Sound perception and production in a foreign language: Does orthography matter?

IDEALAB dissertation by

Katharina Nimz

Table of Contents

Figures

Tables

Abbreviations

AOL	*Age of learning*
CPH	*Critical Period Hypothesis*
CAH	*Contrastive Analysis Hypothesis*
F1	*First formant (etc.)*
FLA	*Foreign language acquisition*
GFL	*German as a Foreign Language*
G-P	*Grapheme-to-phoneme (correspondences)*
GSL	*German as a Second Language*
(G)LMM	*(Generalized) linear mixed model*
H1	*Hypothesis 1 (etc.)*
IPA	*International Phonetic Alphabet/Association*
ISI	*Inter-stimulus interval*
L1	*First language*
L2	*Second language*
L2LP	*L2 Linguistic Perception Model*
LOR	*Length of residence*
PAM-L2	*Perceptual Assimilation Model for L2*
RT	*Reaction time*
SD	*Standard deviation*
SE	*Standard error*
SLA	*Second language acquisition*
SLM	*Speech Learning Model*
VOT	*Voice onset time*
VISC	*Vowel inherent spectral change*

1 Introduction

"If language is a window on the mind, speech is the thin end of an experimental wedge that will pry the window open." (Studdert-Kennedy, 1981)

Learning a foreign language is hard. Many who have learnt German as a Foreign Language (GFL)[1] may have felt challenged by the position of the finite verb in main versus subordinate clauses or the many different plural forms German nouns exhibit. But learning the sounds of German, a language with an unusually high[2] number of 15 vowel phonemes, is just as hard as learning its grammatical structures – possibly even harder (Hirschfeld, 2003: 873). This is especially true for speakers of a language with considerably fewer contrastive vowels, such as Polish. The following utterance by a Polish learner of L2 German exemplifies this issue: *Kann ich den Schrank in R<u>a</u>tten zahlen*? ("Can I pay for this closet in rats?") The Polish learner shared this anecdote in a linguistics seminar at Humboldt-University, Berlin, in order to illustrate the persistent challenge Polish speakers face with German vowels. By producing the troublesome segment in the word *R<u>a</u>ten* [ʁaːtən] ("instalments") too short, the Polish learner found herself in an undesired communication situation, talking about rats while in fact wishing to discuss instalments. Cutler (2015) points out that difficult L2 segmental contrasts are not only problematic for minimal pairs, as was the case in the *Raten-Ratten* example. Another serious problem is that indistinguishable L2 segments can lead to temporary ambiguity among L2 words. This is caused by (spurious) initial overlap of words, which lead to additional processing costs for L2 learners. For example, Dutch learners have difficulties with English /ɛ/ in *pencil* versus /æ/ in *panda*. Cutler et al. (2006)

[1] In the German literature, a rather rigid distinction is made between German as a *Foreign* Language (GFL), with a focus on the acquisition of the language in a classroom setting, and German as a *Second* Language (GSL), with a focus on the acquisition the language by immigrants in a naturalistic setting. In English, both kinds of learning settings are often subsumed under the cover term "second language (L2) acquisition" (Ellis, 1985: 5). In this thesis, too, GFL and L2 acquisition will be used interchangeably. Furthermore, the term GFL refers to an interdisciplinary field of research with ties to areas as diverse as pedagogy, social sciences, literary and cultural studies, and applied linguistics (Barkowski and Krumm, 2010).

[2] Most languages of the world differentiate five to seven vowel phonemes (Maddieson, 1984: 128).

found that when these learners heard *panda*, they were likely to look at the competitor *pencil* too and consequently were slowed down in their language processing. The investigation of L2 speech learning is therefore not only of theoretical interest, but can potentially help learners with serious problems in the acquisition of a foreign language.

This study focuses on the perception and production of German vowels by Polish GFL learners. It draws on different areas of research, most prominently on GFL research and L2 phonetics and phonology. While some "intellectual imperialists" (Pierrehumbert, 1990: 375) have taken the view that phonetics and phonology cannot be integrated, the field of L2 phonetics/phonology does not seem to be concerned with this traditional divide. The reason why the distinction is not as fundamental in L2 research may be that much of the experimental work in the field makes use of phonetic methodologies, such as acoustic analyses, but at the same time takes interest in the sounds that are often also those sounds which are contrastive in the L2 phonological system. The field of L2 sound acquisition can therefore be understood as an area of research where Ohala's call for the integration of phonetics and phonology is put into practice (Ohala, 1990).[3]

As far as GFL research is concerned, Grotjahn (1998) points out the need to incorporate findings from L2 phonetics and phonology into the applied field of GFL pronunciation teaching. He criticizes practices that promote various didactic recipes without taking latest empirical findings into account. Most studies in L2 speech acquisition emphasize the perceptual side of the learning process (Bohn and Munro, 2007: 9), while publications in the field of GFL tend

---

[3] It is worth noting that the terms "phonetic" and "phonological" are not only used to refer to separate disciplines. They also describe different levels of sound representation, as it is for example crucial in generative approaches to phonology (Chomsky and Halle, 1968). The idea is that an underlying *phonological representation*, made up of bundles of phonological features, is transformed by phonological rules into a surface *phonetic representation*, for example the underlying German word form /kɪnd/ ("child") into the surface phonetic form [kʰɪnt], due to the phonological rules of aspiration and final devoicing. This understanding of "phonological" and "phonetic" is applicable when describing phonetic and phonological contrastive analyses. The usage of the different brackets relates to this distinction. Throughout this dissertation, square brackets are used for phonetic representations and slant brackets are used for phonological representations.

to base their findings on production data (e.g. Morciniec, 1990, Slembek, 1995, Müller 2005). It seems therefore worthwhile bringing the findings and hypotheses from both fields together.

Most research in the field of experimental phonetics and L2 phonology has been done on English as an L2 (Hayes-Harb, 2012). Even though some findings on English may be transferable to other languages, experimental research which specifically addresses topics in *German* L2 sound acquisition is still rare (but see Dieling, 1983, Richter, 2008, Darcy and Krüger, 2012, and Darcy et al., 2013). This seems unfortunate since English may not always be the best candidate to investigate current issues in the field. For example, the acquisition of vowel length[4] in L2 English has attracted much attention (e.g. Bohn, 1995, Bogacka, 2004, Cebrian, 2006, Rojczyk, 2011) even though this feature of the vowel system plays only a secondary role in English (Hillenbrand et al., 2000). This is different for German as an L2, especially in the case of lower vowels (Sendlmeier, 1981). For example, in the [ʁatən] ("rats") versus [ʁa:tən] ("instalments") minimal pair, it is primarily the difference in vowel length which differentiates the German phonemes /a/ and /a:/.

Flege (1999: 1275) has pointed out the need to focus on more than one acoustic dimension when investigating L2 sound acquisition. In the case of German vowels, it is not only the length of vowels which is of interest, but also the quality of vowels (i.e. tense versus lax). Hence, both dimensions will be addressed in this study. In the case of the perception experiments, this implies that stimuli will be manipulated both in their length and in their quality. For the production experiment, this means that both vowel duration and formant values are of interest in the acoustic analysis of the present data.

[4] As Fox (2000: 22) has pointed out, the term "length" is often applied indiscriminately in phonetic and phonological sense. The term vowel "duration" is used for a description of phonetic length only. It can be measured in milliseconds and may not have any phonological (i.e. contrastive) implications. In this study too, the term vowel length is used generically. At times where the differentiation between *phonetic* and *phonological* length is important, it will be specified accordingly (s. also Footnote 3 above). The term "quantity" is used for the phonological sense of length only.

In the reprint of Ellis' (2008) monumental work *The Study of Second Language Acquisition* only about 20 pages (out of over 1100 pages) are dedicated to the topic of L2 phonetics and phonology. Yet he states that there has been a considerable growth of interest in L2 phonology in recent years (Ellis, 2008: 103) and various publications of anthologies dedicated to the field of L2 speech learning reflect this trend (e.g. Strange, 1995, James and Leather, 1997, Bohn and Munro, 2007, Edwards and Zampini, 2008). One of the most recent and conspicuous factors of interest in the field is the influence of orthography in the acquisition of an L2 phonology. As early as in 2002, Young-Scholten (2002: 264) noted that "we know very little about the influence of written input on the development of a second phonology. Studies which compare L2 children and adults or L2 adult groups neither control for this input variable nor do they treat it as an independent one". About a decade later, research does begin to take orthography into account. For example, in 2015, *Applied Psycholinguistics* published a special issue on this topic, but findings are not straightforward. While some researchers find orthography to have a positive influence on L2 sound acquisition (Rafat, 2015), others find no results at all (Showalter and Hayes-Harb, 2015), and still others report on negative influences of orthographic input (Bassetti and Atkinson, 2015). The latter's research looks at the influence of orthography in acquiring L2 English, i.e. in a writing system[5] known for its idiosyncrasy. For example, the marking of vowel length is relatively opaque, and both phonetically short and long vowels may be marked by double vowel graphemes, as in *hood* [hʊd] and *food* [fu:d]. In German, vowel length is marked more systematically, for example, by the *lengthening h* in the word *Sahne* ("cream"). However, not every long vowel is explicitly marked, as demonstrated in the above mentioned example *Rate* /ʁa:tə/. This optional but coherent marking of vowel length in German allows for a set-up of an experimental design that would not be possible in English. Most current theories in L2 sound

[5] The term "writing system" has two distinct meanings. On the one hand it refers to the specific system of a given language, thus, there are as many writing systems as there are written languages. On the other hand, it refers to a few types of systems, such as logographic, syllabic, or phonetic writing systems. Here, the term is referring specifically to the English writing system, i.e. the former meaning. The term "orthography" is even more specific in that it refers to the *standardized* variety of a language-specific writing system (Coulmas, 2003: 35).

learning would propose that the quoted Polish GFL learner produced *Raten* as *Ratten* because she perceived it as such (see the following chapter). This study asks the question whether orthography may also play a role in this process, as the absence of the *lengthening h* suggests that the vowel might also be short. In order to investigate this question experimentally, two groups of speakers – 20 Polish GFL learners and 20 native speakers of German – took part both in a perception and a production task involving orthographic marking as an experimental variable. A third discrimination experiment was conducted to investigate the perception of German vowels without any possible influence of orthography.

The study is organized as follows. First, prominent theories of L2 speech learning are reviewed, and current hypotheses regarding L2 vowel length acquisition are discussed. Studies which relate L2 sound acquisition and L2 lexical representation are also reviewed, since the current study uses real German words to investigate the phenomenon under discussion (Chapter 2). Following the general overview, extant studies addressing the role of orthography in L2 phonological acquisition will be presented. In light of this, the orthographic systems of German and Polish will be described in detail (Chapter 3). The next chapter introduces the German and Polish vowel systems. The section on Polish includes two exploratory studies, which investigate Polish vowel duration before singleton versus geminate consonants, as well as Polish vowel quality and average Polish vowel duration (Chapter 4). The following three chapters report the main experiments of the study: a discrimination experiment with manipulated nonce words (Chapter 5), a production experiment with real German words which are either explicitly marked or unmarked for their vowel length (Chapter 6), and an identification experiment which includes the same words as the production experiment (Chapter 7). Chapter 8 summarizes the results, discusses how the three studies relate to each other, and illustrates their relevance for foreign language classroom. The thesis closes with an overall conclusion.

2 Theories of L2 speech learning

A central theme to all theories of L2 speech learning is the role of the native language (L1) in learning a second language. According to Bohn (1998), Polivanov (1931/1974) was the first to describe the influence of the L1 on the perception and production of foreign sounds. In his article on "The Subjective Nature of the Perceptions of Language Sounds" he gives various anecdotal examples of foreign language learners who add or substitute sounds inaccurately (e.g. Japanese speakers producing Russian [tak] ("thus") as [taku] or Russian speakers producing French [lɛ] ("milk") as [le]). Polivanov hypothesized that the learners' deviant productions are due the "subjective nature" of the perception of sounds, which depends on "the complex of language habits attained by every given individual in the process of mastering his mother (native) tongue" (Polivanov, 1974: 231). Later on, Trubetzkoy (1939/1989) coined the term "phonological sieve", which became a widely used metaphor for the idea that our L1 phonology filters out those properties of the L2 speech signal which are not relevant to our L1 phonological system. Because of this, learners of an L2 may perceive and produce foreign sounds erroneously. While neither Polivanov nor Trubetzkoy formulated a comprehensive theory of L2 speech learning, the following sections will present prominent theories and hypotheses in this field, which are relevant for the current study. These include the *Contrastive Analysis Hypothesis* (Lado, 1957), the *Speech Learning Model* (Flege, 1995), the *Perceptual Assimilation Model for L2* (Best and Tyler, 2007), and the *Linguistic Perception Model* (Escudero, 2005). Furthermore, two hypotheses will be presented which address the acquisition of vowel length (Bohn, 1995 and McAllister et al., 2002, respectively). Finally, recent research into L2 lexical-phonological representation will be discussed (i.e. Hayes-Harb and Masuda, 2008, Darcy et al., 2012, Darcy, 2014). While all previous models focus on the acquisition of sounds alone, this last section of the chapter is concerned with the relationship between L2 speech learning and the phonological content of L2 learners' lexicons. This research is important in light of the production and identification tasks of this study (Chapters 6 and 7), as these tap into the nature of L2 lexical representations.

2.1 Contrastive Analysis Hypothesis

Related to the metaphor of the above mentioned "phonological sieve" is a famous idea about how structural similarities and differences in two languages may predict difficulties in an L2. In 1957, Lado put forward his Contrastive Analysis Hypothesis (CAH), which states that "those elements that are similar to [the learner's] native language will be simple for him, and those elements that are different will be difficult" (Lado, 1957: 2). These elements could be phonological, grammatical, lexical, or graphemic in nature.[6] In his chapter on phonological structures, Lado identifies the similarities and differences between L1 and L2 *phonemes* and their distribution as the core task in conducting a phonological contrastive analysis. To him, the easiest phonemes to learn are those sounds which are "physically similar to those of the native language, that structure similarly to them, and that are similarly distributed" (p. 12). While most researchers interpret the CAH as a purely phonemic analysis (e.g. Best and Tyler, 2007, Steinlen, 2009), Lado does, in fact, raise the issue of positional allophones (pp. 13-17). He discusses the acquisition of English /d/ by Spanish learners, and points out that it is important to consider that Spanish employs phonetically different positional variants of /d/. For example, Spanish makes use of a stop variant [d], which is comparable to English [d], and a fricative variant [ð]. Because of the specific distribution of the Spanish /d/ allophones, Lado predicts that Spanish speakers will produce [ð] between vowels in English as their L2. Yet, Lado does not cite any empirical evidence for this prediction.

In his book, Lado stresses the great influence the native phonological system has on learning an L2, and maintains that it is extremely difficult to change anything in that system due to the fact that the native phonology of a language operates as a "system of automatic [...] habits" (p. 11). This wording and his idea that learners transfer their entire native language system when learning a foreign language could explain why contrastive analysis is closely associated with behaviourist theories of L2 learning (Ellis, 2008: 350). Research into the acquisition of L2 *syntactic* structures showed relatively early on that the CAH

---

[6] Lado (1957) even calls for contrastive analyses of cultures. Yet, this aspect of his approach was not taken up, and contrastive analyses focus on the comparison of languages (Lennon, 2008: 51).

had limited powers in explaining errors made in an L2 (see for example Schachter's (1974) research on the acquisition of English relative clauses by learners of different language backgrounds). This is because learners may consciously avoid syntactic structures which are difficult for them, while "there is no such thing as phonological paraphrase, and therefore the avoidance phenomenon is difficult, if not impossible [when it comes to phonology]" (Schachter, 1974: 212). It is therefore not surprising that contrastive analysis is still a prominent methodology in the field of GFL phonetics today (see for example Slembek, 1995 or Hirschfeld, 2005a).

While Lado focused on the comparison between Spanish and English, publications in the field of GFL naturally compare the German sound system with those of other languages. As far as contrastive analyses between German and Polish are concerned, German vowels stand out as a prominent problem for Polish learners of L2 German. Slembek (1995: 43), in her contrastive analysis of the German and Polish phoneme systems, comes to the conclusion that the greatest difficulties for Poles learning German are the German vowels, as Polish does not have any long vowels. Morciniec (1990: 27) also considers the phonological contrast between German short, lax and long, tense vowels a serious learning problem and points out that Polish speakers tend to substitute both vowels of a pair with their Polish equivalent. In his introduction, he briefly mentions that orthography may also play a role in certain phonological substitutions, but he does not specify how. Similarly, Müller (2005) notes that the long, tense vowels are often produced short and open, while the German short, lax vowels may sometimes be produced exaggeratedly long by Polish speakers. Furthermore, she reports that the production of German /e:/ by Poles can be realized as [ej] (Müller, 2005: 13). Hirschfeld (1998: 117) and Dieling (1992: 98) mention the realization of /e:/ as [ej]/[ei] as well, but do not give any suggestions as to why this might be the case. In fact, neither this nor the exaggeration of German short vowels as long segments would be predicted from a phonological contrastive analysis of the two sound systems, which suggests that some researchers in the field of GFL might confound their analyses with errors encountered in the foreign language classroom. While this does not necessarily devalue such works in the context of foreign language teaching, it is difficult to evaluate the theoretical implications of these analyses. Furthermore, it seems problematic that most contrastive analyses in the field of GFL take place on a broad phonemic level, while it may be important to consider more detailed phonetic differences between the languages of interest (e.g. Nimz, 2014).

Neither Lado nor most publications that employ contrastive analyses problematize how orthography might influence the acquisition of L2 phonology. While Lado does dedicate a chapter to the comparison of two writing systems, his focus is on reading and writing and not on phonology per se. In a short passage within his phonology chapter, Lado mentions that it may be possible that Spanish L2 English learners produce English *pine* as /pinɛ/, i.e. English /aɪ/ as /i/. He explains this error by the fact that some words are cognates in Spanish and English (such as *pine* and *pino*), and notes additionally that "spelling is obviously a factor in determining the kind of substitution that is made" (Lado, 1957: 21). Like most researchers in the past and present, he does not elaborate on this idea further.

2.2 Speech Learning Model

The Speech Learning Model (SLM) by James Emil Flege (1995, 2002) is at present considered the most influential model in the field of L2 speech learning (Gut, 2009: 2). Because of its explicit formulation of hypotheses concerning the perception and production of L2 segments, it provides a prominent testing ground for studies in the field. The model was developed to "account for age-related limits on the ability to produce L2 vowels and consonants in a native-like fashion" (Flege, 1995: 237). To this end, it attempts to explain the well-known phenomenon of the earlier, the better, i.e. that the earlier an L2 is learned, the more likely it is for a learner to perceive and produce L2 sounds native-like. The SLM consists of four postulates and seven hypotheses, which are summarized as follows.

Flege's four postulates state that, first, mechanism used in L1 speech learning remain intact over the life span and can be applied to L2 speech learning. Second, speech sounds – or "phonetic categories" – are specified in long-term memory, and third, these phonetic categories established in childhood evolve over the life span. They are affected by all L1 and/or L2 sounds which are identified to belong to the respective category. Fourth, speakers of more than one language try to maintain a contrast between L1 and L2 phonetic categories, which exist in a common phonological space. Based on these postulates and related research conducted in the 80s and early 90s (e.g. Flege, 1980, Flege, 1987, Flege, 1991, Flege, 1993), Flege formulated a number of hypotheses concerning the acquisition of L2 speech by young and adult learners.

The first hypothesis (H1, etc.) of the SLM concerns the perceptual relationship between L1 and L2 sounds. Contrary to most interpretations of the CAH, sounds are assumed in the model to relate on a position-sensitive allophonic level rather than a more abstract phonemic level.[7] Flege cites evidence from Japanese learners of L2 English (e.g. Sheldon and Strange, 1982), who have been found to be more accurate in perceiving and producing the (for Japanese) difficult English /r/-/l/ contrast in word-final rather than in word-initial position. H2 states that L2 learners can establish new L2 phonetic categories as long as the L2 sound is perceived as phonetically different from the closest L1 sound. For example, Flege et al. (1996) found that experienced Japanese learners of L2 English were able to identify English /ɹ/ native-like (unlike /l/) because it was perceptually more different from Japanese /r/ than /l/. Related to H2 is the notion expressed in H3, which proposes that the greater the perceived dissimilarity between the L2 and the closest L1 sound, the more likely it is that the sounds will be differentiated. This implies that L2 sounds which are similar to the respective L1 sounds are the most difficult to discern and to establish a new phonetic category for, as shown for the case of English /l/ for the Japanese speakers in the study by Flege et al. (1996). This is a crucial difference to what has been assumed in the CAH; for Lado, similar sounds are the easiest to learn while for Flege, they are the most difficult to acquire.

According to H4, the likelihood of cross-language phonetic differences being discerned decreases with the age of learning (AOL). In a large-scale study, Flege

[7] Interestingly, Flege (1995: 238) cites Lado (1957) when pointing out that the SLM is less abstract than the CAH. However, Lado does take positional allophones into account, as discussed above, so his level of analysis is not quite as abstract as most researchers assume it to be.

et al. (1995) investigated the degree of perceived foreign accent in Italian L2 English speakers who differed in their AOL. They found that the later in life the Italian speakers started to learn their L2 English, the more strongly their accent was rated as foreign by native speakers of English. Because the relationship was linear, with no sharp discontinuity in pronunciation ability around the age of puberty, Flege (1995, 2002: 238) interpreted the data as proof against a strong version of the Critical Period Hypothesis (CPH)[8]:

> *"According to the SLM [...], L1 vowel and consonant categories become more powerful attractors of the vowels and consonants encountered in an L2 as they develop through childhood and into adolescence. This makes it ever more likely, as the L1 phonetic system develops (and the L2 learner matures, more generally), that category formation will be blocked."*

The mechanism by which category formation for an L2 sound is blocked is called "equivalence classification" in the model. It takes place when a single phonetic category is used to process perceptually linked L1 and L2 sounds, called "diphones". Equivalence classification causes these diphones to be produced similarly (H5). Flege (1987) found evidence for this hypothesis in a study with French and English advanced learners of the other language on the production of their L1 and L2 stop consonants. He found that they produced /t/ (a "similar" sound) in their L2 with mean VOT values that resembled those typical for stops in the L1 or had values that were intermediate to the phonetic norm for VOT in the L1 and the L2. At the same time, VOT values of their L1 stops resembled those values of stops spoken by native-speakers of their L2. When new L2 phonetic categories are eventually established, the SLM postulates that they may differ from those of monolinguals (H6). This could be the case for L2 sounds that are distinguished from other L2 sounds by features not used in the L1, or because L1 features are employed or weighted differently in the L2. As an example, Flege cites Munro (1993), who reported that Arabic speakers of L2 English greatly exaggerated the duration between English lax-tense vowel contrasts. This was most likely due to the fact that there is a

[8] Critical Period Hypothesis was popularized by Lenneberg (1967) who claimed that it is not possible to acquire a native-like level of proficiency in a foreign language after a critical period, normally associated with puberty (Bot et al., 2005: 65).

phonological contrast between short and long vowels in Arabic, hence Arabic learners were assumed to have weighted the length feature differently in their L2.

Finally, H7 concerns the relationship between L2 sound perception and production: "The production of a sound eventually corresponds to the properties represented in its phonetic category representation" (Flege, 1995: 239). Gut (2009) has already pointed out that Flege does not specify what exactly he means by "eventually", nor does he describe the process by which phonetic category formation leads to accurate production. He does, however, propose that perception precedes production, for which he found support in a study with Taiwanese learners of L2 English (Flege, 1988). In this study, learners were either inexperienced (length of residence (LOR) of about 1 year in the U.S.) or experienced (LOR of about 5 years in the U.S.) learners. Both groups were found to speak English with equally strong foreign accents, but differed significantly in their ability to detect foreign accents in their own language group. While experienced learners rated foreign accents by Taiwanese speakers similar to an English native speaker control group, inexperienced learners detected differences to a lesser extent. To Flege, this suggested that perceptual abilities and the knowledge of how the L2 should be pronounced increases more rapidly than the ability to produce it. Furthermore, H7 implies correlations between perception and production data. Flege (1999) discusses various studies where moderate correlations were indeed found (on average about r=0.50). For example, in a study which examined the perception and production of English vowels by highly experienced native Italian speakers of English, Flege et al. (1999) found that the performance of the learners in both domains correlated significantly (r=0.62). Learners who were more accurate in discriminating English and Italian vowels also received higher intelligibility ratings by native speakers of English. The researchers explain the lack of perfect correlation with the possibility that accurate perception may not always be a sufficient condition for improved production. It might be possible that perceptual changes, in some cases, may never be "transported" into production.

As in the case of the CAH, the concept of "similar" versus "different" is central to the SLM. Methodologically, the definition of similarity is not a straightforward undertaking (Bohn, 2002, Strange, 2007). Flege (1997: 17-18) describes a different or "new" sound as a segment that differs acoustically and perceptually from the closest L1 sounds. It is represented by an IPA symbol that is not used for an L1 sound. "Similar" sounds are represented with the same IPA symbol, even though statistical analyses reveal significant and audible differences between the two. However, he does point out that the phonetic symbol criterion is not without problems, as different phonetic transcription criteria are in use. Thus, he maintains that it is necessary to supplement the phonetic symbol test with acoustic data. For example, Bohn and Flege (1992) suggested that a vowel can be considered new only if most of its realizations occupy a portion of the acoustic vowel space that is unoccupied by any of the L1 vowels.

Strange (2007) points out that acoustic measurements may not always lead to correct predictions about *perceived* similarity by L2 learners. For this reason, studies employ similarity ratings on L1 and L2 sounds by naïve L2 speakers in order to predict difficulties L2 learners may have with the respective sounds in the target language (e.g. Oturan, 2002, Bohn et al., 2011). Still, researchers have used acoustic comparisons in order to predict difficulties for L2 learners, as well (e.g. Bohn and Flege, 1992, Flege et al., 1994). Furthermore, recent research by Escudero et al. (2012) suggests that the concept of similarity is after all closely related to the detailed acoustic properties of sounds. For example, Escudero et al. (2012) investigated the effects of regional differences in the L1 Dutch on the perception of the L2 English vowels /ɛ/ and /æ/. Detailed acoustic analyses by means of Linear Discriminant Analysis of North Holland versus Flemish vowel productions predicted differences in the perception of the L2 segments. These were confirmed by the results of a categorization task. Flemish listeners chose Dutch /ɪ/ for English /ɛ/ significantly more often than North Holland listeners, while English /æ/ was categorized as Dutch /ɑ/ more often by North Holland listeners than by Flemish listeners. Steinlen (2009) too showed that acoustic comparisons of L1 vowels by native speakers of the target language can be used to explain L2 English vowel production by L1 Danish and L1 German speakers. In this work she proved that the so-called "arm-chair methods" (Bohn, 2002), that is, contrastive analysis based on phonetic symbols alone, are insufficient in predicting and explaining L2 productions. In the current study, comparable acoustic data are therefore collected to predict and explain the German vowel productions of Polish GFL learners.

2.3 Perceptual Assimilation Model for L2

Along with Flege's SLM, Best's (1995) Perceptual Assimilation Model (PAM) is considered one of the most influential models in current L2 speech research (Bohn, 2002: 196).[9] Since it was primarily aimed at naïve listeners, Best and Tyler (2007) extended the assumptions and predictions of the original PAM model to L2 learning (PAM-L2). As in Flege's model, the notion of similarity/dissimilarity between the L1 and L2 sounds is a central theme, while a focus on *perceived* similarity is crucial for PAM and PAM-L2. Furthermore, and similar to Flege (1995), Best and Tyler (2007: 22) stress that their comparisons between L1 and L2 sounds are not based on phonological distinctions only but consider "non-contrastive *phonetic* similarities and dissimilarities between L1 and non-native/L2 phones, including notions of phonetic goodness of fit, and the relationship between phonetic details and phonological categories and contrasts".

At the heart of PAM(-L2) lies the idea that non-native phonetic segments are perceptually assimilated to the most articulatorily-similar[10] native phoneme. From this premise, Best (1995) and Best and Tyler (2007) describe different assimilation patterns from which various levels of difficulty in L2 sound perception can be predicted. As shown below, certain patterns may be equated to processes already discussed by Flege (1995).

---

[9] Furthermore, Bohn (2002) briefly addresses Major's Similarity Differential Rate Hypothesis (SDRH, Major, 1987, Major and Kim, 1996) and Kuhl's Native Language Magnet model (NLM, Kuhl and Iverson, 1995, Kuhl, 2000). While the SDRH focuses on rate of acquisition and builds upon Flege's SLM, the NLM model focuses on L1 speech acquisition. They will not be discussed in more detail in this study, as they are not considered central accounts in the field of L2 speech learning. Most research in the field tests hypotheses put forward by SLM and PAM(-L2) (e.g. Guion et al., 2000, Oturan, 2002, Bohn et al., 2011, among many others).

[10] Best (1995) and Best and Tyler (2007) stress that PAM(-L2) is founded on the direct-realist premise that speech perception focuses on articulatory gestures rather than acoustic-phonetic cues. Even though this stance differs from Flege's idea of mental representations of phonetic categories, it is not crucial for the main hypotheses of PAM(-L2).

PAM-L2 predicts poor discrimination and identification for an assimilation pattern where two L2 phones are equally good or poor instances of the same native phoneme. This situation is called Single Category (SC) assimilation in PAM-terminology and is reminiscent of the process of "equivalence classification" (SLM). The PAM-L2 assumes that a learner will have great difficulty discriminating two such phones since they are assimilated both phonetically and phonologically to one single L1 category. Success in overcoming this problematic perceptual assimilation will depend on how well the L2 phones fit the L1 category. Measurements of goodness-of-fit can inform this hypothesis. For example, Guion et al. (2000) had naïve Japanese speakers identify English consonants as instances of Japanese categories and rate each sound for its goodness-of-fit to the selected Japanese category (on a scale from 1-7). While this method allows for detailed quantitative measurements of perceived similarity, Best and her colleagues themselves (e.g. Best et al., 2001, Bohn et al., 2011) have used simple orthographic labelling to determine whether two L2 sounds are assimilated into one L1 category. Best and Tyler (2007) assume that, for SC assimilation to be overcome, a learner first needs to establish a new phonetic category for at least one of the two phones, which may be unlikely and depend on the degree of the goodness-of-fit.

A situation where two L2 phones are still assimilated to the same L1 phonological category, but one is a phonetically better example of the native category than the other, is a pattern labelled Category Goodness (CG) assimilation. In this case, discrimination of L2 sounds is predicted to be good, and it is assumed that for the more deviant phone a new category – both on a phonetic and phonological level – can eventually be formed by a learner. While the PAM-L2 (like the SLM) does not make specific claims regarding the exact time-course of a new category formation, Best and Tyler hypothesize that the "bulk of perceptual learning may actually take place fairly early in late-onset SLA" (p. 31).

Other types of assimilation, which predict a very good discrimination between two L2 phones, are the so-called Two Category (TC) assimilation and Uncategorized-Categorized (UC) assimilation. In the case of the latter, only one L2 phone is assimilated into an existing L1 category, which makes discrimination easy. In the TC assimilation, each L2 phone is assimilated to a different L1 category, which implies that the native system of phonological contrasts helps to perceive the L2 distinction, as in the UC case. Research on the perception of Zulu and Tigrinya consonants by English native speakers has supported the predictions of the original PAM model for naïve listeners (Best et al., 2001). Bohn et al. (2011) further tested PAM-L2 predictions for Danish learners of L2 English. The authors found that, in the case of the English /ð/-/v/ contrast, discrimination was difficult. This was surprising as this contrast was categorized as TC assimilation for the Danish speakers as established in a prior labelling task. Bohn et al. (2011) speculate that this may be due to the auditory similarity of the two phones, which contrasts with the PAM stance that it is the articulatory *gestures* which are being perceived.

PAM-L2 has been developed to capture *second* language learning (SLA) rather than *foreign* language acquisition (FLA).[11] Best and Tyler (2007: 19) point out that "FLA is a fairly impoverished context for L2 learning, and perceptual findings for FLA listeners should not be conflated with those for L2 listeners (SLA)". This is a very crucial point, which, for example, Almbark (2014) elaborated on in her work on Arabic foreign language (FL) learners of English. By investigating the perception and production of English vowels, she found that some SC contrasts[12] had excellent identification and discrimination by the FL learners. This result contrasts with what would be predicted by PAM-L2 (as well as the SLM). The author attributed this to the fact that FL learners of her study received *phonemic* classroom instruction which highlighted the distinction between the difficult vowels. Almbark (2014) therefore calls for the need of a "Foreign Language Model", which takes the influence of classroom instruction and the lack of native-speaker input into account. Her model further postulates that perceptual skills of FL learners are reflected in the productions of the learners in that, if differences between sounds are discerned, productions of these segments will be distinct as well. Her data corroborated this assumption.

---

[11] See also Footnote 1 and the differentiation between G*F*L and G*S*L.

[12] For example, FACE and SQUARE vowels were mapped by Arabic speakers onto Arabic /e:/ equally well in a perceptual assimilation task.

While Best and Tyler (2007) do not make any predictions about L2 production, Almbark's hypothesis that production follows perception in FLA reflects what Flege (1995) assumes for SLA.

Like the SLM, the PAM(-L2) does not take into account orthography as an influential variable in L2 speech learning. Still, Best and Tyler (2007: 27-28) briefly mention a case where orthography *could* play a role. They discuss the example of French /r/, which is phonetically realized as a voiceless uvular fricative. English learners of French tend to equate this phoneme with the English liquid /r/, despite the fact that it is phonetically very different from the French category. While the authors emphasize that different phonetic realisations of established phonological categories can be learned, they also point out that the French and the English phoneme /r/ are represented by the same grapheme in the two orthographic systems. This may "contribute to that bias" (p. 28), but the authors do not elaborate on this important observation.

2.4 L2 Linguistic Perception Model

A more recent model in the field of L2 speech learning is Escudero's (2005, 2009) L2 Linguistic Perception Model (L2LP). It makes specific predictions about the acquisition of L2 vowel length, which is an important aspect for the present study. The guiding assumption of this model is that speech perception is a language-specific phenomenon that involves linguistic (phonological) knowledge, and that is why the "perceptual mapping of the speech signal should be modelled within phonological theory" (Escudero, 2009: 153). The model draws on mechanisms formulated in (Stochastic) Optimality Theory (Boersma, 1998, Escudero and Boersma, 2004), and builds on Escudero's general framework of linguistic perception and L1 sound acquisition.[13]

Similar to the previous two models, the L2LP differentiates between *new* (or "different") and *similar* sounds. However, unlike the SLM, it assumes similar sounds to be easier than new sounds. This difference is partly due to

[13] Because this study is concerned with second language learning, the discussion of Escudero's framework is limited to the L2 part of the model. For details of how general perception and L1 acquisition is understood in the model, see Escudero (2005: 41-83) or Escudero (2009: 155-167).

discrepancies between the definitions of what constitutes *new* versus *similar* sounds. In Escudero's model, new sounds are defined as "L2 sounds which are produced with at least one auditory dimension that has not been previously incorporated into the learner's L1 linguistic perception" (Escudero, 2005: 157). New sounds lead the L2 learner to perceive fewer sounds than the ones produced in the L2 environment. This is due to a mechanism Escudero (2005) calls "phonemic equation" of two L2 sound categories with a single L1 category, which is a term that can be equated to Flege's equivalence classification or Best and Tyler's Single Category assimilation. For example, a Spanish L2 learner of Southern British English (SBE) would perceive both English /i/ and /ɪ/ as Spanish /i/. This, according to the model, occurs both at an abstract phonemic level and a "perceptual mapping" level, i.e. most phonetic realizations of the two L2 phonological representations will be perceived as a single L1 category.

Escudero (2005: 155, 173) further refers to the level of lexical representation and assumes that phonemic equation will lead a learner to use the same L1 phoneme for L2 words which usually contain two different L2 phonemes, such as /ʃip/ and /ʃɪp/. This erroneous lexical storage means the learner will have to rely on the semantic and pragmatic context in order to access the correct meaning of the respective L2 word. Drawing attention to the role of non-native sound perception in lexical representation is an important point which is addressed neither in the SLM nor in the PAM-L2.

Similar sounds are defined in the L2LP model as L2 sounds which are equated to two different L1 sounds. This means they are phonemically equivalent but phonetically different. Such a scenario also poses a learning challenge because perceptual mappings have to be adjusted; however, this adjustment is assumed to be easier to master than the creation of new L2 categories as is the case for *new* L2 sounds. This scenario is conceptually similar to the PAM(-L2)'s Two Category assimilation, and both models would predict that these sounds do not pose a great learning challenge. While Flege's SLM would predict the greatest difficulties for the so-called "similar" sounds, it has to be emphasized again that similar sounds are defined differently by Escudero (2005, 2009).

The L2LP assumes that the initial L2 learning task for new sounds consists of two components, namely a perceptual and representational task. In the perceptual task, learners have to create or split perceptual mappings for the new sounds and integrate new auditory cues. In the representational task of the learning process, new abstract phonological categories have to be created. While new sounds involve these two steps, similar sounds only involve the first task of adjusting their perceptual mappings, which is why the L2LP predicts that new sounds are more difficult.

A guiding assumption of the L2LP is the "optimal perception hypothesis", which states that an optimal (native or non-native) listener will prefer auditory cues that reliably differentiate sounds in the production of the L1 or L2. As a consequence, the hypothesis predicts that differences in the productions of two languages (or even language varieties) will lead to differences in the optimal perception of the sounds of the languages. For this reason, it is an important prerequisite of the model to describe how monolingual speakers of a target language perceive and produce it along the respective acoustic dimensions. Furthermore, Escudero compares the L2 perceptual task in learning new sounds with the learning task of an infant beginning to categorize L1 sounds. Along with Maye et al. (2002), she assumes that L1 category formation occurs through distributional learning. This is based on the auditory distributions of phonetic dimensions which have previously not been categorized.

In the case of Spanish learners of L2 English, Escudero compared the F1 and duration values of Spanish and English vowels and established that for monolingual Spanish speakers, vowel duration constraints do not play a role in their L1 perception grammar (unlike in English). The model proposes that this non-previously categorized dimension (vowel duration) is preferred to already-categorized dimensions (F1) when learning to perceive new L2 sounds. Hence, duration is used as a cue by Spanish L2 English learners, as shown by Escudero and Boersma (2004).

To test whether the reliance on duration is a general aspect of Spanish learners' L2 perception, Escudero et al. (2009) investigated the perceptual weighting of vowel duration and vowel spectrum in the identification of Dutch vowel contrasts by L1 Spanish L2 Dutch speakers as well as L1 Dutch and L1 German speakers. Similar to English, some Dutch vowel contrasts differ both in vowel length and vowel quality (i.e. Dutch /aː/ versus /ɑ/), and it was of interest to ascertain how these two dimensions are weighted in the three groups. In order to investigate this question, isolated synthetic vowels were created. Two items represented typical Dutch /a:/ and /ɑ/, while the rest were vowels which differed in equal steps between the two end points in both spectral and durational values. A two-alternative forced choice task was employed in which participant had to categorize these items as either similar to a typical /a:/ or /ɑ/. It was found that Dutch and German speakers relied significantly more on spectral differences than the Spanish L2 Dutch learners did when categorizing the vowels. The inclusion of a different language group (German speakers) was meant to shed light on whether the heavy weighting of duration to categorize the Dutch vowel contrast is a universal strategy for all learners. The results suggested that this is not the case, as L1 German speakers would rely on spectrum rather than duration. This, in turn, was related to their different assimilation patterns of the Dutch vowels to their closest German vowels, which was also investigated in the study. It has to be kept in mind, however, that the German group was naïve to Dutch, while the participants in the Spanish group had medium to high proficiency in Dutch.

Escudero et al. (2009) discuss different approaches for explaining the duration reliance of their L2 learners. They group these into a "developmental approach" (e.g. Bohn, 1995, Escudero and Boersma, 2004) and a "feature approach" (e.g. Kondaurova and Francis, 2008). While the feature approach tries to explain findings primarily with the native language background of the learners, developmental approaches favour more universal explanations. In Escudero's L2LP, these explanations are based on the assumption that learners can form new categories along non-previously categorized dimensions using distributional learning. Bohn (1995), in his Desensitization Hypothesis (see Section 2.5.2), assumes that the high psychoacoustic saliency of vowel duration leads L2 learners to rely on it. While it is not possible to favour one or the other on the basis of their findings, the authors maintain that a feature approach cannot explain the results for the Spanish learners well (but see McAllister et al.'s Feature Hypothesis for L2 Swedish data, Section 2.5.1).

Finally, Escudero et al. (2009) briefly mention the possible influence of Dutch orthography. In monosyllabic words with a closed syllable, Dutch /a:/ is written

<aa> and /ɑ/ as <a>, for example <taak> ("task") and <tak> ("tree branch"). This could play a role in explicit awareness of duration differences in Dutch, but the authors discard this idea as this type of orthographic marking is very restricted. As in the previous models, orthography is not incorporated in this more recent L2 speech learning model. A possible reason for this might be that all models focus on a *second* rather than a *foreign* language-learning context. The authors appear to assume that the primary source of input in this learning context is speech by native speakers (in contrast to written forms), which is unlikely to be influenced by orthography (see for example Alario et al. (2007) and their research on the effects of orthography in L1 spoken word-production).[14] Still, other researchers have addressed the possible influence of orthography in L2 speech learning with studies including orthography as an independent variable, which will be discussed in detail in Chapter 3.

2.5 Specific hypotheses for vowel length acquisition

A number of studies in the field of L2 speech learning have addressed the specific problem of vowel length perception in L2 phonological acquisition (e.g. Bohn, 1995, Bohn and Flege, 1990, McAllister et al., 2002, Bogacka, 2004, Kondaurova and Francis, 2008, Nimz, 2011b, Rojczyk, 2011, Altmann et al., 2012), where also a few of these studies included the investigation of vowel length *production* (i.e. McAllister et al., 2002, Nimz 2011b, and Rojczyk, 2011). Most of these studies investigated L2 English, despite the fact that vowel length has been found to be a secondary cue in native English vowel perception (Hillenbrand et al., 2000). The predominance of English as a target language reflects a general trend in the field of L2 speech learning (Hayes-Harb, 2012), and only a few studies have examined the acquisition of vowel length in other

[14] Native speakers learn to speak before they learn how to read. The idea that orthography is unlikely to influence L1 speech seems therefore justified. However, research in L1 speech perception has shown that orthography can influence how spoken words are processed. For example, Petrova et al. (2011) showed that orthographic consistency (phonological rimes that can be spelled with one versus multiple ways) affects auditory lexical decision latencies. Due to this so-called orthographic consistency effect, inconsistent words produce longer auditory lexical decision latencies and more errors than do words with rimes that could be spelled in only one way.

languages for which vowel length is more prominent than for English (i.e. McAllister et al., 2002 for Swedish, Escudero et al., 2009 for Dutch, and Nimz, 2011b and Altmann et al., 2012 for German).

Nimz (2011b, 2015) investigated the perception and production of German vowels by late Turkish learners of GFL. In her discrimination task, she presented the learners and an age-matched German native speaker control group with nonce word pairs that differed either in vowel quality, vowel length or both. The data demonstrated that Turkish learners could hear length differences in German vowels native-like, while for some vowels, quality differences were more difficult to differentiate for the group. Similarly, the production data of real German words showed that Turkish GFL learners did not have specific problems producing the length of German long vowels while their vowel qualities differed significantly from those produced by German native speakers. Because Turkish speakers have phonetic experience with long vowels in their native language, due to the phenomenon of secondary lengthening after [ɣ] <ğ> (*yumuşak g*) (Kirchner, 1999, Kabak, 2004), it was unclear whether the results Nimz (2011b) obtained were influenced by the native language background of the speakers (see Feature Hypothesis by McAllister et al., 2002 below) or whether vowel length is always easy to acquire, as postulated in the Desensitization Hypothesis by Bohn (1995) (see Section 2.5.2).

Similarly, Altmann et al. (2012) were not able not make a definite claim about their vowel (and consonant) length perception data in Italian and German speakers. The authors found that Italian speakers could perceive vowel length differences as well as German native speakers could, while German native speakers did not perceive consonantal length contrasts as well as Italian speakers did. Italian, unlike German, does not exhibit vocalic length contrasts, while it does show allophonic vowel lengthening in stressed open syllables, which might make it easier for Italian speakers to perceive vowel length contrasts in other languages as well (Kondaurova and Francis, 2008). Since the results for the Italian and German speakers were asymmetric, the authors preferred to interpret their results by differentiating between vowel and consonant perception: vowels carry a higher informational load (suprasegmental features of rhythm and intonation are realized on vowels), they contain information of neighbouring segments in their formant transitions, and convey speaker identity. Altmann et al. (2002) explained the Italians' success in differentiating vowel length with the nature and function of the vocalic acoustic signal. Another possible explanation is offered by the Feature Hypothesis by McAllister et al. (2002).

