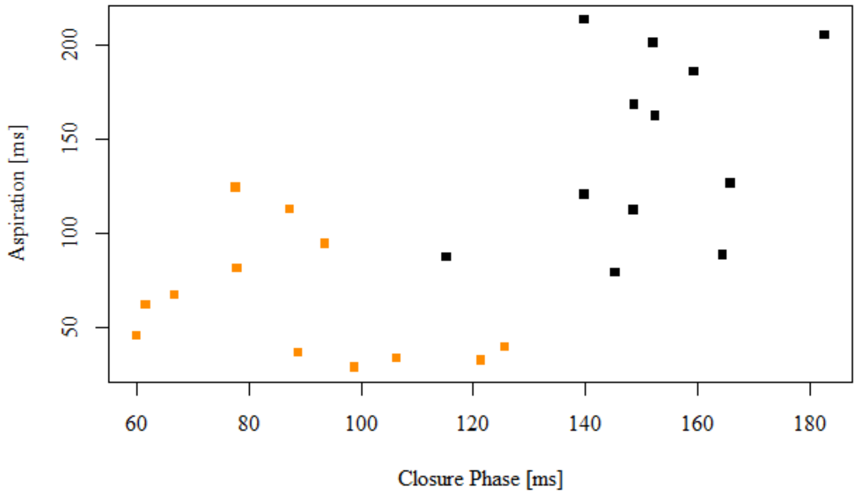
E.4–2. Devoicing scatterplot of speaker #102, corresponding to a Pillai score of .84.



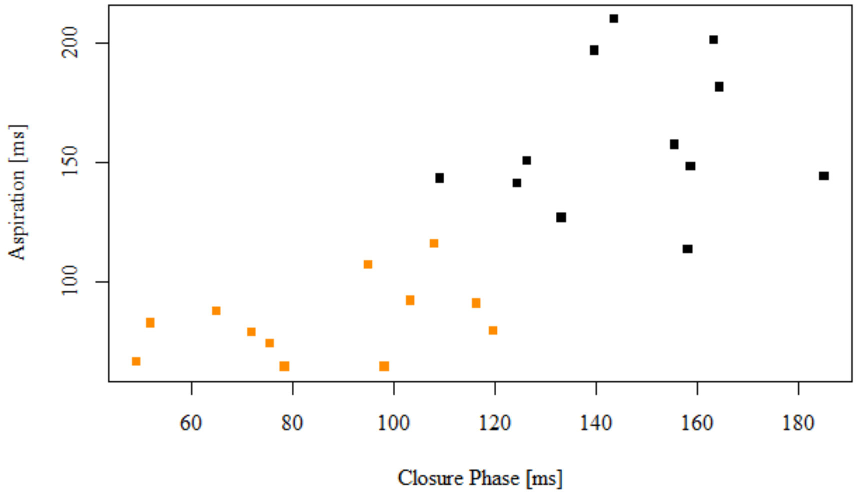
E.4–3. Devoicing scatterplot of speaker #103, corresponding to a Pillai score of .81.



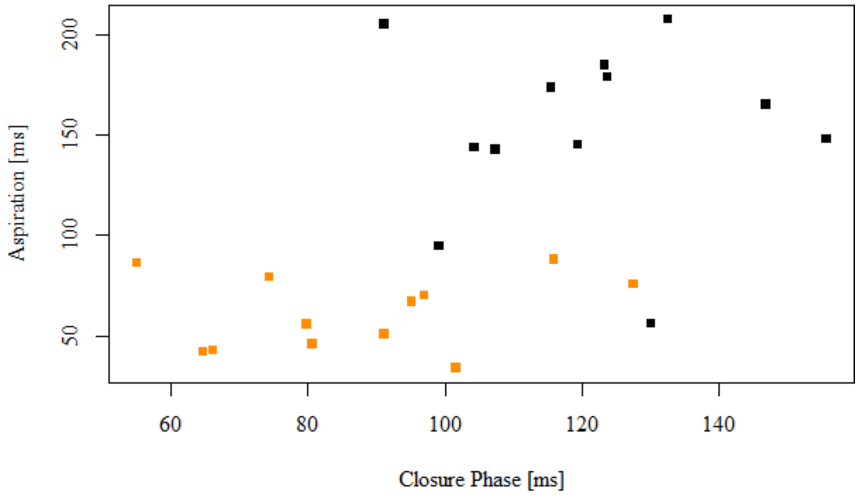
E.4–4. Devoicing scatterplot of speaker #104, corresponding to a Pillai score of .78.