2.5.1 Feature Hypothesis

The Feature Hypothesis by McAllister et al. (2002) is one of two explicitly formulated hypotheses in the field of L2 phonology which concern the acquisition of vowel length. It states that "L2 features [such as duration] not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al., 2002: 230). To test their hypothesis, the authors investigated the perception and production of Swedish quantity distinctions by Estonian, English, and Spanish L2 learners of Swedish. The three learner groups were considered an ideal testing ground as their native phonologies display different degrees of overall prominence of the duration feature. Estonian exhibits both vowel and consonantal length contrasts, English uses vowel length as a secondary cue in vowel perception, and Spanish does not utilize any phonological length contrasts. Hence, the duration feature is most prominent in Estonian L1, followed by English, and then Spanish. It was therefore expected that Estonian speakers can perceive the differences between short and long Swedish vowels better than English or Spanish speakers, while English speakers were expected to outperform the Spanish speakers. The identification task used in the study involved 40 real Swedish words and 40 non-words, which were produced by a native Swedish-speaking phonetician. In the non-words, long vowels were replaced by short vowels (and following long consonants), while short vowels were replaced by long vowels (and following short consonants)[15]. Twenty speakers of Swedish, Estonian, English, and

[15] In Swedish stressed syllables, there is a complementary relationship between the duration of the vowel and the following consonant: a long vowel is followed by a short consonant, and a short vowel is followed by a long consonant (or a cluster).

Spanish were then instructed to judge whether the stimuli they hear were correct or incorrect instances of the respective words. Their accuracy data confirmed their assumptions: Estonian speakers were significantly better than English and Spanish speakers at identifying words correctly, and English speakers were better than Spanish speakers. They further investigated L2 production by means of a definition task accompanied by pictures of the test words used in the perception task. The productions were acoustically analysed for vowel and following-consonant length, and V:/C and V/CC ratios were compared for all vowel and participant groups. The results of the production experiment yielded fewer significant differences between native Swedish speakers and the different learner groups; only Spanish speakers were found to produce smaller length distinctions than the native Swedish speakers, and this was true only for words containing mid vowels. The authors hypothesized that mid vowels might be harder to learn because contrary to non-mid vowels they are not accompanied by smaller spectral differences in Swedish. While the authors do not discuss this further, as they focused on vowel length rather than quality in their study, it would be interesting to investigate the perception and production of both dimensions (length and quality) at the same time. This, among other aspects, is one of the main aims of the current study.

2.5.2 Desensitization Hypothesis

While the Feature Hypothesis stresses the influence of the native language background in L2 speech learning, Bohn's (1995) Desensitization Hypothesis emphasizes the potential importance of language-independent, auditory-based strategies in L2 perception. Based on his results from studies with German, Spanish, and Mandarin native speakers identifying synthetic English vowels, he formulated the following hypothesis: "Duration cues in vowel perception are easy to access whether or not listeners have had experience with them [in their native language]" (Bohn, 1995: 294). In his studies, the participants had to identify English vowels on a *bet* to *bat* (and *beat* to *bit*) continuum as *bet* or *bat* (*beat* or *bit*). The stimuli were synthetic continua of vowels that varied orthogonally in their durational and spectral parameters. Bohn found that German native speakers relied much more on durational differences when identifying stimuli as *bet* or *bat* than native English speakers did. It is well established that German native speakers use vowel duration as an important cue in German vowel identification (Weiss, 1974, Sendlmeier, 1981), which is why the results for this group alone did not lead to the formulation of a language independent hypothesis. More surprisingly, Spanish and Mandarin participants showed a similar pattern to the German participants for the *beat* to *bit* continuum, i.e. they relied more on durational cues than English speakers, which the author explained by means of the Desensitization Hypothesis. Cebrian (2006) conducted similar studies with different groups of L1 Catalan learners of L2 Canadian English. His results supported Bohn's Desensitization Hypothesis in that the Catalan speakers relied on duration as the main cue to the English vowel contrast despite not having experience with duration in their native language. Nimz (2011b), too, seemed to have found preliminary support for Bohn's hypothesis, though it was unclear whether her results may have been influenced by the Turkish participants' experience with secondary long vowels (see above).

While McAllister et al.'s (2002) and Bohn's (1995) hypotheses seem to contradict each other, Tomaschek (2009: 42-44) points out that the two study designs might have tapped into different perceptual modes. McAllister et al. had conducted a perception experiment with real word perception in advanced L2 learners whereas Bohn presented his (relatively) naïve participants with synthetic stimuli. While the task of judging stimuli as "correct" or "incorrect" (McAllister et al.) might have prompted a "phonological" mode of perception, Bohn's task might have allowed for a more "phonetic" mode, as his synthetic stimuli could have been judged without any phonological or lexical knowledge. These open methodological questions are addressed in this study by means of conducting two different perception experiments with the same group of learners: a discrimination (Chapter 5) and an identification task (Chapter 7).

2.6 L2 speech learning and lexical representations

The theories and hypotheses discussed above investigate the acquisition of L2 speech sounds without addressing the consequences for the learners' L2 *lexical representations.* A lexical representation is the "storage of a word in memory, and it contains information about the phonological, morphological, syntactic, semantic and, when available, orthographic structures of words" (Hayes-Harb and Masuda, 2008: 7). It might seem intuitive to assume that a reliable phonetic perception of L2 phones is related to the representation of the phonological content of lexical representations. For example, Pallier et al. (2001) had shown that Spanish-dominant Spanish-Catalan bilinguals treated Catalan word pairs such as /pere/-/perɛ/ ("Peter"-"pear") as homophones in a lexical decision task. The experiment included 24 minimal pairs which were based on phonemic contrasts specific to Catalan, as well as other control pairs, pseudowords, and filler items. In the course of the experiment, one member of each critical Catalan minimal pair appeared and was followed either by itself or by the other item of the minimal pair. Comparisons between repetition effects in identical repetition and minimal pair repetition revealed that Spanish- and Catalan-dominant bilinguals processed the minimal pairs differently. While Spanish-dominant bilinguals showed a spurious repetition effect (facilitation) for the Catalan minimal pairs, Catalan-dominant bilinguals did not. The authors explained this finding by Spanish(-dominant) listeners' difficulty to perceive a phonemic contrast between such vowels as /e/ and /ɛ/, as the distinction does not exist in their L1 phonological system. Hence, words containing these difficult contrasts are represented in the lexicon as homophones. However, recent research has shown that the relationship between L2 perception and L2 lexical representation might in fact not be as straightforward.

In a series of studies which were concerned with lexical processing consequences of phonemic confusions for L2 listeners, Weber and Cutler (2004) and Cutler et al. (2006) found evidence of asymmetric mapping from phonetic to lexical processing. The finding led the researcher to discuss the possible influence of orthography in the construction of lexical representations. In their 2004 study, Weber and Cutler investigated lexical representations of words containing vowels which are difficult to discriminate by Dutch learners of L2 English, namely the vowels /ɛ/ and /æ/ as in *head* and *had*. They hypothesized that lexical competition would be greater for L2 learners due to their difficulties in differentiating the respective sounds – even in words which are only initially confusable, such as *pencil* and *panda*. In order to collect time-sensitive information on the relevant processing mechanisms, they used an eye-tracking design which included pictures of objects that contained the confusable sounds, e.g. a display of a panda and a pencil (and two other distractor items). What they found was that L2 listeners did not equally confuse, i.e. looked at, the confusable items, but rather that their lexical processing was asymmetric regarding the confusable sounds; the Dutch listeners preferred to look at those objects which contained /ɛ/ rather than /æ/, and would look more at the pencil no matter whether they heard *pen-* or *pan-*. In their 2006 paper, Cutler et al. offer two possible explanations for this result. First, Dutch /ɛ/ is phonetically closer to English /ɛ/ (though not the same), thus making English /ɛ/ the dominant category. And second, orthography has a bearing on this result in that the sounds written with <e> in both Dutch and English are pronounced similarly, while that is not the case for sounds written with <a> (i.e. front vowel in English, back vowel in Dutch). This would mean that even though both /ɛ/ and /æ/ are perceived in the input as front vowels, only words containing orthographic <e> would be matched by the two possibilities. While the first interpretation was preferred for their follow-up results with Japanese learners, the role of orthography in L2 phonological acquisition was further investigated by studies which treated orthographic input as the main experimental variable (see Chapter 3).

Hayes-Harb and Masuda (2008) investigated lexical representations of newly-learned L2 words, which contained a novel L2 phonemic contrast. In their study they looked at the perception and production of Japanese consonantal length (e.g. singleton /t/ versus geminate /t:/) by L1 English speakers. Since they were interested in the developmental aspects of lexical encoding, their participants included both naïve speakers and L2 learners of Japanese as well as a native speaker control group. Participants were taught 8 Japanese non-words which differed only in their consonantal length. In an auditory word-picture matching task, they then had to correctly identify the newly-learned words with their respective pictures. The crucial experimental condition consisted of word-picture mismatches that depended on the singleton-geminate contrasts, e.g. the nonce word /pete/ matched with the picture of the nonce word /pet:e/. D' prime analyses revealed that L2 learners were able to detect these mismatched items significantly better than the naïve listeners, while they did not differ significantly from native speakers of Japanese. The authors interpreted this as resulting from the learners' improved ability to encode consonant length in their lexical memory. In the production task, participants had to name the pictures of the newly-learned words. Three independent native speakers of Japanese judged those productions for their accuracy in producing the singleton/geminate consonants. While there were no significant group differences for the singleton productions, the comparisons for the geminate consonants revealed that L2 learners have an improved ability to lexically encode these sounds in comparison to naïve learners. However, the L2 learners were also significantly less successful at this encoding than the Japanese native speakers. The authors drew two conclusions from their results. Firstly, as expected, experience with an L2 influences how phonological contrasts are encoded in the lexical representations of newly learned words. Secondly, L2 learners may not always encode the length feature accurately. While the results of the listening task suggested that L2 learners encoded consonantal length native-like, the production data showed that the encoding can only be partially correct. The authors propose that learners initially represent novel L2 phonemes, such as Japanese geminate consonants, as an unfamiliar version in the form /t*/, where "*" means that this sound is different from /t/, but it is not yet determined that this differences lies in length. This would explain why L2 learners could correctly identify mismatched items in the listening task, but were yet unable to produce the geminate consonants native-like.[16] Hayes-Harb and Masuda further hypothesize that their learners were better able to encode the novel contrast in their memory because Japanese singleton and geminate consonants are spelled differently. Although no orthographic input was given in the task, spelling may still have drawn participants' attention to the contrast in their auditory input. Yet, the authors concede that the nature of the relationship between contrastive spelling and contrastive phonological representation is unclear.

[16] The difference between the perception and the production data could also be discussed in the light of "perception precedes production", as is suggested in Flege's SLM. Furthermore, in his sixth hypothesis, Flege (1995) states that once L2 phonetic categories are eventually established, they may differ from those of native-speakers. This might be the case because L1 features may be employed differently in the L2. Yet, Hayes-Harb and Masuda do not discuss their results in the vein of the SLM.

Darcy et al. (2012) investigated the lexical encoding of front rounded vowels in English learners of L2 French, examining the acquisition of these segments both at the level of phonetic categorization and lexical representations. In an ABX categorization task, four different participant groups (intermediate and advanced English-French learners, French native speakers, and naïve English native speakers) categorized French nonce words involving the contrasts /y/-/u/ and /œ/-/ɔ/. Comparisons of error rates across the groups revealed that the French native speakers made the fewest errors in categorizing the sounds while naïve L1 English speakers showed the highest error rates. The two learners groups were in-between these two groups and did not differ significantly from each other. Both intermediate and advanced L1 English L2 French learners performed significantly more accurately on the /y/-/u/ contrast than on the /œ/-/ɔ/ contrast. Lexical representations of the vowels of interest were investigated using a design similar to that of Pallier et al. (2001). The lexical decision task with repetition priming involved 40 French words which formed 20 minimal pairs that involved the contrasts of interest, as well as the same amount of nonce words and 180 filler pairs. Comparisons of the response times to repeated items and minimal pair items showed that advanced learners did not produce spurious response time facilitations to the /y/-/u/ and /œ/-/ɔ/ pairs. However, the results for the intermediate learners did show priming effects indicative of homophony in that /y/ and /u/ items facilitated the decisions for each other. To the authors, this suggested that the /y/-/u/ pair is not distinguished in lexical representation, which is in line with findings by Pallier et al. (2001). Yet, the results for the /y/-/u/ contrast in *advanced* learners show that this spurious homophony can be overcome. Furthermore, despite considerable categorization errors, the L2 learners were able to lexically encode the difficult /œ/-/ɔ/ contrast. This is a "curious anomaly for standard assumptions according to which the development of new categories is a necessary prerequisite for lexical contrast" (Darcy et al., 2012: 28).

The authors favour a phonological account in explaining their data, which they label "direct mapping from acoustics to phonology" (DMAP). In short, this account argues that lexical representations of phonological contrasts are independent of the attunement of phonetic categories to the L2 input. DMAP assumes that learners detect correlates of phonological features in the raw percepts of the input which may trigger revisions of the interlanguage feature hierarchy. This cue-based learning on the phonological level is, according to DMAP, not related to auditory sensitivity, which means that lexical contrast can precede reliable category formation. While this account has not been probed further in more recent studies, other factors have been addressed by Darcy and colleagues, which may explain the dissociation between phonetic categorization abilities and lexical representations in L2 learners (Darcy et al., 2013). Among these factors are orthography and explicit instruction, as they "might provide first indications to learners to bootstrap the contrast separation lexically" (Darcy et al., 2013: 416-417). All of the above studies briefly referred to orthography as a possible influential variable. Still, none of them had systematically tested its influence.

2.7 Summary: Models and hypothesis in L2 speech learning

Current models of L2 speech learning constitute important starting points for investigating the acquisition of L2 segments. Flege's SLM, Best and Tyler's PAM-L2, and Escudero's L2LP put forward helpful heuristics in identifying, describing, and interpreting phenomena in the process of **L2 sound perception**. As far as **L2 sound production** is concerned, Flege is the only one who explicitly integrates this side of the coin in his model. In his SLM he puts forward the hypothesis that perception precedes production, and cites evidence in favour of this proposition (Flege, 1999). In the introduction to her model, Escudero (2005) discusses further studies which support this notion and points out that "prioritizing the role of perception in explaining the acquisition of L2 sounds seems to be valid and is perhaps the most propitious way of approaching the phenomenon" (p. 3). However, other researchers draw attention to the need of integrating L2 sound production in current models of L2 speech learning. Zampini (2008: 242), for example, argues that models will be incomplete if they do not account for the relationship between L2 sound perception and production. Interestingly, many studies in the field of GFL, which have been mentioned in the context of the Contrastive Analysis Hypothesis above, seem to have gained their insights from *production* errors of

L2 learners, while perception data are hardly integrated. The different foci in the fields of experimental phonetics and GFL teaching call for an interdisciplinary approach which evaluates findings from both fields. The current study aims to do so. Not only does it integrate findings from different fields of research, but it investigates both sides of the coin experimentally: L2 sound perception *and* production.

While both Flege (1995) and Best and Tyler (1997) discuss research on consonants and **vowels**, Escudero's (2005) model is based on research on vowels alone. This limitation does not seem to be a problem, as findings on adults' perception of L2 vowels largely mirror the patterns found for L2 consonants (Best and Tyler, 2007: 20). Because of their diversity, vowels can offer a "wealth of opportunities for researchers in L2 acquisition" (Zampini, 2008: 226), and in the case of Polish GFL learners, this surely applies. As will be shown in Chapter 4, **German and Polish** vowels constitute a fruitful testing ground for theories and hypotheses in the field. Not only have German vowels been identified as a prominent problem for Polish L2 learners (see references in Section 2.1), but German is also a suitable language for evaluating opposing views on vowel length acquisition (Section 2.5). Furthermore, all current models of L2 speech learning (SLM, PAM-L2, and L2LP) stress the need for comparing the sounds systems of interest on a phonetic level. While direct measurements of *perceived* similarity are favoured by some (e.g. Bohn, 2002), Escudero et al. (2012) have shown that the exact acoustic vowel properties of the languages under concern can determine cross-language similarity and, consequently, predict how L2 segments will be perceived and produced. Furthermore, Steinlen (2009) collected both L1 German and L1 Danish acoustic data, with which she could explain L2 production patterns in German and Danish L2 English learners. For this reason, the current study incorporates L1 productions of both language groups as well.

As discussed by Best and Tyler (2007), the PAM-L2 (as well as the SLM) focus on learners who acquire their L2 "naturally"[17] in an environment where the L2 is

[17] Even though the term "natural" is often used for describing the acquisition of the target language in an SLA context, it is not unproblematic. As Apeltauer (1997), among others, has

predominant (i.e. SLA). **Foreign language acquisition** (FLA) is usually characterized as an L2 acquisition context with less exposure to the target language and more formal instruction. While current models may be unable to fully account for findings in FLA (e.g. Almbark, 2014), they are still helpful in understanding data collected from foreign language learners. Darcy et al.'s (2012) participants, for example, were students learning French as a foreign language at an American university. Still, in their discussion they acknowledge that models like the PAM-L2 provide worthwhile insights. Furthermore, Darcy et al. (2013) identify formal instruction as a possible variable in explaining their findings on lexical representations in American GFL learners. L2 learners from an FLA context therefore provide the opportunity to investigate additional factors in L2 speech learning.

Another aspect which has received very little attention in current models of L2 speech research is **orthography**. While most researchers have briefly hinted at possible orthographic influences in L2 phonological acquisition, none of them has integrated orthography as a prominent variable in their models. Recent research at the interface between acoustic and orthographic input in L2 phonology suggests that this might be necessary (Chapter 3). Furthermore, studies which found evidence of perception abilities not always being aligned with lexical-phonological representations (see Section 2.6) often refer to orthography as a source for the unexpected dissociation. The current study sets out to investigate the role of orthography in the context of formal L2 phonological acquisition further.

---

pointed out, FLA is after all not "unnatural" (which would be the logical conclusion when contrasting the two acquisition contexts).

3 L2 phonology and the role of orthography

As was evident in the previous chapter, orthography has not been incorporated into current L2 speech learning models as a possible variable. In this chapter, recent research into the role of orthography in L2 speech learning will be presented which suggests that the inclusion of this factor may be a fruitful venture. Firstly, both perception and production studies in the field of experimental phonetics will be discussed, which will be followed by a review of the status of orthography in GFL phonetics. Lastly, the two orthographic systems of interest – German and Polish – will be described in detail, with special focus on the marking of vowel length in German.

Until recently, orthography and its potential influence on L2 sound perception, production, and/or representation has not been an area of research that has been studied widely. Yet, over the past decade, a growing interest in this factor can be noted (e.g. Erdener and Burnham, 2005, Steele 2005, Escudero et al., 2008, Bassetti, 2009, Silveira, 2009, Escudero and Wanrooij, 2010, Hayes-Harb et al., 2010, Simon et al., 2010, Simon and van Herreweghe, 2010, Dornbusch, 2012, Showalter and Hayes-Harb, 2013, Escudero et al., 2014, Mathieu, 2014). The growing awareness reflects itself in a recent special issue in *Applied Psycholinguistics* on "Second language phonology at the interface between acoustic an orthographic input" (Bassetti et al., 2015). In the editorial of the issue the authors express their surprise that systematic empirical research on the influence of orthography on L2 phonology is a relatively recent enterprise. They hypothesize (p. 2) that the disregard may be due to

> *"[...] a lack of theoretical justification as well as the zeitgeist, because L2 phonological research has been dominated by linguistics' search for universals of language and the primacy of spoken language, and language teaching has been dominated by the communicative approach. Within this context, when researchers came across possible orthographic effects, it was typical to ignore them as irrelevant, inconsequential, or as 'noise in the data'."*

While the authors do not give concrete examples, evidence of tentative reference to the possible influence of orthography (which was not controlled for in the respective studies) can be found in the field. For example Bogacka (2004), in her study with L1 Polish learners of English, had Polish and English participants judge manipulated vowels on a *who'd* to *hood* continuum and found that Polish participants, in contrast to native speakers, judged long stimuli to be *hood* instead of *who'd*. She hypothesized that "the confusion was probably caused by orthography, by a double <o> in the word *hood*" (p. 45). In a similar vein, Cebrian (2006) briefly discusses orthographic interferences for his results with L1 Catalan learners of L2 English. Even though Catalan does not have a short-long vowel distinction, his subjects made extensive use of the duration cue, even more so than English natives. Apart from his reference to Bohn's Desensitization Hypothesis (Section 2.5.2), he hypothesizes that "learners [...] may have equated a double letter grapheme [...] with a long version of the same vowel" (p. 384). Note, however, that Bogacka and Cebrian interpret the influence of English orthography in opposite ways. While Bogacka tries to explain "wrong" length perception, Cebrian explains "unexpected reliance" on the length cue. These opposite conclusions could be explained by the fact that English is described as a language with an opaque (or "deep") orthography (Katz and Frost, 1992), which makes it less consistent in terms of orthographic rules than for example German (see Section 3.4.1). While both Bogacka and Cebrian investigated L2 sound perception, Nimz (2011b) tentatively referred to the influence of orthography in relation to a production experiment she conducted with Turkish learners of L2 German. By means of a picture-naming task, she found that Turkish learners produced German long vowels with similar durations as German native speakers, but only as long as these vowels were marked in the test words' orthography by *lengthening h* (see Section 3.4.1.1 for details on this length marking). As in the other two studies, the spelling of the test words had not been controlled, which is why conclusions could only be provisional.

However, other recent studies have begun to include orthography as an independent variable. Most of these have focused on accessing the influence of orthography in L2 sound perception, while fewer studies investigated production in this context. This reflects a general trend in the field of L2 phonology, where perception studies tend to outweigh those investigating production (Munro and Bohn, 2007: 9).

3.1 Perception studies

Following the results in Weber and Cutler (2004) and Cutler et al. (2006) (see Section 2.6), Escudero et al. (2008) conducted a similar eye-tracking experiment in which they manipulated the orthographic input available to the Dutch learners of L2 English. For this purpose, 50 Dutch participants learned 20 nonce words of which 10 were critical test items which differed in the difficult contrast /ɛ/-/æ/, for example <tenzer> ([tɛnzə]) versus <tandik> ([tændək]). Crucially, half of the participants were presented with auditory input only during the word learning phase, while the other half received orthographic input as well. Fixation proportions during the testing phase suggested the same asymmetric patterns that Cutler and colleagues had found for real words: items containing [ɛ] were fixated more than words containing [æ]. For example, while a word like [tændək] triggered looks to pictures of both /ɛ/- and /æ/-words, a word like [tɛnzə] triggered looks to pictures of words containing /ɛ/ only. Importantly, this was only the case for the group which had been learning the nonce words along with their spelling. Escudero et al. interpreted this finding as evidence that orthographic information is used to establish "*phonological* lexical representations of novel L2 words" (p. 358, emphasis in the original).

Hayes-Harb et al. (2010) investigated how the quality of orthographic input (congruent versus incongruent spelling) may influence lexical-phonological representations for newly-learned words. By means of an auditory word-picture matching task they tested 33 English native speakers on 24 English-like pseudowords such as [kaməd] or [faʃə]. Crucially, one group received *congruent* orthographic input along with the auditory presentation of the new words during the word learning phase (i.e. <kamad> for [kaməd]), one group received *incongruent* input (i.e. <kamand> for [kaməd]: "silent letter" condition, or <faza> for [faʃə]: "wrong letter" condition), while the third group did not receive any orthographic input. During the testing phase, participants had to match correct and incorrect auditory presentations of the newly-learned words with their matching/mismatching pictures. Some of the incorrect auditory stimuli were completely different mismatches in that a word was presented with the wrong picture, while there was also a more "subtle" incorrect condition in which the participant heard [fazə] and saw a picture of an apple, which had been paired during the word learning phase with the auditory form [faʃə] (called "incongruent mismatch items"). While there were no significant group differences in matching accuracy between the group that received congruent orthographic input and the group that did not see any orthographic form, there were significant interactions between word learning group and item type. Participants who learned new words with incongruent orthography performed worse on incongruent mismatch items than participants in the other two groups. The authors concluded that, if orthographic forms of new words are available, they can affect phonological representations of these items. In this specific case, the influence is a negative one in that words which – according to L1 orthographic conventions – are spelled "wrong" (e.g. <faza> for [faʃə]) have a detrimental effect on the correct phonological representations of these words.

Escudero et al. (2014) further investigated the role of the L1 orthographic system in relation to the orthographic system of the L2. The method they used was again a word-picture matching task. Their participants, in contrast to Hayes-Harb (2010), were Spanish speakers who were learning or were naïve to L2 Dutch. The participants learned 12 Dutch pseudowords, of which six were the critical items which were used to form minimal pairs that were either perceptually easy or difficult to discriminate. In the perceptually difficult condition, for example, participants would hear the Dutch pseudoword [pʏx], would be presented with the pictures of a [pʏ̲x] and a [py̲x], and would have to decide which word they just heard. As in the other experimental designs, half of the participants had been presented with the orthographic forms of the pseudowords during word-learning, e.g. <pug> and <puug> for the last example. Besides the two different word-learning conditions (auditory-only and auditory-orthographic input), the grapheme-to-phoneme correspondences of the perceptually difficult contrasts were of interest to the authors. Some of the difficult pairs were classified as "congruent" pairs in that the L1 Spanish orthography and the L2 Dutch orthographic rules led to the same or similar phoneme contrast. For example, /i/-/y/ was considered a congruent pair because the (closest) phonological categories would match with different graphemes in both languages: <i>-<uu> for Dutch and <i>-<u> for Spanish. On the other hand, the former /ʏ/-/y/ contrast was classified as orthographically "incongruent", because in Dutch orthography the phonemes would map onto two different graphemes (<u>-<uu>) while in Spanish both phonological categories would map onto the same Spanish phoneme /u/, which means there would be no orthographic contrast in Spanish: <u>-<u>. Escudero and colleagues found that accuracy in word-picture matching depended not only on the difficulty of the vowel contrasts but also on the grapheme-to-phoneme correspondences across the languages involved. Participants who received both auditory and orthographic input during word learning performed worse than participants in the auditory-only condition on the *incongruent* pairs, while they performed better in the *congruent* pairs. Furthermore, they found that Spanish learners – in contrast to Spanish naïve listeners – performed significantly better on the congruent but not the incongruent pairs. The authors interpreted this finding as evidence for reinforcement of congruent grapheme-to-phoneme correspondences during L2 language learning, while incongruent correspondences persistently interfere with the acquisition of an L2 phonology.

While the previous studies were concerned with the influence of orthographic input in lexical-phonological perception/processing, Escudero and Wanrooij (2010) investigated the effect of orthographic input during Dutch vowel categorization in Spanish and Dutch speakers. In their forced-choice identification tasks, participants had to differentiate Dutch vowel contrasts which are said to pose problems to Spanish learners of L2 Dutch to varying degrees, for example /a/-/ɑ/ and /y/-/ʏ/. While their so-called AUDI identification tasks were purely auditory, their ORTH tasks involved response options which represented the Dutch spelling of the respective vowels, for example <a> for /ɑ/ and <aa> for /a/. While the /a/-/ɑ/contrast was most difficult for the Spanish participants in the AUDI tasks, it became the easiest contrast in the ORTH tasks. The authors explain this finding by referring to the Spanish participants' L1 writing system. Because Spanish is a language with transparent orthography, participants are assumed to develop orthographic representations at the grapheme-to-phoneme level (in contrast to storing and processing larger units, see for example Goswami et al., 2003). This would lead them to use the length information in the Dutch double-vowel grapheme <aa>, which helps perceiving durational differences in the stimuli, as Dutch /a/ is longer than /ɑ/. It is notable that in a later study by Escudero et al. (2014), this contrast would have in fact been classified as "incongruent" and with that more difficult instead of easier. In this later study, however, they investigated word learning instead of vowel identification which, according to the authors, makes it more difficult for Spanish learners of Dutch to rely on the newly acquired length contrast. Furthermore, Escudero and Wanrooij's (2010) study tested the *immediate* effect of orthographic input during vowel perception, while it is difficult to draw conclusions for the role of orthography in long-term phonological representations from that. One of the main implications of their study is therefore of a methodological nature, in that the authors suggest that studies investigating the influence of L1 phonology on L2 sound categorization should not use orthographic response options in their tasks.

Yet another strand of research is concerned with the familiarity and transparency of orthographic systems and how this might influence L2 phonological and lexical representations (e.g. Dornbusch, 2012, Showalter and Hayes-Harb, 2013, Mathieu, 2014, Showalter and Hayes-Harb, 2015). In Dornbusch's (2012) auditory lexical decision task[18], English native speakers and advanced Danish and German L2 English learners had to judge 100 English real and nonce words for their lexical status. Half of the real words were orthographically consistent in that their rimes could only be spelled in one way (for example, /ʌk/ as <uck>), while the other half was orthographically inconsistent, i.e. their rimes could be spelled in multiple ways (for example, /i:p/ can be spelled <eap> or <eep>). Dornbusch found an orthographic consistency effect[19] that was larger in the German L2 English learners than in the Danish L2 English learners and English native speakers, i.e. German participants made more mistakes and reacted slower on auditory presented inconsistent items than the other participants. This finding both supposes the

[18] In his dissertation, Dornbusch (2012) conducted three main experiments. The first two, however, were concerned with phonological awareness, which is why only his third experiment is discussed here.

[19] The orthographic consistency effect refers to the robust finding that words with phonological rimes that could be spelled in multiple ways produce longer auditory lexical decision times and more errors than words with rimes that could be spelled in multiple ways (see also Footnote 14).

existence of a link between orthography and L2 lexical processing, as well as the importance of native orthographic depth in this: German participants coming from a transparent orthographic background are more affected by spelling-to-sound inconsistencies because they rarely occur in their native orthographic system.

The relevance of familiarity with the L2 orthographic system was investigated by Showalter and Hayes-Harb (2015). They explored whether the presentation of an entirely unfamiliar orthography could influence the acquisition of L2 words differentiated by a novel phonological contrast. In their initial experiment, 30 native English speakers learned 12 Arabic nonce words which would form six minimal pairs, contrasted by the velar-uvular contrast /k/-/q/ as in [kubu] versus [qubu]. Half of the participants learned the words together with their randomly assigned object pictures, while the other half was additionally exposed to their written form in Arabic script. D-prime analyses of their performance in their picture-matching accuracy (after sufficient training) did not show any effects of learning condition, i.e. the presence of the unfamiliar script did not seem to help learners in acquiring new lexical-phonological representations. Positive effects were still absent when the orthographic learning conditions were simplified by providing explicit teaching or minimizing talker variability. Similarly, Mathieu (2014) investigated the acquisition of Arabic voiceless pharyngeal and uvular fricatives by English native speakers, and, using varying degrees of script unfamiliarity, did not find positive effects on target-like phonological representations either.

Studies on script familiarity therefore suggest that L2 learners are unable to use written input in a beneficial manner when the orthographic forms are presented in an entirely novel script. Furthermore, differences in orthographic depth between native and target language may influence orthographic effects in L2 phonological processing, as in Dornbusch (2012). Simon et al. (2010) also discuss the role of orthographic depth when, in their experiment with L1 English learners of L2 French, they found that orthographic input did not help learners acquire the new French /u/-/y/ contrast. They hypothesize that “L1 English listeners may be less likely to rely on spelling to create distinct phonological categories than speakers of a language with a more transparent orthographic system" (Simon et al., 2010: 391-392).

Taken together, perception studies suggest that orthographic input may be helpful in establishing lexical-phonological representations when grapheme-to-phoneme correspondences between native and target language are congruent, and languages are similar in orthographic depth. As will be shown in Section 3.4, the Polish and German orthographic systems can both be classified as shallow orthographies, even though German incorporates specific rules to mark vowel length.

3.2 Production studies

One of the first studies to investigate orthographic effects in L2 production is a study by Erdener and Burnham (2005). The authors presented 32 native Turkish speakers (transparent L1 orthography) and 32 native Australian speakers (opaque L1 orthography) with Spanish (transparent orthographic system) and Irish (opaque orthographic system) stimuli. The 48 nonwords (for each language) appeared in different experimental conditions, among them the conditions auditory-only and auditory-orthographic, which are used in most perception studies as well.[20] Participants were asked to repeat the words upon presentation and their productions were recorded and scored for phoneme errors. The authors found that Turkish speakers made fewer errors than English speakers when orthographic information was present and when the language was Spanish, i.e. transparent. However, when the orthographic information given was opaque, i.e. Irish, Turkish participants' performance was significantly worse than that of the Australian participants, while the performance of the Australian was almost equivalent for Spanish and Irish. The results suggest that speakers with a transparent native orthographic system are more affected by the (L2) orthographic system and are more likely to be misled by orthography if it does not match the L2 phonology in a straightforward way. On the other hand, speakers of languages with opaque orthographic systems may represent

[20] Erdener and Burnham were also interested in the influence of other visual input, which is not of interest here and therefore not reported.

lexical items in a more "picture-orthographic" way and therefore have weaker connections between orthography and phonology, i.e. make fewer mistakes when presented with another opaque language, but also use orthographic information less where it could help. These conclusions match well with results from other perception studies (Dornbusch, 2012, Simon et al., 2010).

Rafat (2015) investigated how the presence of orthographic input affects the production of Spanish assibilated rhotics in native English speakers with no prior knowledge of Spanish, and how acoustic input may modulate possible orthographic effects. Similar to Escudero et al. (2008), participants went through a training phase in which half of the participants were exposed to acoustic input only, while the other half received both acoustic and orthographic input. Of interest were six words (out of a total of 108 as part of a larger study on L2 production) which contained an assibilated rhotic word-finally. Three of the six words showed a higher degree of assibilation in their acoustics, which allowed for tentative conclusions as to how acoustic properties might modulate the effect of orthographic input. The training phase consisted of the auditory and visual presentation of the test words' images, while each word was assigned a new meaning via pictures of common picturable words. During training, participants in the auditory-only group heard each word three times, while the auditory-orthographic group was also presented with the written forms of the words. All participants were instructed to produce the test words directly after they had been presented with them three times. During testing, the auditory-only group saw only the picture of the word they had to produce, while the images for the auditory-orthographic group included written words as in the training phase. The results of the acoustic analyses revealed significant differences in the production patterns of the two groups in that the auditory-orthographic group produced significantly more assibilated and approximant rhotics than the auditory-only group. Furthermore, it was found that those words with higher degree of assibilation in the input triggered more assibilated rhotics productions by the auditory-orthographic group. The author ascribed this result to the idea that assibilation rather than rhoticity is the more salient feature in assibilated rhotics, but orthography helps in making the less salient rhoticity feature more salient; hence, the auditory-orthographic group produced significantly more (assibilated and approximant) rhotics. This effect, according to the author, seemed to be modulated by the acoustic characteristics of the input. Furthermore, while the production of more assibilated rhotics in the auditory-orthographic group can be viewed as a positive effect of orthographic input, the higher production of approximant rhotics in the same group is less desirable for an English learner of L2 Spanish, as the (English) approximant will likely be perceived as heavily accented. This two-sided result goes to show that it is difficult to evaluate the effect of orthography as generally positive or negative, as it depends on the specific type of grapheme-to-phoneme correspondence and how similar or different they are in the native and target language, as found in Erdener and Burnham (2005) or Escudero et al. (2014). It is not clear however whether the results can be generalized to natural productions of English L2 Spanish learners. It is possible that the results are specific to the reading process, as written input had been provided during the production task.

Young-Scholten (2004) and Young-Scholten and Langer (2015) analysed the productions of three American students who spent a year at a German secondary school. By means of monthly production tasks, Young-Scholten (2004) collected longitudinal production data over the course of a whole year. The tasks included spontaneous conversation and a series of tasks which did not include reading, such as translation tasks and picture card activities. She auditorily analysed and transcribed pronunciation errors concerning German final devoicing (as in for example [kint] for <kind>) as well as the German allophones [ç] and [x], which are both written as <ch>. The results showed higher rate of target-likeness for the velar and palatal fricatives than for the neutralization of the voiced final stops. Young-Scholten hypothesized that orthographic input over the course of second language learning is the reason for continuous erroneous voicing of devoiced consonants because the voiced plosives exist in both languages and are spelled the same in German and English. With regard to the fricatives she argues that they are "new" phonemes and, with reference to the SLM by Flege (1995), easier to acquire, and the influence of orthographic input might therefore be weaker. Young-Scholten and Langer (2015) transcribed as well as acoustically analysed word-initial <s>, which is realized as [z] in German but as [s] in English. Their finding compares to the result for final devoicing: The students applied their native grapheme-to-phoneme-correspondences to German and devoiced the fricative where it should be voiced, for example *[si:] for <sie> "she" [zi:]. Although the authors do not explain their results with reference to "spelling pronunciation", one might argue that pronouncing *[kind] for <kind> and *[si:] for <sie> is exactly that, as the term is used for the pronunciation of a word according to its spelling (Neuman, 2013). Interestingly, "spelling pronunciation" is not used in any of the research articles discussed in this chapter, while it is a well-acknowledged phenomenon in the field of pronunciation teaching. For example, Polish L2 English learners, for example, are found to produce sounds that would normally be silent, e.g. the <e> in the past-tense suffix <-ed> (Sobkowiak, 2001).

Studies by Bassetti (2006) and Bassetti and Atkinson (2015) could also be described as evidence for spelling pronunciation. In her 2006 study, Bassetti investigated L2 phonological representations by means of a phoneme counting task as well as a production task. Eighteen English first-year students of Chinese-as-a-Foreign-Language were asked to count the number of phonemes in Chinese syllables which differed in their pinyin (Chinese Latin alphabet) transcription, e.g. [iou] may be spelled <iu> (as in *liù* "six") or <you> (as in *yòu* "right side"). Participants tended to count one vowel less in syllables whose pinyin spelling omits one vowel, even though the participants were presented with the hanzi (Chinese logographic system) characters. A smaller group of 5 participants were further recruited to pronounce the same list of hanzi characters to see whether the finding extended beyond phonological awareness. The participants both read all test words and pronounced all the segments in each syllable one by one. The results showed that learners omitted the same phonemes in production that they seemed to omit in the phoneme awareness task due to the irregular spelling of the respective syllables in hanzi.

Bassetti and Atkinson (2015) conducted a series of four experiments to investigate the effects of spelling on the pronunciation of known words in experienced Italian learners of L2 English. In their first experiment they investigated silent letter-induced epenthesis in 14 Italian native-speaking high-school learners of English, for example in the word *[wɔlk] <walk>. The participants had to first read the eight test words as they were presented on a computer screen (reading-aloud task) and then repeat the same words in the second task where the orthographic input was shown only temporarily and was then presented auditorily (word repetition task). They found that in 85% of the cases, participants pronounced the silent letters in the reading-aloud task, while in the repetition task fewer participants made this mistake, although the amount was still high (56%). It is unclear, however, which task may reflect best the actual L2 phonological representations of the words. Both tasks used an orthographic prompt (even if it disappeared in the second task), while in the repetition task one cannot rule out the possibility that participants may have been affected by the traces of the native speaker's phonological input. The same problem applies to their experiment with 12 pairs of homophonic words, which are spelled differently, for example <sun> and <son>. They again found that in the reading task the majority of the pairs were pronounced differently (57%), while in the repetition task 23% of the pairs were produced non-homophonically (e.g. [sʌn] for <sun> and [sɔn] for <son>). The authors concluded that the different realizations of the same phonological form are due to the application of one or a combination of incorrect grapheme-to-phoneme correspondences.

In their third experiment, Bassetti and Atkinson (2015) collected production data on the past tense marker <ed> by means of a verb paradigm-production task. Five of the 21 target verbs had past tense markers that should correctly be produced as /Vd/ (e.g. in *painted*), six should be produced as /d/ (e.g. in *called*), and ten should be produced as /t/ (e.g. in *booked*). They found that participants produced the past tense markers to varying degrees with a /Vd/-ending, even those endings which would be produced with a voiceless stop by native speakers. The voiced productions of /t/ can be well explained by the fact that <d> represents /d/ both in Italian and English (with only a few exceptions). This result is comparable to Young-Scholten's (2004) finding that English learners of L2 rarely produce voiceless stops when they are spelled with the voiced counterpart.

The authors further investigated the production of vowel duration as a function of orthographic marking through vowel digraphs as in <seen> versus <scene>. This experiment relates to tentative conclusions by Bogacka (2004) and Cebrian

(2006) (see above). Both authors had hypothesized that their L2 English learning participants might use vowel length information in the form of double letter graphemes in that they would equate these graphemes with long vowels. In their experiment, Bassetti and Atkinson (2015) investigated seven English word pairs where the two words contained the same target long vowel, spelled with either a double letter grapheme or not. Productions were elicited by means of a reading-aloud task[21], in which the target words had to be produced three times in a carrier phrase. Vowel duration was analysed acoustically and statistical analyses were performed with the mean values for each target vowel. The authors found that vowels spelled with digraphs were significantly longer than vowels spelled with single letters by on average 14%. Unfortunately it is not clear whether the same results would have been obtained if direct orthographic input had been absent. It might be the case that durational differences can only be found in a reading task, while there are no durational differences in the lexical-phonological representations of the learners after all. Furthermore, it would have been of interest to compare the L2 learners' productions with those of English native speakers. It is possible that the findings are not specific to L2 learners, as orthographic influences on phonological processes have been observed in native speakers as well (Damian and Bowers, 2003). The other three experiments conducted by Bassetti and Atkinson (2015) can easily be interpreted as evidence for spelling pronunciation in L2 learners who apply their L1 grapheme-to-phoneme-correspondences incorrectly as for example in *[wɔlk] for <walk> or in *[askɛd] for <asked>.[22] These insights might not be as surprising as mispronunciations of this kind are "well-known effects" (Cutler, 2015: 125). The case of double vowel letters, however, could be considered a more intriguing example. In Italian, double vowel letters are pronounced as quickly rearticulated vowels, hence there is no

[21] In this experiment they did not use a repetition task. It would have been interesting to see whether the same task-induced differences as in the other experiments would have shown up for vowel duration as well.

[22] In the given examples, it is actually unclear whether learners are applying their L1 or L2 rules as they lead to the same results: <l> corresponds to /l/ in both English and Italian, and <d> corresponds to /d/ in both German and English. Young-Scholten and Langer's (2015) findings on the production of German <s> suggest that it is the L1 rules that are applied incorrectly.

direct grapheme-to-phoneme-correspondence between a digraph and a long vowel in Italian. The marking of length in English must therefore be interpreted by means of more abstract grapheme-to-phoneme-correspondences ("double vowel letter means long vowel") and cannot be explained by spelling pronunciation based on L1 orthographic rules. Unfortunately, English orthography is not very consistent in this regard. As Bogacka (2004) had already discussed in her study on English vowel length perception, participants may have been led astray by the fact that the word *hood* is spelled with a digraph, yet contains a short vowel. Still, English is the most studied language in the field. Almost every study reported – be it perception or production – involved English as either native or target language.

Experimental studies in the field appear to focus on perception rather than production. While Erdener and Burnham's (2005) study on non-native production of languages with different orthographic depths seems to support findings in the field of perception (e.g. Dornbusch, 2012, Simon et al., 2010), there is still need to investigate further how the L1 and/or L2 orthographic system influences L2 perception and production, and how and if the two domains interact. None of the studies reported systematically investigated both learners' perception *and* production, while it would be important to verify findings in either field by doing so.

Most production studies (excluding those investigating spelling pronunciation) as well as most reported perception studies investigated (near)[23] pseudowords. As the authors of word learning studies with pseudowords point out themselves, it is unclear whether newly learned words are comparable to those types of lexical-phonological representations which learners establish for familiar words (Hayes-Harb et al., 2010: 380). Bassetti and Atkinson's (2015) study on vowel duration is one of the few which investigates production of known words in experienced learners. Unfortunately, as discussed above, they used a reading task, which might not truly reflect the learners' actual lexical representations

[23] Rafat (2015) did use real words, however, they were matched with "new" pictures, meaning that the experiment was, in effect, similar to that of those studies investigating nonwords.