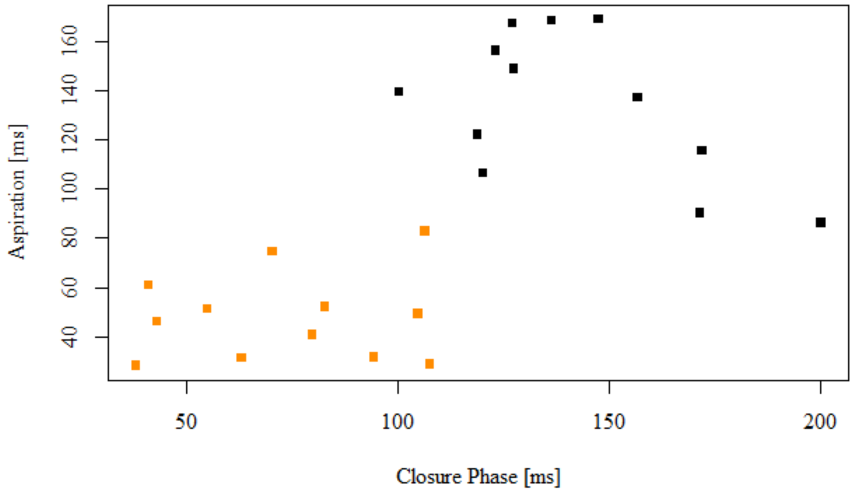
E.4–5. Devoicing scatterplot of speaker #105, corresponding to a Pillai score of .79.



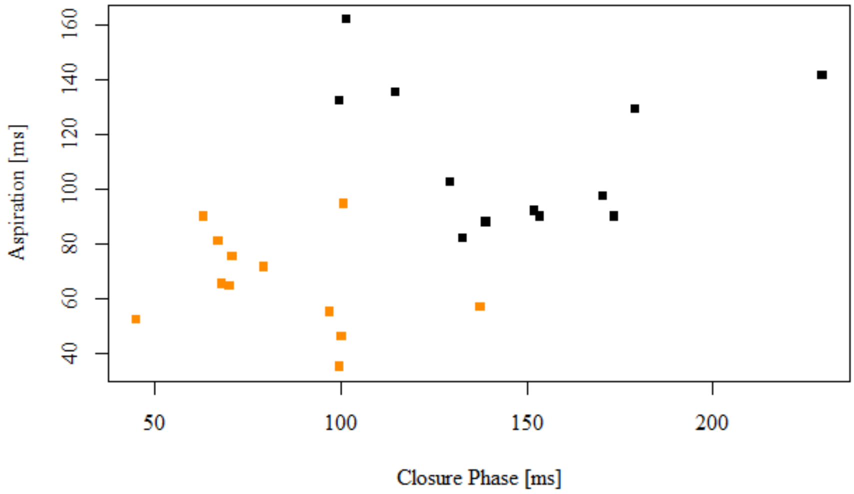
E.4–6. Devoicing scatterplot of speaker #106, corresponding to a Pillai score of .79.



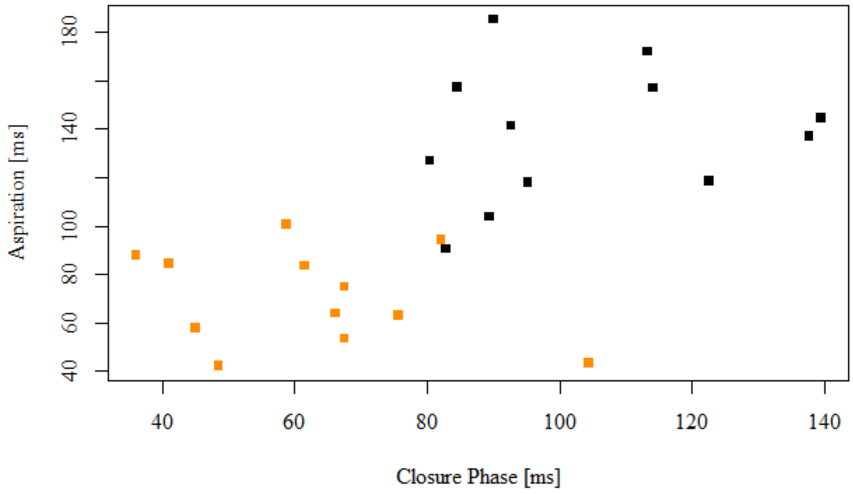
E.4–7. Devoicing scatterplot of speaker #107, corresponding to a Pillai score of .73.