(which is why McAllister et al., 2002, for example, refrained from using a reading task).

Furthermore, most studies in perception and production have investigated only a small amount of words. Word learning studies are usually limited in how many words a participant can learn in one experimental session, which means that perception studies hardly have more than 10 critical items. Similarly, production studies have investigated very few items. While Bassetti and Atkinson (2015) investigated seven pairs, Rafat (2015) reported acoustic analyses of only six words.

Taken together, experimental studies on orthography and L2 perception and production point to the need of investigating further how more **real words** and their phonological forms are learned by L2 learners. Furthermore, none of the studies has collected data on **both perception and production** from the same group of subjects in order verify results in either domain. Since English is a language with an opaque orthography and since studies have shown that this might interfere with whether and how orthography plays a role (e.g. Dornbusch, 2012, Erdener and Burnham, 2005, Mathieu, 2014), it seems important to study **languages other than English**, both as L1 and L2. German and Polish present a promising combination, as both languages are considered to have rather shallow orthographies, with German being an interesting example for making vowel length in its orthography. A number of didactic papers have addressed theoretically how the German orthographic system might play a role in the acquisition of German as a foreign language. This question will be outlined in the following section.

3.3 Orthography and phonetics in GFL research

In GFL phonetics research, many publications are concerned with the prediction and analysis of production errors. Mostly, these analyses are based on phonological contrastive analyses between the native language of the learners and the target language German (e.g. Hirschfeld, 2005a). There is a general consensus in the field that German long vowels pose a prominent problem for Polish learners of GFL (see Section 2.1). Yet, how orthography may play a beneficial or adverse effect in this regard is rarely addressed. In fact, even phonetics[24] alone – without its relation to orthography – is a rather neglected area of research in GFL and has often been referred to as the “poor cousin” of other research in the field (Hirschfeld, 2003: 873; for a similar claim for L2 English see Derwing and Munro, 2005). Within the last decade, however, there seems to be a growing interest in promoting phonetics in the GFL classroom, although what role orthography may play in the development of an L2 phonology is not a central theme (e.g. Hirschfeld and Kelz, 2005, Bunk, 2005, Dieling and Hirschfeld, 2007).

Only very few papers exist which address how orthographic instruction may be relevant for GFL. Eisenberg (1995), for example, notes that orthography is hardly an issue in the field of GFL, hence, in most course books this subject is ignored. He argues that the basic rules of the German graphemic system should be taught in the GFL classroom, as it would enable students to write words correctly, even when encountered for the first time. This, of course, would assume that a learner of GFL would *hear* the new word correctly in the first place, which – especially when talking about the acquisition of German long vowels – is unlikely to be the case. In the end, Eisenberg (1995) is not concerned with how orthography may interact with the L2 phonological system, but rather how the teaching of the graphemic system is necessary for *writing* correctly in the foreign language.

With her contrastive analysis of German and Turkish, Neumann (1981) is one of the few authors who combine a phonetic with an orthographic analysis. She, like many others who have conducted phonological contrastive analyses on Turkish, predicts that Turkish learners of GFL would have problems perceiving and producing German long, tense vowels. In relation to this, she discusses how the German orthographic system may give clues regarding the length of German vowels. Short vowels may be marked by the doubling of the following consonant; long vowels may either be marked by the doubling of vowel letters

[24] In more applied approaches, the term phonetics is often used interchangeably with “pronunciation teaching”. This is the case for many publications cited in the current chapter, but for reasons of continuity the term phonetics will be used in this chapter.

(<aa, ee, oo>) or by adding <e> to <i> or <h> to <a, e, i, o, u, ü, or ä>. Hirschfeld (2005b: 15), too, in a short didactic closing of her contrastive analysis of German and Spanish (a language which does not contrast short and long vowels either), suggests that the doubling of consonants as a marker for short vowels and the role of *lengthening h* should be taught to GFL learners. Dieling (1992) gives similar didactic suggestions for Polish learners of GFL and draws attention to the fact that interference from the L1 orthographic system can often result in phonetic errors (pp. 14-16). For example, German <ch> [ç/x] may be produced by a French speaker as [ʃ], based on his native grapheme-to-phoneme correspondences. While this is an example of a negative influence of orthographic marking, she also points out that German *lengthening h* can be an important aid – for the *informed* learner (p. 43, emphasis KN).

In her contrastive analysis of German an Turkish, Rolffs (2005) hypothesizes that some Turkish GFL learners may produce the (silent) German *lengthening h* as a glottal or palatal fricative, and that German double consonants in medial position may be articulated by Turkish GFL learners as geminates (as would be the case in Turkish).[25] These kinds of production errors would be typical examples of "spelling pronunciation" (see Section 3.2 and studies by Young-Scholten and Bassetti), while the current study focuses on the more abstract influence of orthographic length marking.

Dieling and Hirschfeld (2007: 65) point out that very few GFL textbooks broach the issue of spelling-to-sound relationships in German. In most specialized didactic texts on GFL phonetics, the marking of vowel quantity through *lengthening h* and double consonant letters (as well as <ie> for /i:/ and double vowel letters) is, however briefly, mentioned (Stock and Hirschfeld, 1996: 21, Bunk, 2005: 14, Dieling and Hirschfeld, 2007: 197, Hirschfeld et al., 2007: 163).[26]

[25] The errors predicted by Rolffs (2005) did not arise in an experimental study with medium-advanced Turkish GFL learners (Nimz, 2011a). The same is true for the Polish speakers of this study.

[26] Stock and Hirschfeld (1996) as well as Hirschfeld et al. (2007) also refer to the syllable structure in determining vowel length: If a stressed vowel appears in an open syllable, it is long, e.g. <reden> [re:.dən] ("to speak"). The same is true if a word ends with only one consonant, but yields on open syllable once it is inflected, such as for example in <Weg> [ve:k] ("way") →

However, the fact that these didactic suggestions exist does not yet prove that orthographic cues really help learners in acquiring new words more correctly. This is one of the research questions which will be addressed in the main experimental part of this study.

3.4 The German and Polish orthographic systems

Both Polish and German use alphabetic writing systems which are based on the Latin alphabet. Additionally, both languages use symbols which diverge from the ordinary Roman script, mostly by means of diacritics. For Polish, this is mainly the case for its many different fricative sounds (e.g. <ź> for the voiced palatal fricative /ʑ/ or <ć> for the voiceless alveolopalatal affricate /t͡ɕ/ (Rubach, 1984). In German, we find the modified vowel letters <ü> for the rounded high front vowels /ʏ, yː/, <ö> for the rounded mid front vowels /œ, øː/[27], and <ä> for the unrounded mid vowels /ɛ, ɛː/ (Eisenberg, 2005). We also find combinations of consonantal letters such as <sch> for the voiceless postalveolar fricative /ʃ/, or /ß/ for the voiceless alveolar fricative /s/ (after long vowels and diphthongs). The main focus of this study is the marking of vowel quantity.

3.4.1 German orthography and the marking of vowel length

As will be described in detail in Section 4.1 ("The German vowel system"), German contrasts 15 vowel phonemes, of which 8 are long vowels. German long vowels are generally tense, while German short vowels are lax. In his thorough description of the German graphemic system, Eisenberg (2013) lists the following grapheme-to-phoneme correspondences for German (here without schwa):

<Wege> [veː.gə] ("ways"). There are exceptions to this rule, which will be discussed in more detail in the following chapter.

[27] The German rounded front vowels pose a special problem for Polish learners. Even though they are an interesting case, they will not be investigated in this study, as the focus lies in vowel length and the lax-tense contrast. They have been investigated thoroughly by Hentschel (1982).

German grapheme	German phoneme
<ie>	/iː/
<i>	/ɪ/
<e>	/eː/
<e>	/ɛ/
<a>	/aː/
<a>	/a/
<o>	/oː/
<o>	/ɔ/
<u>	/uː/
<u>	/ʊ/
<ü>	/yː/
<ü>	/ʏ/
<ö>	/øː/
<ö>	/œ/
<ä>	/ɛː/

Table 1: German G-P correspondences according to Eisenberg (2013) (the diacritic /:/ for marking vowel length is not used in the original)

From this overview, it becomes apparent that, with the exception of /iː/ and /ɛː/, pairs of the long/tense and short/lax vowels are mapped onto the same grapheme, for example /eː/ and /ɛ/ are both mapped onto <e>. Still, there are other ways of marking vowel length in German, which will be described in what follows.

3.4.1.1 Marking of long vowels

Although the German orthographic system is more transparent (or "shallow") than English, it is also not quite as transparent as Polish. This is because correct orthographic writing or correct pronunciation in German cannot be achieved by adherence to the G-P correspondences alone, but is further guided by the so-called syllabic writing principle [28], which includes the use of *lengthening h* and the writing of double consonant letters (Fuhrhop, 2006: 13-25). Based on the phonology of the German syllable structure, German stressed,

[28] Another important principle is the morphological writing principle (Fuhrhop, 2006: 25-32), which states that related morphemes are supposed to be written the same (when possible), even if the writing may not be motivated by the phonetic form (for example <Kind> ("child") for [kɪnt] because of [kɪndɐ] <Kinder> "children").

open syllables must always contain a long vowel, e.g. [roː.zə] (“rose”), while a phonological restriction like that does not exist in Polish, e.g. [vɔ.da] (“water”). For this reason, it could be claimed that German does not need to mark vowel length explicitly, as it is implied in its syllable structure (Eisenberg, 2005: 71).[29]

The case is different for stressed syllables with one consonant in coda position, as here both long and short vowels can appear, e.g. <Flut> [fluːt] (“flood”) and <Busch> [bʊʃ] (“bush”). For such cases, Eisenberg (2005: 73) lists two rules for deciding whether a vowel is long or short:

(1) If a one-syllable word has only one grapheme in coda position, the vowel is long, e.g. <schön> [ʃøːn] (“beautiful”).

(2) If the syllable is part of a two-[or more-]syllable word and ends with one grapheme in coda position, the vowel is short, e.g. <Kante> [kan.tə] (“edge”).

The problem with these rules is that there are many counterexamples (Ramers, 1999a). German function words which are often spelled with only one consonant but are pronounced with a short vowel (for example <an> “on”, <in> “in”, <man> “one (PRON)”, etc.) and words which are derived from other languages, such as for example <Bus>, <Pop>, <Kap>, etc. While these are counterexamples to rule (1), Ramers (1999a) further lists a large number of words from the native lexicon which do not adhere to rule (2), e.g. <Adler> [glottal aː.dlɐ] (“eagle”), <Wüste> [vyːs.tə] (“desert”), <Kloster> [kloːs.tɐ] (“abbey”), among others. Still, the German orthographic system has means to mark vowel length explicitly.

The *lengthening h* is a silent letter[30] which precedes the sonorant sounds [r, l, m, n]; this is a necessary, but not a sufficient condition in that we find words with

---

[29] This of course is speaking from the perspective of a German native speaker. If a speaker does not pronounce short, lax vowels in open syllables anyway, she does not depend on a distinction in the orthography. The case is different for Polish L2 learners, who do not have a comparable phonological restriction in their native language.

[30] The letter <h> is not only used to mark vowel length. It also functions as the grapheme <h> corresponding to the voiceless glottal fricative /h/ word-initially. Ternes (2012: 189) writes that sometimes words such as for example *Ehe* [eːə] (“marriage”) may be pronounced [eːhə] by native

long vowels written as <Bohne> ("bean") or <Lohn> ("wages"), but also words without *lengthening h* such as <Schwan> ("swan") or <Strom> ("electricity"). Eisenberg (2013: 303) calls the *lengthening h* an "aid in reading" as, for reasons captured in the Sonority Sequencing Principle[31], it is likely that another consonant may follow the respective sonorants. If that was the case, the two consonant letters would then be a misleading hint for a short vowel (see rule (2) above).

Furthermore, the doubling of vowels is used as a means to mark vowel length. Similar arguments to those for *lengthening h* apply to the doubling of the vowel graphemes <a>, <e>, and <o> in words such as <Paar> ("pair") or <Meer> ("sea"). However, the marking of vowel length through double vowel letters is considerably less common than the marking through *lengthening h*. Primus (2000) reports data from a corpus of native words only, which shows that less than 1% of all long vowels are marked by means of double vowel letters. On the other hand, <e>, <o>, and <a> are followed by lengthening h about 12% of the time.

Although Eisenberg (2005, 2013) and Ramers (1999a) do not formulate their arguments and rules explicitly for GFL learners, it seems important to make learners aware of the meaning of the *lengthening h* as a marker for vowel length. Accordingly, we find mention of this marker in papers concerned with GFL phonetics and orthography (e.g. Neumann, 1981, Dieling, 1992, Hirschfeld, 2005b). Similarly, we find mention of the doubling of consonants as a marker for short vowels in the few GFL publications concerned with this issue.

---

speakers in a hyper-correcting fashion, influenced by orthography. Furthermore, <h> it is part of the di- and trigraphs <ch> and <sch>, which correspond to the phonemes /ç/ and /ʃ/.

[31] The Sonority Sequencing Principle states that each syllable is organized around the syllable nucleus (the sonority peak of the syllable) in that preceding and/or following segments increase in sonority the closer they are to the syllable nucleus (Hall, 2011: 231). The sonority values of the segments are captured by the sonority hierarchy: open vowels are more sonorous than closed vowels than liquids than nasals than voiced fricatives than voiceless fricatives than plosives (Kohler, 1995: 74). For the case at hand, sonorant sounds such as [r, l, m, n] make it possible that another consonant could be following, especially in a language like German, in which heavy consonant clusters are possible (see Section 4.1.3).

3.4.1.2 Marking of short vowels

The gemination of consonantal letters in the German writing system is one of the most conspicuous orthographic features (Eisenberg, 1999: 343). The graphemic rules that apply to them have caused quite a debate, especially in the context of the latest German orthographic reform in 1996.

In the so-called "accent-based approach" (Ramers, 1999a), the rule which accompanies consonantal letter gemination in German is the following: if the stressed vowel in a word stem is followed by only one consonant, a short vowel is to be marked by the doubling of the following consonant (Ramers, 1999b: 53). The "syllable-based approach" (Eisenberg, 1999, 2005, 2013) stresses the function of the consonant as a *Silbengelenk* ("ambisyllabic consonant") in the phonological structure of the word. Eisenberg (2005: 77) points out that the doubling of consonantal graphemes does not have its origin in the marking of short vowels, but in the marking of ambisyllabicity. Because ambisyllabic consonants only occur after short vowels, geminate consonants in the orthography appear only after short vowels. To Eisenberg (2005), it is wrong to maintain that the short vowel alone is the reason for the doubling the consonant.

In the linguistic debate on German graphemics, the question of which approach explains the appearance of geminate letters best is hotly debated (Eisenberg, 1999, Ramers, 1999a, Ramers, 1999b). As far as teaching the phenomenon in the foreign language classroom is concerned, it is quite clear that the accent-based approach is favoured (e.g. Neumann, 1981, Rolffs, 2005). This is not only the case because it seems "easier"[32], but also because this is the way it is taught to native speakers, due to the decisions made during the last orthographic reform (Ramers, 1999b).

As already partially implied in rule (2) above, short vowels could also be said to be "marked" if the vowel is simply followed by more than one consonant. However, as Ramers (1999a: 54-55) points out, there is quite a large number of counterexamples to this rule, e.g. <Mond>, [mo:nd] ("moon"), <Papst> [pa:pst]

[32] It could be said that this approach is easier to teach, because it is not necessary to refer to suprasegmental levels of representation, such as the syllable structure.

(“pope”), or <Obst> [o:pst] (“fruit”).[33] Ramers (1999a) further points out that after consonants which are spelled with more than one letter (and which therefore do not appear as geminates in the writing), for example <ng> for [ŋ] or <ch> for [x]/[ç], rules of the syllable-based approach are again inconsistent. Before <ng> vowels are always short, while in front of <ch> we can find both long and short vowels. The same is true for vowels preceding <sch> for [ʃ] (e.g. <Dusche> /du:ʃə/ “shower”, but <Büsche> /bʏʃə/ “bushes”). These idiosyncrasies are not addressed in GFL textbooks, and the simplified rule of “double consonants indicate short vowels” therefore seems to be sufficient (e.g. Neumann, 1981, Hirschfeld, 2005b, Rolffs, 2005).

3.4.2 Polish orthography

Since Polish does not contrast short and long vowels, the issue of vowel length marking does not exist for Polish. All six Polish vowels (excluding the nasal vowels, for more details see 4.2) are symbolized by one corresponding grapheme, with the exception of /u/, which may be represented in the Polish orthography as <u> or <ó>[34] (Tworek, 2012).

Polish grapheme	**Polish phoneme**
<i>	/i/
<y>	/ɨ/
<e>	/ɛ/
<a>	/a/
<o>	/ɔ/
<u> or <ó>	/u/

Table 2: Polish G-P correspondences

The letter <i> has a special role in Polish, as it not only represents the phoneme /i/, but also signals the palatalization of the consonants /ʑ/, /ɕ/, /d͡ʑ/, and /t͡ɕ/ (Sadowska, 2012: 6). This is only the case when the palatal sounds precede a vowel (e.g. in <ziajać> [ʑajat͡ɕ] “to pant” or <dzieciak>

[33] See Section 3.4.1.1 for counterexamples in two-syllable words.

[34] In Old Polish, <ó> represented a different phoneme, but over time the quality of this sound developed into that of /u/. At times, this historical origin is preserved in the writing (Stieber, 1973).

[d͡ʑet͡ɕak] “child”), otherwise they are represented as <ź>, <ś>, <dź>, and <ć> (e.g. in <śruba> [ɕruba] “screw” or <ćma> [t͡ɕma] “moth”).

While there is no necessity for an equivalent to the German *lengthening h* in Polish, the letter <h> is still present in the Polish writing system in that it corresponds to the phoneme /x/ in borrowed words such as <historia> (“history”) (Skibicki, 2007: 2). It is also part of the digraph <ch> which represents the phoneme /x/, as well, but in native words (e.g. <chaber> “cornflower”). Its function is therefore very different from that of the German *lengthening h*.

Similarly, double consonant letters exist in Polish but, again, have a different function. As will be discussed in more detail in Section 4.2.4.1, Polish exhibits geminate consonants, which are represented in the orthography by doubling the respective consonantal grapheme, e.g. <lekki> for [lɛk:i] (“light ADJ”). Unlike in German, this orthographic marking is not related to the Polish syllable structure or Polish vowel length.

With few exceptions such as for example the representation of the phoneme /u/ by both <u> and <ó>, G-P correspondences in Polish are very consistent, which is why it is generally classified as a language with transparent or shallow orthography (Kaminska, 2003). German may be classified as slightly less transparent than Polish in that speakers/readers cannot rely on G-P correspondences alone. Most importantly, German makes use of additional ways of marking vowel length, such as *lengthening h* or double consonant letters (see above). While the function of these markings is specific to German, these explicit ways of marking vowel length may help L2 learners in acquiring some German words more correctly. This should especially be true for Polish learners, as they are used to relying on information in the spelling from their L1. This possibility forms the basis for some of the main hypotheses of this study.

4 The German and Polish vowel system

While Polish is considered a “consonantal” language with over 30 consonant phonemes (Jassem, 2003), German exhibits an unusually high [35] number of vowel phonemes. Most researchers will agree that there are 15 contrastive vowels in German, excluding the diphthongs [aɪ, aʊ, ɔɪ] and the two German schwa sounds [ə] and [ɐ] (Pompino-Marschall, 2009, Hall, 2011, Ternes, 2012)[36]. Eight of these 15 vowel phonemes are considered long, as they are on average at least twice as long as their short counterparts (Antoniadis and Strube, 1984). Polish, on the other hand, is described as a language with only six vowel phonemes, not including the two nasal vowels /ɛ̃/ and /ɔ̃/ (Gussmann, 2007, Hentschel, 1986, Sadowska, 2012). These considerable differences in vowel inventories make the two languages an ideal testing ground for the research questions of this study. In the following, the two vowel systems will be described in more detail and, towards the end of this chapter, they will be directly compared using acoustic data which were collected in an exploratory study prior to the main experiments.

4.1 The German vowel system

Only about 20% of the world’s languages exhibit vocalic quantity distinctions, i.e. durational differences in the productions of vowels (Maddieson, 1984). German is one of them, though for most of the contrastive vowel pairs there is a complex interplay between vowel length and vowel quality: Long vowels are usually tense, while short vowels are lax. One exception is the vowel pairs /a/-/aː/ and /ɛ/-/ɛː/, even though the functional load of /ɛː/ is questionable. It is commonly asserted that /ɛː/ is often substituted with /eː/ (Wängler, 1974), and Ternes (2012) points out that many speakers, especially from Northern Germany, do not produce [ɛː] at all. For this reason, /ɛː/ was not investigated in this study.

---

[35] Most languages contrast between five to seven vowel phonemes (Maddieson, 1984).

[36] In his phonological account of the German vowel system, Becker (1998) postulates only 8 vowel phonemes, as in his analysis he emphasizes the role of German syllable structure in determining whether a vowel is short or long.

Because of their distribution, stressed[37] German vowels have traditionally been divided into two groups (Ramers, 1988):

Group (a): long, tense vowels /iː yː uː eː ɛː øː oː aː/

Group (b): short, lax vowels /ɪ ʏ ʊ ɛ œ ɔ a/

Vowels in group (a) can appear in both open and closed syllables, for example *Lied* [liːt] ("song"), *Lieder* [liː.dɐ] ("songs"). Vowels in group (b) are restricted to closed syllables, i.e. *Bett* [bɛt] ("bed"), *mit* [mɪt] ("with"), etc.

From a phonological point of view, it is of interest which vocalic feature may be the primary one in distinguishing the two groups: vowel length or vowel quality. While it is not important to make a definite claim for the current study, the differentiation between the two dimensions "length" and "quality" will play an important role in the design of the experiments to follow (Chapters 5-7).

4.1.1 Vowel quality

Most vowel systems can be exhaustively described using the three parameters tongue height, tongue backness, and lip rounding (Ladefoged and Maddieson, 1990). Depending on the position of the tongue and the configuration of the lips, different vowel sounds with their respective vowel qualities will be produced. A well-established way of presenting vowel systems schematically is the use of vowel quadrilaterals. Hence, in its description of various vowel systems of the languages of the world, the International Phonetic Association (IPA) makes use of this way of depiction. Figure 1 shows the German vowel quadrilateral as published in the *Handbook of the IPA*. It is meant to symbolize a mid-sagittal section of the part of the vocal tract in which vowel articulation takes place.

[37] The description is restricted to vowels in stressed syllables only. In unstressed syllables, tense vowels are shortened, for example *vielleicht* [fiˈlaɪçt] (Hall, 2011: 69). Investigating both stressed and unstressed vowels is beyond the scope of the present study.

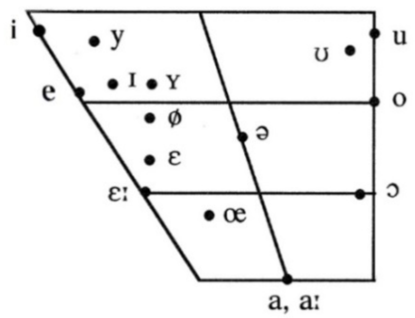


Figure 1: German vowel quadrilateral (Kohler, 1999: 87)

The position of each vowel symbol is supposed to refer to the highest point of the tongue during articulation of the respective vowel. The positioning of the symbols relates to Daniel Jones' cardinal vowel system (Jones, 1917), who defined extreme vowels as reference points by both articulatory and auditory means. While the vowels [i], [u], and [ɑ] were described in articulatory terms (for example, [i] is the vowel with the highest, most fronted tongue position), all other cardinal vowels were defined by Jones (1917) as equal auditory steps between the three point vowels (Pompino-Marschall, 2009: 222). (For a slightly modified version of the cardinal vowels including centralized vowels see Appendix I, "Vowels"). It is customary today to refer to vowel qualities by means of articulatory terms. This is the case despite the fact that x-ray studies have shown that the articulatory parameters do not always describe the relative tongue position correctly (Wood, 1982). Attempts to use other terms, such as for example "acute" for front vowels or "grave" for back vowels, have failed (Ladefoged, 2006: 88). Hence, the parameters tongue height and tongue backness are commonly used. For example, German [u:] is described as the vowel for which the highest point of the tongue is the highest and most back. Lip rounding is not explicitly marked in the IPA depiction above, as it can usually be inferred from the fact that unmarked front vowels are unrounded and unmarked back vowels are rounded. Since German exhibits typologically rare rounded front vowels ([y:], [ʏ], [ø:], [œ]), they are explicitly marked as such in other depictions (e.g. Pompino-Marschall, 2009: 266). Even though they are an interesting case, they are not investigated further in the current study.[38]

From an acoustic point of view, the terms height and backness are related to the first (F1) and second formant (F2) of a vowel. If F1 and F2 are plotted on adjusted axes, they can approximately reflect the respective tongue position of a vowel, even though F2 is affected by lip rounding, i.e. it is lowered (Ladefoged, 2006). In the experimental part of the study, F1 and F2 are used to depict the vowels as they are produced by the participants of this study, as this is still considered the "most useful representation of the vowels of a language" (Ladefoged, 2001: 39).

Phonological descriptions of German either postulate three (e.g. Becker, 1998), or four distinctive vowel heights (e.g. Ternes, 2012). As far as vowel backness is concerned, phonologists either differentiate front and back vowels (Becker, 1998), or front, central, and back vowels (Meinhold and Stock, 1982). Table 3 shows a more phonetically-oriented summary of the German vowels, adapted from Morciniec (1990). In the original depiction, the low vowels are classified as front ([a]) and back ([ɑ:]). Here, the two vowels are represented as central vowels, which is in line with the descriptions of most researchers (e.g. Kohler, 1999, Pompino-Marschall, 2009, Ternes, 2012).

	Front	Central	Back
High	**i:** y:		**u:**
Open-high	**ɪ** ʏ		**ʊ**
Mid	**e:** ø:	**ə**	**o:**
Open-mid	**ɛ(:)** œ		**ɔ**
Low		**a a:**	

Table 3: The German vowel system partially modified on the basis of Morciniec (1990) (non-bold symbols represent the typologically rare rounded front vowels)

[38] As will be laid out in Section 6.3, the choice of experimental items was limited. For this reason, the study focuses on the vowels [a:, a, e:, ɛ, o:, ɔ].

By differentiating high/high-open and mid/mid-open vowels, the difference in tenseness for example between /iː/ and /ɪ/ is also indirectly captured. Ladefoged and Maddieson (1990) list the opposition between *tense* and *lax* as one of the many secondary features in describing vowels in the world's languages, though for German it is a central aspect. Even though the term tenseness could suggest that tense vowels are produced with more muscular tension, this assumption has not been verified (Becker, 1998: 47). Other phonetic correlates could be found in the form of articulatory (Mooshammer, 1998) and acoustic (Jørgensen, 1969) data, which showed that lax vowels are more centralized than tense vowels. This means that the tense high front vowel /iː/ is both higher (as captured in Table 3) and more front than lax /ɪ/. The low central vowels are an exception, as they are generally assumed to be produced with the same vowel quality (see above). For this reason, tenseness alone does not suffice to divide all German vowels into the two afore mentioned groups, while vowel length does.

4.1.2 Vowel length

The fact that German differentiates short and long vowels is undisputed and unanimously described in phonetic and phonological descriptions of German (e.g. Kohler, 1995, Pompino-Marschall, 2009, Hall, 2011). As mentioned above, long high and mid vowels are usually tense, while the short ones are lax. From a phonological point of view, it is of interest to decide which of the two (tenseness or length) is the primary feature in distinguishing the vowel pairs. While Kohler (1995: 142) argues for vowel quality to be the decisive feature due to findings in Northern German speakers who did not produce durational differences between some of the vowel pairs, Ternes (2012: 92-94) comes to the conclusion that vowel length should be considered the primary feature, as the pairs /a/-/aː/ and /ɛ/-/ɛː/ cannot be differentiated by vowel quality alone. Early perception studies with German native speakers (Weiss, 1974, Sendlmeier, 1981) came to the conclusion that it is not possible to uniformly say whether quality or length is the more important dimension in the German vowel system. In his study with German native speakers, Sendlmeier (1981) manipulated 28 German minimal pairs such as *Schiff–schief* ("ship"–"crooked"), which covered all German lax-tense vowel pairs. The manipulation was such that long tense vowels were shortened to the average length of their corresponding short counterparts, while short lax vowels were lengthened to the average length of their corresponding long counterparts. The participants than had to choose from a list of possible answers which word they heard. On the one hand, Sendlmeier found that shortened high tense vowels (i.e. [i], [y], and [u]) were most often still heard as the same long vowels, which would support the view that vowel *quality* is more important in native vowel perception. On the other hand, he found that the shortened tense mid vowel [e] was mostly heard as [ɪ] (68% of time), and shortened [o] was often heard as [ʊ]. The reverse was also true for the lengthened short vowels. This result not only suggests that length is more important for the mid vowels but, further, that the quality differences between, for example, [e:] and [ɛ] seem to be larger than for example between [e:] and [ɪ]. Furthermore, all shortened long [ɑ]-vowels[39] were heard as short [a] (100%), and almost all lengthened short [a:]-vowels were heard as [ɑ:] (98%). For the low central vowels it therefore seems to be clear that length is the important dimension, which is why Sendlmeier comes to the conclusion that in fact the same vowel symbols should be used, i.e. [a] for the short and [a:] for the long vowel.

German long vowels are on average twice as long as their short counterparts (Antoniadis and Strube, 1984, Nimz, 2011a). Antoniadis and Strube (1984) had measured the specific vowel durations spoken by three male native speakers. The vowels appeared in three different consonantal contexts ([p, t, k]) in the form [CVCə] and were spoken in a carrier phrase, with 10 repetitions for each vowel, i.e. 90 productions of each vowel. Table 4 shows the average vowel duration values from Antoniadis and Strube (1984) for 5 German vowel pairs[40].

[39] Like Morciniec (1990), Sendlmeier initially used the symbol /ɑ:/ for the long vowel.

[40] In the following, the vowel pairs which have traditionally been grouped together will be referred to as, for example, the "u-pair" (for the /u:/-/ʊ/ pair). According to Sendlmeier (1985), this common grouping (which is also reflected in the same graphemes for the vowels of a pair) is due to the phonetic similarity of the two sounds. Yet, as Sendlmeier (1981) had pointed out himself, this is not necessarily true for every pair (i.e. the e-pair, where /e:/ is also close to /ɪ/).

"Pair"	Long vowel	Short vowel	Ratio
i	137 ms	62 ms	2.2
u	141 ms	63 ms	2.2
e	155 ms	81 ms	1.9
o	155 ms	76 ms	2.0
a	184 ms	78 ms	2.4

Table 4: Vowel duration values in Antoniadis and Strube (1984)

For [e:]/[ɛ], the ratio was the smallest, while the ratio for the a-pair was the highest. Yet, all ratios indicate that long vowels are about twice as long as their short counterparts. Their long vowels measured on average 154 ms and their short vowels measured about 72 ms (average ratio: 2.1). As can be seen more clearly in the long vowel data, they further found that vowel quality had a significant effect on vowel duration in that higher vowels were generally shorter than lower ones. This observation is consistent with universal tendencies of intrinsic vowel durations (Kohler, 1995).

In her unpublished Master's thesis, Nimz (2011a) had measured the average vowel length productions of the point vowel pairs /u:/-/ʊ/, /i:/-/ɪ/, and /a:/-/a/ by eight native speakers of German (and Turkish). Each vowel appeared in three different monosyllabic German words, which were produced three times in random order. Consonantal context could not be controlled for as the main prerequisite for the picture-naming task was imageability; hence, the choice of test words was limited. The average length of the long vowels was 148 ms (SD=39 ms, 216 data points), and the average length of the short vowels was 85 ms (SD=23 ms, 216 data points). With that, her average vowel duration ratio was slightly smaller than that of Antoniadis and Strube (1984), namely 1.7.

In a recent study, Weirich and Simpson (to appear) investigated the influence of speaker sex on German vowel durations. They had 5 female and 6 male German native speakers produce 5 lax-tense contrasts in accented and unaccented sentence position. Accented tense vowels showed a tendency to be longer in female speakers than in male speakers (in the vowels [a:], [e:], and [o:]). In the unaccented condition, vowel contrasts (expressed as the duration difference between long and short vowels) were also significantly larger in female speakers.

Port and O'Dell (1985) investigated vowel duration as a function of underlying word-final voicing. Even though German is well known for its neutralization of the voicing contrast in syllable-final obstruents ("Auslautverhärtung"), the researchers investigated several acoustic parameters which suggested that this neutralization may be incomplete. In their study, 10 German native speakers produced 10 German minimal pairs which differed only in their underlying word-final voicing (e.g., /rad/ "wheel" versus /rat/ "advice"). The researchers found that, among other parameters, vowels before underlying voiced consonants were significantly longer than those before voiceless consonants by about 15 ms. By averaging over the reported mean values for each word, vowels before voiceless obstruents measured on average 227 ms.[41] In relative terms this means that vowels before underlying voiced consonants were on average 7% longer. In order to investigate whether this difference can be used as a cue for word identity by German listeners, they further conducted a perception experiment with the tokens produced in the previous production experiment. The 10 subjects were able to distinguish the voiced and voiceless pairs with an accuracy of 59%, which was significantly different from chance. However, the authors did not propose that the slight differences in acoustic cues serve a communicative function.

4.1.3 Syllable structure

Because the level of the syllable plays a role in the analysis of the production data (Chapter 6), a brief description of this suprasegmental level is included as well. As in many other languages, the syllable is the most obvious and salient prosodic unit in German (Wiese, 2000: 33). While it has been asserted that a phonetic definition of the syllable as such is impossible (Ladefoged and Maddieson, 93-94), it is an important unit in phonological descriptions. For example, as mentioned above, the syllable is the domain of final devoicing in German: /kɪnd.lɪç/ → [kɪnt.lɪç]. As is usually the case in the languages of the

[41] This duration seems rather long in comparison to the values reported in the previous studies. This may be related to the fact that vowel durations were taken from oscillograms only, which for some test words meant that the measurement included the following sonorant as well. Hence, absolute vowel duration values need to be interpreted with caution in this study.

world, vowels form the most common syllable nuclei in German too, while the syllabic consonants (nasals and [l̩]) may at times also serve the function of the syllable nucleus (e.g. *laufen* [laʊ.fn̩] "to run" or *Hagel* [haːgl̩] "hail").

German syllable structure can be relatively complex and together with the Slavic languages it is described as one of the more "extreme" cases among the languages of Europe (Ternes, 2012: 188). Not taking morpheme boundaries into account, a German one-syllable word may be as complex as CCCVCCCC, for example in the word *strolchst* [ʃtrɔlçst] ("to vagabond-2SG.PRS"), where each consonantal phoneme takes a C slot and the vowel [ɔ] the V slot. The simplest structure, namely V, is exemplified in the word *Ei* /aɪ/ ("egg") (Ternes, 2012: 186). However, this is only true for a phonological analysis which does not assume the glottal stop to be a phoneme of German, as phonetically speaking every syllable-initial, stressed German vowel is preceded by [ʔ], e.g. [ʔaɪ] (Hall, 2011: 236). Furthermore, this example assumes the diphthong to be *one* segment, while in some phonological analyses the second part of the diphthong is analysed as a separate segment filling the coda position of the syllable (Ramers and Vater, 1995, Eisenberg, 2013).

Not every consonantal sound may take any onset position in a structure that is more complex than CV. In a structure such as C_1C_2V, C_1 has to be an obstruent while C_2 can either be a sonorant or an obstruent (for a more detailed discussion of the restrictions see Hall, 2011: 237-240). Most German clusters adhere to the Sonority Sequencing Principle (SSP) (see Footnote 31), with exceptions such as [ʃp] and [ʃt] in for example *Spiel* ("play") or *Stuhl* ("chair"). As far as the syllable coda is concerned, a German syllable may consist of zero to four consonants. In a structure such as VC, C may be filled with any consonant except [h] or voiced obstruents, as they are subject to final devoicing. When more than one consonant constitutes the coda cluster, the SSP again plays an important role. If all obstruents (whether fricatives or plosives) are grouped into one class, as in Hall (2011), it can be observed that all combinations of two consonants are possible in the coda as long as the sonority decreases from nucleus to the last consonant (for exceptions see Hall, 2011: 241-244). Words like *Koks* ("coke") and *Keks* ("cookie") seem to be exceptions, and so do some words with more than two consonants in coda position, for example

Obst ("fruits") or *Herbst* ("autumn"). Instances like these are often explained with the concept of extrasyllabicity, namely the idea that some segments are not assigned to any syllable at all (Wiese, 2000: 47-49). If the syllable is stressed and the coda position is not filled with any consonant, the vowel has to be long. Stressed open syllables with a short vowel do not exist in German (Ternes, 2012: 189).

Syllables may be stressed or not. In German, stressed syllables are louder and longer (Mengel, 2000: 176), while in other languages F0 may be more important in identifying lexical stress (e.g. in Polish). German is usually described as a language with free stress placement with a tendency to stress the word stem (Bußmann, 2008: 22), for example *sichtbar* ("visible"). Stress may be placed on any syllable, however, and may move from the stem to an affix (e.g. *unsichtbar* "invisible"). Hall (1992: 24) tries to formulate a stress rule for German which states that the final syllable of a word is stressed if its coda is filled. If the syllable does not contain a sound in the coda position, the penultimate syllable is stressed. If the penultimate syllable does not contain a sound in the coda position either, the antepenultimate syllable is stressed. The example above is one of many exceptions. Thus, Grzeszczakowska-Pawlikowska (2007) describes German lexical stress rules as rather complex and views this as a general feature of the rhythmic class to which German belongs. It is unanimously described as a stress-timed language. [42] This means that unstressed syllables may be compressed to fit into the given time interval between stressed syllables. This especially affects the vocalic syllable nuclei, most prominently schwa [ə], which may at times be omitted entirely (Pompino-Marschall, 2009: 248).

[42] The differentiation between stress- versus syllable-timed languages goes back to Pike (1945), who postulated that in a language like English, it is stress that reoccurs at regular intervals (hence "stress-timed"), while in other languages like for example Spanish it is the unit of the syllable which divides time into equal portions (hence "syllable-timed"). Phonetic measurements have challenged the "isochrony theory", as empirical evidence does not always seem to support this classification (Dauer, 1983). Still, the terminology is widely used and reinterpreted into other phonological correlates such as for example syllable complexity or vowel reduction (Ramus et al., 2003).

4.2 The Polish vowel system

4.2.1 Vowel quality

Polish is usually described as a language with six contrastive oral vowels and the two nasal vowels /ɔ̃/ and /ɛ̃/ (Wójtowicz, 1981, Morciniec, 1990, Skibicki, 2007, Sadowska, 2012). Jassem (2003), in his IPA article on Polish, does not list the nasal vowels. Hence, his overview as shown in
Figure 2 includes only the six phonemes /i/, /ɨ/, /e/, /a/, /o/, and /u/, as they appear in his examples in (1).

(1) a. /bitɨ/ *bity* "beaten"[43] d. /bati/ *baty* "whips"

b. /bɨtɨ/ *byty* "entities" e. /botɨ/ *boty* "women's high-boots"

c. /betɨ/ *bety* "bedding (coll.)" f. /butɨ/ *buty* "shoes"

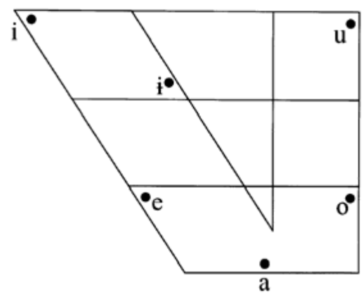


Figure 2: Polish vowels quadrilateral (Jassem, 2003: 105)

The "vexed question of the Polish nasal vowels", as Gussmann (2007: 2) calls it, is not problematic for the current study because this feature is absent from German. The phonological status of nasality in Polish is not as straightforward as for example in French, as both /ɔ̃/ and /ɛ̃/ may be pronounced either as (lightly) nasalized[44] vowels, as an oral vowel, or as a combination of an oral vowel plus nasal consonant, depending on the position within the word and the consonantal context (Morciniec, 1990: 22-24). Yet, since neither the

[43] Depending on which phonological assumptions form the basis of this analysis (see below), one might transcribe the bilabial consonant preceding /i/ as palatalized /bʲ/ instead. However, both for Jassem (2003) and for this study, this question is not of importance.

[44] While Morciniec (1990) speaks of nasalized and "lightly" nasalized vowels, Gussmann (2007) describes the realization of some nasal vowels as an oral vowel followed by a nasalized labio-velar glide, e.g. [gew̃stɨ] for *gęsty* ("thick").

phonological status of nasality in Polish nor its phonetic realizations are related to the research questions of this study, the following descriptions will not be concerned with the nasal vowels any longer.

Another theoretical issue is the phonological status of Polish [ɨ]. Bethin (1992: 32) for example considers this vowel a positional variant of the phoneme /i/, as it appears after non-palatalized consonants only. The sound [i] is found after palatalized consonants and word-initially (examples in (2) are taken from Gussmann, 2007: 33-34):

(2) a. [pʲiw] *pił* "he drank" – [pɨw] *pył* "dust"
b. [mʲiwɨ] *miły* "nice" – [mɨwɨ] *myły* "they washed"
c. [iɕtɕ] iść "go" – Ø

From a phonological perspective, it may be an elegant solution to subsume [ɨ] under the phoneme /i/, as the system could then be reduced to five instead of six vowel phonemes. Yet, structuralist arguments such as this are not relevant for phonetic descriptions of Polish, hence, most studies which take a more applied approach grant /ɨ/ its full phoneme status (e.g. Biedrzycki, 1974, Morciniec, 1990, Jassem, 2003).

In terms of tongue height and backness, Morciniec (1990: 20) summarises the Polish vowel system as shown in Table 5 (here excluding nasality for reasons described above):

	Front	Central	Back
High	**i**	**ɨ**	**u**
Mid	**ɛ**		**ɔ**
Low		**a**	

Table 5: The Polish vowel system partially modified on the basis of Morciniec (1990)

The symbols Morciniec and many others use for the mid vowels differ from those Jassem (2003) uses (see
Figure 2). Even though the choice of symbols does not change the analysis from a phonological point of view, it is still assumed that the more open symbols [ɛ]

and [ɔ] do more justice to the phonetic realization of the respective vowels. Biedrzycki (1974: 60), for example, asserts that Polish [ɛ] is similar to the short, lax German vowel [ɛ] in Bett [bɛt] ("bed"), and Polish [ɔ] to the short, lax German vowel [ɔ] in Post /pɔst/ ("mail"). While Biedrzycki's analysis is based on subjective auditory judgements[45], Hentschel (1986) provides a more objective assessment by means of an acoustic comparison between German and Polish. Even though he remarks that the data were collected using two different experimental tasks, they are still an important reference, because they constitute the only source for contrastive German-Polish formant frequencies.

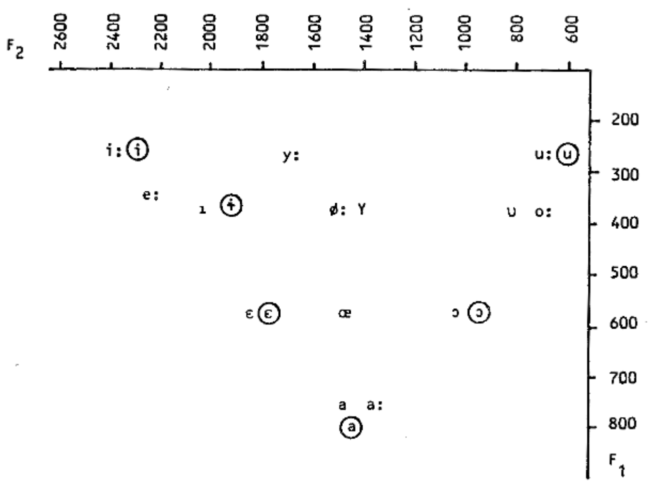


Figure 3: Acoustic comparison of the Polish and German vowels (Hentschel, 1986)

As can be inferred from Figure 3, the two Polish mid vowels are almost identical in their quality to the corresponding German short, lax vowels /ɛ/ and /ɔ/, as was postulated by Biedrzycki (1974). Furthermore, Polish /i/ is very similar in its vowel quality to tense German /i:/, and Polish /u/ seems to be of the same quality as tense German /u:/. Polish /ɨ/ is similar to German /ɪ/, while being slightly more central. Polish /a/ is also very similar to German /a/ and /a:/. The German vowels /e:/, /o:/, and /ʊ/ do not seem to have a clear counterpart in Polish. The quality of German /e:/ seems to be between Polish /i/ and /ɨ/, while German /o:/ and /ʊ/ are between Polish /ɔ/ and /u/ in the

[45] He does not specify this in his introduction, but it can be fairly safely assumed, as he does not mention any other (experimental) techniques.

acoustic plain.[46] Unfortunately, Hentschel did not provide any information as to how many data points are presented in the graph, nor does he specify number, age, and sex of the speakers.

In his study, Hentschel (1986) further set out to investigate the perceptual assimilation patterns in 35 naïve Polish listeners of the German vowels to the six Polish. The 15 German vowels were presented auditorily in various consonantal contexts and the Polish participants were ask to label these German vowels as one of the six Polish categories (or as "foreign", if they found them to be too different). By investigating Polish speakers who were naïve to the language of investigation, Hentschel, in 1986, already adhered to an important prerequisite in perceptual similarity studies today (Strange and Shafer, 2008). The acoustic similarities established above matched his perceptual results in that the vowels that seemed to be almost identical acoustically were clearly mapped onto one native category, such as for example German /i:/ to Polish /i/, or German /a/ and /a:/ to Polish /a/. Hentschel's criterion for a German category to be "clearly" mapped was that the Polish equivalent was chosen in 90% of the cases, while alternatives were chosen less than 5% of the time. Hence, German /ɪ/, /e:/, /o:/, and /ʊ/ were *not* clearly mapped onto one Polish equivalent, but instead were assimilated to two different Polish vowels. German /ɪ/ was perceived as Polish /ɨ/ in 59% of the cases, while it was categorized as Polish /i/ in 37% of the cases.[47] German /e:/ was categorized as Polish /i/ 74% of the time and as Polish /ɨ/ 23% of the time. German /o:/ was perceived as Polish /u/ (73%) and Polish /ɔ/ (26%), and German /ʊ/ was also predominantly perceived as Polish /u/ (86%) and less so as Polish /ɔ/ (13%).

[46] The relationship between the German front rounded vowels and Polish vowels will not be discussed in detail, as they are assumed to pose a special problem for learners. See also Footnote 27.

[47] The numbers do not add up to 100% because in some instances (less than 5% in all cases) the vowels were assimilated to other Polish vowel categories.

German vowel →	Polish vowel
/iː/	/i/
/ɪ/	/ɨ/, /i/
/eː/	/i/, /ɨ/
/ɛ/	/ɛ/
/aː/	/a/
/a/	/a/
/ɔ/	/ɔ/
/oː/	/u/, /ɔ/
/ʊ/	/u/, /ɔ/
/uː/	/u/

Table 6: Perceptual assimilation patterns of German to Polish vowels by Polish speakers without knowledge of German as found by Hentschel (1986)

For reasons described in Section 6.3, the current study focuses on the vowel pairs /aː/-/a/, /eː/-/ɛ/, and /oː/-/ɔ/. Applying the terminology of Best and Tyler's (2007) Perceptual Assimilation Model for L2 speech (2.3) to the perceptual assimilation data above, the /aː/-/a/ pair can be classified as an instance of Single Category assimilation. This implies that the discrimination of the (prototypical) pair would be very difficult, as the two German vowels are perceived as the same Polish category. On the contrary, since the German vowels /eː/ and /ɛ/ are mapped onto different Polish categories (Two Category assimilation), discrimination should be very easy for this pair. Discrimination for the /oː/-/ɔ/ would also be predicted to be easy, as their predominant mappings are onto different Polish categories, as well. It could be argued that this pair is slightly more difficult than the /eː/-/ɛ/ pair, as they share the same Polish category, namely /ɔ/, 26% of the time. Hence, it could also be understood as a case of Category Goodness assimilation (i.e. "good" discrimination according to Best and Tyler, 2007). Even though the discrimination task of the study (Chapter 5) primarily sets out to answer open questions regarding vowel length and quality perception separately, the testing of the predictions above is a side effect. In fact, in his dissertation, Hentschel (1986) had conducted a discrimination task which included – among others – these three pairs. As would be predicted by the PAM-L2, he found that /eː/-/ɛ/

and /o:/-/ɔ/ were discriminated perfectly by Polish speakers (100% and 99% correct, respectively), while the low pair /a:/-/a/ posed a perceptual problem, in that learners could discriminate them in only 59% of the cases. The fact that Polish learners discriminated the mid-vowel pairs equally well supports the classification of both pairs as cases of Two Category assimilation.

Furthermore, Hentschel (1986) investigated the perception of diphthongization of German vowels by Polish native speakers. As was mentioned in Section 2.1, GFL researchers have reported that German /e:/ may be realized as [ei] or [ej] by Polish L2 German learners. Hentschel too was aware of similar observations, and further reported that German /o:/ may also be produced as [ɔu]/[ɔw], quoting Prędota's (1978) work on Polish-German pronunciation interferences. Hentschel postulated a reinterpretation of vowel length as a cause for diphthongization, which led him to further conduct a modified identification task with the same German vowels as in his first identification task. In this modified version, he instructed the participants to indicate whether the vowel they had just heard was a normal instance of a Polish vowel ([V]), a longer version of a Polish vowel ([V:]), or a diphthongized version of a Polish vowel with either [i] ([Vi]) or [u] ([Vu]) as the second element. Despite the fact that his instructions may have influenced the perception of the participants[48], the high percentages of diphthong (and length) identifications led him to make statements regarding favourable conditions for diphthong perceptions. Table 7 shows that only high vowels are subject to diphthong perception, even monophthongs are sometimes perceived as diphthongs (/ʊ/ 22% of the time and /ɪ/ 18% of the time). Hentschel explains this – with reference to Donegan's (1978) Natural Phonology account of vowels – by the fact that monophthongs have a natural tendency to be replaced by diphthongs when they are high, tense, and long.

[48] By giving the options "long vowel" or "diphthongized vowel", the researcher implies that some of the vowels *are* long or diphthongized (which is not the case for the latter). This may have influenced the participants to judge some items as diphthongized, which they might otherwise not have done.

German V	[Vi] (%)	[Vu] (%)	[V:] (%)	[V] (%)
/e:/	**42**	**-**	50	6
/o:/	**-**	**41**	53	5
/u:/	**-**	**33**	59	6
/i:/	**30**	**-**	61	8
/ʊ/	**-**	**22**	-	72
/ɪ/	**18**	**-**	-	77
/ɛ/	-	-	-	96
/ɔ/	-	-	5	91
/a:/	-	-	81	17
/a/	-	-	-	95

Table 7: Identification patterns of German vowels by Polish native speakers in Hentschel's (1986) "diphthong test" (values below 5% are not included; diphthong identifications are highlighted in bold)

All long, high vowels are largely (between 30% and 42% of the time) perceived as diphthongs, which seems to support Hentschel's assumption that German vowel length is reinterpreted perceptually by Polish listeners. Furthermore, those vowels are more likely to be perceived as two segments which are assimilated into two Polish categories, namely /e:/ and /o:/. Even though these perception data are an important (and the only) starting point for explaining some of the later findings of the current study (Chapter 6), it would have been of interest to compare the results to those of native speakers of German. Would these have similar (natural) tendencies to perceive long vowels partially as diphthongs when given the same response options? Even though it seems unlikely, questions like these point to the importance of including a control group, which will be the case for the experimental part of this study.

4.2.2 Vowel length

While long vowels were present in Old Polish (Stieber, 1973), modern Polish is unanimously described as a language without contrastive vowel length (Jassem, 2003, Gussmann, 2007, Tworek, 2012, among many others). Biedrzycki (1974: 59) writes that all vowels are produced relatively short (Biedrzycki, 1974: 59). His book is an introduction to Polish phonetics for German native speakers; it is not an experimental phonetic account. He impressionistically compared and described German and Polish vowels, and equates the length of German short vowels with that of Polish (short) vowels. Hentschel (1986: 135-136) too assumes the average Polish vowel length to be similar to that of the short German vowels, but does not provide any empirical evidence. Since Polish does not differentiate between short and long vowels, there are only few studies which investigated Polish vowel length experimentally. Exceptions are Frąckowiak-Richter (1973), Keating (1984), Slowiaczek and Dinnsen (1985) and Slowiaczek and Szymanska (1989).

Frąckowiak-Richter (1973) had investigated the duration of Polish vowels by measuring them in one- and two-syllable nonce words spoken by 10 native speakers in varying consonantal contexts. She found that vowels in monosyllables were significantly longer than in disyllables, as can be seen from her measurements in Table 8.

Vowel	Monosyllables	Disyllables
/**u**/	113 ms	88 ms
/**i**/	119 ms	78 ms
/**ɨ**/	127 ms	90 ms
/**ɔ**/	135 ms	110 ms
/**ɛ**/	143 ms	111 ms
/**a**/	151 ms	124 ms

Table 8: Mean durations of Polish vowels in mono- and disyllabic words as reported in Frąckowiak-Richter (1973)

In monosyllabic words, Polish vowels in monosyllables measured on average 131 ms. Vowels in disyllables were on average 100 ms long. Furthermore, Table 8 reflects a universal influence of vowel quality on vowel duration, which was also reported for German: high vowels are generally shorter than low vowels. Within the disyllables, Frąckowiak-Richter also investigated the influence of voice in the postvocalic consonant on the duration of the preceding vowel. She found that vowels are consistently longer before voiced consonants, on average 16%. This result supports Chen (1970), who postulated that vowel lengthening before voiced consonants is a language universal phenomenon.

Other studies investigated Polish vowel duration as a function of the following consonant as well. Keating (1984) measured vowel durations before voiced and voiceless plosives in the Polish words *rata* ("rate") and *rada* ("advice"). Twenty-four native speakers of Polish were recorded reading this pair in isolation, and the durations of the stressed vowel were measured from the oscillographic display. Keating reported a mean duration of [a] before [t] of 167 ms, and for [a] before [d] of 169 ms. These results are interesting in two important aspects. Firstly, they contrast with what Frąckowiak-Richter (1973) had found. While Keating did not find a significant duration difference for vowels preceding voiced versus voiceless consonants, Frąckowiak-Richter did. Secondly, the overall durations Keating measured seem rather long for vowels which are described as "relatively short". In Antoniadis and Strube (1984), for example, short German [a] was on average 78 ms long, while the long counterpart measured 184 ms on average. This comparison would imply that Polish vowels are more similar in length to German long vowels than German short vowels. The scarceness of consistent and comparable data led to the measurement of Polish vowel length (and quality) in the exploratory part of the study below.

Slowiaczek and Dinnsen (1985) further investigated whether Polish vowel durations may differ preceding voiceless and *underlying* voiced consonants. Like German, Polish exhibits final devoicing. All word-final obstruents are assumed to be produced as voiceless, for example /karb/ <karb> ("notch") is pronounced [karp]. Because it was found that the underlying voicing distinction is sometimes phonetically preserved in other languages, the researchers examined 15 Polish minimal pairs which contrasted in the underlying voicing of the word-final consonant only (e.g., /log/ "logarithem" versus /lok/ "curl"). Five Polish native speakers produced these 30 words in randomized order and their productions were analysed acoustically for preceding vowel duration as well as consonant and glottal pulsing duration. These phonetic parameters had been examined in other studies investigating word-final devoicing, but vowel duration turned out to be the most important parameter for Polish. Slowiaczek and

Dinnsen found that vowels preceding voiceless consonants were on average 10% shorter than those preceding voiced consonants.[49]

Slowiaczek and Szymanska (1989) tested whether the findings by Slowiazek and Dinnsen (1985) might mean that Polish listeners make use of vowel duration as a perceptual cue in distinguishing items that differ in their underlying final voicing. In their perception study, vowel length preceding voiced and voiceless consonants differed by 55%. Despite this comparably large difference (in comparison to the 10% difference found in the production study), subjects did not attend to the difference in vowel durations and did not perform significantly above chance in identifying underlying voiced items. Their result supports the general view that Polish does not make use of vowel duration as a cue to vowel or following consonant identity.

On the other hand, Polish does make use of *consonantal* duration to contrast singleton and geminate consonants, as indicated by double consonant letters in the orthographic forms (e.g., <leki> [lɛki] "medicines" versus <lekki> [lɛkːi] "light (ADJ)"). Geminate consonants are produced significantly longer than singleton consonants. For the nasal consonants, for example, Rojczyk and Porzuczek (2014) found a geminate/singleton ratio of 2.9. Polish geminates are sometimes subject to double-articulation, meaning they are rearticulated phonetically (Thurgood, 2001). Even though this is in conflict with Ladefoged and Maddieson's (1996) definition of geminates[50], researchers still describe Polish as a language which exhibits true geminate consonants (Thurgood, 2001, Thurgood and Demenko, 2003, Tworek, 2012, Rojczyk and Porzuczek, 2014). The reasons for occasional double-articulations are not clear. Tworek (2012: 139) assumes that they are idiolectal in nature.

Even though Polish contrasts consonantal length, transfer to vowel perception and production seems to be very unlikely. Flege (1995: 267) discusses findings of

---

[49] The absolute values of their vowel durations are unfortunately not very informative as their segmentation criteria implied that for some words the sonorant preceding or following a vowel was included in the measurement of vowel duration.

[50] According to Ladefoged and Maddieson (1996: 92), true geminates of any kind may not be separated by an epenthetic vowel or other interruption. Fake geminates are sequences of the same consonant as in, for example, the English word *book-case*.

Flege and Port (1981) in the light of "free feature recombination" for the voiceless feature in stop consonants by Arabic speakers of L2 English. They found that it was not possible for the L2 learners to transfer the voiceless feature of their /t/ and /k/ to /p/. Since it does not seem to be possible to transfer features within one natural class, it is highly unlikely that it is possible to transfer a feature used for consonants to vowels. Pajak and Levy (2014)'s study might challenge this view as they showed that Vietnamese and Cantonese speakers could perceive Polish consonantal length contrasts better than Mandarin speakers, even though they had native experience with vowel length contrasts only. However, Altmann et al.'s (2012) results with German speakers cannot be aligned with Pajak and Levy's proposition that speakers are able to abstract from vowels to consonants. They found that German subjects without experience with Italian were clearly worse at perceiving non-native consonantal length contrasts than German L2 Italian learners, even though they all have experience with *vowel* length contrasts in their native language. Only with L2 experience did German speakers improve their perception of consonantal length contrasts. If feature abstraction was the underlying principle in the perception of L2 length contrasts, monolingual German participants should have performed comparably well to the German L2 Italian learners.

What may be of interest regarding consonantal length is the length of vowels preceding singleton versus geminate consonants. It has been shown for Italian, for example, that the duration of the vowel is shortened preceding a geminate consonant and that this serves as secondary cue in singleton versus geminate perception (Pickett, 1999). This timing compensation, however, is not a language universal feature (Port et al., 1980). In Japanese, the effect seems to be reversed and vowels appear to be slightly longer preceding geminates (Han, 1994). Smith (1995) explains this with reference to the different prosodic properties Japanese and Italian exhibit, namely mora- and syllable-timing respectively. In her Articulatory Phonology account, she could interpret the data by proposing that gestures are coordinated differently in time in the two languages.

Hardly any data exist on Polish vowel duration in relation to consonantal length. In a pilot study, Thurgood (2002) investigated vowel duration *following* geminates in Polish minimal pairs both in perception and production.[51] Most vowels seemed to be longer following geminates (on average by a multiple of 1.4). Yet, in almost a quarter of all items measured, vowels were longer in the singleton items. As far as perception goes, her data revealed that vowel length did not help participants perceive the phonetically long fricatives as geminates, which suggests that vowel length following geminate consonants is not a reliable secondary cue for Polish native speakers.

A recent study, Rojczyk and Porzuczek (2014) investigated the acoustic properties of Polish singleton and geminate nasal consonants as well as their vocalic surroundings.[52] 26 Polish native speakers were asked to produce the words *pana* [pana] ("gentleman") and *panna* [pan:a] ("maiden") in the same carrier phrase. They found a significant difference between singleton and geminate duration in that nasal singletons were on average 58 ms long, while nasal geminates measured on average 167 ms (i.e. they were 2.9 times longer). The duration of the following vowels did not differ significantly in this study, while the researcher did find a small but significant difference for vowels preceding the consonants. Vowels preceding singletons were on average 73 ms long, while vowels preceding geminates were 12 ms longer. The ratio between the first [a] in *panna* versus the first [a] in *pana* is therefore only 1.16, which is even less than what Thurgood (2002) found for vowels *following* singleton/geminate consonants. It therefore seems unlikely that Polish listeners make use of preceding vowel duration as a secondary cue to singleton/geminate perception. Additionally, in an exploratory study of the Polish vowels, acoustic measurements of vowel durations before singleton and geminate consonants did not show any significant differences (see below).

[51] It is unclear why Thurgood chose to measure the duration of the following and not the preceding vowels (or both). Measuring the preceding vowel is generally more common in geminate studies. This is because geminates contribute weight to the preceding syllable, which means that, in a language like, for example, Malayam, vowels preceding geminates have to be shortened in order to adhere to the language-specific moraic structure (Broselow et al., 1997).

[52] They also measured the duration of fake geminates across word boundaries, but these data are not of immediate interest to this study.

4.2.3 Syllable structure

Like German, Polish allows for complex syllable structures. A one-syllable word may consist of one vowel, as for example *a* ("but"), or as many as eight segments, with a vowel constituting the syllable nucleus (e.g. *skąpstw* [skɔmpstf] "avarice-GEN.PL", rare example, taken from Bethin, 1992: 22). Polish is especially known for its long onset clusters and may show clusters as complex as five consonants within a phonological word, e.g. [ˈspstrɔŋ] "with a trout" in *z pstrągiem* (Jassem, 2003: 103). Lexeme-initially, four consonants are not unusual, for example *zdźbło* [ʑd͡ʑbwɔ] "stalk" or *wzgląd* [vzglɔnt] "respect". These examples show that Polish allows complex sequences of obstruents, which do not seem to follow language universal tendencies captured in the Sonority Sequencing Principle (for in-depth phonological analyses of this issue see Bethin, 1992 or Gussmann, 2007). The fact that both Polish and German allow for relatively complex syllable structures makes the two languages a suitable pair for studying L2 vowel acquisition. This is because the syllable structure of the target language is unlikely to pose additional problems for Polish L2 German learners, which might otherwise be the case (Ternes, 1978).

As far as accent placement is concerned, Polish differs from German in that its predominant stress pattern is penultimate stress. In contrast to German, it is irrelevant whether the penultimate syllable constitutes the stem or not, for example *pański* ("gentlemanly", where *pan* means "man") but *panowie* ("men"). Exceptions are posed by verbs with conditional endings, e.g. *czytać* ("to read") but *czytałbym* ("read-1P.COND"). Other exceptions can be found in borrowed words or clitics (for details see Biedrzycki, 1974: 126-134). In Polish, stress is mainly indicated by a change in fundamental frequency (Tworek, 2012: 212), even though Igras and Ziółko (2013) found that stressed Polish vowels also showed higher intensity and are slightly longer (by about 5%). Since unstressed Polish vowels are hardly reduced, Grzeszczakowska-Pawlikowska (2007: 8) comes to the conclusion that Polish should be classified as syllable- rather than stress-timed. Other researchers classify Polish as a rhythmically "mixed" language due to the combination of a very complex syllable structure with a lack of vowel reduction (Gut, 2003). Perceptual studies seem to support this view, at least as far as it is difficult to clearly classify Polish within one of the main rhythmic classes (Ramus et al., 2003). Because of the differences in vowel reduction in relation to stress placements, empirical studies have found that Polish learners have problems reducing German vowels native-like (Gut, 2003, Grzeszczakowska-Pawlikowska, 2007, Richter, 2008). In this study, the investigated vowels are all stressed, which is why the lack of reduction in unstressed syllables is not problematic for the current study.

Like German, Polish devoices final obstruents. However, the domain of this process is not the syllable but the word (Gussmann, 2007: 289). In the test items of this study, all syllable-final voiced obstruents are also word-final, e.g. *Weg* [ve:k] ("way").

4.2.4 An exploratory study of Polish vowels

Very few studies have investigated vowel duration before singleton and geminate consonants in Polish and/or provide general vowel length and quality data which are comparable to those collected in the main study of this dissertation. For this reason, a pilot production experiment was conducted in at a high school in Lublin, Poland, with young Polish adults. With the help of a native speaker [53], the productions of the six Polish vowels in different consonantal contexts were recorded in a quiet classroom. Furthermore, productions of minimal pairs containing singleton and geminate consonants were collected. Each word was produced twice by each speaker, while the singleton/geminate minimal pairs were produced three times. In all, the productions of 21 speakers were recorded, of which seven male recordings were chosen for further analysis.[54] The average age of the analysed speakers was 17.6 years (SD=0.5).

[53] It was decided to have a native speaker interview and instruct the participants to ensure that the speakers would be in the desired language mode (Grosjean, 2001). The interview consisted of questions about the speakers' educational and language learning background.

[54] Since this part of the study was only exploratory, not all recordings were segmented. Originally it was planned to use the Munich Automatic Segmentation System (MAUS), which allows for automatic alignment of orthographic input with the acoustic signal (Schiel et al., 2011). Unfortunately, it turned out that this technique could not segment most vowels reliably, which is why extensive manual corrections were still necessary.

The participants were instructed to produce words which were presented to them on a computer screen in orthographic form. In the first run of the experiment, 34 words were presented in random order, while in the second part the same words were presented in a different random order. Of these 34 words, ten words formed the following five Polish minimal pairs:

(3) a. *buda* ("doghouse") – *Budda* ("Buddah")
b. *Grecy* ("Greeks") – *greccy* ("Greek")
c. *leki* ("medicines") – *lekki* ("light (ADJ)")
d. *pana* ("gentleman") – *panna* ("maiden")
e. *uczę* ("teach-1SG.PRS") – *uczczę* ("celebrate-1SG.FUT")

The remaining 24 words were two-syllable words which contained the six Polish vowels /i, ɨ, ɛ, a, ɔ, u/ in three consonantal contexts (bilabial, alveolar, and velar), in order to balance the influence of the consonantal environment on the vowels of interest (Hillenbrand et al., 2001). For example for the vowel /a/, the three words *papa* ("tar paper"), *tata* ("daddy"), *skakać* ("to jump") were recorded twice from each speaker. Of interest was always the vowel in the first syllable, which was stressed in all test words. Furthermore, 2x3 additional words with palatal consonantal contexts were collected for the vowels /a/ and /ɛ/, as Jassem (2003: 106) asserted that there is little contextual allophony except for these vowels in palatal contexts. In the final run of the experiment, the participants produced the minimal pair words again in the carrier phrase *Słowo ... jest na liście* ("The word ... is in the list") and were asked to speak faster than they had in the first two runs, in order to check whether this may have an effect on vowel and consonant realizations as a function of consonant type. A list of all test words of this pilot study can be found in Appendix II.

The productions of all participants were recorded with a Beyerdynamic Opus 54.16/3 headset and a Marantz PMD 660 solid state recorder at 44.1 kHz. After seven recordings were pre-segmented with the help of MAUS (see Footnote 54), manual corrections to the exact beginning and end of each vowel of interest were carried out in PRAAT (Boersma and Weenink, 2014). The same segmentation criteria applied as for the data of the main production experiment (for details see Appendix III).

4.2.4.1 Pilot study 1: Vowel duration before singletons/geminates

A considerable amount of the geminates in this data set was produced in the form of double articulations: as much as 82.7% (in all N=105) were rearticulated. The distribution of single versus double articulations in the pilot study depended on the run of the experiment, in that double articulations became less common in the third run, in which participants were asked to produce the words faster and in the context of a sentence. Figure 4 shows the type of geminate realisation as a function of run.

Figure 4: Distribution of single and double articulations of Polish geminates in the pilot study

A generalized linear mixed model was fit to the binomial realisation data (single and double) in R (R Core Team, 2014) with RUN (*first* versus *second* versus *third*) as fixed factor and participants and words as random factors.[55] The model revealed that, in the third run, geminates were produced significantly more as single articulations than in the first run (z=-2.23, p=0.03), while the second run did not differ significantly from the first (z=-0.71, p=0.48). It therefore seems to be the case that the type of realisation of Polish geminates depends on tempo and context rather than idiolectal differences, as was suggested by Tworek (2012). The fact that a substantial number of geminates were rearticulated might challenge the general view that Polish exhibits true geminates. Despite this concern, vowel duration was still measured as a function

[55] *Model.realisation = glmer(Realisation ~ Run + (1|Participant) + (1|Pair), data= data.realisation, family="binomial")*

of the following consonant type, as it was important to make sure that Polish does not pattern similar to Italian.

In the following duration measurements, single and double articulations are not analysed separately, as Thurgood and Demenko (2003) did not report significant length differences between singly and doubly articulated geminate consonants. As expected, geminate consonants are considerably longer than singleton consonants, in these data by about 92%. Geminate consonants were on average 235.6 ms (SD=50.2 ms) long, while singletons measured on average 123.0 ms (SD=47.5 ms). Figure 5 shows the duration of singleton versus geminate consonants as a function of consonant type and run.

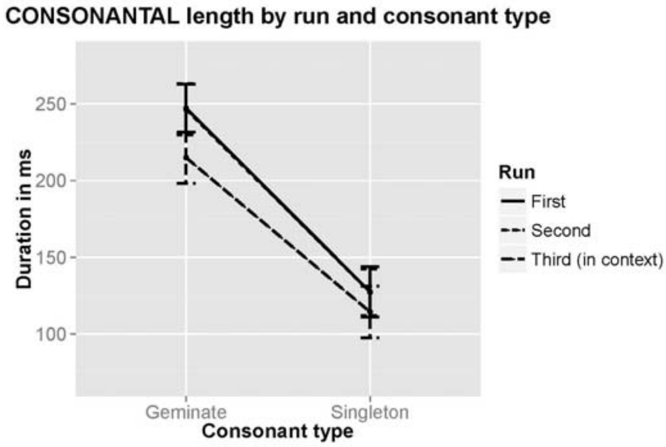


Figure 5: Duration of singleton versus geminate consonants in Polish (error bars show 2 SE)

As can be clearly seen, durations do not differ in the first and second run, which varied only in presentation order.[56] Consonant durations in the third run seem to be shorter, which would be expected considering participants were asked to produce the test words in a carrier phrase and speak faster than they did before. A linear mixed model with CONSONANT TYPE (*singleton* versus *geminate*) and CONTEXT (*context-yes* versus *context-no*) as fixed factors and participants and word-pairs as random factors[57] revealed that the difference between singletons

---

[56] The lines of the first and second run completely overlap. For this reason, they were analysed as one factor level ("context-no") versus the third run ("context-yes").

[57] *Model.consonats = lmer(Consonant_ duration ~ Context + Consonant_ type + (1|Participant) + (1|Pair), data=data.duration, REML=FALSE)*

and geminates was highly significant (t=36.28, p<0.001), as was the effect of CONTEXT on consonant duration (t=-6.84, p<0.001).

Of particular interest in this pilot study was the duration of vowels before geminate and singleton consonants. As can be seen in Figure 6, the effect consonant type on vowel duration was minimal.

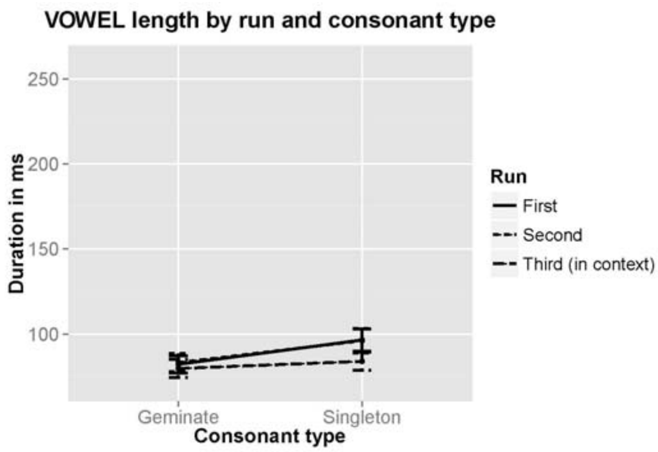


Figure 6: Duration of vowels before singleton and geminate consonants (error bars show 2 SE)

On average, vowels before geminate consonants were 81.7 ms (SD=15.5 ms) long, 92.1ms (SD=19.5ms) before singleton consonants. This means that, in this data set, vowels were on average 1.1 times longer before singletons than before geminates. To investigate whether this small difference is significant, a linear mixed model was fit to the data which also included the factor REALISATION (*single* or *double articulation*) as a control variable. This was motivated by the pattern that vowel durations showed before double articulations (only possible in geminate consonants) and single articulations (both in geminate and singleton consonants), as shown in Figure 7.

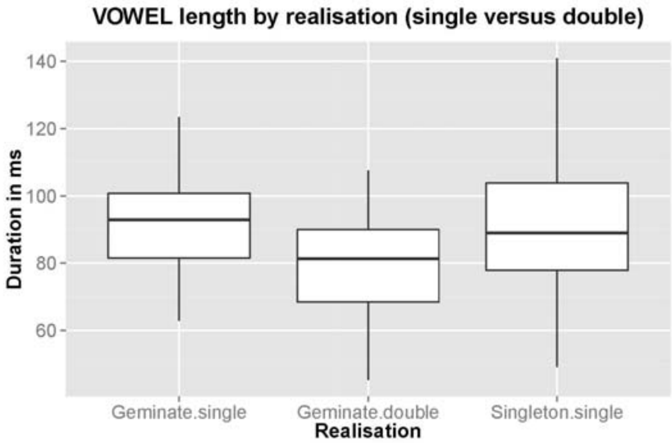


Figure 7: Vowel duration as a function of realisation (the upper and lower hinges correspond to the first and third quartile, while the whiskers extend to the highest and lowest value within 1.5 times the distance between the first and third quartiles)

A linear mixed model was fit to the data to predict vowel duration by the fixed factors REALISATION (*single* versus *double*), CONSONANT TYPE (*singleton* versus *geminate*), and CONTEXT (*context-yes* versus *context-no*).[58]

The model revealed that the factor REALISATION approached significance only (t=-1.97, p=0.051). Importantly, vowels were not significantly longer before singleton consonants (t=1.03, p=0.30). As expected, vowels were significantly shorter in the third run of the experiment (t=-3.52, p<0.001).

Contrary to what Rojczek and Porzuczek (2014) found, vowels did not differ significantly as a function of consonant type in this study. While their vowel duration ratio before geminate/singleton consonants was also only 1.16 (though in the opposite direction in that vowels in his data set were shorter preceding singletons), it may be the case that this small difference reached significance because they did not take into account that vowel duration may be influenced by the specific articulation (single versus double) of the geminate. Rojczek and Porzuczek tested group difference by means of an ANOVA, for which the inclusion of control factors is not as common as it is for linear mixed models.

[58] *Model.vowels = lmer(Vowel_duration ~ Realisation + Consonant_type + Context + (1|Participant) + (1|Pair), data=data.duration, REML=FALSE)*

In any case, both Rojczek and Porzuczek's data and the current study taken together suggest that vowel duration in Polish is hardly influenced by consonantal duration. It is therefore concluded that Polish, in contrast to German, is a language in which differences in vowel length do not play a role – be it on a phonological or phonetic level.

4.2.4.2 Pilot study 2: Polish vowel quality and vowel length

Since only Hentschel (1986) has thus far compared Polish and German acoustically, it was of interest to collect additional data, which are comparable to the acoustic data analysed in the main production part of the study (Chapter 6). For this purpose, the first (F1) and second (F2) formants of a small sample of Polish vowels (N=126 + 42 vowels in palatal context) were extracted with a PRAAT script, which was found to be most reliable for measuring even the high back vowel /u/.[59] This was necessary because the weak F2 of high back vowels is often missed in automatic measurements, and, as a consequence, F3 instead of F2 is picked up by the algorithm (Remijsen, 2004). Even with this adjusted script, some of the respective formants were still missed and had to be corrected by hand or labelled as missing values where it was not possible to track a clear formant structure (N=8). Table 9 summarizes the mean frequencies of F1 and F2 of the six Polish oral vowels spoken by 7 native speakers. Figure 8 shows a plot of the mean of each category as well as each single data point by means of the R-package *phonR* (McCloy, 2015).

	/i/	/ɨ/	/ɛ/	/a/	/ɔ/	/u/
F1	283	392	548	637	504	347
F2	2134	1675	1559	1251	992	829

Table 9: Mean Polish formant frequencies for 7 male speakers (in Hz)

[59] Formant frequency measurements were done by means of an LPC analysis with 10 coefficients, 25 ms analysis window in 5 ms steps. Target frequencies were calculated as the mean formant frequencies between the 40% and the 60% point of vowel duration in the resulting PRAAT formant object.

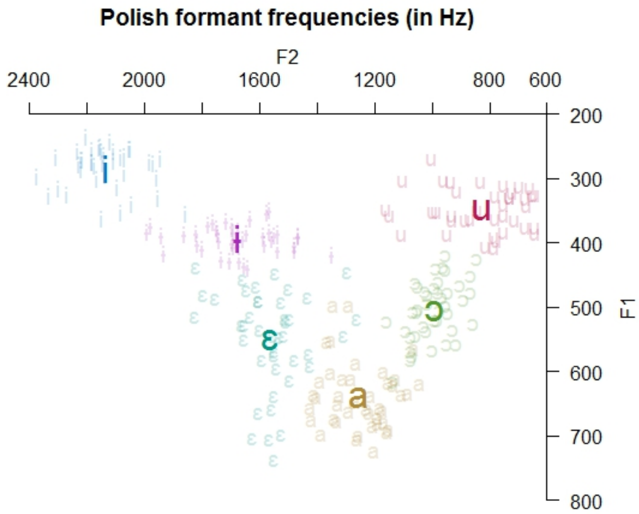


Figure 8: Formant frequencies for each token of the data set (large symbols represents grand mean; N(tokens)=118)

A visual comparison with Hentschel's (1986) data (Figure 3, page 77) reveals that the relative positions of the six means in the acoustic vowel space look very similar, although the exact formant values are not the same. While Hentschel did not specify this in his work, it might be the case that his speakers were female, as his F1 and F2 values are higher than those in the current data set (Simpson and Ericsdotter, 2007). This underlines the importance of collecting acoustic data which are comparable to the vowels which were collected in the main part of this study, i.e. same speaker-sex and same age.

Figure 9 represents the same vowels as above as well as [a] and [ɛ] in palatal contexts, for example in the test words *ziajać* [ʑajat͡ɕ] ("to pant") or *dzieciak* [d͡ʑet͡ɕak] ("child")[60]. As Jassem (2003) predicted, vowels in this context seem to be considerably fronted and, in the case of /ɛ/, raised as well.

[60] Recall from Section 3.4.2 that the palatal consonants /ʑ/, /ɕ/, /d͡ʑ/, and /t͡ɕ/ are represented in orthography as <ź, ś, dź, ć>, when they precede a vowel as <zi, si, dzi, ci>, where <i> functions as a sign for palatalization.

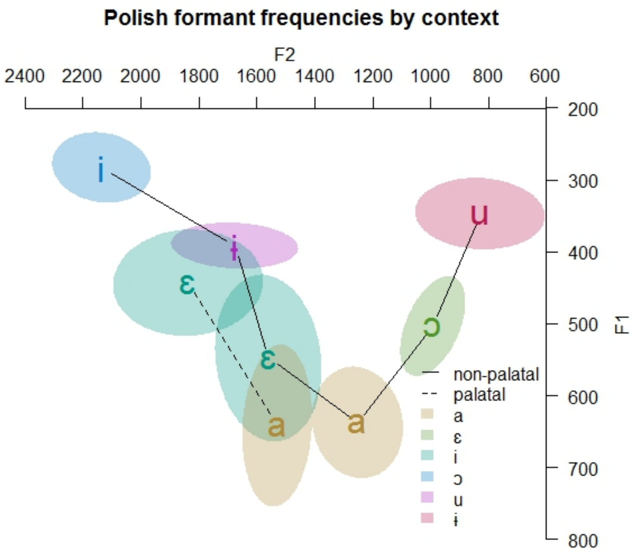


Figure 9: Mean formant frequencies of Polish vowels by context (palatal versus non-palatal; ellipses correspond to a confidence level of ±1 standard deviation from the bivariate mean)

As far vowel length is concerned, the average duration of all Polish vowels in this study measured 87.0 ms (SD=25.9). This finding corroborates measurements reported by Frąckowiak-Richter (1973) rather than Keating (1984). Since Rojczyk and Porzuczek (2014) also reported that vowels measured on average only 86 ms, it appears that Polish vowels are indeed "relatively short" and are more comparable in their length to German short vowels. Yet, when comparing the exact values to those of comparable German vowels (Section 4.1.2), it appears that Polish vowels are on average slightly longer than the German vowels. Antoniadis and Strube (1984) had measured German vowels in disyllabic words and these measured on average 72 ms. Comparing this value to the average Polish vowel length found in this study (or in Frąckowiak-Richter, 1973 or in Rojczyk and Porzuczek, 2014) it seems clear that Polish vowels in disyllabic words are (at least) about 14 ms longer than German vowels in a similar context.[61] This finding helps to formulate more informed hypotheses about vowel length productions which will be developed in Chapter 6.

[61] Similar observations can be made for vowels in monosyllabic words, even though only Frąckowiak-Richter (1973) and Nimz (2011a) are comparable in this case. For vowels in monosyllabic words, Frąckowiak-Richter measured an average duration of 131 ms, while in Nimz (2011a) vowels in monosyllabic words were considerably shorter and measured only 85 ms.

5 Experiment 1: Discrimination

Even though the discrimination task was administered as the last of the three main experiments (i.e. first: production, second: identification, third: discrimination), it will be described first. This is because its results will be relevant for the interpretation of the production and identification data. It was decided to conduct the discrimination task last, so as not to draw the participants' attention to the main focus of the study, namely vowel length and vowel quality. All participants[62] took part in all three experiments and were further asked to fill out a questionnaire on their language learning background (see Appendix IV). Additionally, each Polish participant took part in two post-tests, which were administered to evaluate the participants' orthographic knowledge.

All participants were scheduled to meet with the experimenter on two different days. On the first day, each participant took part in the production experiment (Chapter 6), followed by the identification experiment (Chapter 7). Each Polish participant was then asked to take the first post-test, which consisted of writing each word that had been part of the previous two experiments as well as indicating on a scale from 1 ("I don't know this word") to 7 ("I know this word well") how familiar the participant was with each word (see Appendix V). On average, this first session lasted about 1.5 hours for the Polish participants and 45 minutes for the German participants.

The second session was scheduled on another day and included the participation in the discrimination task as well as the second post-test, which consisted of questions regarding the participants' metalinguistic knowledge of the German vowels (only for Polish participants) and a short orthographic exercise in which the participants had to name and, if possible explain, the length of vowels in a small sample of German nonce words (see Appendix VI). The second session was shorter than the first and lasted about 30 minutes for both the German and

[62] Except four participants: Two German participants did not come back on the last day of the study, and two Polish participants' language learning backgrounds were too different from the rest of the group, so that they were only asked to participate in the discrimination task (though their data were never included in the analysis).

the Polish participants: 15 minutes for the discrimination experiment and 15 minutes for the post-test.

5.1 Hypotheses

The discrimination experiment was conducted to address two main research questions, which were motivated by research on vowel length perception laid out in detail in Section 2.5. Firstly, Bohn (1995) had formulated in the Desensitization Hypothesis that duration cues in vowel perception are always easy to access for L2 learners, while McAllister et al.'s (2002) Feature Hypothesis would predict that German vowel duration differences are difficult to perceive for Polish L2 German learners. This is because Polish does not use vowel length phonologically or phonetically, for example as an additional cue in geminate perception (Section 4.2.4.1). Secondly, McAllister et al. (2002) had found that mid vowels were harder to perceive (and to produce) by learners than non-mid vowels. The authors attributed this finding to the fact that mid vowels are not accompanied by vowel quality differences. This implies that quality differences might be easier to discern for L2 learners than pure length differences. Since the authors never tested this possibility, the discrimination task was designed in such a way that would allow for separate testing of the perception of vowel length and vowel quality.

While Bohn (1995) and McAllister et al. (2002) had investigated L2 English and L2 Swedish, the current study looks at the perception of German as an L2. Like Swedish and English, German is a Germanic language and vowel length is highly correlated with vowel tenseness, as laid out in detail in Section 4.1. Unlike in English, vowel length is a primary cue in vowel perception, specifically in the low vowels (Sendlmeier, 1981, Weiss, 1984). Unlike in Swedish, vowel length does not correlate with the length of the following consonant; hence, German is a more suitable language to study the perception of vowel length. The following hypotheses originate from the seemingly contradictory findings by Bohn (1995) and McAllister et al. (2002) about vowel length perception in L2 learners. Additionally, they test assumptions put forward by researchers in the field of GFL, who have predicted that German vowel length is problematic for Polish L2 German learners (Section 2.1).

Hypothesis 1: Polish GFL learners will be less accurate than German native speakers at perceiving pure length differences in German vowels.

Hypothesis 2: Polish GFL learners will be equally good as German native speakers at perceiving vowel quality differences in German vowels.

5.2 Participants

The participants of the study were recruited at a Polish high school in Warsaw, Poland (experimental group) and at a German high school in Dortmund, Germany (control group). Because of two unexpected German drop-outs, two more participants were recruited in the same area of Germany, who did not visit the same high school as the other German participants but matched in age and educational background.

5.2.1 Polish group

The Polish participants were all students at a Polish high school with special emphasis on German as a Foreign Language (GFL). At this school, students who choose the so-called bilingual branch will go through an extra preparation year before they begin their regular high schooling from 10th to 12th grade. During this preparation year, students receive 18 hours of GFL per week, which includes grammar, vocabulary training, German culture and media, and presentation skills.[63] They are taught both by German native speakers and highly-advanced Polish L2 German speakers. After the end of the preparation year, students take a language test (level B1 according to the Common European Framework of Reference for Language (CEF) (Council of Europe, 2001), which, if they pass, classifies them as “intermediate” speakers. This qualifies them to move on into their bilingual high schooling. During their three regular high school years, they receive an average of 10 hours of German per week, of which 6 hours are GFL lessons and the remaining are geography,

[63] In a preparative interview with the head of the school, it was mentioned that students also receive a few hours of phonetic instruction; however, in the questionnaires hardly any of the students mentioned phonetic training (2 out of 22). Most likely the phonetic instruction was not very extensive.

history, and cultural studies in German. At the end of their 12th year, the students either have the option to take the German high school diploma (*Abitur*), for which they need a CEF level of C2, or they can take the *Deutsche Sprachdiplom II* ("German Language Diploma II"), which certifies their C1 language level. The participants of the study were recruited from both the 11th and 12th grade of the German bilingual branch. This means that all of them had received at least two years of intensive GFL lessons and can be classified as medium-advanced speakers of German (B2/C1).

22 Polish students took part in the experiment, of which two were excluded from further analysis because their language learning background was not comparable to that of the other students. Subject P21 had lived in Germany from the age of 9 until 16, which makes her a *second* language speaker rather than a *foreign* language speaker, and subject P22 spoke Bulgarian as her native language.

All other Polish participants were late foreign language learners, that is, none of them had received any intensive (more than 6 hours per week) GFL teaching before they entered high school at around the age of 15. All of them spoke English as their first foreign language, some of them also spoke a third or fourth foreign language, but not of them more advanced then German or English. None of them spoke German (or any other language) at home or reported to speak a distinct Polish dialect. None of them reported any hearing or learning problems. The average age of all participants (excluding P21 and P22) was 18.5 years (SD=0.6); 4 of them were male. For more detailed information about the language learning background of the participants see Appendix IVa.

5.2.2 German group

Apart from subjects G22 and G23, all German participants were recruited at a German high school in Dortmund, Germany (Western Germany), while G22 and G23 were from the same region but had recently completed the same level of German high schooling ("Gymnasium"). All participants spoke Standard German, while G1 reported to have a Brazilian mother who speaks German with a foreign accent. The subject herself spoke German without a foreign accent, and her results did not differ significantly from the other German subjects, which is why she was still included in the study. Just like the Polish participants, all German subjects spoke English as their first foreign language. Most of them spoke French as their second foreign language, while some spoke other languages such as Spanish, Chinese, or Latvian as their second or third foreign language. Three of them reported different kinds of minor language problems such as (former) stammering or lisping, which were not considered problematic for the study. Furthermore, one male participant reported to be dyslexic, which is why he was excluded from the other two experiments that investigated the effects of orthographic marking. However, he was still included in the analysis of the discrimination task. One female participant (G22) reported after the discrimination task that she was not sure whether she understood the instructions correctly, which is why she was excluded from the current analysis. The average age of the remaining 20 German participants was 17.9 years (SD=1.1); 6 of them were male. More detailed information of the German participants' background can be found in Appendix IVb.