E.4–8. Devoicing scatterplot of speaker #108, corresponding to a Pillai score of .87.



E.4–9. Devoicing scatterplot of speaker #109, corresponding to a Pillai score of .73.



E.4–10. Devoicing scatterplot of speaker #110, corresponding to a Pillai score of .76.

Appendix F – Vowel Length and Vowel Length Variation by the Speakers of the Tokens

F.1 – British English

- F.1–1 KIT
- F.1–2 DRESS
- F.1–3 TRAP
- F.1–4 LOT

F.2 – American English

- F.2–1 KIT
- F.2–2 DRESS
- F.2–3 TRAP
- F.2–4 LOT

F.1 – British English

F.1–1 KIT

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
hit / hid	107.732	135.975	26.22	-20.77
lit / lid	56.54	73.681	30.32	-23.26
rip / rib	107.129	150.602	40.58	-28.87
lip / lib	71.092	107.979	51.89	-34.16
pick / pig	60.691	131.081	115.98	-53.70
brick / brig	118.359	195.666	65.32	-39.51

F.1–2 DRESS

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
bet / bed	12.5874	18.0843	43.67%	-30.40%
wet / wed	104.331	146.76	40.67%	-28.91%
ep / ebb	95.924	162.128	69.02%	-40.83%
dep / deb	128.057	186.716	45.81%	-31.42%
peck / peg	74.077	132.764	79.22%	-44.20%
beck / beg	113.095	197.576	74.70%	-42.76%

F.1–3 TRAP

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
Matt / mad	135.921	221.737	63.14%	-38.70%
bat / bad	133.404	232.113	73.99%	-42.53%
cap / cab	121.517	176.199	45.00%	-31.03%
tap / tab	120.506	177.339	47.16%	-32.05%
lack / lag	112.008	207.374	85.14%	-45.99%
back / bag	134.987	268.815	99.14%	-49.78%

F.1–4 LOT

Word pair	Length preceding voiceless coda	Length preceding voiced coda	VLV increase (%)	VLV decrease (%)
knot / nod	130.014	196.362	51.03%	-33.79%
pot / pod	90.475	133.774	47.86%	-32.37%
cop / cob	136.143	183.012	34.43%	-25.61%
mop / mob	109.991	188.516	71.39%	-41.65%
lock / log	124.478	183.138	47.12%	-32.03%
dock / dog	164,903	241,911	46,69%	-31,83%

F.2 – American English

F.1–1 KIT

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
hit / hid	127.22	168.984	32.83%	-24.71%
lit / lid	106.19	143.886	35.50%	-26.20%
rip / rib	85.29	136.147	59.63%	-37.35%
lip / lib	77.798	123.728	59.04%	-37.12%
pick / pig	69.154	141.398	104.47%	-51.09%
brick / brig	78.488	130.457	66.21%	-39.84%

F.1–2 DRESS

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
bet / bed	115.313	178.381	54.69%	-35.36%
wet / wed	102.173	160.177	56.77%	-36.21%
ep / ebb	101.485	193.444	90.61%	-47.54%
dep / deb	118.768	230.249	93.86%	-48.42%
peck / peg	123.121	205.055	66.55%	-39.96%
beck / beg	126.665	185.165	46.18%	-31.59%

F.1–3 TRAP

Word pair	Length preceding voiceless coda (ms)	Length preceding voiced coda (ms)	VLV increase (%)	VLV decrease (%)
Matt / mad	149.204	254.622	70.65%	-41.40%
bat / bad	127.899	185.638	45.14%	-31.10%
cap / cab	149.67	235.569	57.39%	-36.46%
tap / tab	118.674	235.164	98.16%	-49.54%
lack / lag	161.74	237.664	46.94%	-31.95%
back / bag	152.824	224.159	46.68%	-31.82%