5.3 Experimental design

In their study on Swedish vowel length perception, McAllister et al. (2002) had used an identification task (for details see Section 2.5.1) and discussed how it would have been of interest to conduct a discrimination experiment as well. For this reason it was decided to conduct two types of perception task: an identification task similar to the one in McAllister et al. *and* a discrimination task. The latter was designed as a speeded same-different task with nonce words that contained the vowels of interest (see below). Technically, the inter-stimulus interval (ISI) was 0ms, though effectively the silent interval between two items was 150 ms, as every item was preceded by a bilabial stop which was manipulated in a way that the sound wave of the stop silence was set to an amplitude of 0. The actual ISI of 150 ms was chosen because there is evidence that an ISI of 0 may in fact decrease discrimination performance (Pisoni, 1973). In general, a relatively short ISI was chosen because the experiment was designed to tap into a phonetic rather than a phonemic level of processing in order to contrast with the identification task (Werker and Tees, 1984). A short ISI is thought to minimize memory load and enable L2 listeners to differentiate items which they might not be able to differentiate in a more demanding, real-world context (Strange and Shafer, 2008). For the same reason, nonce words were chosen as stimuli, so as to prevent influences from a high-order lexical processing level, which was to be tested later in the identification experiment (Chapter 7).

5.3.1 Stimuli manipulation

Because the main experiments were designed to be comparable in that they were investigating the same vowels, the vowel pairs /a:/-/a/, /e:/-/ɛ/, and /o:/-/ɔ/ served as experimental items in all three experiments. The restriction to these vowel pairs was due to orthographic factors which had to be met in the other two experiments rather than in the discrimination task. As described in Section 4.1, most German vowels differ both in vowel length and in vowel quality. To avoid the problems of McAllister et al.'s (2002) study, i.e. the confounding of vowel length and vowel quality, it was decided to manipulate the vowels in a way that would allow for the differentiation between the two dimensions. This was done by means of a design similar to that used by Sendlmeier (1981) with German native speakers. In this study, vowels were either shortened or lengthened to the average length of is corresponding counterpart (see Section 4.1.2 for the results of his study).

For the purpose of the current study, a female German native speaker produced the vowels of interest in the bilabial consonantal frame [b_p], as this context involves the least amount of tongue movement, so as little co-articulation as possible was expected. Since the main experiment was to be preceded by a practice phase, four additional nonce words in the same context including the vowels [ɪ], [i:], [ʊ], and [u:], were recorded as well. The speaker produced five versions of each nonce word in the context *Ich hab einen [...] gesehen* ("I have seen a [...]") in randomized order. The recordings were made in a sound-proof booth at a sampling rate of 48k Hz (16 bit) at the Centre for General Linguistics (ZAS) in Berlin. Of the five productions of each vowel, those productions were chosen for further analysis which were closest to the speaker's mean values of the first and second formant of the respective vowel. These most prototypical (at least as far as their *quality* goes) items were then used for further manipulation with the help of PRAAT (Boersma and Weenink, 2014).

In order to manipulate the items in their length dimension, the average duration in milliseconds was calculated for each vowel category. The average ratio of each vowel pair corresponded well to vowel duration data collected elsewhere, in that the long vowels were on average twice as long as their short counterparts (Antoniadis and Strube, 1984). In order to create appropriate items for the experimental conditions described below, each prototypical vowel was shortened or lengthened to the average length of its counterpart (short or long) in order to create the items for the *length* and *quality* condition. Where the quality-wise prototypical items did not automatically match their own prototypical length, they were also slightly shortened or lengthened to the group's average length for the *proto* condition. Lengthening and shortening was achieved through the *Dur* ("duration") function in the PRAAT manipulation settings, which allows for duration manipulation while maintaining original pitch and vowel quality. Hence, from the 10 prototypical vowels the same amount of manipulated items was created, e.g. from the long tense vowel [e:] the short tense vowel [e]. Because pitch contours differed slightly for each item, it was decided to take the average pitch of each vowel pair and normalize each token within each group (e.g. group "a", "e", and "o") to its respective average pitch. This was achieved by means of the *Pitch* function in the PRAAT manipulation settings. In this way, each item within a pair had the same F0 so that differences could not have been detected on the basis of different pitch height or contour.[64] Finally, items were matched for three experimental conditions.

5.3.2 Experimental conditions

Given the hypotheses formulated in 5.1, items were matched for the conditions *length* and *quality*, where the manipulated items either differed in length or in quality only (s. below). Furthermore, the condition *proto* was included. In this condition, items differed both in quality and length because they constituted

[64] Pitch values for the experimental pairs were 160 Hz for the a-pairs, 167 Hz for the e-pairs, and 173 Hz for the o-pairs.

prototypical items (hence “proto” condition). This condition was included in order to be able to compare the results of the manipulated items with those for unmanipulated items, as they appear in real communication situations.

- Condition **quality**: a prototypical long tense vowel matched with a lengthened lax vowel *or* a prototypical short lax vowel matched with a shortened tense vowel, for example [e:]-[ɛ:] or [ɛ]-[e] (note: condition is called “quality” because items only differ in quality)
- Condition **length**: a prototypical long tense vowel matched with a shortened tense vowel *or* a prototypical short lax vowel matched with a lengthened lax vowel, for example [e:]-[e] or [ɛ]-[ɛ:] (note: condition is called “length” because items only differ in length)
- Condition **proto**: a prototypical long, tense vowel was matched with its short, lax counterpart[65], for example [e:]-[ɛ]

5.3.3 Procedure

The experiment was run in PRAAT. Participants were instructed to answer as correctly and fast as possible whether the two sounds they just heard were the “same” or “different”. Instructions to both groups were given in German. Each participant went through a short practice trial of 8 different practice pairs consisting of nonce words that were similar in structure but did not contain the vowels that were of interest in the study. Participants did not receive any feedback after the completion of the practice trial.

Each vowel pair was judged 8 times in each condition for being “same” or “different”, while in the *quality* and *length* condition these 8 times were technically 2 x 4 times each of the possible item combinations (e.g. 4 times [e:]-[ɛ:] and 4 times [ɛ]-[e]). Hence there were 72 (8 x 3 x 3) potentially “different” pairs, plus the same amount of “same” filler pairs. These were created by repeating each of the prototypical long and short vowels 12 times (12 x 6 = 72 “same” trials). Furthermore, a control condition of 12 pairs that were clearly

[65] It is worth keeping in mind that from what has been established in Section 4.1, the German low central vowels are not expected to show a great difference in vowel quality. Results will corroborate this assumption.

different (i.e. 2 times [a:]-[ɔ], 2 times [a]-[e:], 2 times [ɛ]-[o:], 2 times [e:]-[a], 2 times [o:]-[ɛ], 2 times [e:]-[a]) was included in order to be able to check whether participants paid enough attention to the task. In all there were 156 pairs to be judged, which were presented by permuting the first and second item evenly and, overall, randomly for each subject. There were three breaks in the experiment and subjects were free to choose as long a break as they wanted to take. The experiment was run on an Acer Timeline Laptop, the stimuli were presented over high-quality Sennheiser headphones, and participants gave their answers via mouse click on marked squares on the screen indicating "same" or "different". On average, the experiment lasted about 15 min. The accuracy data were automatically collected by PRAAT.

5.4 Results

In all, 6240 data points were analysed (40 participants x 156 pairs). Of these 6240 data points, 480 judgements belonged to the control condition, i.e. the condition in which vowels were clearly different. Before the main statistical analysis, it was checked whether all participants had understood the task and paid enough attention by judging the clearly different control pairs correctly as different (e.g. [ba:p]-[bɔp]). It was found that two participants had judged more than one control pair incorrectly (German G16 had made two wrong judgements; Polish P15 had made three wrong judgements). Since this was still considered relatively low (at least 75% correct of all control pairs) and because these two participants belonged to both the German and Polish group, none of the participants was excluded due to their performance on the clearly different control pairs.

Another way of determining whether participants paid enough attention to the task was to investigate their judgement of the filler pairs which had to be identified as "same". Since these pairs were created by repeating the same token of each word (e.g. [ba:p] paired with the identical [ba:p]) it was not expected that participants would judge these pairs incorrectly. Descriptive investigations of the "same" pairs (2880 data points) revealed that each participant was more than 90% accurate in judging these filler pairs. Furthermore, inferential statistics were run in R (R Core Team, 2014). A generalized linear mixed model

(GLMM) was fit to the binomial data with LANGUAGE as fixed factor and participants as random factor[66], which revealed that there was no significant difference between the groups in judging the same-pairs correctly (z=-0.51, p=0.61). For this reason, none of the participants were excluded due to their responses to the filler pairs either.

Because only very few same-pairs were judged incorrectly, it was decided to use the raw accuracy data for further statistical analysis instead of d' scores[67]. Even though this measure is sometimes used in the analysis of perception data (e.g. Altmann et al., 2012, Showalter and Hayes-Harb, 2015), the raw accuracy data were judged to be a suitable unit, as they are widely used as well (e.g., Escudero et al., 2009, Weber et al., 2011).

Of interest to the study were the "different" stimuli, as they formed the three experimental conditions *length*, *proto*, and *quality*. In all, there were 2880 data points for three different vowel groups. As is evident in Figure 10, results for the a-pairs clearly differed from those for the mid-vowel pairs (e- and o-pairs). For this reason, the low vowel pairs were analysed separately from the mid-vowel pairs (to be statistically justified later).

---

[66] *Model.same = glmer(Correct ~ Language + (1|ID), data=data.same, family="binomial")*

[67] This measure reflects a participant's sensitivity in detecting differences in the stimuli, as it is based on hits (correctly identifying different-pairs) and false alarms (incorrectly identifying same-pairs) (Macmillan and Creelman, 2005). In this data set, almost half of the participants (48%) did not make *any* false alarms.

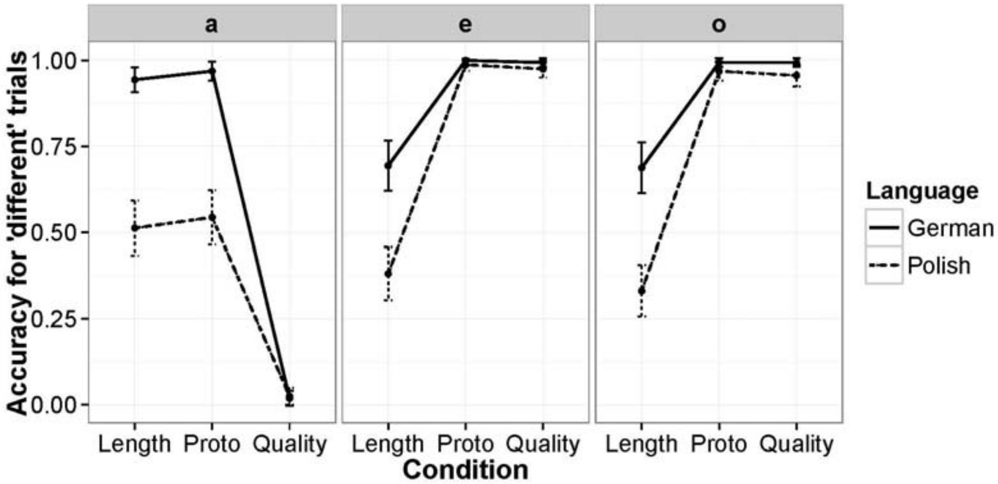


Figure 10: Correct responses for the three vowel groups by condition and language group (N=20 per language group; error bars show 2 SE)

Two separate generalized linear mixed models (GLMM) were fit for the two groups "a" and "e+o" (*Model.a* and *Model.eo*). For the a-group, a GLMM was fit to the accuracy data (dependent variable) which contained LANGUAGE (*German* versus *Polish*) and CONDITION (*length* versus *proto* versus *quality*) as fixed factors (independent variable) and random intercepts for participants. To control for an effect of presentation order, a likelihood ratio test examined whether the inclusion of this additional fixed effect improved model fit (Cunnings, 2012, Winter, 2013). It indicated that presentation order did not influence the accuracy of participants ($\chi2(1)=0.71$, $p=0.40$). Since it was of interest to compare Polish and German speakers within each experimental condition, custom comparisons were conducted by the final model[68], which revealed that German native speakers were significantly better than the Polish GFL learners in detecting pure length differences between the a-group stimuli ($p<0.001$). Table 10 summarizes the detailed statistics of the model.

---

[68] *Model.a = glmer(Correct ~ Combined + (1|ID), data=data_a, family="binomial")* (COMBINED is the combination of the fixed factors LANGUAGE (2 levels) and CONDITION (3 levels) into a new factor with 6 levels, e.g, level *a* (Polish-quality), level *b* (Polish-length), etc. This re-coding was necessary in order to conduct custom comparisons within the groups of interest, for example Polish-quality versus Polish-length.)

	Estimate	Std. error	z-value	p-value (>|z|)	Significance
(Intercept)	-0.24	0.35	-0.673	0.50	n.s.
P-G length	**3.97**	**0.82**	**4.84**	**<0.001**	***
P-G quality	-0.04	1.05	-0.03	0.97	n.s.
P-G proto	**4.43**	**0.88**	**5.02**	**<0.001**	***
P: length-proto	-0.23	0.30	-0.76	0.45	n.s.

Table 10: Summary of fixed effects of *Model.a* (significant comparisons in bold; N(observations)=960; N(participants)=40)

While German native speakers were accurate in their discrimination performance 94.4% ± 1.8% (SE) of the time, the Polish participants performed at about chance with 51.3% ± 4.0% (SE). The two groups were equally unsuccessful at detecting quality differences in the a-group (1.9% ± 1.1%(SE) and 2.5% ± 1.2%(SE) for German and Polish speakers, respectively, p=0.97). This clearly reflects that the German low-vowel pair is differentiated by length only, which has been suggested by other researchers too (see Section 4.1). Furthermore, the German group could differentiate prototypical (unmanipulated) pairs significantly better than the Polish group (p<0.001): In 96.9% ± 1.4% (SE) of the cases they judged the prototypical vowel pairs as correct, while the learners again only slightly performed above chance (54.4% ± 4.0% (SE)).

While the comparisons of Polish and German participants in the *length* and in the *quality* condition directly address Hypotheses 1 and 2 above, the comparison within the *proto* condition serves as a reference point for the discrimination of unmanipulated vowels as they appear in real communication situations with native speakers. As reported above, Polish learners were slightly more correct in the *proto* condition (54.4%) than in the *length* condition (51.3%). Another pairwise comparison within in the final model revealed that this difference was not significant (p=0.45).

As described in the section on stimuli manipulation, items in the *length* and *quality* condition had been manipulated both by lengthening the short vowels and by shortening the long vowels. To investigate whether the type of manipulation influenced the results, another GLMM was fit to a subset of the previous data which only included the manipulated items. The basic model structure was identical to *Model.a* above. Likelihood ratio tests were again used to examine whether the inclusion of the factor MANIPULATION TYPE improved the model. Neither as a main effect (χ2(1)=0.42, p=0.52) nor in interaction with LANGUAGE (χ2(1)=0.42, p=0.52) or CONDITION (χ2(1)=1.02, p=0.31) did MANIPULATION TYPE show significant improvement of the model. It was therefore justified to not include this factor in the final model.

Like the model for the a-pairs, the model for the mid-vowel pairs included the fixed factors LANGUAGE (*German* versus *Polish*) and CONDITION (*length* versus *proto* versus *quality*) as well as random intercepts for participants.[69] As in the other group, the control factor PRESENTATION ORDER was added to the model and, by means of likelihood ratio tests, it was found that this factor did again not improve model fit ($\chi^2(1)=0.11$, p=0.74). Hence it was not included. The same was the case for manipulation type. It was further checked via the same procedure whether the factor VOWEL (*o* versus *e*) should be added to the model. As was already apparent in Figure 10, accuracy was not influenced by whether the stimuli were in the "e" or in the "o" group, neither when VOWEL was added as a main effect ($\chi^2(1)=1.97$, p=0.16) nor in interaction with the two main experimental factors ($\chi^2(5)=3.12$, p=0.68). It was therefore justified to combine the two vowel groups into one model.

The final model revealed that German native speakers were significantly better than the Polish GFL learners at detecting pure length differences between the mid vowel stimuli (p<0.001). Table 11 summarizes the detailed statistics of the model.

---

[69] *Model.eo = glmer(Correct ~ Combined + (1|ID), data=data.eo, family="binomial")* (COMBINED is the combination of the fixed factors LANGUAGE (2 levels) and CONDITION (3 levels) into a new factor with 6 levels, e.g, level *a* (Polish-quality), level *b* (Polish-length), etc. This re-coding was necessary in order to conduct custom comparisons within the groups of interest, for example Polish-quality versus Polish-length.)

	Estimate	Std. error	z-value	p-value	Significance
(Intercept)	3.50	0.31	11.38	<0.001	***
P-G length	**1.79**	**0.44**	**4.11**	**<0.001**	***
P-G quality	1.17	0.90	1.31	0.19	n.s.
P-G proto	1.35	1.16	1.16	0.25	n.s.
P: quality-proto	**-5.69**	**0.51**	**-11.10**	**<0.001**	***

Table 11: Summary of fixed effects of *Model.eo*, significant comparisons in bold (N(observations)=1920; N(participants)=40)

Still, discrimination in the *length* condition by German native speakers was not perfect either. While the Polish learners detected a difference in only 35.6% ± 2.7% (SE) of the cases, German natives were correct only 69.1% ± 2.6% of the time. The groups did not differ significantly in the other two experimental conditions. Both Polish and German participants performed at ceiling when discriminating *quality* and *proto* pairs. In the quality pairs, native speakers were correct 99% ± =0.4% of the time and learners a little less with 97% ± 1% of the time (p=0.19). In the proto-pairs, native speakers were 99.7% ± 0.3% correct and learners 97.8% ± 0.8% (p=0.25). This means that Polish speakers performed significantly better in the *proto* condition than in the *length* condition, where they had only been correct 35.6% of the time (p<0.001). For the a-pairs, this had not been the case. As will be discussed in what follows, this finding can be related well to the overall difference between the German low and mid vowels.

5.5 Discussion

The experiment set out to answer two main hypotheses which were concerned with vowel length and vowel quality perception in Polish GFL learners in comparison to a native speaker control group. For this reason, stimuli had to be manipulated in order to differentiate between the dimensions of interest. The hypotheses were motivated by both research in experimental phonetics and predictions made by GFL researchers. Hypothesis 1 stated that Polish GFL learners will be less accurate than native speakers at perceiving length differences in German vowels. This hypothesis could be confirmed in all three vowel groups. Polish speakers were significantly less accurate in discriminating manipulated vowels which differed solely in vowel length. Interestingly, German native speakers did not discriminate mid vowels in the length condition perfectly either: while for the low vowel pairs they were correct 94% of the time, for the mid-vowel pairs they were correct only 69% of the time. This corroborates findings by Sendlmeier (1981), who had investigated German natives' perception of manipulated minimal pairs (Section 4.1.2). He too had found that vowel length was more important for the perception of low vowels than for mid or high vowels. Since, in the mid vowels, Polish learners could hear quality differences just as well as German native speakers, their performance in differentiating prototypical, unmanipulated pairs was also native-like (at ceiling). Apparently, for the e- and o-pairs, quality differences are sufficient in differentiating these vowels; hence, length only plays a secondary role for both native speakers and L2 learners in perceiving differences between these vowels. Still, Polish GFL learners are significantly less accurate at perceiving pure length differences than native speakers, which is crucial for the perception of differences in the low vowel pairs.

The low vowels /a/ and /a:/ mainly differ in length, which is why many researchers decide to represent the two vowels with the same symbol (but see Dudenredaktion, 2005 or Morciniec, 1990). For this reason, both German natives and Polish learners do not perceive differences in the quality condition of the a-pairs (accuracy of only 1.9% and 2.5%, respectively), where vowels are only *supposed* to differ in quality, as length differences are neutralized. Before the experiment, it was not clear that the participants would be unable to perceive the slight quality differences between short and long /a/, as the items in the experiment were not identical in their F1 and F2 values. The stimuli based on the speaker's most prototypical /a/ measured about 830 Hz for F1 and 1350 for F2 (at vowel midpoint), while /a:/ measured about 910 for F1 and 1390 Hz for F2. With that, F2 values differed by only 40 Hz between the items, while F1 showed a considerable difference of 80 Hz between short and long /a/. This difference is twice as large as what had been used by Escudero (2009) as just-noticeable threshold in her perception grammar. Still, in almost all cases, German and Polish participants considered the length-manipulated vowels to be the same. Consequently, Hypothesis 2 can be corroborated in that Polish learners did indeed perceive quality differences in German vowels comparably to native speakers, although it has to be kept in mind that even native speakers do not attend to quality differences in the *low* vowels.

Since the quality differences in the a-pairs do not seem to play a role in perception, length is the primary cue in these pairs. This is problematic for Polish GFL learners. While in the mid-vowel pairs their "length-blindness" can be compensated for by the quality differences in the e- and o-pairs, this is not the case for the a-pairs. In the *length* condition they perform at chance and, crucially, the same is true for the *proto* condition. Hence, in real communication situations, when the difference between an /a/ and an /a:/ is not discerned, a *Rate* ("instalment") may sound like a *Ratte* ("rat"). On the other hand, the discrimination results for the mid-vowel pairs would predict that Polish GFL learners would not make the mistake of perceiving a word like *Polen* ("Poland") as *Pollen* ("pollen"). This assumption will be tested in the second perception experiment, in which the perception of *real* words will be investigated (see Chapter 7).

Regarding the perception of the prototypical pairs only, the experiment further replicated findings by Hentschel (1986). He too had found that /e:/-/ɛ/ and /o:/-/ɔ/ were discriminated perfectly by Polish speakers, while /a:/-/a/ were only discriminated by naïve Polish participants 59% of the time (see Section 4.2.1). Using Best and Tyler's (2007) terminology, the low vowel pair is difficult to discriminate because it is a clear case of *Single Category assimilation* for Polish speakers. In Flege's (1995) terms, /a:/ and /a/ undergo *equivalence classification* because they are similar sounds. For Escudero (2005), /a:/ is a *new* sound for Polish natives, as it is produced with an additional auditory dimension (vowel length), which has not been previously incorporated into the learners' L1 perception. Because of this, /a:/ and /a/ are subject to what she terms *phonemic equation*. While Escudero (2005) and Escudero et al. (2009) assume that learners can form new categories along non-previously categorized dimensions, this assumption is not corroborated by the current findings. If, as stated by the L2LP, distributional learning leads L2 learners to incorporate a new dimension into their perceptual grammar easily, medium-advanced Polish learners should be using length difference to differentiate the German vowels at hand. This however was not the case. It could be argued that this is due to the fact that the learners of the study are *foreign* language learners, who are not exposed to as much native-speaker input as *second* language learners may be. While this argument might work for the L2LP, the findings still contrast with Bohn's (1995) Desensitization Hypothesis. According to him, duration cues are *always* easy to access for L2 learners.

McAllister et al.'s (2002) Feature Hypothesis seems to explain best why Polish GFL learners are less accurate than German native speakers in perceiving length differences. Since the learners do not use vowel length in their L1, neither phonologically nor phonetically, they have difficulties perceiving (and producing) a phonological contrast based on this L2 feature. Furthermore, McAllister et al. assumed that quality differences in vowels could help L2 learners perceive differences between long and short vowels in Swedish non-mid vowels. The same seems to be true for German, where *mid* vowels are accompanied by quality differences and differentiated well in the *proto* condition.

The current results further help interpret findings with Turkish GFL leaners by Nimz (2011b, Nimz, 2015). In a similar perception experiment, Turkish learners, who were comparable to the Polish learners in L2 proficiency, were able to hear length differences well. While in Nimz (2011b, 2015) it was unclear whether this was due to the L1 background of the Turkish learners (experience with vowel length on a phonetic level) or language universal perception abilities (Bohn, 1995), the finding with Polish GFL learners help resolve this uncertainty.

The results further draw attention to the fact that generalizations over the German long-short vowel contrasts are not possible. While Polish learners of L2 German seem to have severe problems differentiating the German low vowels, the differences between German mid vowels can be perceived well. This may have consequences for L2 lexical representations which involve these vowels. Furthermore, it is of interest to investigate the production of these same vowels. This will be addressed in the following chapter.

6 Experiment 2: Production

The production experiment consisted of a picture-naming task similar to the one used in Tsukada et al. (2005). However, in contrast to their study with Korean speakers, it was decided not to use an auditory prompt in the first run of the experiment because, as the authors themselves discussed, the possibility that the presence of a native speaker model augmented the accuracy with which participants produced the items could not be ruled out. Furthermore, it was considered crucial not to present the participants with the written forms of the test items, as it has been shown that the presence of this cue can significantly influence results (Bassetti and Atkinson, 2015). Two productions of each experimental item were recorded from each participant, which yielded 3648 data points (38 subjects x 48 words x 2 runs). Of interest were the stressed vowels in each target word, which were later, after their segmentation in PRAAT, analysed for the main dimensions of interest: vowel length and vowel quality. As has been laid out in Section 4.1, both vowel length and vowel quality play into the differentiation of most contrastive German vowel pairs. Hypotheses and the respective acoustic and statistical analyses therefore need to address both dimensions.

6.1 Hypotheses

6.1.1 Vowel length

To the knowledge of the author, no other experimental study has investigated German vowel length productions by Polish L2 German learners acoustically. Darcy and Krüger (2012) and Nimz (2011b) did conduct acoustic measurements of L2 German vowels, but their study group were Turkish L2 German speakers. Furthermore, as was shown by Nimz (2014), Turkish speakers had considerable phonetic experience with long vowels in their native language. This is not the case for Polish native speakers. As was established in the exploratory study of the Polish vowels (see Section 4.2.4), Polish vowel durations do not differ significantly before geminate and singleton consonants, which could have been a confounding factor if this had been the case.

Most current models of L2 speech learning focus on perception rather than production of L2 sounds. Flege's SLM is the only model that specifically addresses sound production in that it proposes that sound production follows sound perception. In the field of GFL, various researchers (e.g. Morciniec, 1990, Slembek, 1995, Müller, 2005) predict that Polish learners of German as a Foreign Language produce German long vowels too short. Even though most GFL researchers agree that German long vowels pose a special problem for Polish L2 German learners, empirical evidence is lacking. One of the few empirical studies which addresses vowel length production is McAllister et al. (2002). As discussed in Section 2.5.1, their Feature Hypothesis states that, if vowel length is not used to signal phonological contrast in the native language, L2 learners will have difficulties producing contrasts based on this feature. For Polish learners of L2 German, this means that they will not produce length differences between German long and short vowels. This leads to the following hypothesis:

Hypothesis 3: German long vowels will be produced shorter by Polish GFL learners compared to German native speakers.

Furthermore, as has been laid out in detail in Chapter 3, orthographic input may play a critical role in L2 vowel length acquisition. For example Bassetti and Atkinson (2015) had found that English vowels are produced significantly longer by Italian L2 English learners when these vowels are spelled with digraphs. German too makes use of orthographic ways of marking long vowels, most prominently by means of *lengthening h* (Section 3.4.1.1). This explicit way of marking vowel length may help Polish learners in acquiring some German words more correctly. Nimz (2011b) had already hypothesized that Turkish GFL learners produce German long vowels more native-like when they are explicitly marked through *lengthening h*. However, this hypothesis has – until now – not been experimentally tested.

Hypothesis 4: Orthographically marked German long vowels will be produced longer by Polish GFL learners than those vowels which are not marked.

Since Polish vowels have been found to be slightly longer than German short vowels (see Section 4.2.4.2), the following hypothesis is stated as regards the production of German short vowels.

Hypothesis 5: German short vowels will be produced longer by Polish GFL learners compared to German native speakers.

Müller (2005) also addressed vowel length production of German short vowels by Polish GFL learners. She predicted that short vowels will *sometimes* be produced too long, but she did not give any theoretical or empirical reasons for this assumption. A possible explanation could be that the orthographic marking of German short vowels additionally modulates duration. In German, double consonant letters explicitly mark short vowels, but this is not the case for all short vowels (see Section 3.4.1.2). It is therefore additionally predicted that there may be differences in the productions of vowel durations between orthographically marked versus unmarked short vowels.

Hypothesis 6: Orthographically unmarked German short vowels will be produced longer than marked short vowels by Polish GFL learners.

6.1.2 Vowel quality

Only minimal data exist which compare German and Polish vowel qualities experimentally. The most extensive study is Hentschel's (1986) work on Polish speakers' perception of German vowels, which included a small set of production data as well. Furthermore he collected a large amount of perceptual assimilation data. Even though Hentschel did not provide information about the experimental design, number of participants, or speakers' age and sex, his production data provide a starting point for the formulation of hypotheses concerning the productions of German vowel quality by Polish L2 German learners. At the same time, the collection of comparable acoustic data in the exploratory part of the study (Section 4.2.4) allows for a more reliable discussion of the results of the main experiment. As has been pointed out in Section 4.2.1, it seems from the limited production data from Hentschel (1986) that the vowel qualities of German /a/, /ɛ/, and /ɔ/ are identical to those of Polish /a/, /ɛ/, and /ɔ/. The perceptual assimilation patterns of naïve Polish speakers support this observation. Long German /aː/ is equally identical in vowel quality to Polish /a/. For German /e:/ and /o:/ the case is different, as there are no comparable Polish counterparts. The quality of German /e:/ seems to lie between the qualities of Polish /i/ and /ɨ/, and the quality of German /o:/ seems to lie between the qualities of Polish /u/ and /ɔ/. Again, the perceptual assimilation patterns of Polish speakers support these observations. For this reason, the following hypotheses are formulated, which concern the vowel quality productions of Polish L2 German learners:

Hypothesis 7: Polish GFL learners will produce the vowel qualities of German /a/, /ɛ/, /ɔ/, and /a:/ native-like.

Hypothesis 8: The vowel quality productions of German /o:/ and /e:/ by Polish GFL learners will deviate from the productions of German native speakers.

A number of GFL researchers have specifically commented on the production of German /e:/ (see Sections 2.1 and 4.2.1), and report that this vowel is often diphthongized. To the knowledge of the author, none of the authors support their own perceptual impressions with experimental data. In order to investigate these impressionistic observations further, the final hypothesis of the production study concerns German /e:/.

Hypothesis 9: Polish GFL learners will show formant movements in their productions of German /e:/ which deviate from those of German native speakers.

6.2 Participants

The same participants as in the discrimination task took part in the production experiment, with the exception of the two additional German participants who were recruited for the discrimination task and the two Polish participants who had a very different language learning backgrounds compared to all other participants (see 5.2. for details). Altogether, 21 German subjects participated in the production part of the study, of which only 20 were analysed acoustically because one dyslexic male participant was excluded from all analyses. The average age of these 16 female and 4 male German participants was 17.9 years (SD=1.1). Of the 20 recorded Polish participants, two female participants did not know at least 75% of the test words used, which was established as the cut-off point before the acoustic analysis. The average age of the remaining 18 participants was 18.6 years (SD=0.6); 4 of them were male.

6.3 Experimental items and conditions

The primary prerequisite for the experimental items of the production and identification task was picturability, as a simple reading task would not allow for differentiation between orthographic effects in reading and actual phonological representation (see also Hentschel, 1986: 54). Secondly, words had to be familiar to all participants. This prerequisite was naturally more constrained by the Polish speakers, who spoke German as an L2 and would therefore be less familiar with some German words. Because it was assumed that a simple frequency measure might not reflect the actual use of words in a foreign language classroom, it was decided to use those words as experimental items which were judged by the German teachers of the school as words which are very familiar to the students.

For this purpose, a word list with picturable one- and two-syllable words taken from various GFL textbooks was given to three teachers prior to the actual fieldwork. The teachers were asked to rate on a scale from 1 ("Very unlikely that the students know this word") to 7 ("Very likely that all students know this word") how familiar they think their students are with the test items. Only items which reached an average familiarity of at least 5 were included in the final items list. This measure might have been more reliable if the students had been asked directly, but since there were not many students at the school who could be classified as medium-advanced learners of German, it was best not to influence some of the potential participants with a prior familiarity rating task.

The vowels investigated in this study were the long vowels /eː/, /aː/, /oː/ and their short counterparts /ɛ/, /a/, /ɔ/. These were chosen because they yielded enough test items which are marked in their orthography by *lengthening h*. Primus's (2000) analysis of the overall distribution of various length markers in German shows that, for example for /uː/, the use of *lengthening h* is considerably less common than for the three long vowels chosen.[70] For each vowel in the experiment, eight test items were chosen, of which four were explicitly marked, while the remaining were not. Half of all test items were one-syllable words and the other half were two-syllable words. Long vowels were considered "explicitly marked" when they were spelled with *lengthening h*. Short vowels were considered "explicitly marked" when double consonant letters followed the short vowel in their orthographic representation. This difference in marking constituted the main experimental factor ORTHOGRAPHIC MARKING, with the two factor levels *marked* versus *unmarked*. It could be argued, as outlined in Section 3.4.1, that even those items classified as *unmarked* are in fact marked for their quantity, as the phonology of German syllable structure can predict which vowels must be long, at least as concerns open syllables. However, hardly any of the didactic GFL texts reviewed in Section 3.3 refer to the level of the syllable when discussing the marking of German vowel quality (rare exceptions are Stock and Hirschfeld, 1996 and Hirschfeld et al., 2007). For this reason, it seemed appropriate to still consider those words spelled without *lengthening h* or double consonant letters as unmarked. A short orthography post-test (Appendix VI) showed that this classification was justified: none of the learners knew that vowel length could, in some cases, be deduced from the syllable structure of a word.

While it would have been desirable to control for segmental context, achieving the main criteria picturability, high-familiarity, and explicit marking resulted in a variation in consonant context, as can be seen in Table 12. In all there were 48 test items (6 vowels (2 x 3 vowel pairs) x 8 words) plus 8 additional items which served as practice tokens in the identification experiment (see Chapter 7).

[70] Only 3% of all native words containing /u:/ are marked by means of *lengthening h*, while for the vowels chosen for the experiments the relative occurrence is about three times higher.

	Long - marked	Long - unmarked	Short - marked	Short - unmarked
/a:/ + /a/	die Sahne ("cream")	die Gabel ("fork")	das Wasser ("water")	lachen ("to laugh")
	fahren ("to drive")	die Tafel ("blackboard")	der Schatten ("shadow")	die Tasche ("bag")
	der Zahn ("tooth")	der Tag ("day")	der Kamm ("comb")	der Wald ("forest")
	die Bahn ("train")	das Schaf ("sheep")	nass ("wet")	die Wand ("wall")
	der Hahn ("cock")	der Schal ("scarf")	das Blatt ("leave")	
/e:/ + /ɛ/	der Fehler ("mistake")	Nebel ("fog")	das Wetter ("weather")	der Becher ("cup")
	der Lehrer ("teacher")	geben ("to give")	der Sessel ("armchair")	das Fenster ("window")
	die Zehn ("ten")	der Weg ("way")	das Bett ("bed")	das Geld ("money")
	das Mehl ("flour")	der Keks ("cookie")	nett ("nice")	die Welt ("world")
			der Teller ("plate")	rechnen ("to calculate")
/o:/ + /ɔ/	die Kohle ("coal")	der Boden ("floor")	der Sommer ("summer")	die Woche ("week")
	wohnen ("to live")	der Monat ("month")	die Sonne ("sun")	die Wolke ("cloud")
	der Sohn ("son")	das Rot ("red")	voll ("full")	das Loch ("hole")
	der Lohn ("salary")	hoch ("high")	Gott ("god")	der Koch ("cook")
	der Kohl ("cabbage")	Polen ("Poland")	das Schloss ("castle")	

Table 12: Experimental items for both the production and identification experiment (grey items served as practice items in the identification task)

Attention was further paid to exclude German-Polish cognates or words with two consonants in coda position following a long vowel.[71] Unfortunately, choices for the one-syllable, unmarked /e:/-words were limited, which is why the word *Keks* ("cookie") was included in the final items list despite the fact that the word exists in Polish as well, meaning "cake". An alternative word on the prior items list may have been *Steg* ("pier"); however, this word received an average familiarity rating of 1. It was decided to include *Keks* despite its problematic characteristics and to check by statistical means whether the inclusion/exclusion of this word would change the results.

[71] As already mentioned in Section 3.4.1.2, two consonants in coda position could be understood as a marker for vowel shortness, though Ramers (1999a) lists a large amount of counterexamples. This might be a reason why such a "rule" is not encountered in GFL textbooks either.

6.4 Procedure

Pictures of the test items were presented on the same Acer Timeline laptop that was used for the discrimination task. The same random order was used for each participant. Subjects were instructed to say the word which they thought would be represented by the picture they saw. If a different word than the target was produced, the German experimenter would ask the participant to name another possible word until the correct one was recorded. Whenever a participant did not know a word, the experimenter described the respective item in more detail without using the word itself, so as to prompt an authentic production even if the picture could not be named. Whenever a participant still did not know the respective word, the next item was presented.

In the second run of the production experiment, the same pictures were presented in a different random order and the participants were asked to name the pictures in the same way they had done in the previous run. Once a word was encountered which was not known in the first run, the experimenter would produce the word for the participant and ask him or her to repeat the word twice. Sometimes a participant would remember a word in the second run which could not be named in the first run. In this case, the participant was asked to produce the word twice so that, in the end, two authentic productions of each item could be used for further analysis. Recordings were made in a quiet classroom at a sampling rate of 44.1 kHz (16-bit resolution) by means of a KORG MR-2 high resolution mobile recorder with an integrated high-quality microphone.

6.5 Acoustic analysis

The productions of 38 participants were analysed using PRAAT. Each test word and its (stressed) vowel were carefully segmented by hand, following the guidelines in Appendix III. As mentioned above, some participants' productions were based on repetitions of the model productions of the experimenter. These unknown, repeated words constituted 8% of the whole data set. Furthermore, 61 vowels had to be labelled as missing data due to background noise, incorrect utterances, unnatural (exaggerated) productions, unreliable segmentation, or interruptions.

In order to at least partially check the reliability and objectivity of the author's acoustic measurements, a student assistant was hired to segment about 20% of the same production data. The assistant, with experience in PRAAT, was given all male recordings (4 German and 4 Polish participants) and was instructed to segment the vowels according to the given segmentation criteria. A Pearson's correlation between the durations of all segmented vowels of the first and second annotator revealed a correlation of r=0.84, which is classified as "strong" and can therefore be understood to indicate acceptable reliability (Anderson-Hsieh et al., 1992). The coefficient was even higher (r=0.87) when those vowels were excluded which were followed by a plosive (27% of the data set). This was tested because the author found that the student assistant tended to include the closure phase of the following plosive into the duration of the preceding vowel. More detailed comparisons, which differentiated between the different consonantal contexts (i.e. plosives, fricatives, nasals, and liquids), revealed that the fricative and nasal contexts were the most reliable to segment (both r=0.92), while the liquid contexts were the most problematic (r=0.66). This quantitative difference was mirrored in the qualitative inspection of problematic cases. For example, the vowels in *Geld* and *Wald* were frequently difficult to segment due to the following lateral consonant.[72] In such problematic cases, the author discussed the segmentation with two other phoneticians. If it was not possible to agree on the boundary between the vowel and the lateral, the segment was labelled as a missing data point.

6.6 Results

The following graphs and statistical models are based on all available data points excluding the repeated words and those words which, in a post-test addressing the writing skills of the Polish participants, were written incorrectly

[72] This is a known problem in acoustic segmentation. Turk et al., 2006, in their discussion of relative segmentability, list lateral approximants as the hardest segments to segment from vowels and suggest avoiding them altogether. Because of the very limited choice of potential test words, this consonantal context could unfortunately not be avoided.

in a way which could confound with the main experimental variable "orthographic marking" (e.g. *<Schaff> for [ʃa:f], correctly spelled <Schaf>). While 143 instances had to be repeated, only 14 words were spelled incorrectly. In all, 10% of all the data were excluded, including those items which had to be labelled as missing data (see above). Statistical analyses showed that the inclusion of the problematic word *Keks* did not change the results, which is why the word remained part of the analysis.

As the respective hypotheses differentiate between short and long vowels, the data for the analysis of vowel length are divided into two separate data sets, namely data set *long* (1639 data points) and data set *short* (1647 data points). Furthermore, vowel length and vowel quality are addressed separately, mirroring the order of the hypotheses in Section 6.1 above.

6.6.1 Vowel length

The dependent variable in the analysis of vowel length is the duration of the (stressed) vowels of the test words in milliseconds. Because the analyses are carried out by means of linear mixed models, which for example allow for varying intercepts per participant, the fact that some speakers may speak slightly slower or faster than others is accounted for. Other possible ways of normalization were found inappropriate for this data set, as for example Polish voiced plosives are more voiced (and therefore longer) than German speakers', which makes conceivable normalization procedures such as vowel/preceding consonant (or even word) unsuitable. Though possible, normalization by standard z-scores of the duration data would also be misleading, as important information about the absolute length of the data (per subject and group) get lost in this procedure. Finally, it is common practice to analyse and report raw duration data (e.g. Elsendoorn, 1984, Bohn and Flege, 1992, Bassetti and Atkinson, 2015). The two main independent variables in the experiment were language (German versus Polish) and orthographic marking (marked versus unmarked). As laid out in Section 3.4.1.1, German long vowels are considered marked when *lengthening h* is following the vowel, as this is an explicit marker for vowel length. German short vowels are considered explicitly marked when double consonant letters are following the vowel (see Section 3.4.1.2).

6.6.1.1 Long vowels

Figure 11 shows the average duration of long vowels produced by Polish GFL learners and German native speakers for orthographically marked and unmarked vowels.

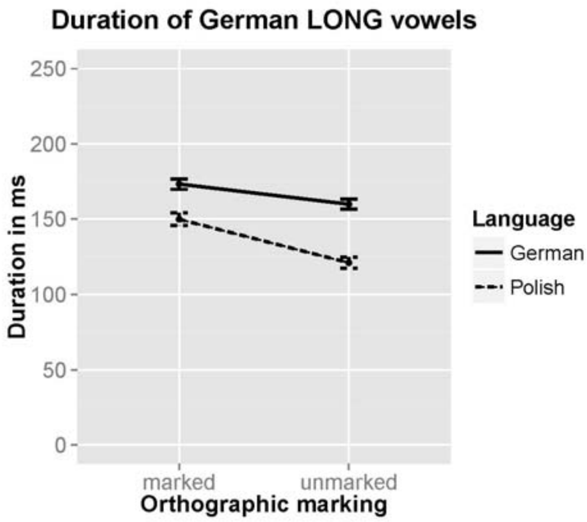


Figure 11: Duration (in ms) of German long vowels by language and orthographic marking (N(Polish)=18, N(German)=20; error bars show 2 SE)

It seems to be clear from the visual inspection of the graph that Polish speakers produce shorter long vowels than German speakers (Hypothesis 3). There also seems to be a less clear tendency in the Polish group to produce orthographically marked vowels relatively longer than German native speakers (Hypothesis 4). In order to test these differences statistically, a linear mixed model (LMM) was fit to the duration data which included LANGUAGE (*German* versus *Polish*) and MARKING (*marked* versus *unmarked*) as fixed factors, random intercepts for participants and items, as well as by-participant random slopes for the effect of orthographic marking and by-item random slopes for the effect of language. The random slopes account for the fact that participants and items vary with regards to how sensitive they are to the experimental variables.[73] Adding by-participant random slopes for marking implies that some

---

[73] This is why random slopes can only be added to repeated measures fixed effects. Here, participants produce both marked and unmarked items (i.e. repeated measures for marking by participants) but they can only be Polish *or* German (non-repeated). Items on the other hand are produced both by Polish and German speakers (i.e. repeated measures for language by item), but are either marked *or* unmarked in their orthography (non-repeated).

participants are more affected by the marking of a long vowel than others. By-item random slopes for the effect of language mean that the average item duration depends on which language group produced it. It is crucial to add random slopes for all experimental variables as they decrease Type I error rates, i.e. rejecting the null hypothesis (in this case: no effect of orthographic marking) when it is true (Barr et al., 2013). Winter (2013) advises to add them for all main experimental factors (in this case LANGUAGE and MARKING). While in the previous models of the study random slopes did not improve model fits, they did in this model; hence they should be included.

Furthermore, control variables such as SYLLABLE (*one*- versus *two*-syllable test words), SEX (*male* versus *female*), VOWEL (*a* versus *e* versus *o*), VOICING of the following consonant (*voiced* versus *voiceless*), and RUN (*first* versus *second* run) were added to the final model. Due to the experimental set-up and design of the production study, these variables could not be controlled for beforehand, but are controlled for statistically by including them in the model. The variable SYLLABLE was added because German and Polish belong to different rhythmic classes and might therefore reduce/shorten vowels in one- or two-syllable words differently (see Sections 4.1.3 and 4.2.3 for German and Polish, respectively). The control variable SEX was added to the model because (Weirich and Simpson, to appear) had found that female native speakers had a tendency to produce German vowels longer than male native speakers. The variable VOWEL was added because research on vowel perception had shown that low and mid vowels are perceived differently (Weiss, 1974, Sendlmeier, 1981). This had also been confirmed in the discrimination experiment above (Chapter 5). Furthermore, the VOICING of the following consonant was considered an important control variable because in most languages, vowels are longer before voiced consonants, but it was unclear whether German and Polish would differ in this respect (Chen, 1970). Lastly, it was decided to add the variable RUN instead of averaging over the first and second production cycle. This way, "pseudoreplication" is controlled for (Winter, 2011), while being able to keep all data points. In order to verify whether the inclusion of these additional variables was justified, likelihood ratio tests were performed whereby each control factor was separately added to the basic model as well as its interaction with the factor LANGUAGE. The following final production data model includes all significant control factors and/or their interaction with the factor language[74]:

*Model.long = lmer(Duration ~ Language*Marking + Syllable*Language + Post_voicing:Language + Vowel:Language + Sex + Run + (1+Marking|ID) + (1+Language|Word), data=long, REML=FALSE)*

The current version of the R-package *lme4* (Bates et al., 2014) does not automatically provide p-values along with the given t-values, because it is not yet clear how to calculate the appropriate degrees of freedom when random slopes are included in the model. (This is irrelevant for the earlier calculated binomial models, as for them the z-distribution is relevant.) Baayen et al. (2008) points out that, for large data sets like this one, a simple way to assess significance at the 5% level is to check whether the absolute value of t exceeds 2. With this guideline, it becomes clear from the inspection of the t-statistics in Table 13 that orthographic marking is in fact not significant, neither as a main effect nor in interaction with language (Hypothesis 4). On the other hand, the main experimental factor LANGUAGE has the largest significant influence on vowel duration.

[74] A simple interaction between a control factor and LANGUAGE is denoted by a column between the two; an interaction between a control factor and language including a main effect of the control factor is denoted by a star.

	Estimate	Std. error	t-value
(Intercept)	207.16	7.54	11.38
Language (Polish)	**-54.36**	**7.46**	**-7.28**
Marking (unmarked)	**-5.95**	**7.84**	**-0.76**
Sex (male)	-14.01	5.16	-2.72
Syllable (two)	-37.95	6.65	-5.71
Run (second)	14.0	1.22	12.29
Lang. (Pol.):Marking (unm.)	**-9.33**	**6.71**	**-1.39**
Lang. (Ger.):Post_v (voiceless)	-12.98	9.68	-1.34
Lang. (Pol.):Post_v (voiceless)	-22.89	9.22	-2.48
Lang. (Ger.):Vowel (e)	-34.14	6.59	-5.18
Lang. (Pol.):Vowel (e)	16.37	6.36	2.57
Lang. (Ger.):Vowel (o)	-25.22	6.59	-3.83
Lang. (Pol.):Vowel (o)	-7.41	6.27	-1.18
Lang. (Pol.):Syllable (two)	16.43	5.51	2.98

Table 13: Summary of fixed effects of *Model.long* (main experimental factors in bold; N(observations)=1639; groups: 24 words and 38 participants)

Exact p-values can be obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. This way, a value of $p<0.001$ is obtained for the effect of LANGUAGE (Hypothesis 3). Vowels produced by German native speakers are on average 166.5 ms ± 1.2 ms (SE) long, while Polish vowels measure on average 135.0 ms ± 1.5 ms (SE). Hence, Polish vowels are on average more than 30 ms shorter than German vowels. German *marked* long vowels are on average 173.0 ms ± 1.8 ms (SE), and German *unmarked* long vowels are on average 159.8 ms ± 1.6 ms (SE). There is a similar difference for this factor in the Polish group: 149.8 ms ± 2.1 ms (SE) for *marked* vowels and 120.8 ms ± 1.8 ms (SE) for *unmarked* vowels. However, these difference did not turn out to be significant, both as a main effect ($\chi^2(1)=2.28$, $p=0.13$), and in interaction with LANGUAGE ($\chi^2(1)=1.88$, $p=0.17$). This is due to the fact that not all participants are affected similarly by the experimental factors (as represented by the random slopes within the model), and the addition of control factors which, if they had not been included, may have been mistaken for an effect of MARKING.

All of the control factors display t-values whose absolute values clearly exceed 2. To illustrate the estimates of the model, three significant control factors will briefly be discussed: First, the model output shows significant interaction for the factors LANGUAGE and VOICING (of the following consonant). Figure 12 illustrates this interaction.

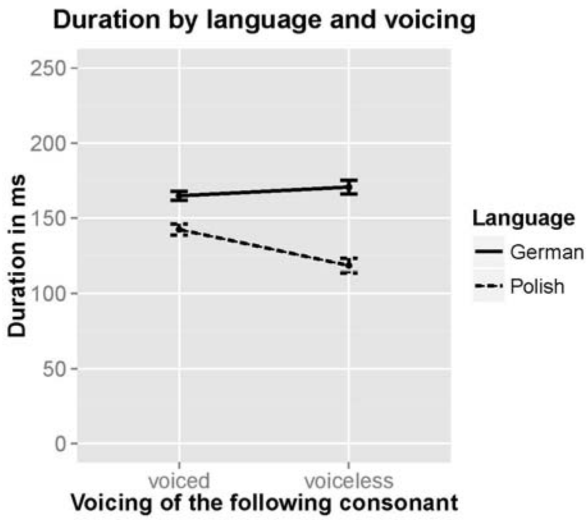


Figure 12: Duration (in ms) of German long vowels by language and voicing of the following consonant (N(Polish)=18, N(German)=20; error bars show 2 SE)

While vowels for Polish speakers are on average 24.0 ms longer before voiced consonants, this is not the case for German native speakers. As far as the Polish data are concerned, the result corroborates findings by Slowiaczek und Dinnsen (1985) (see Section 4.2.2). The results for the German group, however, were unexpected because Port (1985) had found that German speakers even produced slightly longer vowels before *underlying* voicing consonants (see Section 4.1.2). This underlines the importance of including control factors, as the effect of VOICING in the Polish speakers could have been mistaken as an effect of MARKING, as there are less vowels followed by a voiced consonant in the *unmarked* condition. Second, as can be seen in Figure 13, vowels in two-syllable words were relatively longer in Polish speakers than in German natives.

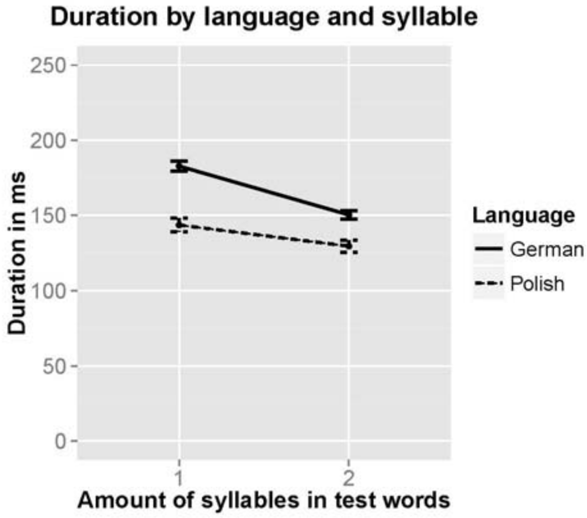


Figure 13: Duration (in ms) of German long vowels by language and amount of syllables in test words (N(Polish)=18, N(German)=20; error bars show 2 SE)

German speakers have a much stronger tendency to shorten German long vowels in words with two syllables. Vowels by Polish speakers are – relatively – longer, even though they are still slightly shorter in two-syllable words. German speakers shorten their vowels in two-syllable on average by 32.6 ms, while Polish speakers only do so by on average 14.1 ms This interaction between LANGUAGE and amount of SYLLABLES supports the idea that syllables in stress-timed languages, such as German, are compressed to fit into a given time interval. Even though researchers disagree on whether Polish should be classified as syllable-timed or "mixed" (see Section 4.2.3), it seems clear from the data at hand that Polish speakers reduce vowels less than German speakers in words which consist of more than one syllable. This supports Grzeszczakowska-Pawlikowska's (2007) classification of Polish as a syllable-timed language. Third, the data show an interesting interaction between VOWEL and LANGUAGE, as exemplified in Figure 14.

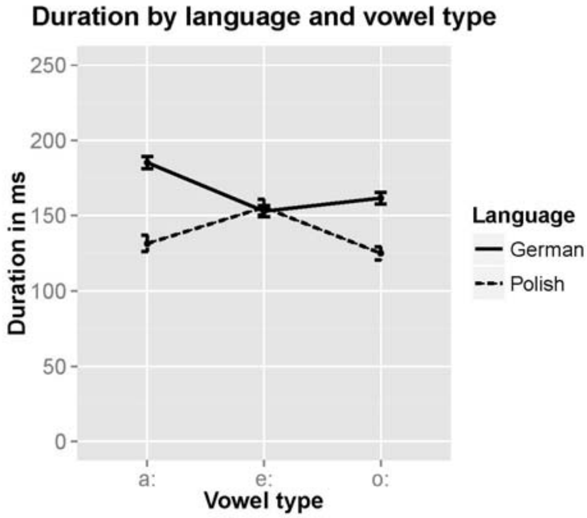


Figure 14: Duration (in ms) of German long vowels by language and vowel type (N(Polish)=18, N(German)=20; error bars show 2 SE)

While it is true that Polish learners produce shorter long vowels when averaged over *all* vowels, the vowel /e:/ is a notable exception. In Polish speakers it is on average 155.4 ms ± 2.6 ms long and in German speakers 152.8 ms ± 1.9 ms. The fact that /e:/ is produced longer by Polish speakers relates well to the results of the vowel quality analysis below.

It may be possible that an interaction of language and orthographic marking was not found because the Polish GFL learners were not aware of the meaning of *lengthening h*. In one of the post-tests, the orthographic knowledge of the learners was tested by presenting them with German nonce words that were written with different length markings (*lengthening h* and double consonant letters versus no explicit markings). The participants were asked to identify whether – and possibly why – the vowels of the nonce words were long or short (see Appendix VI). By identifying some participants as learners who had explicit knowledge of *lengthening h* (i.e. those which mentioned it in the test), the data could be subset to a learner group which only consisted of Polish participants who knew the rules (7 of 18). The orthography test was also administered to the German participants; all of them were aware of the meaning of the *lengthening h*. The LMM was again run with this subset of participants (1225 data points). Still, MARKING did not turn out to be significant, either as main effect ($\chi^2(1)=1.87$, $p=0.17$) or in interaction with LANGUAGE ($\chi^2(1)=2.78$, $p=0.10$).

Three more variables were further examined, which were only relevant for the Polish group: *ambition* to sound like a native speaker, *familiarity* with the test words, and the *class* the learners were in. The participants were recruited from the two highest German classes, namely the third and fourth (last) year of high schooling. It may be possible that length of formal instruction has an influence on how well vowel length is produced. Similarly, more familiar words may be produced more correctly (i.e. with longer vowel durations), or more ambitious students may produce long vowels more native-like.

An LMM was fit to the data which included the main experimental factor MARKING as well as all control factors which had been found to be significant in the previous model (i.e. VOWEL, SEX, RUN, number of SYLLABLES, and VOICING of the following consonant). Furthermore, AMBITION was added as a main effect as well as in interaction with orthographic MARKING. This factor was coded as a continuous variable from 1 ("no ambition to sound like a native speaker") to 10 ("high ambition to sound like a native speaker"). FAMILIARITY with the test words was also operationalized as a continuous variable on an interval scale from 1 ("not familiar with the word") to 7 ("very familiar with the word"), and was added as main effect and in interaction with MARKING. The additional factor CLASS included the two factor levels *third grade* and *fourth grade*, and was added in the same way as the other two variables. The model also specified random intercepts for participants and by-participant random slopes for the effect of marking, as these improved model fit.[75] Table 14 summarizes the statistics of the model.

[75] *Model.Polish = lmer(Duration ~ Marking*Ambition + Marking*Class + Marking*Familiarity + Vowel + Sex + Run + Syllable + Post_voicing + (1+Marking|ID) + (1|Word), data=long.Polish, REML=FALSE)*

	Estimate	Std. error	t-value
(Intercept)	88.45	18.83	4.70
Marking (unmarked)	22.28	20.01	1.11
Ambition	**4.87**	**1.56**	**3.13**
Class (fourth)	**14.64**	**8.54**	**1.72**
Familiarity	**2.0**	**1.41**	**1.42**
Post_voicing (voiceless)	-22.36	9.12	-2.45
Vowel (e)	17.21	6.34	2.71
Vowel (o)	-6.09	6.25	-0.97
Sex (male)	-7.15	7.06	-1.01
Syllable (two)	-21.32	6.21	-3.44
Run (second)	16.38	1.98	8.25
Marking (unm.):Ambition	**-2.42**	**0.97**	**-2.48**
Marking (unm.):Ambition	**-3.19**	**5.11**	**-0.62**
Marking (unm.):Ambition	**-2.51**	**2.49**	**-1.01**

Table 14: Summary of fixed effects of *Model.Polish* (factors of interest in bold; N(observations)=689; groups: 24 words and 18 participants)

Looking at the t-values, it becomes apparent that AMBITION (and its interaction with MARKING) is the only factor that exceeds the absolute value of 2 (i.e. a significance level of 5%). All previous control factors remain significant, with the exception of speaker SEX. In this smaller data set, this factor ceases to be significant, which implies that the effect of this control factor is generally not as strong as those of the other factors. Since AMBITION seemed to play a role in the Polish group in that more ambitious speakers produced slightly longer vowels and were influenced positively by orthographic marking, it was decided to run the original model again with only those participants who had indicated an ambition to speak German native-like of at least 8 (on scale from 0 to 10). Five Polish participants were therefore excluded from the re-run of *Model.long*.

The same control factors as in the first run of the model remained significant after the exclusion of the most unambitious participants. More importantly, the main effect of MARKING ($\chi^2(1)=1.72$, $p=0.19$) and its interaction with LANGUAGE ($\chi^2(1)=2.75$, $p=0.10$) remained non-significant. Hypothesis 4, which states that Polish learners produce longer vowels when they are orthographically marked, therefore remains to be rejected on the basic of the data at hand.

6.6.1.2 Short vowels

In order to test Hypotheses 5 and 6, the durations of the German short vowels were examined as well. From the visual inspection of the 1647 data points of the short vowel productions, it becomes apparent that Polish speakers produce the German short vowels longer than the German control group (see Figure 15).

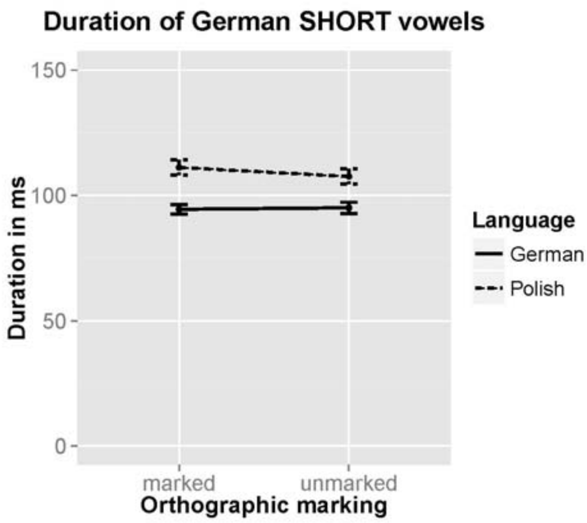


Figure 15: Duration (in ms) of German short vowels by language and orthographic marking (N(Polish)=18, N(German)=20; error bars show 2 SE)

Once again, an effect of orthographic marking does not seem to be present. To examine this statistically, an LMM was fit to the data which included LANGUAGE and MARKING as fixed factors, random intercepts for participants and items, as well as by-item random slopes for the effect of LANGUAGE. By-participant random slopes for the effect of MARKING did not improve model fit ($\chi^2(2)=0.56$, $p=0.76$), which is why they were not included in the final model of the short vowels.

Due to the likelihood ratio test results for the short vowel data, SEX, RUN, and amount of SYLLABLES in test words were included as main effects and VOWEL was included in interaction with the factor LANGUAGE.[76] In contrast to the model for the long vowels, SYLLABLE was not added in interaction with LANGUAGE, because for the short vowel data it did not improve model fit ($\chi^2(1)=0.64$, $p=0.43$). This can be explained by the fact that the duration of the short vowels cannot be shortened as much as in the long vowels; hence, the stronger effect in long vowels for German speakers is limited by the fact that short vowels are already relatively short. The same may be true for the effect of VOICING, which did not significantly improve the fit of *Model.short*, so it was not included as a control factor in this model.

	Estimate	Std. error	t-value
(Intercept)	100.80	4.70	21.46
Language (Polish)	**22.62**	**4.73**	**4.78**
Marking (unmarked)	**0.68**	**3.91**	**0.18**
Sex (male)	-12.54	3.49	-3.60
Syllable (two)	-19.03	3.27	-5.82
Run (second)	9.31	0.90	10.35
Lang. (Pol.):Marking (unm.)	**-4.77**	**3.72**	**-1.28**
Lang. (Ger.):Vowel (e)	0.83	4.78	0.17
Lang. (Pol.):Vowel (e)	-9.95	4.48	-2.22
Lang. (Ger.):Vowel (o)	2.12	4.78	0.44
Lang. (Pol.):Vowel (o)	-3.92	4.49	-0.88

Table 15: Summary of fixed effects of *Model.short* (factors of interest in bold; N(observations)=1647; groups: 24 words and 38 participants)

As can clearly be seen from the t-values for the effect of MARKING (t=0.18) and its interaction with LANGUAGE (t=-1.28), this experimental factor does not play a role in the production of the duration of German short vowels either. To obtain exact p-values for this, likelihood ratio tests were performed as usual: the

[76] *Model.short = lmer(Duration ~ Language*Marking + Vowel:Language + Sex + Syllable + Run + (1|ID) + (1+Language|Word), data=short, REML=FALSE)*

main effect of MARKING was not significant at $p=0.56$ ($\chi^2(1)=0.34$) and the interaction between MARKING and LANGUAGE was not significant at $p=0.21$ ($\chi^2(1)=1.59$). Yet, language had a significant effect on vowel duration ($p<0.001$). On average, Polish speakers produced German short vowels 14.7 ms longer than German native speakers (average Polish length: 109.3; average German length: 94.6 ms). This supports Hypothesis 5, which was based on acoustic measurements of comparable Polish and German vowels. However, the duration of short vowels was not additionally modulated by orthography (Hypothesis 6). As for the long vowel data above, it was again double-checked whether this result might be due to the fact that some learners did not know the orthographic rules for short vowels (i.e. that double consonant letters indicate that the preceding vowel is short). By checking the results of the orthography test for the short vowels, it was found that only 4 Polish speakers were aware of this rule. *Model.short* was run again with this subset of Polish speakers and all German native speakers (1108 data points). Still, MARKING remained non-significant (main effect: $\chi^2(1)=0.10$, $p=0.75$; interaction with LANGUAGE: $\chi^2(1)=0.43$, $p=0.51$). Furthermore, the additional control factors for the Polish group (AMBITION, FAMILIARITY, and CLASS) did not turn out to be significant predictors in the Polish-only model. Hence, *Model.short* was not re-run with a subset based on any of those factors.

As can be deduced from the comparison of the estimates of the control factors in Table 13 (long vowel data) and Table 15 (short vowel data), male speakers generally produce shorter vowels than female speakers, which supports findings by Weirich and Simpson (to appear). The control factor RUN also points into the same direction for both long and short vowels: vowels are generally longer when produced a second time. Even though this may seem counter-intuitive at first glance, it could be explained by the level of stress of the speakers in the first run. When they are first presented with the pictures, they might be concerned with whether they will know the up-coming words or not. In the second run, they already know the test items and as a consequence might be more relaxed and speak slightly slower, hence, their vowels are shorter.

An interesting difference between the control factors in the long and short vowel models is the interaction between VOWEL and LANGUAGE. While /ɛ/ was relatively shorter than the other vowels (negative algebraic sign for the estimate in the Polish group in *Model.short*), /eː/ was relatively longer. In order to explain why Polish speakers may produce the duration of [eː] differently than in all other vowels, it will be of interest to look at the analysis of the vowel quality data below.

6.6.1.3 Discussion

Four hypotheses had been put forward for the dimension of vowel length. Half of them were corroborated by the data: Hypothesis 3 stated that German long vowels will be produced shorter by Polish GFL learners compared to German native speakers. This was clearly the case and vowels produced by Polish speakers were on average 32 ms shorter than those same vowels produced by German speakers. Furthermore, German short vowels were produced *longer* by Polish speakers by about 15 ms. This fits in well with acoustic comparisons of monolingual data of the two languages, where it was found that Polish vowels are on average at least 14 ms longer than German short vowels (Section 4.2.4.2).

Yet, the fact that Polish speakers' long vowels are shorter, and short vowels are longer than those of German native speakers does not mean that Polish speakers do not differentiate between long and short vowels at all. To support this statistically, a final LMM was fit to the Polish data with the factor VOWEL LENGTH (*short* versus *long*) as fixed factor, random intercepts for participants and items, and by-participant random slopes for the effect of VOWEL LENGTH. The previous significant control factors were also added as main factors.[77] The model revealed that Polish GFL learners produced German short vowels significantly shorter than German long vowels ($\chi^2(1)=13.43$, $p<0.001$). Polish speakers' short vowels were on average 109.4 ms ± 1.1 ms (SE) long, while their long vowels were on average 136.2 ms ± 1.5 ms long. The learners therefore contrasted German short and long vowels in their productions, but did not do so native-like.

---

[77] *Model.short-long = lmer(Duration ~ Vowel_length + Post_voicing + Sex + Syllable + Run + (1+Vowel_length|ID) + (1|Word), data=Polish, REML=FALSE)* (VOWEL was not added as it was inherent to the factor VOWEL LENGTH.)

As far as orthographic marking is concerned, neither Hypothesis 4 nor Hypothesis 6 could be corroborated. Hypothesis 4 stated that orthographically marked German long vowels will be produced longer by the Polish speakers than those vowels which are not marked, while this would not be the case for the German native speakers. On the basis of the data at hand, this hypothesis needs to be rejected as there was no significant interaction between MARKING and LANGUAGE. Even when the data were subset to those participants who knew the orthographic rule for *lengthening h* explicitly or to those who were the most motivated to speak native-like, marking remained non-significant. The same was true for orthographic marking in the short vowels. Polish speakers did not produce orthographically unmarked German vowels longer than marked vowels. Hypothesis 6 therefore needs to be rejected as well.

Even though they were not of main interest to the study, several control factors were found to have a significant influence on German vowel duration. For example, vowels were generally longer in female speakers and in the second run of the experiment. Furthermore, the amount of syllables in the test words was found to be an influential factor in predicting vowel duration. In words with two syllables, vowels were shorter in both groups, while for the long vowels, German speakers showed a stronger effect than Polish speakers. This can be explained well with the rhythmic classes the two languages belong to, as a stress-timed language like German would be expected to show more compression than a non-stress-timed language like Polish. A significant effect of the voicing of the following consonant was found for the Polish speakers in the long vowels, as well. This demonstrates that it is important to include significant control factors, as the effect of voicing could have easily been mistaken as an effect of orthographic marking. Lastly, the control factors showed an interesting effect of vowel type for long /eː/, as this vowel was produced relatively longer by the Polish speakers. As a number of GFL researchers have already pointed out, Polish L2 German learners tend to diphthongize this vowel. It might be the case that /eː/ was therefore longer, which will be examined in the next section.

6.6.2 Vowel quality

The investigation of vowel quality was operationalized by measuring the first (F1) and second formant (F2) by means of a PRAAT script[78] at vowel midpoint, as this is common practice for describing vowel qualities in the field (e.g. Bohn and Flege, 1992, Flege et al., 1997, Darcy and Krüger, 2012, or Escudero et al., 2012, among many others). Furthermore, F1 and F2 were measured at 25% and 75% of the vowels' duration, following Steinlen (2009), who had investigated formant movements in L2 learners of English. This was done in order to provide data for testing Hypothesis 9, which is concerned with possible diphthongization of German /e:/ by Polish L2 German learners. Speaker sex is a confounding factor in formant measurements as female formants tend to be higher than those of male speakers (Simpson and Ericsdotter, 2007). For this reason, the data for female and male speakers were analysed separately. By doing so, it was possible to directly compare the 7 male Polish formant values from the exploratory study (Section 4.2.4.2) with the formant values of the 4 male German native speakers of the main production experiment. This was done first in order to check whether Hentschel's (1986) production data used for the formulation of the hypotheses above provided a reliable basis.[79] After the descriptive inspection of the male formant data, the data of 30 German and Polish female speakers will be statistically analysed.

6.6.2.1 Descriptive comparison of German and Polish vowels

The average age of the 7 male Polish speakers who had produced the six Polish vowels in the earlier exploratory study was 17.6 years (SD=0.5). With that, they are very well comparable to the 4 German speakers of the main production experiment, whose average age was 17.8 years (SD=0.4). The productions of the

[78] Formant frequencies were measured by means of an LPC analysis. The analysis window of the script was 25 ms (5 ms time steps). The maximum formant value for female voices was set at 5500 Hz; for male voices at 5000 Hz (5 formants were tracked). Different values were used for /ɔ/ and /o:/: the maximum formant value was set at 3300 Hz for female voices; for male voices at 3000 Hz (3 formants were tracked). This was done because F2 is comparably low in the mid back vowels and was often missed by the original settings.

[79] It was not possible to base the main hypotheses on the study's exploratory data from the start, as the German data were collected after the Polish L2 German data.

Polish speakers had been elicited by means of a reading task (singly presented words on a computer screen), while the production of the German speakers had been elicited by means of a picture-naming task. Even though the experimental set-up was slightly different, it is still assumed that the data provided valid comparisons. To the knowledge of the author, it has not been shown that vowel qualities differ depending on whether a word is read or a picture is named. Both tasks elicited words one at a time in a situation that required participants to speak with equal (normal) speed and clarity. The data in both groups had been collected with high-quality field recorders in quiet classrooms. Even though both the German and the Polish speakers spoke English as their first L2, the speaking mode in both settings was in the participants' native language. This was done in order to prompt a monolingual mode for all speakers (Grosjean, 2001). While the German participants had been instructed by the author of this study (a German native speaker), the Polish participants had been instructed by a student assistant who spoke Polish as her native language. As has been described above, all test words appeared in a variety of consonantal contexts and none of the vowels was restricted to a particular context. Figure 16 shows the mean formant frequencies of the German vowels investigated in this study in comparison to the six Polish vowels.

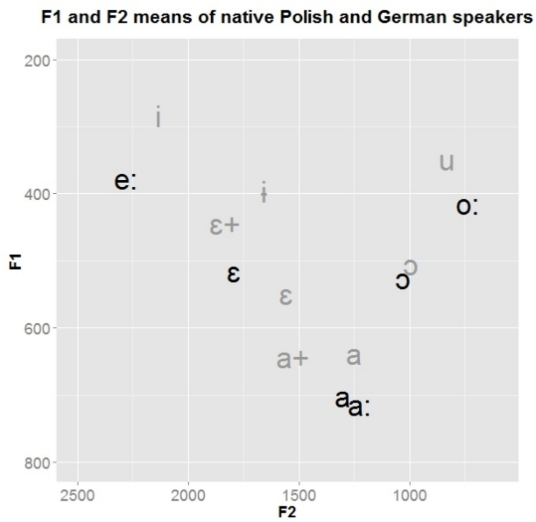


Figure 16: Average Polish (grey) and German (black) vowel qualities as spoken by native speakers; Polish vowels in palatal context are marked by "+" (N(Polish)=7, N(German)=4)

All measured vowels were stressed vowels, and mean frequencies were calculated over all speakers and tokens. For example, for Polish /i/ 7 (speakers) x 3 (word) x 2 (repetitions) tokens served the analysis. For German /e:/, 4 (speakers) x 8 (words) x 2 (repetitions) items were measured. When compared to Hentschel's acoustic data (Figure 3, page 77), some important similarities and minor differences can be made out. First, both Hentschel's data and the current data show that German /ɔ/ and Polish /ɔ/ are almost the same, while German /o:/ is closest in its quality to Polish /u/. In this data set, Polish /a/ seems to be located slightly higher in the acoustic vowel space, both in palatal and non-palatal contexts. While this relationship was reversed in Hentschel's data (Polish /a/ seemed to be located slightly lower), the conclusion is the same: The closest German vowel to Polish /a/ is German /a/, even if it is acoustically slightly different. As far as Polish /ɛ/ is concerned, the current data set provides evidence that its quality changes substantially according to its consonantal context (palatal versus non-palatal), but the different allophones are always closest to German /ɛ/. In Hentschel's graph, it looks like the quality of Polish /ɛ/ is almost the same as German /ɛ/. This would have been the case in the current data set as well, had the vowels not been plotted separately for the palatal and non-palatal context. For this reason, it could be speculated that Hentschel had collected vowels in different kinds of contexts including palatal consonants, but did not analyse them separately. Even though Polish /ɛ/ is considerably raised in palatal contexts, it is still closer to German /ɛ/ than German /e:/. German /e:/ seems closest to Polish /i/ in both Hentschel's and the current data. At the same time the next-closest Polish vowel would be Polish /ɛ/ in palatal context, although in the current data set Polish /ɨ/ is close as well. With that, both German /o:/ and /e:/ emerge as the vowels which may cause the largest difficulties for Polish L2 German learners, as their assimilation to German vowels is arbitrary (Hypothesis 8). At the same time, German /a(:)/, /ɛ/, and /ɔ/ find clear counterparts in Polish, which are relatively close to the respective German vowels and are therefore likely to be produced native-like – at least as far as their quality is concerned (Hypothesis 7). To test these hypotheses, statistical analyses were conducted with the female participants' data of the main picture-naming task.

6.6.2.2 Statistical analysis of female vowel qualities

For the investigation of vowel quality, 14 female Polish speakers and 16 female German speakers were analysed. The average age of the Polish group was 18.5 years (SD=0.6) and the average age of the German group was 17.6 years (SD=0.9). The data were subset in the same way as the length data, i.e. repetitions and incorrectly spelled words were not included in the analysis. Furthermore, items for which either F1 or F2 could not be measured reliably were not included in the analysis. In all, 2477 data points served the vowel quality analysis of the female data. Figure 17 shows the mean formant frequencies of the six German vowels of interest for the Polish L2 German learners and the German native speakers.

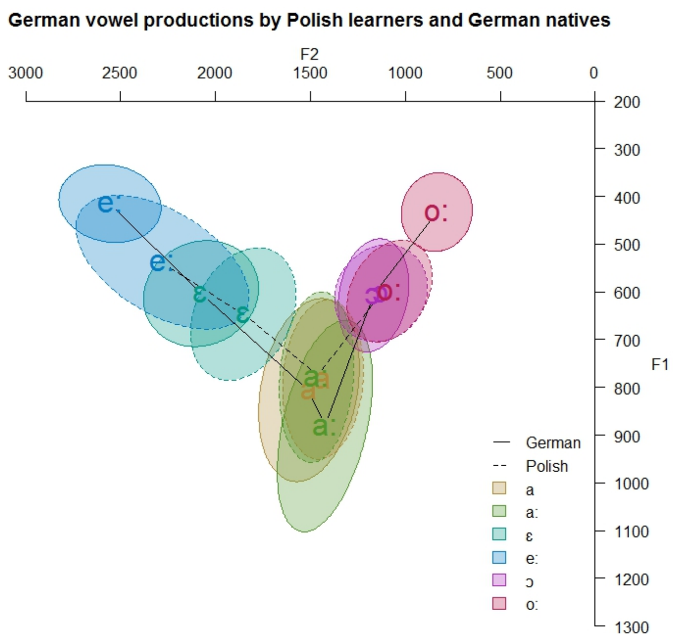


Figure 17: Mean formant frequencies for the German vowels /a, a:, ɛ, e:, ɔ, o:/ produced by Polish GFL learners and German native speakers. Ellipses correspond to a confidence level of ±1 standard deviation from the bivariate mean (N(Polish)=14, N(German)=16)

From the graphical display it becomes obvious that German /o:/ is produced by Polish GFL learners with vowel qualities identical to German /ɔ/. This means that the Polish group does not differentiate German /o:/ and /ɔ/ quality-wise. While this is problematic for German /o:/ (Hypothesis 8), German /ɔ/ is produced native-like by the Polish speakers (Hypothesis 7). To test this statistically, separate LMMs were fit to the F1 and F2 data for all vowel categories with LANGUAGE as fixed factor, random intercepts for participants and items, and by-item random slopes for the effect of language, if these improved model fit (this was the case for 9 out of the 12 models). Table 16 Table 16 summarizes the p-values for the effect of LANGUAGE on F1 and F2 for all vowels, which were obtained via likelihood ratio tests.

	/oː/	/ ɔ/	/aː/	/a/	/eː/	/ɛ/
F1	**$p<0.001$ ($\chi^2(1)=29.07$)**	$p=0.75$ ($\chi^2(1)=0.10$)	**$p<0.001$ ($\chi2(1)=11.65$)**	$p=0.53$ ($\chi2(1)=0.40$)	**$p<0.001$ ($\chi2(1)=14.70$)**	$p=0.051$ ($\chi2(1)=3.81$)
F2	**$p<0.001$ ($\chi2(1)=45.85$)**	$p=0.15$ ($\chi2(1)=2.12$)	$p=0.34$ ($\chi2(1)=0.91$)	$p=0.14$ ($\chi2(1)=2.16$)	**$p=0.002$ ($\chi2(1)=9.80$)**	**$p<0.001$ ($\chi2(1)=13.14$)**

Table 16: p-values of likelihood ratio tests for the factor LANGUAGE on F1 and F2 for all vowels of interest (significant results in bold)

While F1 and F2 of the vowel /oː/ are significantly different for the Polish speakers, the values for /ɔ/ do not differ significantly. Hence, the visual impressions find support in the statistical analyses of the formant values.

The Polish and German productions of German /a/ also overlap almost entirely and thus do not differ significantly in either F1 or F2. As far as long /aː/ is concerned, the German native speakers produce a slight quality difference in comparison to short /a/ in that /aː/ lies a bit lower in the acoustic vowel space. Since the Polish speakers produce both German /aː/ and /a/ identical, their productions of the long vowel differ significantly in F1. Still, there is a large overlap between the productions of /aː/ by the German and Polish group, just as there is between the Germans' productions of /aː/ and /a/. It is therefore assumed that the statistical difference does not necessarily result in an auditory difference. This assumption is supported by the fact that in the discrimination task of the study, German native speakers did not hear the quality difference between short and long /a/, even though the F1 values of the stimuli differed too, i.e. by about 80 Hz (see Section 5.5).

German /ɛ/ is produced slightly more open and further back by Polish GLF learners than by German natives. While only the F2 values differ significantly, F1 group differences are approaching significance (p=0.051). It is difficult to say whether the productions of /ɛ/ by the Polish group would still be considered native-like by German native speakers, but the fact that the ellipses largely overlap would support this idea. The relative location of the Polish production furthermore suggests that it is the non-palatal variant which predicts the variance best. This is understandable, as the German test words did not provide palatal contexts.

The most interesting case is /e:/. While the German group produced a clear difference between /e:/ and /ɛ/, the productions by the Polish group span over German /e:/ and /ɛ/. Additionally, the mean value of /e:/ as produced by the learners lies right between the two German vowels. Both F1 and F2 for this vowel category differ significantly between the groups. The large span of the Polish ellipsis may be related to formant movements within this vowel category, which have been postulated earlier (Hypothesis 9).

In order to investigate this further, F1 and F2 measurements at 25% and 75% of each vowel were plotted for /e:/. Figure 18 shows the movements for each item (403 data points), with the beginning of each arrow being the 25%-measurement and the arrowhead the 75%-measurement.

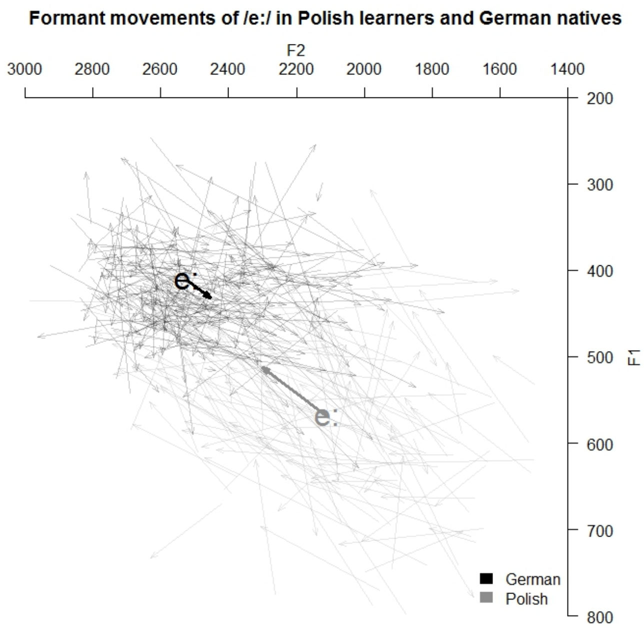


Figure 18: Formant movements for the German vowel /e:/ in Polish GFL learners and German native speakers. Overall group means at 25% and 75% in bold (N(Polish)=14, N(German)=16)

While there is obvious variation within the speakers and items, two important aspects become visible in the above graph. First, Polish speakers show a much greater formant movement within /e:/. Second, the direction of the movement is diametrically opposed to the German speakers' movement. For the Polish speakers, F1 decreases on average by 9.8% and F2 increases by about 9.1%. In the German group, F1 increases by about 5.1% and F2 decreases by 2.7%. The formant changes in F1 are significant in both groups (Polish group: $p<0.001$ ($\chi2(1)=52.16$); German group: $p = 0.002$ ($\chi2(1)=10.06$)). The changes in F2 are only significant in the Polish group ($p<0.001$ ($\chi2(1)=59.88$); German group: $p=0.08$ ($\chi2(1)=3.13$)). In her study of L2 learners' productions of English vowels, Steinlen (2009) had used a benchmark of 10% to characterise significant formant movement, i.e. vowel-inherent spectral change (VISC). With this benchmark, the average German production of /e:/ is far from showing VISC. The average Polish production of German /e:/ however shows F1 and F2 movements that are close to the 10% benchmark (F1: 9.8%; F2: 9.1%). This of course does not mean that every single token produced by a Polish GFL learner shows significant VISC, or that tokens do not exceed a 10% change. Importantly, none of the 75% measurements of /e:/ (the arrowheads) by Polish speakers lie higher in the acoustic vowel space than any German production of /e:/. This implies that the diphthong would more correctly be represented as [εe] rather than [ei], as has been suggested by, for example, Hirschfeld (1998) or Dieling (1992).

6.6.2.3 Discussion

The comparison of the monolingual Polish and German vowel quality data largely confirmed Hentschel's (1986) measurements. However, it proved insightful to investigate the Polish vowel /ε/ separately for palatal and non-palatal contexts. While according to Hentschel German /ε/ and Polish /ε/ are identical in their quality, it seems that according to the new data the quality of Polish /ε/ is more centralized and slightly more open. Because of this, Polish L2 German speakers tend to produce German /ε/ more central as well. Still, German /ε/ produced by Polish and German speakers largely overlaps and it would have to be investigated further as to whether German speakers could hear the slight differences between German /ɛ/ and a more Polish, centralized /ɛ/.

As far as /a:/ is concerned, the significant differences in F1 between the Polish and the German speakers are unlikely to be audible. Polish speakers produce German /a:/ with the same quality as German /a/. Since the female group produced /a:/ significantly lower than the short counterpart (a less clear tendency could be seen in the male data too), Polish speakers produced /a:/ slightly higher than the native speakers. Since it has been found in the discrimination task that German speakers do not differentiate /a/ and /a:/ in their quality, it can be assumed that the lack of differentiation in the Polish group is not problematic. German /a/ and /ɔ/ are produced native-like by Polish speakers as well, which is entirely supported by the statistical analyses. Hypothesis 7, which stated that German /a(:)/, /ɛ/, and /ɔ/ are produced native-like, is therefore considered to be largely corroborated. Only slight precaution is proposed for /ɛ/.

Hypothesis 8 stated that the Polish vowel quality productions of German /o:/ and /e:/ will deviate from those of German native speakers. This is clearly borne out by the current data. The productions of /o:/ and /ɔ/ entirely overlap in Polish speakers, hence the Polish productions of /o:/ differ significantly from those of German native speakers. The monolingual data show that this cannot be explained by the fact that German /o:/ is closest to Polish /ɔ/. In fact, German /o:/ is relatively close to Polish /u/. This acoustic impression is supported by Hentschel's (1986) perceptual assimilation results, where German /o:/ was perceived as Polish /u/ 73% of the time and as Polish /ɔ/ 26% of the time (p. 78). It seems likely that this is because the Polish GFL learners equate the grapheme <o> with their phoneme /ɔ/, as this is the grapheme-to-phoneme correspondence in their native language.

The other problematic case for Polish L2 German learners is the vowel quality of /e:/. Their productions clearly differ from those of German natives, and their mean formant frequencies lie between those of German /ɛ/ and /e:/. Furthermore, the variance spans over the two German vowel categories, which suggested that there is larger variation in formant movements in the Polish productions of /e:/. This was supported by F1 and F2 measurements at 25 and

75% of the vowel. The changes were significantly different from those in the German group and suggested that there is considerable diphthongization of /e:/ for Polish speakers. This corroborates Hypothesis 9. Yet, the latter part of the diphthong did not reach formant values as high as those of /i/, as some learners have suggested.[80] The fact that Polish GFL learners tend to show VISC (only) for German /e:/ is quite a puzzle. As far as perception is concerned, Hentschel (1986) had explained Polish speakers' perception of German /e:/ as [ei] (and /o:/ as [ɔu]) with Donegan's (1978) Natural Phonology account of vowels. According to him, vowels have a natural tendency to be replaced by diphthongs when they are high, tense, and long, which is why listeners may compensate for this in perception. Unfortunately, this account cannot fully explain the production data at hand, as considerable spectral changes were only found for Polish speakers' /e:/ but not for /o:/ [81], even though Hentschel's (1986) data showed that these two vowels were equally often perceived as diphthongs.

An alternative approach could lie in a combination of an orthographic and a perceptual account. As /o:/, /e:/ is represented with a grapheme which maps onto a Polish vowel: the grapheme <e> corresponds to Polish /ɛ/, like <o> corresponds to Polish /ɔ/. While German /o:/ is still somewhat close to Polish /ɔ/ (as mentioned above, it is perceived as /ɔ/ 26% of the time), this is not the case for /e:/. In Hentschel's (1986) perceptual assimilation task, /e:/ was perceived as Polish /i/ 74% of the time and as Polish /ɨ/ 23% of the time, but as /ɛ/ only 3% of the time. It could therefore be assumed that the diphthong reflects a process whereby a learner tries to override an initial orthographic production. While <e> would automatically lead to the pronunciation of Polish /ɛ/, the perceptual input opposes this interpretation. What a Polish speaker perceives when being presented with a German /e:/ is most likely Polish /i/, but hardly Polish /ɛ/. Hence, when a Polish GFL learner acquires the new

[80] It could be argued that this is due to the fact that the point of measurement was at 75% of the vowel and not later. For this reason, those instances were inspected closer which had been specifically labelled during the analysis as sounding like a diphthong to the author. Even in these conspicuous examples, formant values at the very end of the vowel did not reach higher values than those for /e:/ as produced by German natives.