F.1–4 LOT

Word pair	Length preceding voiceless coda	Length preceding voiced coda	VLV increase (%)	VLV decrease (%)
knot / nod	148.057	292.087	97.28%	-49.31%
pot / pod	139.59	244.76	75.34%	-42.97%
cop / cob	88.415	209.573	137.03%	-57.81%
mop / mob	115.734	304.278	162.91%	-61.96%
lock / log	126.785	213.514	68.41%	-40.62%
dock / dog	173.217	303.593	75.27%	-42.94%

Appendix G – Zusammenfassung in deutscher Sprache

Die vorliegende Arbeit befasst sich mit dem Erwerb von subphonemischer Variation bei deutschen Lernern des Englischen. Dies ist bis heute, mit Ausnahme der verschiedenen Realisationen von /t/, weitestgehend unerforscht. Die Relevanz von subphonemischer Variation bei der Verarbeitung von Sprache ist jedoch in der jüngsten Forschung deutlich geworden und daher auch für den Zweitspracherwerb höchst relevant.

Die Dissertation erforscht die Produktion und Perzeption von Vokallängenvariation in den vier englischen Vokalen KIT, DRESS, TRAP und LOT vor den sechs Plosiven /b, p, t, d, k, g/ an insgesamt 66 deutschen Lernen des Englischen und 20 Muttersprachlern. Es werden die beiden großen englischen Varietäten *British English* und *American English* in Betracht gezogen. Im Englischen sind Vokale länger, wenn ihnen ein stimmhafter Konsonant folgt als wenn die Koda von einem stimmlosen Konsonanten gebildet wird. So ist zum Beispiel im Minimalpaar *bad* / *bat* der Vokal im ersten Fall länger. Im Rahmen der Datenerhebung wurde ein Produktionsexperiment durchgeführt, bei dem die Sprecher beim Vorlesen von vier Leseaufgaben mit abfallender Formalität (Minimalpaare, Wortliste, Sätze und ein Text) aufgenommen wurden. Bei dem folgenden Perzeptionsexperiment wurden den Teilnehmern 48 Sätze vorgespielt, bei denen Minimalpaare vorkommen, die sich nur durch die Vokallänge unterscheiden. Hier wurde getestet, ob die Teilnehmer des Experiments Vokallänge wahrnehmen können. Es zeigte sich im Folgenden ein starker Zusammenhang zwischen Perzeption und Produktion, wobei nur solche Sprecher Vokallängenvariation produzieren, die sie auch wahrnehmen können. Dies unterstützt die in der Forschung weitgehend vertretene Meinung, dass die Perzeption der Produktion vorausgeht.

Ein weiterer großer Teil der Arbeit beschäftigt sich mit der Frage, inwieweit phonologisches Können eine Voraussetzung für den Erwerb von subphonemischer Variation ist. Hierbei wurde phonologischer Erfolg daran bemessen, ob die deutschen Lerner a) separate Phonemkategorien für die beiden englischen Vokale DRESS und TRAP hatten und b) Auslautverhärtung zeigten. Diese beiden Merkmale sind für häufig auch für sehr fortgeschrittene deutsche Lerner des Englischen sehr schwierig. Gleichzeitig sind sie für die Vokallängenvariation höchst relevant, da besonders bei einem Zusammenfall der beiden Vokale und gleichzeitiger Auslautverhärtung Minimalpaare im Englischen nicht mehr unterschieden werden können, was zu großen Schwierigkeiten in der Kommunikation führen kann. Die Analyse führt zu dem Schluss, dass diejenigen Lerner des Englischen Vokallängenvariation auf muttersprachlichem Niveau erreichen, die beide Vokale distinktiv aussprechen und keine Auslautverhärtung zeigen. Dies ist direkt relevant für unser Verständnis der kognitiven Repräsentation von Lauten, bedeutet es doch, dass distinktive und subphonemische Merkmale werden gebündelt erlernt werden.

Die Dissertation umfasst fünf Kapitel. Kapitel 1 ist eine Einführung in das Thema und gibt Auskunft darüber, wie wichtig es ist, zu erforschen, warum manche Lerner ein muttersprachliches Niveau erreichen, während dies anderen nicht möglich scheint. Kapitel 2 gibt einen Überblick über den bisherigen Forschungsstand. Hierbei wird auf die Vokallänge und Vokallängenvariation im Englischen und im Deutschen eingegangen. Weiterhin werden Faktoren besprochen, die eine Auswirkung auf die Vokallänge und Vokallängenvariation haben, soweit sie für die vorliegende Arbeit relevant sind. Außerdem wird der Vorgang der Perzeption und Produktion besprochen und es wird erklärt, welche Einflüsse hierbei für Lernende von Bedeutung sind. Speziell bei der Perzeption werden vier Modelle vorgestellt, die erläutern, wie Lernende Phoneme einer Fremdsprache erwerben. Zuletzt werden wichtige phonetische Signale, sogenannte *cues*, diskutiert. Es wird dargestellt, wie spektrale Informationen und die Lautdauer von Vokalen von Muttersprachlern erworben werden und wie deren Gewichtung bei Muttersprachlern verschiedener Sprachen ausfällt. Die Forschungsfragen, die im Rahmen dieser Dissertation bearbeitet wurden, runden Kapitel 2 ab.

Kapitel 3 zeigt die Methodik auf, die in der Studie verwendet wurde. Die Teilnehmer werden vorgestellt und kategorisiert. Im Anschluss werden im Detail das Produktions- sowie das Perzeptionsexperiment erklärt. Danach gibt das Kapitel einen Überblick über die Daten, die mittels eines Fragebogens erhoben wurden. Es wird dargestellt, wie Vokallänge und Vokallängenvariation, die Produktion der Vokale TRAP und DRESS, das Ausmaß der Auslautverhärtung und die Wortfrequenz der in der Studie vorkommenden Wörter gemessen wurden.

Die Resultate und Analyse der Dissertation erfolgt schließlich in Kapitel 4. Nach der Erhebung der Daten wurden diese akustisch vermessen und dann qualitativ und quantitativ untersucht. In diesem Sinne erfasst und erklärt Kapitel 4 zuerst die statistischen Methoden, auf die sich die Analyse stützt. Im Folgenden werden die Resultate des Produktionsexperiments detailliert besprochen und Faktoren aufgezeigt, die zu einem erfolgreichen

Erwerb der Vokallängenvariation führen. Die Resultate des Perzeptionsexperiments schließen sich an. Auch hier werden Erfolgsfaktoren diskutiert. Es zeigt sich, dass Faktoren für die Produktion deutlich leichter festzulegen sind als für die Perzeption, die vor allem durch das Alter beim ersten Kontakt mit der Fremdsprache beeinflusst scheint. Im weiteren Verlauf des Kapitels wird das Verhältnis von Produktion und Perzeption diskutiert und es wird dargestellt, welch großen Einfluss der erfolgreiche Erwerb von phonematischen Unterschieden auf den Erwerb von subphonemischer Varianz hat. Am Ende des Kapitels befindet sich eine Zusammenfassung der Resultate.

Die Diskussion findet in Kapitel 5 statt. Hier werden die Resultate im Hinblick auf die Forschungsfragen analysiert. Die Dissertation kommt zu dem Schluss, dass Vokallängenvariation ein wichtiges Merkmal für deutsche Lerner des Englischen ist und unbedingt erlernt werden sollte. Da die Ergebnisse der Dissertation sowohl für die Soziophonetik und Perzeptionsstudien von großem Interesse sind als auch für die Sprachlehrforschung, wird an mehreren Beispielen aufgezeigt, zu welchem Zeitpunkt und auf welche Weise ein Erwerb der Vokallängenvariation geschehen kann. Ebenso wird die politische Dimension, vertreten durch den „Muttersprache + 2“-Ansatz der Europäischen Kommission, der besagt, dass jeder europäische Bürger zusätzlich zu seiner Muttersprache noch zwei weitere europäische Sprachen beherrschen sollte, in Betracht gezogen. Es wird klar, dass Deutschland von diesem Ziel noch weit entfernt ist und es weiterer Forschung bedarf, die sich auf den (Schul-)Unterricht und die Bedürfnisse von Lernenden konzentriert. Bei den abschließenden Bemerkungen werden Anreize zu weiteren Forschungsmöglichkeiten gegeben.