[81] The spectral changes for /o:/ were even smaller than those for /e:/ in the German group. F1 decreased by 1.2% and F2 increased by 3.8%.

German sound /e:/ by both reading *and* listening to it, she may try to correct her initial spelling pronunciation and, in the process, goes from [ɛ] to a more closed variant. While this variant may still not reach the quality of Polish [i], it is significantly higher and more front. The acoustic measurements of the current data suggest that the Polish diphthong would best be represented as [ɛe].

6.7 Synopsis

Five out of seven of the hypotheses formulated for the production study have been borne out. As far as vowel quality production is concerned, all hypotheses were largely corroborated. The monolingual data of the study showed that vowel quality predictions based on Hentschel (1986) were valid with the exception of /ɛ/. In his data it seemed that Polish /ɛ/ is identical to German /ɛ/, but in the exploratory data collected for the present study it was apparent that Polish /ɛ/ by monolingual speakers is more centralized and open. Hence, the productions of German /ɛ/ by Polish learners were also significantly centralized. The question remains whether this difference would be perceived by German native speakers. The fact that the comparison of the exploratory data helped explain findings in the main production experiment underlines the importance of collecting comparable monolingual data, if these do not yet exist.

The quality productions of German /o:/ and /e:/ by Polish GFL learners deviated greatly from those of German native speakers. This was expected, as the closest Polish vowels are very different in their qualities and are perceptually assimilated to two different Polish vowel categories. As concerns /o:/, Polish speakers produce it identically to the qualities of Polish /ɔ/ (and, with that, German /ɔ/ as well). From perceptual assimilation patterns in naïve Polish listeners it would have been expected that /o:/ is more likely to be assimilated to /u/ than /ɔ/. It is suggested that orthography can explain this deviation, as the two languages use the same grapheme <o> for two very different vowel qualities. A similar explanation is proposed for the Polish productions of the vowel /e:/, though slightly more complex.

The data provide the first acoustic evidence for the observation of GFL researchers that /e:/ is often diphthongized by Polish L2 German learners. The reasons for these formant movements are still unknown. While it is possible that /eː/ is *perceived* as a diphthong, this explanation is problematic as the same should be true for /oː/. Both /oː/ and /eː/ are perceptually closest to Polish vowels which are very different from the actual German qualities of the two vowels but orthographically they are represented with graphemes which find equivalents in Polish. Only for /oː/ does the orthographic match partially overlap with Polish perceptual assimilation patterns, i.e. the native G-P correspondence suggests Polish /ɔ/, which is also perceptually close to German /oː/ for Polish listeners. For /eː/, orthographic and perceptual facts contradict each other. The grapheme <e> suggests Polish /ɛ/, but /eː/ is rarely perceptually assimilated to this sound. Thus, the diphthong may be an attempt of the learners to combine orthographic and perceptual interferences.

This diphthongization of /eː/ explains findings for the dimension of vowel length. On average, German long vowels were produced shorter by Polish GFL learners, which corroborates the first hypothesis regarding vowel length productions. However, it was also found that the Polish speakers produce /eː/ relatively longer than all other vowels. This finding is surprising when considering the duration data only, but makes sense when integrating both length and quality results. Diphthongs are naturally longer than monophthongs, hence /eː/, or rather [ɛe], is produced longer. Furthermore, it could be shown that the Polish *short* vowel productions are significantly longer than those of German native speakers. This could be predicted based on the Polish vowel length data collected in the exploratory study.

The two hypotheses which need to be rejected on the basis of the data at hand were concerned with the influence of orthographic marking on vowel length production. Even when the data were subset to those participants who explicitly knew the rules for *lengthening h* and double consonant letters, the duration differences between *marked* and *unmarked* vowels did not turn out to be significant. The effect of orthographic marking might therefore not be very strong for vowel *length* marking. Yet, the vowel *quality* data suggest that orthography might still matter. The effect of orthographic length markings was further investigated in the following identification experiment.

7 Experiment 3: Identification

Most research in L2 phonetics and phonology investigates the perception of segments without real-word lexical context (e.g. Ylinen et al., 2005, Escudero and Wanrooij, 2010, Altmann et al., 2012) or by means of one single minimal pair (e.g. Bohn, 1995, Cebrian, 2006, Rojczyk, 2011). McAllister et al. (2002) are among the few who manipulated a large number of real words and investigated the perception of vowel length by different learner groups (see Section 2.5.1 for more details). While the discrimination experiment in Chapter 5 was designed to tap into a phonetic level of processing (short ISI, nonce words), the identification experiment of this study involved the phonological representation of real words. By conducting two perception experiments which are assumed to cover different levels of processing, both theoretical and methodological questions can be addressed. On the one hand, it can be investigated whether Polish learners of GFL are able to perceive German vowel length and quality differences both on a phonetic and lexical-phonological level. On the other hand, results can be interpreted more reliably because two different experiments were designed to investigate the same phenomenon.

7.1 Hypotheses

Even though the results of the discrimination experiment above have been reported before those of the identification experiment, the design of the experiments was done in parallel. Hence, the results of the discrimination task could not serve in formulating hypotheses for the current experiment. As has been discussed earlier (Sections 2.5 and 5.1), studies by Bohn (1995) and McAllister (2002) had yielded contradicting results as regards the acquisition of L2 vowel length. Furthermore, McAllister et al. (2002) had hypothesized that vowel quality differences accompanying vowel length differences may help learners in perceiving L2 vowels. Because it was of interest to differentiate between vowel length and vowel quality (as in the other two experiments), hypotheses for the identification experiments also addressed these two dimensions. Since reaction time data allow for a more sensitive measurement of relative difficulty, they were collected in the identification experiment as well. It is generally assumed that longer reaction times indicate increased difficulty

(Strange and Shafer, 2008), so predictions about reaction times were added to the accuracy prediction accordingly. Similar to the discrimination task, the following hypothesis is put forward in the context of the identification experiment:

Hypothesis 10: Polish GFL learners will be more accurate and faster at judging items which are manipulated in their quality than in their length.

Furthermore, GFL researchers have postulated that Polish learners of L2 German substitute German long, tense vowels by their short, lax counterparts (Morciniec, 1990, Slembek, 1995, Müller, 2005, see Section 2.1). This would imply that vowels which are wrong in *both* quality and length (long, tense vowels presented as short, lax vowels) might be even harder to identify as incorrect as this is what they might be (erroneously) represented as in Polish GFL learners. Hence, the following hypothesis is put forward as well:

Hypothesis 11: Polish GFL learners will be less accurate and slower at judging items which are incorrect in *both* vowel length and vowel quality than those items which are incorrect in just one dimension.

Furthermore, it was investigated whether explicit orthographic marking of vowel length may help Polish learners in perceiving length-manipulated German long vowels more correctly. Escudero et al. (2010) had found that Spanish learners of Dutch were better at categorizing Dutch vowels when orthographic input was given during the task. The researchers hypothesized that the length information in the Dutch double-vowel grapheme <aa> helped the learners perceive durational differences between Dutch long /a/ and short /ɑ/. However, in a later study, Escudero et al. (2014) investigated the perception of difficult Dutch contrasts in a word-learning experiment and found that this contrast is in fact rather difficult to learn, despite orthographic input. The authors suggested that, in the context of lexical learning, Spanish participants will find it difficult to rely on the newly acquired length contrast; hence, orthography does not help in this case. Because it is unclear whether and how orthography might play a role in Polish GFL learners acquiring German long vowels, the last hypothesis of this study was formulated in relation to the hypotheses of the production part of this study. As it was assumed that orthography helps in vowel length production (Hypothesis 4), it was also hypothesized that orthography will help in perception.

Hypothesis 12: Polish learners will be more accurate and faster in identifying length-manipulated German long vowels as incorrect when they are orthographically marked.

7.2 Participants

The same participants as in the production task took part in the identification experiment. As in the production experiment, one German participant was excluded from the analysis due to dyslexia and two Polish participants due to their insufficient knowledge of the test words. The average age of the 20 analysed German participants (4 males) was 17.9 years (SD=1.1), and the average age of the 18 Polish participants (4 males) was 18.6 years (SD=0.6).

7.3 Experimental design

The task in this perception experiment was to identify auditorily presented stimuli as correct or incorrect instances of a given item, which was presented visually in the form of a picture. The test words were the exact same ones as those encountered during the production experiment (see 6.3). This was convenient for two reasons: Firstly, all participants had been equally familiarized with the picture of a given test word, as in the production experiment each subject encountered and named each picture twice. Secondly, words which were not known to participants during the production task could easily be excluded from the analysis of the identification data.

7.3.1 Stimuli manipulation

The stimuli of the identification task were manipulated similar to the nonce words of the discrimination task. Since it was of interest to disentangle vowel quantity and quality, items had to be manipulated accordingly. For this purpose, a female German native speaker (the author) produced all 24 long vowel items as well as their nonsense counterparts with a short lax vowel, e.g. [n<u>eː</u>bəl] (real word meaning "fog") and [n<u>ɛ</u>bəl] (nonsense counterpart).

Furthermore, the speaker produced all correct and incorrect filler items. All words followed the phrase *Ich sage [...]* ("I say [...]"), so that listeners could adjust to the speaking rate of the model, as Gottfried et al. (1990) had shown that sentential speaking rate plays an important role in the identification of vowels which are differentiated by duration.

All stimuli were recorded three times in randomized order in a sound-attenuated booth at Newcastle University at a sampling rate of 48k Hz (16-bit resolution). The first production of each item was segmented in PRAAT and analysed for vowel duration and quality (F1 and F2), in order to verify that the speaker produced the test items in the desired manner. Long /a:/-vowels (in real words) were on average 2.4 times as long as their short counterparts, long /e:/-vowels were on average 2.1 times as long as their short counterpart, and long /o:/-vowels were on average 2.2 times as long as their short counterpart. These values correspond well with long/short vowel duration proportions collected in other studies (Antoniadis and Strube, 1984, Nimz, 2011a). Because vowel length may be influenced by consonantal context (Chen, 1970), it was decided to lengthen and shorten the respective stimuli according to the length of their corresponding nonsense counterpart. For example, the vowel in the prototypical real word [bo:dən] ("floor") was 185 ms long, while the vowel in its corresponding nonce word [bɔdən] was 89 ms short. The shortened, manipulated item (for the so-called *length* condition) would then be the real word containing the tense vowel [o] shortened to an average of 89 ms. The duration manipulations were, as in the discrimination experiment, achieved via the *Dur* ("duration") function in PRAAT. All real-word/nonce-word ratios were checked for their respective values and whether they were comparable to the overall mean ratios of their given category. All ratios fell within the range of 1.5 standard deviations from the category mean ratio. For some pairs this had not been the case for the tokens from the first production. In these cases, tokens were taken from the second and third production run. None of the ratios between long real word and short nonce word were smaller than 1.6 or bigger than 2.4.

All tokens were extracted including the carrier phrase *Ich sage […]*, and were normalized to an intensity of 65 dB. The manipulated items as well as the unmanipulated nonce words belonged to three experimental conditions.

7.3.2 Experimental conditions

While the prototypical tokens of the real words (e.g. [bo:dən] "floor") had to be judged in the experiment as well, the following three experimental conditions yielded the stimuli of interest:

- Condition "**length**": a long, tense vowel in a real word shortened to the average length of its nonsense counterpart, for example [hox] ("high") as shortened from [ho:x] → *length* is incorrect
- Condition "**quality**": a short, lax vowel in a nonsense counterpart lengthened to the average length of the real word, for example [hɔ:x] as lengthened from [hɔx] → *quality* is incorrect
- Condition "**both**": the unmanipulated nonsense counterpart of a real word [hɔx] → *both* length and quality are incorrect

7.3.3 Procedure

The experiment was run in DMDX, which allowed for the collection of both accuracy data and reliable reaction time data. A game pad was used to collect the participants' responses. The upper right button of the devise was marked with a sign reading "correct", while the upper left button was marked with a sign reading "incorrect". Pictures were presented on an Acer Timeline laptop, while the auditory stimuli were presented over high-quality Sennheiser headphones. Visually and auditory stimuli appeared simultaneously; the picture was visible for 2 seconds and participants could make their judgement until 6 seconds after stimulus onset. After this time-out, the next item was presented.

Each participant went through a short practice trial consisting of 8 items, which represented all possible experimental and filler conditions the participants would encounter throughout the experiment. The practice items had been included in the prior production experiment solely for the purpose of having additional words for the practice phase of the identification task. Participants did not receive any feedback after the practice trial, but had time to ask questions if there were any.

During the main part of the experiment, 192 items had to be judged for their correctness. Half of the items were experimental items (words with long vowels), while the other half were filler items (words with short vowels). The 24 experimental items appeared in four conditions, of which three were the experimental conditions *quality*, *length*, and *both* plus the additional condition *proto*, which was included so that each experimental item appeared once as (definitely) "correct". Half of the 24 filler items were (completely) incorrect utterances of the respective word in that their stressed vowel was replaced by another vowel which was perceptually not close (i.e. [ʃʊtən] for the word <Schatten>; correct: [ʃatən]). The other half of all filler items were correct renditions of the given word. Correct and incorrect versions were not equally divided among the filler items. One quarter of all fillers appeared four times as correct, one quarter appeared once as incorrect and three times as correct, one quarter appeared three times as incorrect and once as correct, and one quarter appeared always as incorrect. This was done so as to not lead participants to develop a false strategy of assuming that each word in the experiment must be judged twice as correct and twice as incorrect. None of the incorrect filler items yielded a possible real word.

The stimuli were presented in four blocks of 48 test items; each word appeared only once in each block in one of the four conditions. The presentation of blocks and order within each block was randomized for each subject. Participants could take breaks after each block, i.e. three breaks overall, if they needed to. On average, the experiment lasted 20 minutes.

7.4 Results

The statistics for the accuracy and reaction time data will be presented separately.

7.4.1 Accuracy data

Of interest were those words which appeared in the three experimental conditions *length*, *quality*, and *both* (24 items x 3 conditions x 41 participants).

The data of three participants were excluded for reasons explained in Section 7.2. Furthermore, the data points of those words and participants were excluded which had not been known during the production task (70 instances x 3 conditions), as well as those tokens which had been written incorrectly (10 x 3 conditions). Lastly, the judgements of the *quality* condition for long /a:/ were excluded, because the results of the discrimination task had shown that all participants (both native Germans and Polish learners) could not hear any quality differences between /a/ and /a:/. Hence, the quality-manipulated words including /a:/ were not comparable to the other test items. In all, 2224 data points were analysed. Figure 19 shows the correct responses to the three experimental conditions in both language groups.

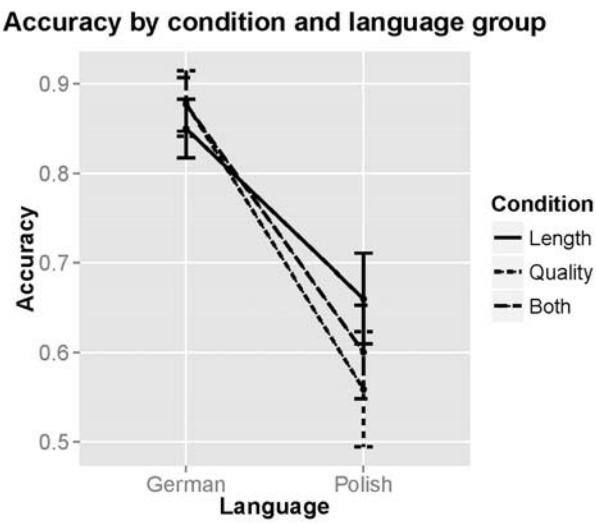


Figure 19: Correct responses to the manipulated items by German and Polish participants (N(Polish)=18, N(German)=20; error bars show 2 SE)

A generalized linear mixed model (GLMM) was fit to the binomial accuracy data which contained LANGUAGE (*German* versus *Polish*) and CONDITION (*length* versus *quality* versus *both*) as fixed factors and random intercepts for participants and items. Three control factors were added to the model in order to examine whether the inclusion of these additional fixed effects improved model fit. First, it was examined whether it was justified not to include VOWEL as a separate factor. By means of likelihood ratio tests (Cunnings, 2012, Winter, 2013) it was found that VOWEL did not improve model fit, either as a main effect ($\chi^2(2)=0.22$, $p=0.90$) or in interaction with the two main experimental factors ($\chi^2(8)=6.34$, $p=0.61$). Second, PRESENTATION ORDER was examined. This control factor was not significant either (as main effect: $\chi^2(1)=0.01$, $p=0.90$; in interaction: $\chi^2(5)=2.21$, $p=0.82$). Finally, it was also examined whether the LEXICAL STATUS of manipulated words in the *both* condition may have had an effect on the participants' responses. For instance, the word *Boden* [bo:dən] ("floor") became the (real) word [bɔdən] ("shallow bay"), or the word *Wohnen* [vo:nən] ("to live") became [vɔnən] ("delights"). Even though all words were presented with their respective pictures (which could not be mistaken for the other meaning) it was still double-checked whether the lexical status of some words could influence the accuracy results. As with the other two control factors, this was not the case for LEXICAL STATUS, either as main effect ($\chi^2(1)=1.65$, $p=0.20$), or in interaction with the main experimental variables ($\chi^2(5)=8.72$, $p=0.12$).

From Figure 19 it already becomes apparent that German speakers perform significantly better than Polish speakers. The overall accuracy mean of German speakers is 88.2% ± 1.0% (SE), while Polish GFL learners are on average only 63.2% ± 12.8% correct. This difference is significant at $p<0.001$. Since it was of interest to compare the three conditions *within* each group, custom comparisons were conducted by the final model[82]. These revealed that there were no significant differences between the conditions in the German group, but there were in the Polish learner group. In order to test Hypothesis 10, the model compared the accuracy performance in the *length* condition versus the *quality* condition for both groups. While the differences in accuracy per condition were not significant in the German group (85.0% ± 1.6% correct in the *length* condition versus 87.8% ± 1.8% correct in the *quality* condition), Polish speakers performed significantly better in the *length* condition ($p=0.01$). In 66.0% ± 2.5% of the time they correctly judged items which were too short as "incorrect",

[82] *Model.identification = glmer(Correct ~ Combined + (1|ID) + (1|Word), data=identification, family="binomial")* (COMBINED is the combination of the fixed factors LANGUAGE (2 levels) and CONDITION (3 levels) into a new factor with 6 levels, e.g. level *a* (Polish-quality), level *b* (Polish-length), etc. This re-coding was necessary in order to conduct custom comparisons within the groups of interest, for example Polish-quality versus Polish-length.)

while they judged items which were too lax correctly only 55.9% ± 3.2% of the time. Table 17 summarizes the detailed statistics of the model.

	Estimate	Std. error	z-value	p-value	Significance
(Intercept)	1.16	0.05	21.41	<0.001	***
German vs. Polish	1.46	0.11	13.39	<0.001	***
G-quality vs. G-length	0.24	0.21	1.13	0.26	n.s.
G-length vs. G-both	-0.23	0.19	-1.22	0.22	n.s.
P-quality vs. P-length	-0.43	0.17	-2.48	0.01	*
P-length vs. P-both	0.26	0.16	1.64	0.10	n.s.

Table 17: Summary of fixed effects of *Model.identification* (N(observations)=2224; groups: 24 words and 38 participants)

It was further hypothesized that Polish learners would perform worst in the *both* condition. Hence, the model also tested whether the difference between the *length* and the *both* condition was significant within the two experimental groups. As can be seen in Table 17, this was not the case for either group (Polish: p=0.10, German: p=0.22).

It was then investigated whether the orthographic marking of vowel length may have helped the learners in judging the items in the *length* condition more correctly (Hypothesis 12). From Figure 20 it already becomes obvious that this was clearly not the case.

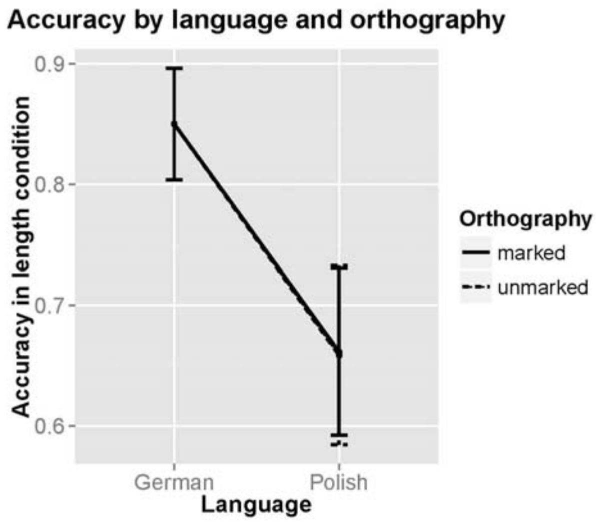


Figure 20: Correct responses in the *length* condition for orthographically marked and unmarked items (N(Polish)=18, N(German)=20; error bars show 2 SE)

Neither for the Polish nor for the German speakers did it matter whether long vowels were written with *lengthening h* or not. An LMM was fit to the data of the length condition (833 data points) with LANGUAGE, ORTHOGRAPHIC MARKING and their interaction as fixed factors and random intercepts for participants and items[83]. It revealed that orthographic marking clearly had no effect, either as main effect ($\chi^2(1)<0.001$, p=0.98) or in interaction with language ($\chi^2(1)<0.001$, p=0.98). The fact that the lines in Figure 20 almost completely overlap makes this statistical investigation rather needless.

7.4.1.1 Summary

As far as accuracy is concerned, none of the hypotheses could be corroborated. However, other interesting effects were found. It was assumed that Polish listeners would be better at identifying items which were incorrect in their quality than in their length (Hypothesis 10). This was not the case, and in fact Polish GFL learners were significantly better at identifying those items as incorrect which were incorrect in their *length*. This finding is surprising, as the discrimination task had shown that Polish speakers were significantly worse at perceiving differences in vowel length rather than vowel quality. A possible

[83] *Model.accuracy.orthography = glmer(Correct ~ Language*Orthography + (1|ID) + (1|Word), data=identification.orthography, family="binomial")*

explanation could have been the orthographic marking of half of the test words. Hypothesis 12 stated that Polish speakers may be better at identifying length differences in words which are orthographically marked for their length via the *lengthening h*. This was clearly not the case, so this hypothesis must be rejected on the basis of the accuracy data. Hypothesis 11 must be rejected as well. The Polish GFL learners were not worse (or better) at identifying items which were incorrect in *both* vowel length and vowel quality. The analysis of the reaction times was thought to allow for a more fine-grained differentiation between the conditions.

7.4.2 Reaction times

Those data points were excluded from the analysis of the reaction times (RTs) for which participants responded faster than 200 ms and slower than 2000 ms. While the lower limit has been suggested by Baayen (2008), the upper limit is more generous, so as to not undergo unnecessary data loss.[84] Darcy et al. (2012) too had used a similar upper limit (2200 ms) in their study with comparable L2 learners. While they had measured latencies from the onset of the word, it was decided to measure reaction times from the offset of each vowel within the test words. This way, items containing shortened vowels did not automatically (and falsely) measure shorter reaction times.

Since the raw RT data were not normally distributed (as is usually the case for this type of data), statistics were performed on the logarithmic (log) reaction times. Furthermore, only those RT data were analysed which the participants had responded to correctly (overall about 77% of the data). Lastly, the data of the quality condition for all words including /a:/ were excluded for the same reasons as in the accuracy analysis. In all, 1639 data points served the analysis of the RT data. Figure 21 shows the log-transformed RTs of both language groups to words containing long vowels which were manipulated in their length, their quality, or both.

[84] Since the task is more difficult for L2 leaners than for native speakers (as was evident in the accuracy data), it can be assumed that learners take on average more time than native speakers.

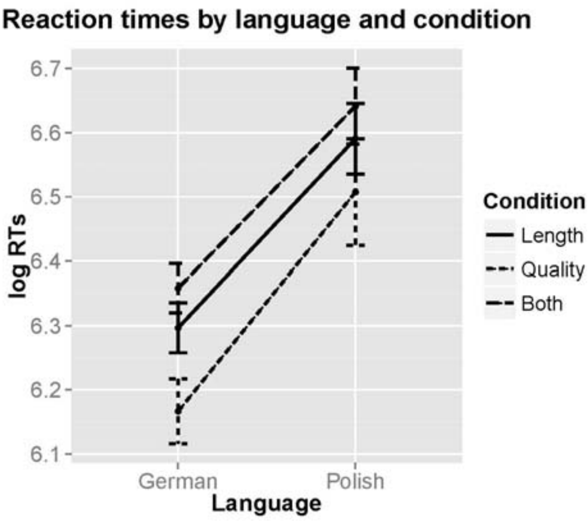


Figure 21: Log-transformed reaction times of Polish and German speakers for correctly identified items (N(Polish)=18, N(German)=20; error bars show 2 SE)

As was already evident from the accuracy data, Polish speakers have more difficulty identifying items correctly, as shown by generally longer reaction times. Interestingly, the pattern for length-manipulated and quality-manipulated items seems to be reversed: participants (both German and Polish) take longer when judging vowels which are too short than vowel which are too lax. This is in line with Hypothesis 10. Furthermore, the condition in which *both* vowel length and vowel quality are incorrect seems to be the most difficult – for both groups (Hypothesis 11). In order to investigate these visual impressions statistically, an LMM was fit to the log RT data with LANGUAGE, CONDITION, and their interaction as fixed factors and random intercepts for participants and items. By-item random slopes for the effect of condition were added as well, as these improved model fit significantly ($\chi^2(5)=19.07$, $p=0.002$)[85]. As in the accuracy model, it was checked whether the inclusion of any of the three control variables PRESENTATION ORDER, LEXICAL STATUS of words in the *both* condition or VOWEL type would improve the model. As in the previous model, this was not the case. Table 18 shows the summary of the final RT model.

[85] *Model.RT = glmer(logRT ~ Language*Condition + (1|ID) + (1+Conodition|Word), data=data.RT, REML=FALSE)*

	Estimate	Std. error	t-value
(Intercept)	6.31	0.05	123.18
Language (Polish)	**0.29**	**0.07**	**3.98**
Condition (Quality)	**-0.15**	**0.04**	**-4.05**
Condition (Both)	0.04	0.03	1.22
Lang. (Pol.):Cond. (Quality)	0.06	0.05	1.31
Lang. (Pol.):Cond. (Both)	0.02	0.04	0.46

Table 18: Summary of fixed effects of *Model.RT* (significant differences in bold; N(observations)=1639; groups: 24 words and 38 participants)

As concerns the RT data, there are significant main effects of LANGUAGE ($\chi^2(1)=16.23$, $p<0.001$) and CONDITION ($\chi^2(2)=23.41$, $p<0.001$), but no significant interaction between the two factors ($\chi^2(2)=1.7$, $p=0.42$). The significant effect of CONDITION is driven by the difference between the *length* and the *quality* condition in that both groups take longer to judge length-manipulated items than quality-manipulated items. This also means that items which are manipulated in *both* dimensions are more difficult for participants than quality-manipulated items, as participants took longest to judge items in the *both* condition. The difference between the *length* and the *both* condition is not significant, as can be seen in the summary of the table above.

Even though Polish participants (like German natives) take longest in the *both* and *quality* conditions – which goes in hand with Hypotheses 10 and 11 – this result needs to be viewed with caution. As could be seen in the analysis of the accuracy data, the means of the Polish speakers in all conditions are below 70% accuracy. In their RT analysis of L1 and L2 speakers in a lexical decision task, White et al. (2010) excluded all participants who performed with overall accuracies below 70%. Darcy et al. (2012) used an even higher margin of 75%. Even if mean accuracies were only computed over the *length* and the *both* condition (highest accuracies), none of the Polish participants would reach a mean accuracy higher than 75%. If 70% was taken as the cut-off point, only two Polish participants would be included. As far as the German participants are concerned, all participants were accurate more than 75% of the time. For this reason, the RT results for the German group may be considered more meaningful than those for the Polish group.

It was further investigated whether the orthographic marking of the test words had an effect on the reaction times in the *length* condition (Hypothesis 12). From Figure 22 it becomes apparent that orthographic marking does not have a significant effect, even though it seems that native speakers may be slightly faster in judging marked items correctly.

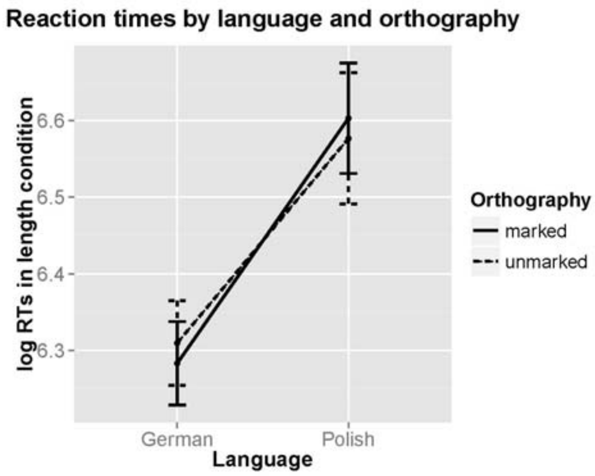


Figure 22: Log-transformed reaction times in the length condition by language and orthographic marking (N(Polish)=18, N(German)=20; error bars show 2 SE)

An LMM (for 639 data points) with LANGUAGE, ORTHOGRAPHY, and their interaction as fixed effects and participants and items as random intercepts[86] revealed that orthographic marking did not have a significant effect on the log RTs. Neither as a main effect ($\chi^2(1)=0.05$, $p=0.83$), nor in interaction with LANGUAGE ($\chi^2(1)=0.65$, $p=0.42$) did this factor turn out to be significant. With that, the RT data corroborate the accuracy results.

7.4.2.1 Summary

In contrast to the accuracy data, the reaction times seemed to corroborate Hypothesis 10, in that Polish speakers were faster in judging items in the *quality* than in the *length* condition. However, it has to be kept in mind that

---

[86] *Model.RT.orthography = glmer(logRT ~ Language*Orthography + (1|ID) + (1|Word), data=data.RT.orthography, REML=FALSE)*

the Polish participants were on average only 56% correct in the *quality* condition, which suggests that they performed only slightly above chance. Other researchers exclude participants who show an accuracy of below 70%-75% (e.g., White et al., 2010, Darcy et al., 2012). All German participants had an overall accuracy above 75%, which is why the reaction time data for this group are more valid. The German native speakers showed longer RTs in the *both* and the *length* condition, while they were significantly faster in the quality condition. The main effect of CONDITION was driven by a significant difference between the *length* and the *quality* condition; the difference between the *length* and the *both* condition did not reach significance. Even though this result is considered uninterpretable for the Polish group, the differences that emerged in the German group are an interesting finding and will be discussed below. Finally, the RT data, too, suggest rejecting the hypothesis that orthographic marking could help participants in judging the length-manipulated items.

7.5 Discussion

The results of the identification experiment showed an interesting, unexpected effect of manipulation-type in the Polish GFL learners. The accuracy data revealed that length-manipulated items could be judged significantly more correct than quality-manipulated items. This runs counter to Hypothesis 10, which stated that quality-manipulated items would be judged more correctly by the Polish group. Hypothesis 12 put forward the idea that orthographic marking could positively influence the identification of items which are too short. This could have explained why Polish speakers might be more correct at identifying the length-manipulated items. This hypothesis, however, could not be corroborated by the present data.

An alternative way of explanation may lie in the Polish learners' metalinguistic knowledge. In a questionnaire, which was administered after the main experiments, all participants affirmed that they were aware of the existence of long and short vowels in German, but none of them knew that vowels differ in their quality as well.[87] Darcy et al. (2013) too had briefly addressed the possibility that explicit instruction may explain findings of dissociations between perception abilities and lexical representations. As had been laid out in Section 2.6, a number of studies found asymmetric effects in their studies with L2 learners. They suggested that metalinguistic information, in the form of orthographic knowledge or otherwise, may be the cause for these surprising effects (e.g. Weber and Cutler, 2004, Hayes-Harb and Masuda, 2008, or Darcy et al., 2013). The current findings rule out the idea that orthographic knowledge may be the cause of these surprising effects and suggest that metalinguistic awareness – as mediated by explicit instruction – may present an alternative interpretation. It needs to be qualified though that, if metalinguistic knowledge does help, it is still not the case that accuracy for length-manipulated items is very high in the present data, i.e. 66.0% accuracy for length-manipulated items versus 55.9% for quality-manipulated items. Still, the difference is significant.

Since the average accuracy for each Polish participant never exceeded 75%, it was decided to not interpret the reaction time data of the Polish group any further (even though they pointed in the predicted directions). The fact that participants, in some conditions, often performed only slightly above chance level suggested that reaction times may not be very insightful.[88] This is different for the German group. For them, the effect of different manipulation-types can be interpreted as follows. The RT data showed that the *quality* condition was responded to fastest by the German natives. Reaction times were longest in the *both* condition, while reaction times for the *length* condition were in between. Items which are incorrect in the *both* condition may be the most difficult (longest reaction times), because vowels that are manipulated in both length and quality yield vowels which are in fact present in the phonology of the native speaker (e.g. /eː/ becomes /ɛ/). It may therefore be harder to reject words

---

[87] One participant had written that the short vowels are "stronger" and the long vowels are "softer". While this may be considered at least partial knowledge of different vowel qualities, it should still be considered unhelpful. If these terms were used at all, it would probably be the tense vowels that should be termed "strong" and the lax vowels that should be termed "soft".

[88] This is because those conditions which are very hard might be judged faster: if a participant just guesses, this guessing could arguably be performed quicker than a genuine judgement.

which contain a vowel that is – by itself – not incorrect, even though the word as a whole needs to be rejected. In a similar vein, it was expected that Polish learners would find it more difficult to identify the *both* condition as incorrect (Hypothesis 11), because their phonological lexical representations might in fact resemble items in this condition. This hypothesis was, however, not corroborated.

Lastly, the reaction time data showed that the *quality* condition is significantly more difficult to reject for the German participants than the *length* condition. This finding can be related well to the results of the discrimination task. In Chapter 5 it was found that German speakers are significantly worse at perceiving length differences than quality difference in the mid vowels. Hence, German natives may take longer to reject real-word items which are manipulated in the length condition. The relationships between the results of the three main experiments will be discussed further in the final chapter of this study.

8 Triangulation: Discrimination, production, identification

The present study drew on different areas of research, most prominently on research in the field of German as a Foreign Language and experimental studies in L2 phonetics and phonology. It set out to experimentally test hypotheses which were partially motivated by observations made by GFL researchers, but also by unanswered research questions in the field of L2 perception. This interdisciplinary approach led to a fruitful area of investigation, which is summarized graphically in Figure 23.

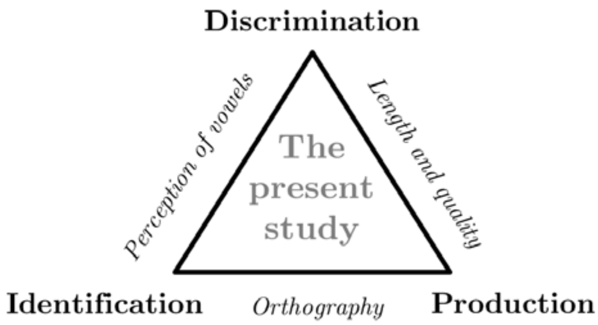


Figure 23: Graphical summary of the present study

On the one hand, the diagram highlights the three main experiments conducted for this study: a discrimination, an identification, and a production study. On the other hand, it shows that these approaches are related and, together, address topics which were found to be understudied and presented exciting areas for further research, such as the role of orthography or the perception of vowel length. Importantly, the present study explored these topics with a combination of languages which has not yet received much attention in the field of L2 speech learning: German and Polish. Most experimental approaches have studied English as a native or second language, while studies in the field of GFL often lack empirical evidence. This study set out to fill the gaps on both sides. In the following, the overall findings of this study are summarized and the results of the three main experiments are discussed in relation to each other. Furthermore, the relevance of the findings for the foreign language classroom is addressed. The dissertation ends with an overall conclusion.

8.1 Overall summary and discussion

Before the conduction of the three main experiments, important exploratory data on the Polish vowels were collected in the context of a monolingual pilot study (Section 4.2.4). On the one hand, the Polish vowel length data allowed for the formulation of more directed hypotheses for the main experimental part of the study. On the other hand, the vowel quality data of an age-matched control group made possible a more valid interpretation of the vowel production data collected in the main production experiment. Not only did the pilot study explore average Polish vowel length and vowel qualities, it also investigated Polish vowel length as a function of the following consonant type (singleton versus geminate). It was found that, unlike in Italian for example, vowel length does not differ systematically before singleton and geminate consonants. This was an important prerequisite, as other studies investigating Italian speakers (e.g. Altmann et al., 2012) could not rule out the possibility of an influence of allophonic experience with vowel length. Similarly, Nimz (2011b) could not give a definite explanation for her discrimination results with Turkish L2 German learners. She found that Turkish learners were equally good as German native speakers at perceiving length differences in German vowels. It was not clear, however, whether this was due to the possibility that vowel length differences are generally easily perceived by L2 learners (Bohn, 1995), or because Turkish speakers have experience with allophonic vowel length in their L1 (McAllister et al., 2002).

The discrimination task of the study set out to shed light on this question, and provide data on the phonetic perception abilities of Polish L2 German learners. The two hypotheses formulated in the context of this discrimination task were borne out. First, Polish speakers were significantly less accurate than German native speakers in discerning pure length differences in German vowels. Second, German native speakers and Polish GFL learners differentiated pure quality differences between the vowels equally well. This provides evidence against Bohn's (1995) Desensitization Hypothesis as well as Escudero's (2005) L2 Linguistic Perception Model (see Sections 2.5.2 and 2.4, respectively). Even in a discrimination task with a low memory load and a high stimulus certainty, Polish speakers have difficulties discerning length differences native-like. It may be possible that Escudero's (2005) predictions do not apply to this specific group of L2 learners who predominantly learned German in the foreign language classroom. This learning context is considered fairly impoverished as far as L1 input is concerned (Best and Tyler, 2007). However, the learners of this study did receive native-speaker input in the classroom as well, even though not all teachers were L1 German teachers. Hence, the results of the discrimination task appear to support assumptions made by McAllister et al. (2002), who hypothesize that features, such as duration, not present in the L1 phonology will be difficult to perceive for L2 learners. Along with Flege (1995), they further propose that this perceptual problem will be reflected in the productions of the learners. The current study also tested this hypothesis.

The acoustic analysis of over 3500 vowels revealed that Polish GFL learners produce shorter German long vowels than native speakers do. On average, their productions of German long vowels were 32 ms shorter than the same vowels produced by native speakers, whose vowels were – on average – 167 ms long. Consequently, the productions of Polish speakers are about 20% shorter, which is a substantial difference considering that, for example, Dutch speakers are able to notice vowel duration changes of only 6% in a forced-choice identification task (Nooteboom and Doodeman, 1980).

Furthermore, the production study revealed that Polish GFL learners produce German short vowels significantly *longer*, namely by about 16% (average in Polish group: 109.3; average in German group: 94.6 ms). This result was predicted based on the Polish vowel length data collected in the exploratory study. Still, German short and long vowels produced by Polish GFL learners differ from each other, and Polish speakers produce the long vowels significantly longer than the short vowels. This implies that the Polish GFL leaners do differentiate between short and long vowels, the question, however, remains whether this difference is enough to be noticed by German native speakers. The results of the identification experiment suggest that German speakers are sensitive to incorrect vowel durations, even if they show a certain "tolerance" in their judgements.

The identification experiment explored the perception of real German words, which were manipulated in their vowel length and in their quality. German speakers judged 85.0% of the length-manipulated items and 87.8% of the quality-manipulated items as incorrect. The length-manipulation entailed the shortening of the long vowels by about half their length, which is still considerably more than the average shortening in Polish speakers' productions. For the low-vowel contrast /a/-/aː/, Tomaschek et al. (2011) had found that, for German listeners, there is a sharp categorical boundary at about 106 ms (in the disyllabic nonce word /tatə/). From this, it can be inferred that the shortening of the long vowels by about 20% could already be crucial, depending on the number of syllables and other modulating factors such as vowel type and context. In order to be able to make a definite claim about the perceptual consequences of the Polish shortening of long vowels (as well as the lengthening of Polish short vowels), it would be necessary to conduct further experiments with German native speakers.

The production data were collected by means of a picture-naming task. Crucially, no orthographic input was given during elicitation. This was important because three out of the twelve hypotheses tested in this study addressed the role of orthographic marking in vowel length production and perception. Further, it was of interest to investigate the influence of orthography on phonological *representations* rather than its effect in the process of reading; hence, orthographic input was avoided. For the long vowels, it was hypothesized that orthographic marking by means of *lengthening h* would help Polish GFL leaners produce German vowel length more correctly. For the short vowels, it was proposed that the marking of short vowels through double consonant letters would influence the learners positively as well. However, state-of-the-art statistical analyses by means of linear mixed models did not show any significant effects of orthographic marking, either for the long vowels or for the short vowels. Even when the group of Polish participants was subset to the most motivated learners or those who could explicitly name the orthographic rules in a post-test, orthographic marking remained a non-significant factor. However, other (control) factors were found to influence vowel length substantially. For example, the number of syllables in a test word demonstrated an interesting interaction with the language groups: German learners shortened German long vowels in two-syllable words significantly more than Polish GFL learners. This relates well to the classification of the two languages as stress-timed (German) and syllable-timed (Polish) (Sections 4.1.3 and 4.2.3, respectively). Furthermore, an unexpected interaction of voicing of the following consonant with the speakers' native language was found. Long vowels in Polish speakers were significantly lengthened before voiced consonants, while vowels in German speakers were not. This was only relevant for intervocalic consonants, because both Polish and German exhibit final-devoicing. The finding that orthographic marking did not affect the phonological representations of the Polish participants of the study was corroborated by the results of the identification task. Polish (and German) listeners were not significantly better at identifying a length-manipulated item as incorrect when it was marked in its orthography by means of *lengthening h*.

The statistical analysis of the long vowel productions further discovered an interesting effect of vowel type in that German /eː/ was produced relatively longer by the Polish speakers in comparison to all other vowels. This finding fits in well with the analysis of the vowel *quality* productions of the participants, which were explored in this study as well. As regards /eː/, it was found that the productions of Polish speakers show conspicuous formant movements which are not present in the German productions. On average, /eː/ is produced by the Polish participants as a vowel which moves from the quality of /ɛ/ towards /e/, i.e. [ɛe]. The data provide the first acoustic evidence for the observation made by GFL researchers that Polish L2 German learners tend to diphthongize this vowel. Interestingly, this is only the case for /eː/ but not for /oː/. The reasons for this idiosyncratic effect are unknown, but it is proposed that this specificity is due to interferences from both perception and orthography. For example, Rafat (2015), in her investigation of English speakers' productions of Spanish rhotics, had found that orthographic and acoustic input *together* modulate L2 productions. She found that rhotics with a higher degree of assibilation in the input triggered more assibilated rhotics in the productions of those participants who had received orthographic input as well (Section 3.1). While the design and learning context of the current study is different from Rafat (2015), it is still hypothesized that both acoustic and orthographic input lead to the diphthongization of /eː/ in Polish learners. This is because the phoneme is represented by the grapheme <e>, which, in the Polish writing system, corresponds to /ɛ/. Acoustically, however, this vowel is closer to Polish /i/. It may be the case that Polish GFL learners incorporate the orthographic and perceptual interferences by starting out with the Polish spelling production of <e>, but satisfy their auditory input by moving towards the quality of a higher vowel (even if they do not reach the quality of /i/).

As for German /o:/, Polish GFL learners produce this vowel identically in its quality to the Polish vowel /ɔ/ (and, with that, to German /ɔ/). Both Polish /ɔ/ and German /o:/ are represented in the respective orthographies as <o>. As far as the perception of /o:/ by Polish speakers is concerned, previous research has shown that this vowel is assimilated to Polish /u/ rather than /ɔ/. It is proposed that orthography effectively modulates the production of /o:/ in that Polish GFL learners apply their native grapheme-to-phoneme correspondence to their L2. This is different from the production of /e:/, as Polish /ɛ/ is perceptually too different for a pure spelling pronunciation to occur. The acoustic measurements of the exploratory part of the study support this claim.

Based on earlier acoustic comparisons of the German and Polish vowels (Hentschel, 1986), it was predicted that the German vowel qualities of /a/, /ɛ/, /ɔ/, and /a:/ would be produced native-like by Polish GFL learners. This hypothesis could largely be corroborated. Only the productions of /ɛ/ were slightly more centralized by the Polish group. It is possible that this significant difference in the acoustic data may not be audible to German native speakers, but this would have to be investigated further in future studies. This slight centralization can be explained well with reference to the exploratory monolingual data. In the pilot study, /ɛ/ was analysed in both palatal and non-palatal context. While previous studies did not seem to take this context into account, it could be shown by the present study that Polish /ɛ/ in non-palatal contexts is indeed more centralized than German /ɛ/. The fact that careful acoustic comparisons of monolingual data can help predict and explain bilingual data supports findings of other studies which have conducted similar acoustic comparisons (Flege et al., 1994, Steinlen, 2009, Escudero et al., 2012). The monolingual data further support the results of the discrimination task, i.e. the result that the vowel qualities of the German pairs /o(:)/-/ɔ/ and /e(:)/-/ɛ/ can be differentiated well. The German vowels are acoustically closer to different Polish vowels and so are assimilated to different Polish categories. As a consequence, these vowels could be differentiated native-like by Polish GFL learners, which goes in hand with Best and Tyler's (2007) predictions for the so-called Two Category cases (Section 2.3). In this sense, the predictions of the PAM-L2 also hold true for *foreign* language learners, even though the authors stress that the model was originally designed for *second* language learners.

The results of the discrimination and identification experiments show an interesting dissociation as far as the perception of vowel length is concerned. While the discrimination task showed that Polish GFL learners have difficulties perceiving length differences native-like, the task involving *real* German words provided evidence that length-manipulated items are more correctly identified than quality-manipulated items. This unexpected result mirrors findings by other researchers who investigated both phonetic perception and lexical representations in L2 learners (Section 2.6). For example, Darcy et al. (2012) had found that despite considerable categorization errors, English learners of L2 French were able to lexically encode the difficult /œ/-/ɔ/ contrast. A possible explanation for these unexpected findings may be the influence of orthography – in the case of German, for example, the signalling of vowel length through *lengthening h*. However, it was found that orthographic marking did not have a significant effect on the Polish GFL learners' identification performance. This was not expected, as this specific group of learners (i.e. *foreign* language learners) had extensive exposure to written forms in the foreign language classroom. Furthermore, the learners' L1 orthography is a very transparent writing system, which would make it even more likely for them to rely on information in the spelling (Simon et al., 2010). A possible explanation for this unexpected finding may lie in the first L2 of the learners, i.e. English. It is conceivable that the experience with an opaque L2 might have made the Polish learner less prone to rely on orthographic markings, even in a more transparent language like German. Lastly, research in aphasia has shown that phonological and orthographic word forms can be activated autonomously: in an impaired patient, *written* identification of a word does not automatically entail the same *oral* identification of a word and vice versa (Miceli et al., 1997). From a neuropsychological perspective, this implies that phonological and orthographic components in lexical activation are not as strongly connected as is often assumed. For L2 research, this could imply that the influence of the L2 orthographic system in L2 phonological acquisition is limited. Even though neuropsychology and L2 speech research naturally focus on different aspects of lexical representation, the parallel is still a noteworthy observation.

By means of a post-test, it was found that all learners were aware of the long-short contrast in German vowels; however, none of them was introduced to the fact that German vowels differ in vowel quality as well. It is therefore assumed that metalinguistic awareness, as mediated by explicit instruction, can help learners to integrate difficult contrasts into their phonological representations at the lexical level. It seems clear that this hypothesis deserves further investigation, as other researchers have also hinted at the influence of explicit instruction (e.g. Cebrian, 2006). The issue of explicit (versus implicit) instruction is also of great interest to the broader field of SLA, as is evident in a recent special issue on this topic (Andringa and Rebuschat, 2015).

8.2 Relevance for the foreign language classroom

The primary focus of this study was to investigate whether the explicit marking of vowel length in German may help learners of L2 German in perceiving and producing German vowels – and, with that, *words* – more correctly. As had been laid out in Section 3.3, a number of GFL researchers have suggested that the *lengthening h* and double consonant letters may help learners to develop a more native-like pronunciation, but this hypothesis has – until now – not been experimentally tested. Contrary to expectations, evidence for the positive influence of these length markers could not be found. The fact that the learners investigated in this study were foreign language learners, i.e. learners who have extensive written input, seems to make this finding even more definite. In a post-test addressing the participants' explicit knowledge of these length markings, it was found that only 7 out of 20 were aware of the meaning of *lengthening h* and only 5 out of 20 were aware of the meaning of double consonant letters. This suggests that the instruction of orthographic rules had not been focused on. Furthermore, through a general questionnaire about the learners' experience with specific pronunciation training, it became clear that explicit phonetic instructions had not been very extensive either. In an early experimental study on L2 speech by Spanish and Laotian GFL learners, Dieling (1983: 182) had commented on *lengthening h* and her impression that the learners were simply ignoring the information it carries. While she did not specifically investigate orthographic length markings, she still recommended a stronger "cognitive" focus in pronunciation teaching. This means that the instruction of phonological and orthographic structures and rules (i.e. metalinguistic knowledge) needs to be promoted, a position that other GFL researchers in the field of pronunciation teaching advocate as well (Hirschfeld, 2003, Settinieri, 2010). Since there is evidence that the participants of this study did not receive extensive phonetic instruction, it is still possible that orthographic length marking could prove helpful in a more cognitive teaching environment where the metalinguistic awareness of the students is promoted.

The fact that some learners were better than others at discerning length differences in the discrimination task did not imply that these learners performed better in the *length* condition of the identification task ($r=-0.04$, $p=0.79$), or that they produced long vowels more correctly, i.e. longer ($r=-0.05$, $p=0.73$).[89] This suggests that mere perceptual training of minimal pairs might not be enough to improve learners' L2 speech. On the other hand, Dieling (1992: 28-29) points out that minimal pairs are important because they draw the learners' attention to the problem, for example, the fact that a difference in vowel length can change the meaning of a word. This, in turn, is the starting point for metalinguistic awareness, which might prove a crucial variable in successful L2 phonological acquisition.

[89] It needs to be noted that the measurements available from the production experiment (i.e. vowel duration in ms) might not be the most adequate unit to correlate with the length-discrimination data. For example, Flege et al. (1999) or Richter (2007) had measured L2 production performance by means of native speaker ratings. As will be pointed out in the conclusion, it would be highly interesting to complement the current findings with German native speaker ratings of naturally produced Polish L2 German vowels in future studies.

However, vowel length is not the only feature that is problematic for Polish GFL learners. The German vowel pairs /eː/-/ɛ/ and /oː/-/ɔ/ (and also other German pairs not investigated in this study) differ in vowel quality as well. Even though the vowel quality differences in these vowels are discriminated native-like by Polish L2 German learners, it does not mean that they are produced native-like. The Polish vowel quality productions of /oː/ completely overlap with the qualities of (German or Polish) /ɔ/. It is proposed in this study that this is partially due to interference from the Polish native orthographic system. A similar explanation is given for /eː/, even though its diphthongization in Polish productions hints at a more complicated interplay between orthography and perception. The simple instruction that learners need to be aware that their L1 G-P correspondences cannot simply be applied to the L2 could help avoid considerable pronunciation problems, as easy as this clue may seem to be.

The unexpected findings of the identification task supported the idea that metalinguistic knowledge can help learners in representing L2 features that are particularly difficult to perceive. All of the Polish participants were aware that German vowels may differ in their length, yet none of them was aware of crucial quality differences between the vowels. The results of the identification experiment can be interpreted with reference to the phonetic instruction the learners had received, which apparently was limited to the length dimension. This draws attention to the need for more informed pronunciation teaching in the foreign language classroom, possibly with reference to tongue positions and other articulatory movements during production. Hirschfeld (2001: 874) points out though that often teachers themselves are not adequately educated in the field of phonetics. This is unfortunate, as previous research has shown that phonetic instruction improves the comprehensibility and accentedness of L2 learners (Derwing et al., 1998), and may possibly be a crucial factor in attaining native-like speech (Bongaerts et al., 1997).

8.3 Conclusion

In all, the study tested twelve hypotheses, of which seven were corroborated by the present findings. The predictions made were motivated both by open questions in the field of L2 phonetics and phonology and by observations made by GFL researchers with regards to Polish L2 German learners. The fact that not all hypotheses were borne out highlights the importance of collecting empirical evidence for theoretically-motivated hypotheses. This is particularly crucial in an applied field of research such as German as a Foreign Language, where it is the ultimate goal to improve foreign language learning and teaching.

The role of orthography in the acquisition of an L2 phonology has received considerable attention over the past decade. Findings so far have not been straightforward in that some studies report a beneficial effect of orthographic input, while others report a negative or non-existent influence of the writing system. Similarly, in this study, results provide more than one answer to the question: Does orthography matter? On the one hand, the investigation of the influence of vowel length markers in German suggests that GFL learners do not make use of length information provided by the L2 orthographic system. This might be surprising in view of the fact that the learners of this study had extensive exposure to written forms of German. It could be that the role of the *L2* orthographic system in L2 phonological acquisition has been overestimated. On the other hand, the data at hand give reason to assume that the *L1* orthographic system plays a crucial role in the acquisition of a foreign language. As was evident in the productions of the German vowels /e:/ and /o:/, Polish learners seem to be negatively influenced by their native grapheme-to-phoneme correspondences. This finding is crucial in light of current L2 speech models. All of these models stress the importance of L2 perception and how it is influenced by the L1 phonological system. None of them incorporate the L1 orthographic system as a modifying variable. Even though this study investigated *foreign* language learners, it is reasonable to assume that second language learners might just as well be influenced by their L1 orthographic system. In future research, this should be acknowledged, both theoretically and methodologically.

By conducting two different types of perception experiments, the study could show that difficulties in phonetic perception are not straightforwardly related to L2 lexical-phonological representations. Similar to other recent studies addressing the lexical level, it was found that a difficult phonological feature (i.e. length) could be identified more correctly than was expected based on the results of the discrimination experiment. It is proposed that metalinguistic knowledge, as mediated by explicit instruction, helped the learners in the phonological acquisition of this difficult feature. The experimental investigation of this new variable promises to be a fruitful venture.

Finally, this study investigated the perception and production of both vowel length and vowel quality. The experimental design of the experiments entailed the need to manipulate and analyse the two dimensions of interest separately. This meant that the native speakers of this study judged items which were specifically tailored to the research questions at hand. In future research, it would be of interest to have German native speakers identify items that have been naturally produced by Polish GFL learners. While the rigorous experimental approach was the appropriate way to address the hypotheses of the present study, it is still worthwhile to take the next step.

References

Alario, F.-X.; Perre, L.; Castel, C.; Ziegler, J. C. (2007): The role of orthography in speech production revisited. In *Cognition* (102), 464–475.

Almbark, R. A. (2014): Perception and production of SSBE vowels by foreign language learners: Towards a foreign language model. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech: Concordia Working Papers in Applied Linguistics* (5), 3–18.

Altmann, H.; Berger, I.; Braun, B. (2012): Asymmetries in the perception of non-native consonantal and vocalic length contrasts. In *Second Language Research* (28), 387–413.

Anderson-Hsieh, J.; Johnson, R.; Koehler, K. (1992): The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. In *Language Learning* (42), 529–555.

Andringa, S.; Rebuschat, P. (2015): New directions in the study of implicit and explicit learning. In *Studies in Second Language Acquisition* (37), 185–196.

Antoniadis, Z.; Strube, H. (1984): Untersuchungen zur spezifischen Dauer deutscher Vokale. In *Phonetica* (41), 72–87.

Apeltauer, E. (1997): *Grundlagen des Erst- und Fremdsprachenerwerbs. Eine Einführung*. Berlin: Langenscheidt.

Baayen, R.; Davidson, D.; Bates, D. (2008): Mixed-effects modeling with crossed random effects for subjects and items. In *Journal of Memory and Language* (59), 390–412.

Baayen, R. (2008): *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Barkowski, H.; Krumm, H.-J. (2010): *Fachlexikon Deutsch als Fremd- und Zweitsprach*e. Tübingen: Francke.

Barr, D. J.; Levy, R.; Scheepers, C.; Tily, H. J. (2013): Random effects structure for confirmatory hypothesis testing: Keep it maximal. In *Journal of Memory and Language* (68), 255–278.

Bassetti, B. (2006): Orthographic input and phonological representations in learners of Chinese as a Foreign Language. In *Written Language and Literacy* (9), 95–114.

Bassetti, B. (2009): Orthographic input and second language phonology. In Piske, T.; Young-Scholten, M. (Eds.): *Input Matters in SLA*. Bristol: Multilingual Matters, 191–206.

Bassetti, B.; Atkinson, N. (2015): Effects of orthographic forms on pronunciation in experienced instructed second language learners. In *Applied Psycholinguistics* (36), 67–91.

Bassetti, B.; Escudero, P.; Hayes-Harb, R. (2015): Second language phonology at the interface between acoustic and orthographic input. In *Applied Psycholinguistics* (36), 1–6.

Bates, D.; Maechler, M.; Bolker, B.; Walker, S. (2014): *lme4: Linear mixed-effects models using Eigen and S4*. R-package version 1.1-7.

Becker, T. (1998): *Das Vokalsystem der deutschen Standardsprache*. Frankfurt am Main: Peter Lang.

Best, C. T. (1995): A direct realist view of cross-language speech perception. In Strange, W. (Ed.): *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore: York Press, 171–204.

Best, C. T.; McRoberts, G. W.; Goodell, E. (2001): Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. In *The Journal of the Acoustical Society of America* (109), 775–794.

Best, C. T.; Tyler, M. D. (2007): Nonnative and second-language speech perception. Commonalities and complementarities. In Bohn, O.-S.; Munro, M. J. (Eds.): *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company, 13–34.

Bethin, C. Y. (1992): *Polish Syllables. The Role of Prosody in Phonology and Morphology*. Columbus: Slavica Publishers.

Biedrzycki, L. (1974): *Abriß der polnischen Phonetik*. Leipzig: VEB.

Boersma, P. (1998): *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. Dissertation. The Hague: Holland Academic Graphics.

Boersma, P.; Weenink, D. (2014): *Praat: doing phonetics by computer* [Computer program]. Version 5.4.04.

Bogacka, A. (2004): On the perception of English high vowels by Polish learners of English. In Daskalaki, E.; Katsos, N.; Mavrogiorgos, M.; Reeve, M. (Eds.): *CamLing 2004. Proceedings of the University of Cambridge Second Postgraduate Conference in Language Research in association with the*

Cambridge Institute of Language Research. Cambridge: Cambridge Institute of Language Research, 43–50.

Bohn, O.-S. (1995): Cross-language speech perception in adults. First language transfer doesn't tell it all. In Strange, W. (Ed.): *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore: York Press, 279–304.

Bohn, O.-S. (1998): Wahrnehmung fremdsprachlicher Laute: Wo ist das Problem? In Wegener, H. (Ed.): *Eine Zweite Sprache lernen. Empirische Untersuchungen zum Zweitspracherwerb*. Tübingen: Narr.

Bohn, O.-S. (2002): On phonetic similarity. In Burmeister, P.; Piske, T.; Rohde, A. (Eds.): *An Integrated View of Language Development. Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, 191–216.

Bohn, O.-S.; Best, C. T.; Avesanic, C.; Vayrad, M. (2011): Perceiving through the lens of native phonetics: Italian and Danish listeners' perception of English consonant contrasts. In Lee, W.-S.; Zee, E. (Eds.): *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, University of Hong Kong, 336–339.

Bohn, O.-S.; Flege, J. E. (1990): Interlingual identification and the role of foreign language experience in L2 vowel perception. In *Applied Psycholinguistics* (11), 303–328.

Bohn, O.-S.; Flege, J. E. (1992): The production of new and similar vowels by adult German learners of English. In *Studies in Second Language Acquisition* (14), 131–158.

Bohn, O.-S.; Munro, M. J. (Eds.) (2007): *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company.

Bongaerts, T.; van Summeren, C.; Planken, B.; Schils, E. (1997): Age and ultimate attainment in the pronunciation of a foreign language. In *Studies in Second Language Acquisition* (19), 447–465.

Bot, K. de; Lowie, W.; Verspoor, M. (2005): *Second Language Acquisition. An Advanced Resource Book*. London: Routledge.

Broselow, E.; Chen, S.-I.; Huffman, M. (1997): Syllable weight: convergence of phonology and phonetics. In *Phonology* (14), 47–82.

Bunk, G. (2005): *Phonetik aktuell - Kopiervorlagen*. Ismaning: Hueber.

Bußmann, H. (Ed.) (2008): *Lexikon der Sprachwissenschaft*. 4th edition. Stuttgart: Kröner.

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

Chen, M. (1970): Vowel length variation as a function of the voicing of the consonant environment. In *Phonetica* (22), 129–159.

Chomsky, N.; Halle, N. (1968): *The Sound Pattern of English*. New York: Harper & Row.

Coulmas, F. (2003): *Writing Systems. An introduction to their linguistic analysis*. Cambrige: Cambridge University Press.

Council of Europe (2001): *Common European Framework of Reference for Languages. Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cunnings, I. (2012): An overview of mixed-effects statistical models for second language researchers. In *Second Language Research* (28), 369–382.

Cutler, A. (2015): Representation of second language phonology. In *Applied Psycholinguistics* (36), 115–128.

Cutler, A.; Weber, A.; Otake, T. (2006): Asymmetric mapping from phonetic to lexical representations in second-language listening. In *Journal of Phonetics* (34), 269–284.

Damian, M. F.; Bowers, J. S. (2003): Effects of orthography on speech production in a form-preparation paradigm. In *Journal of Memory and Language* (49), 119–132.

Darcy, I. (2014): They sound the same, but I know they are different. Dissociated mechanisms for phonetic and lexical learning in a second language. *Workshop on Linguistic and Psycholinguistic Studies in L2 Phonology*, 22nd of November. Stuttgart.

Darcy, I.; Daidone, D.; Kojima, C. (2013): Asymmetric lexical access and fuzzy lexical representations in second language learners. In *The Mental Lexicon* (8), 372–420.

Darcy, I.; Dekydtspotter, L.; Sprouse, R. A.; Glover, J.; Kaden, C.; McGuire, M.; Scott, J. H. (2012): Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. In *Second Language Research* (28), 5–40.

Darcy, I.; Krüger, F. (2012): Vowel perception and production in Turkish children acquiring L2 German. In *Journal of Phonetics* (40), 568–581.

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

Derwing, T. M.; Munro, M. J. (2005): Second Language Accent and Pronunciation Teaching: A Research-Based Approach. In *TESOL Quarterly* (39), 379–397.

Derwing, T. M.; Munro, M. J.; Wiebe, G. (1998): Evidence in favor of a broad framework for pronunciation instruction. In *Language Learning* (48), 393–410.

Dieling, H. (1983): *Zur Perzeption und Produktion von Vokalen im Fremdsprachenunterricht Deutsch (Anfänger).* Dissertation. Universität Halle-Wittenberg, Halle.

Dieling, H. (1992): *Phonetik im Fremdsprachenunterricht Deutsch.* Berlin: Langenscheidt.

Dieling, H.; Hirschfeld, U. (2007): *Phonetik lehren und lernen.* Berlin: Langenscheidt.

Donegan, P. J. (1978): *On the Natural Phonology of Vowels.* Dissertation. University of Ohio, Columbus.

Dornbusch, T. (2012): *Orthographic Influences on L2 Auditory Word Processing.* Dissertation. Technische Universität Dortmund.

Dudenredaktion (Ed.) (2005): *Duden. Die Grammatik.* Mannheim: Dudenverlag.

Edwards, J. G. H.; Zampini, M. L. (Eds.) (2008): *Phonology and Second Language Acquisition.* Amsterdam: John Benjamins Publishing Company.

Eisenberg, P. (1995): *Die deutsche Orthographie und Deutsch als Fremdsprache: Analoge Strukturierung von System und Erwerb.* In *Jahrbuch Deutsch als Fremdsprache* (21), 171–184.

Eisenberg, P. (1999): Vokallängenbezeichnung als Problem. In *Linguistische Berichte* (179), 343–360.

Eisenberg, P. (2005): Phonem und Graphem. In Dudenredaktion (Ed.): *Duden. Die Grammatik.* Mannheim: Dudenverlag, 19–94.

Eisenberg, P. (2013): *Das Wort. Grundriss der deutschen Grammatik.* 4th edition. Stuttgart: Metzler.

Ellis, R. (1985): *Understanding Second Language Acquisition.* Oxford: Oxford University Press.

Ellis, R. (2008): *The Study of Second Language Acquisition.* 2nd edition. Oxford: Oxford University Press.

Elsendoorn, B. (1984): Production and perception of English vowel duration by Dutch speakers of English. In van den Broecke, M. P. R.; Cohen, A. (Eds.):

Proceedings of the 10th International Congress of Phonetic Sciences. Dordrecht: Foris Publications, 673–676.

Erdener, V. D.; Burnham, D. K. (2005): The role of audiovisual speech and orthographic information in nonnative speech production. In *Language Learning* (55), 191–228.

Escudero, P. (2005): *Linguistic Perception and Second Language Acquisition. Explaining the attainment of optimal phonological categorization.* Dissertation. Utrecht: LOT.

Escudero, P. (2009): The linguistic perception of SIMILAR L2 sounds. In Boersma, P.; Hamann, S. (Eds.): *Phonology in Perception.* Berlin: Mouton de Gruyter, 151–190.

Escudero, P.; Benders, T.; Lipski, S. C. (2009): Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. In *Journal of Phonetics* (37), 452–465.

Escudero, P.; Boersma, P. (2004): Bridging the gap between L2 speech perception research and phonological theory. In *Studies in Second Language Acquisition* (26), 551–585.

Escudero, P.; Hayes-Harb, R.; Mitterer, H. (2008): Novel second-language words and asymmetric lexical access. In *Journal of Phonetics* (36), 345–360.

Escudero, P.; Simon, E.; Mitterer, H. (2012): The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. In *Journal of Phonetics* (40), 280–288.

Escudero, P.; Simon, E.; Mulak, K. E. (2014): Learning words in a new language: Orthography doesn't always help. In *Bilingualism: Language and Cognition* (17), 384–395.

Escudero, P.; Wanrooij, K. (2010): The effect of L1 orthography on non-native vowel perception. In *Language and Speech* (53), 343–365.

Flege, J. E. (1980): Phonetic approximation in second language acquisition. In *Language Learning* (30), 117–134.

Flege, J. E. (1987): The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. In *Journal of Phonetics* (15), 47–65.

Flege, J. E. (1988): Factors affecting degree of perceived foreign accent in English sentences. In *The Journal of the Acoustical Society of America* (84), 70–79.

Flege, J. E. (1991): The interlingual identification of Spanish and English vowels: Orthographic Evidence. In *The Quarterly Journal of Experimental Psychology* (43A), 701–731.

Flege, J. E. (1993): Production and perception of a novel, second-language phonetic contrast. In *Journal of the Acoustical Society of America* (93), 1589–1608.

Flege, J. E. (1995): Second language speech learning. Theory, findings, and problems. In Strange, W. (Ed.): *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore: York Press, 233–277.

Flege, J. E. (1997): English vowel productions by Dutch talkers: more evidence for the "similar" versus "new" distinction. In James, A.; Leather, J. (Eds.): *Second-Language Speech. Structure and Process*. Berlin: Mouton de Gruyter, 11–52.

Flege, J. E. (1999): The relationship between L2 production and perception. In Ohala, J. J. (Ed.): *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco: University of California, 1273–1276.

Flege, J. E. (2002): Interactions between the native and second-language phonetic systems. In Burmeister, P.; Piske, T.; Rohde, A. (Eds.): *An Integrated View of Language Development. Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, 217–244.

Flege, J. E.; Bohn, O.-S.; Jang, S. (1997): Effects of experience on non-native speakers' production and perception of English vowels. In *Journal of Phonetics* (25), 437–470.

Flege, J. E.; MacKay, I. R.; Meador, D. (1999): Native Italian speakers' perception and production of English vowels. In *Journal of the Acoustical Society of America* (106), 2973–2987.

Flege, J. E.; Munro, M. J.; Fox, R. A. (1994): Auditory and categorial effects on cross-language vowel perception. In *Journal of the Acoustical Society of America* (95), 3623–3641.

Flege, J. E.; Munro, M. J.; MacKay, I. R. A. (1995): Factors affecting strength of perceived foreign accent in a second language. In *Journal of the Acoustical Society of America* (97), 3125–3134.

Flege, J. E.; Port, R. F. (1981): Cross-language phonetic interference: Arabic to English. In *Language and Speech* (24), 125–146.

Flege, J. E.; Takagi, N.; Mann, V. (1996): Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. In *The Journal of the Acoustical Society of America* (99), 1161–1173.

Fox, A. (2000): *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.

Frąckowiak-Richter, L. (1973): The duration of Polish vowels. In *Speech Analysis and Synthesis* (3), 87–115.

Fuhrhop, N. (2006): *Orthografie*. 3rd edition. Heidelberg: Winter.

Goswami, U.; Ziegler, J. C.; Dalton, L.; Schneider, W. (2003): Nonword reading across orthographies: How flexible is the choice of reading units? In *Applied Psycholinguistics* (24), 235–247.

Gottfried, T. L.; Miller, J. L.; Payton, P. E. (1990): Effect of speaking rate on the perception of vowels. In *Phonetica* (47), 155–172.

Grosjean, F. (2001): The bilingual's language modes. In Nicol, J. (Ed.): *One mind, two languages. Bilingual language processing*. Malden, MA: Blackwell, 1–22.

Grotjahn, R. (1998): Ausspracheunterricht: Ausgewählte Befunde aus der Grundlagenforschung und didaktisch-methodische Implikationen. In *Zeitschrift für Fremdsprachenforschung* (9), 35–83.

Grzeszczakowska-Pawlikowska, B. (2007): Probleme beim Rhythmuserwerb. Ausgangssprache Polnisch und Zielsprache Deutsch. In *Zeitschrift für Interkulturellen Fremdsprachenunterricht* (12), 1–23.

Guion, S. G.; Flege, J. E.; Akahane-Yamada, R.; Pruitt, J. C. (2000): An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. In *The Journal of the Acoustical Society of America* (107), 2711–2724.

Gussmann, E. (2007): *The Phonology of Polish*. Oxford: Oxford University Press.

Gut, U. (2003): Non-native speech rhythm in German. In Solé, M. J.; Recasens, D.; Romero, J. (Eds.): *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Universitat Autònoma de Barcelona, 2437–2440.

Gut, U. (2009): *Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt am Main: Peter Lang.

Hall, T. A. (1992): *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Max Niemeyer.

Hall, T. A. (2011): *Phonologie. Eine Einführung*. 2nd edition. Berlin: de Gruyter.

Han, M. S. (1994): Acoustic manifestations of mora timing in Japanese. In *The Journal of the Acoustical Society of America* (96), 73–82.

Hayes-Harb, R. (2012): Orthographic input and the acquisition of second language forms. *Second Language Acquisition of Phonology (SLP)*, July 6th. York.

Hayes-Harb, R.; Masuda, K. (2008): Development of the ability to lexically encode novel second language phonemic contrasts. In *Second Language Research* (24), 5–33.

Hayes-Harb, R.; Nicol, J.; Barker, J. (2010): Learning the phonological forms of new words: Effects of orthographic and auditory input. In *Language and Speech* (53), 367–381.

Hentschel, G. (1982): *Die Perzeption der vorderen, labialen Vokale des Deutschen durch Native-Speaker des Polnischen.* Frankfurt am Main: Peter Lang.

Hentschel, G. (1986): *Vokalperzeption und Natürliche Phonologie. Eine kontrastive Untersuchung zum Deutschen und Polnischen.* München: Otto Sagner.

Hillenbrand, J. M.; Clark, M. J.; Houde, R. A. (2000): Some effects of duration on vowel recognition. In *Journal of the Acoustical Society of America* (97), 3099–3111.

Hillenbrand, J. M.; Clark, M. J.; Nearey, T. M. (2001): Effects of consonant environment on vowel formant patterns. In *The Journal of the Acoustical Society of America* (109), 748–763.

Hirschfeld, U. (1998): Eine Schwerpunkte für die Arbeit an der Aussprache bei polnischen Deutschlernenden. In *Glottodidactica* (26), 113–121.

Hirschfeld, U. (2003): Vermittlung der Phonetik. In Bausch, K.-R. (Ed.): *Handbuch Fremdsprachenunterricht.* 4th edition. Tübingen: Francke, 872–879.

Hirschfeld, U. (Ed.) (2005a): *Phonetik International. Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache.* 4th edition. [CD-ROM]. Waldsteinberg: Popp.

Hirschfeld, U. (2005b): Spanisch. In Hirschfeld, U. (Ed.): *Phonetik International. Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache.* 4th edition. [CD-ROM]. Waldsteinberg: Popp, 1–16.

Hirschfeld, U.; Kelz, H. P. (2005): Didaktische Empfehlungen für die Arbeit an der Aussprache im Unterricht Deutsch als Fremdsprache. In Hirschfeld, U. (Ed.): *Phonetik International. Von Afrikaans bis Zulu. Kontrastive Studien* *für Deutsch als Fremdsprache.* 4th edition. [CD-ROM]. Waldsteinberg: Popp, 1–12.

Hirschfeld, U.; Reinke, K.; Stock, E. (2007): *Phonothek intensiv. Aussprachetraining.* Berlin: Langenscheidt.

Igras, M.; Ziółko, B. (2013): The influence of prosodic features on the prominence of accent and sentence boundaries in spoken Polish. *APAP Conference*, 21st–23rd of June. Lublin.

James, A.; Leather, J. (Eds.) (1997): *Second-Language Speech. Structure and Process.* Berlin: Mouton de Gruyter.

Jassem, W. (2003): Polish. In *Journal of the International Phonetic Association* (33), 103–107.

Jones, D. (1917): *An English Pronouncing Dictionary.* London: Dent.

Jørgensen, H. P. (1969): Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen. In *Phonetica* (19), 217–245.

Kabak, B. (2004): Acquiring phonology is not acquiring inventories but contrasts: The loss of Turkish and Korean primary long vowels. In *Linguistic Typology* (8), 351–368.

Kaminska, Z. (2003): Little Frog and Toad: Interaction of orthography and phonology in Polish spelling. In *Reading and Writing* (16), 61–80.

Katz, L.; Frost, R. (1992): The reading process is different for different orthographies: The Orthographic Depth Hypothesis. In Frost, R.; Katz, L. (Eds.): *Orthography, Phonology, Morphology, and Meaning.* Amsterdam: North-Holland, 67–84.

Keating, P. (1984): Universal phonetics and the organization of grammar. In *Working Papers in Phonetics* (59), 35–49.

Kirchner, M. (1999): Türkisch ğ, Deutsch r. In Johanson, L.; Rehbein, J. (Eds.): *Türkisch und Deutsch im Vergleich.* Wiesbaden: Harrassowitz, 139–151.

Kohler, K. (1995): *Einführung in die Phonetik des Deutschen.* 2nd edition. Berlin: Schmidt.

Kohler, K. (1999): German. In IPA (Ed.): *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press, 86–89.

Kondaurova, M. V.; Francis, A. L. (2008): The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. In *Journal of the Acoustical Society of America* (124), 3959–3971.

Kuhl, P. K. (2000): A new view of language acquisition. In *Proceedings of the National Academy of Sciences* (97), 11850–11857.

Kuhl, P. K.; Iverson, P. (1995): Linguistic experience and the "perceptual magnet effect". In Strange, W. (Ed.): *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore: York Press, 121-154.

Ladefoged, P. (2001): *Vowels and Consonants*. Malden: Blackwell.

Ladefoged, P. (2006): *A course in phonetics*. 5th edition. Boston: Thomson Wadsworth.

Ladefoged, P.; Maddieson, I. (1990): Vowels of the world's languages. In *Journal of Phonetics* (18), 93–122.

Ladefoged, P.; Maddieson, I. (1996): *The Sounds of the World's Languages*. Oxford: Blackwell.

Lado, R. (1957): *Linguistics Across Cultures*. Ann Arbor: The University of Michigan Press.

Lenneberg, E. H. (1967): *Biological Foundations of Language*. New York: Wiley.

Lennon, P. (2008): Contrastive analysis, error analysis, interlanguage. In Gramley, S. (Ed.): *Bielefeld Introduction to Applied Linguistics. A Course Book*. Bielefeld: Aisthesis, 51–62.

Macmillan, N. A.; Creelman, C. D. (2005): *Detection Theory. A User's Guide*. 2nd edition. Mahwah: Lawrence Erlbaum.

Maddieson, I. (1984): *Patterns of Sounds*. Cambridge: Cambridge University Press.

Major, R. C. (1987): Phonological similarity, markedness, and rate of L2 acquisition. In *Studies in Second Language Acquisition* (9), p. 63.

Major, R. C.; Kim, E. (1996): The Similarity Differential Rate Hypothesis. In *Language Learning* (46), 465–496.

Mathieu, L. (2014): *The Influence of Unfamiliar Orthography on L2 Phonolexical Acquisition*. Dissertation. University of Arizona, Tucson.

Maye, J.; Werker, J. F.; Gerken, L. (2002): Infant sensitivity to distributional information can affect phonetic discrimination. In *Cognition* (82), B101–B111.

McAllister, R.; Flege, J. E.; Piske, T. (2002): The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. In *Journal of Phonetics* (30), 229–258.

McCloy, D. (2015): *phonR: Tools for phoneticians and phonologists.* R-package version 1.0-3.

Meinhold, G.; Stock, E. (1982): *Phonologie der deutschen Gegenwartssprache.* 2nd edition. Leipzig: VEB Bibliographisches Institut.

Mengel, A. (2000): *Deutscher Wortakzent. Symbole, Signale.* München: Phorm-Verlag.

Miceli, G.; Benvegnù, B.; Capasso, R.; Caramazza, A. (1997): The independence of phonological and orthographic lexical forms: Evidence from aphasia. In *Cognitive Neuropsychology* (14), 35–69.

Mooshammer, C. (1998): *Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen.* Dissertation. Institut für Phonetik, München.

Morciniec, N. (1990): *Das Lautsystem des Deutschen und des Polnischen.* Heidelberg: Julius Groos Verlag.

Müller, U. (2005): Polnisch. In Hirschfeld, U. (Ed.): *Phonetik International. Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache.* 4th edition. [CD-ROM]. Waldsteinberg: Popp, 1–17.

Munro, M. J. (1993): Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. In *Language and Speech* (36), 39–66.

Neuman, Y. (2013): Graphophonemic assignment. In Khan, G. (Ed.): *Encyclopedia of Hebrew Language and Linguistics.* 2nd volume. Leiden: Brill, 135–145.

Neumann, R. (1981): Sprachkontrast Deutsch/Türkisch im Bereich von Aussprache und Rechtschreibung. In *Deutsch lernen* (2), 3–23.

Nimz, K. (2011a): *Vokalperzeption und -produktion türkischer DaF-Lerner. Eine empirische Untersuchung unter besonderer Berücksichtigung der Dimensionen Quantität und Qualität.* Master's Thesis. Humboldt-Universität, Berlin.

Nimz, K. (2011b): Vowel perception and production of late Turkish learners of L2 German. In Lee, W.-S.; Zee, E. (Eds.): *Proceedings of the 17th International Congress of Phonetic Sciences.* Hong Kong: Department of Chinese, Translation and Linguistics, University of Hong Kong, 1494–1497.

Nimz, K. (2014): Predicting vowel length production in an L2: Phonetic versus phonological contrastive analyses. In Cyran, E.; Szpyra, J. (Eds.): *Crossing phonetics-phonology lines.* Newcastle upon Tyne: Cambridge Scholars Publishing, 313–325.

Nimz, K. (2015): The Perception of vowel quality and quantity by Turkish learners of German as a Foreign Language. In Raimy, E.; Cairns, C. E. (Eds.): *The Segment in Phonetics and Phonology*. Hoboken: John Wiley, 253–266.

Nooteboom, S. G.; Doodeman, G. J. N. (1980): Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America* (67), 276–287.

Ohala, J. J. (1990): There is no interface between phonology and phonetics: A personal view. In *Journal of Phonetics* (18), 153–171.

Oturan, K. (2002): *Vokalperzeption und -produktion in der Zweitsprache. Eine Untersuchung mit türkischen Muttersprachlern*. Aachen: Shaker.

Pajak, B.; Levy, R. (2014): The role of abstraction in non-native speech perception. In *Journal of Phonetics* (46), 147–160.

Pallier, C.; Colomé, A.; Sebastián-Gallés, N. (2001): The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. In *Psychological Science* (12), 445–449.

Petrova, A.; Gaskell, M. G.; Ferrand, L. (2011): Orthographic consistency and word-frequency effects in auditory word recognition: New evidence from lexical decision and rime detection. In *Frontiers in Psychology* (2), 1–11.

Pickett, E. R.; Blumstein, S. E.; Burton, M. W. (1999): Effects of speaking rate on the singleton/geminate consonant contrast in Italian. In *Phonetica* (56), 135–157.

Pierrehumbert, J. B. (1990): Phonological and phonetic representation. In *Journal of Phonetics* (18), 375–394.

Pike, K. L. (1945): *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Pisoni, D. B. (1973): Auditory and phonetic memory codes in the discrimination of consonants and vowels. In *Perception and Psychophysics* (13), 253–260.

Polivanov, E. D. (1931): La perception des sons d'une langue étrangère. In *Traveaux du Cercle Linguistique de Praque* (4), 79-96. Cited from: Leont'ev, A. A. (Ed.) (1974): *E. D. Polivanov: Selected Works. Articles on General Linguistics*. The Hague: Mouton, 223–238.

Pompino-Marschall, B. (2009): *Einführung in die Phonetik*. 3rd edition. Berlin: Walter de Gruyter.

Port, R.; Al-Ani, S.; Maeda, S. (1980): Temporal compensation and universal phonetics. In *Journal of Phonetics* (37), 235–252.

Port, R. F.; O'Dell, M. L. (1985): Neutralization of syllable-final voicing in German. In *Journal of Phonetics* (13), 455–471.

Prędota, S. (1978): *Die Polnisch-Deutsche Interferenz im Bereich der Aussprache*. Wrocław: Zakład Narodowy Imienia Ossolińskich.

Primus, B. (2000): Suprasegmentale Graphematik und Phonologie: Die Dehnungszeichen im Deutschen. *Linguistische Berichte* (181), 9-34.

R Core Team (2014): *R: A Language and Environment for Statistical Computing* [Computer program]. R Foundation for Statistical Computing, Vienna.

Rafat, Y. (2015): The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative rhotics. In *Applied Psycholinguistics* (36), 43–66.

Ramers, K. H. (1988): *Vokalquantität und -qualität im Deutschen*. Tübingen: Max Niemeyer.

Ramers, K. H. (1999a): Vokalquantität als orthographisches Problem. Zur Funktion der Doppelkonsonanzschreibung im Deutschen. In *Linguistische Berichte* (177), 52–64.

Ramers, K. H. (1999b): Zur Doppelkonsonanzschreibung im Deutschen. Eine Rereplik. In *Linguistische Berichte* (179), 350–360.

Ramers, K. H.; Vater, H. (1995): *Einführung in die Phonologie*. Hürth: Gabel.

Ramus, F.; Dupoux, E.; Mehler, J. (2003): The psychological reality of rhythm classes: perceptual studies. In Solé, M. J.; Recasens, D.; Romero, J. (Eds.): *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Universitat Autònoma de Barcelona, 337–342.

Remijsen, B. (2004): Acoustic measurements on prosody using Praat. *Experimental Prosody Research Workshop*, 7th–9th of October. Leipzig.

Richter, J. (2007): Perzeption vor Produktion in der L2? In Doff, S; Schmidt, T. (Eds.): *Fremdsprachenforschung heute. Interdisziplinäre Impulse, Methoden und Perspektiven*. Frankfurt am Main: Peter Lang.

Richter, J. (2008): *Phonetische Reduktion im Deutschen als L2. Eine empirische Querschnittsstudie*. Baltmannsweiler: Schneider-Verlag Hohengehren.

Rojczyk, A. (2011): Overreliance on duration in nonnative vowel production and perception: The within lax vowel category contrast. In Wrembel, M. (Ed.): *Achievements and Perspectives in SLA of Speech: New Sounds 2010*. 2n volume. Frankfurt am Main: Lang, 239–250.

Rojczyk, A.; Porzuczek, A. (2014): Acoustic properties of nasal geminates in Polish. In Cyran, E.; Szpyra, J. (Eds.): *Crossing Phonetics-Phonology Lines*. Newcastle upon Tyne: Cambridge Scholars Publishing, 347–364.

Rubach, J. (1984): C*yclic and Lexical Phonology. The Structure of Polish*. Dordrecht: Foris Publications.

Sadowska, I. (2012): *Polish. A Comprehensive Grammar*. London: Routledge.

Schachter, J. (1974): An error in error analysis. In *Language Learning* (24), 205–214.

Schiel, F.; Draxler, C.; Harrington, J. (2011): Phonemic segmentation and labelling using the MAUS technique. *Workshop on New Tools and Methods for Very-Large-Scale Phonetics Research*, January 28th–31st. University of Pennsylvania.

Sendlmeier, W. F. (1981): Der Einfluß von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. In *Phonetica* (38), 291–308.

Sendlmeier, W. F. (1985): Die Beschreibung der deutschen Vokale in betonter Stellung. Ein forschungshistorischer Überblick. In Borowsky, V.; Köster, J.-P. (Eds.): *Neue Tendenzen in der Angewandten Phonetik I*. Hamburg: Helmut Buske, 167–198.

Settinieri, J. (2010): Ausspracheerwerb und Ausspracheverмittlung. In Krumm, H.-J.; Fandrych, C.; Hufeisen, B.; Riemer C. (Eds.): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*. 2nd volume. Berlin: de Gruyter, 999–1008.

Sheldon, A.; Strange, W. (1982): The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. In *Applied Psycholinguistics* (3), p. 243.

Showalter, C. E.; Hayes-Harb, R. (2013): Unfamiliar orthographic information and second language word learning: A novel lexicon study. In *Second Language Research* (29), 185–200.

Showalter, C. E.; Hayes-Harb, R. (2015): Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. In *Applied Psycholinguistics* (36), 23–42.

Silveira, R. (2009): Investigation the role of orthography in the acquisition of L2 pronunciation: A case study. In Watkins, M. A.; Rauber, A. S.; Baptista, B. O. (Eds.): *Recent Research in Second Language Phonetics/Phonology*. Newcastle: Cambridge Scholars Publishing, 270–290.

Simon, E.; Della Chambless; Alves, U. K. (2010): Understanding the role of orthography in the acquisition of a non-native vowel contrast. In *Language Sciences* (32), 380–394.

Simon, E.; van Herreweghe, M. (2010): The relation between orthography and phonology from different angles: Insights from psycholinguistics and second language acquisition. In *Language and Speech* (53), 303–306.

Simpson, A.; Ericsdotter, C. (2007): Sex-specific differences in f0 and vowel space. In Barry, W. J.; Trouvain, J. (Eds.): *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Universität des Saarlandes, 933–936.

Skibicki, M. (2007): *Polnische Grammatik*. Hamburg: Buske.

Slembek, E. (1995): *Lehrbuch der Fehleranalyse und Fehlertherapie. Deutsch hören, sprechen und schreiben. Für Lernende mit griechischer, italienischer, polnischer, russischer und türkischer Muttersprache*. 2nd edition. Heinsberg: Agentur Dieck.

Slowiaczek, L. M.; Dinnsen, D. A. (1985): On the neutralizing status of Polish word-final devoicing. In *Journal of Phonetics* (13), 325–341.

Slowiaczek, L. M.; Szymanska, H. J. (1989): Perception of word-final devoicing in Polish. In *Journal of Phonetics* (17), 205–212.

Smith, C. (1995): Prosodic patterns in the coordination of vowel and consonant gestures. In Connell, B.; Arvaniti, A. (Eds.): *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 205–222.

Sobkowiak, W. (2001): *English Phonetics for Poles*. Poznań: Wydawnictwo Poznańskie.

Steele, J. (2005): Assessing the role of orthographic versus uniquely auditory input in acquiring new L2 segments. *Septièmes Journées Internationales du Réseau Français de Phonologie*, 2nd–4th of June. Aix-en-Provence.

Steinlen, A. (2009): Phonetic input in second language acquisition: Contrastive analysis of native and non-native sounds. In Piske, T.; Young-Scholten, M. (Eds.): *Input Matters in SLA*. Bristol: Multilingual Matters, 219–237.

Stieber, Z. (1973): *A Historical Phonology of the Polish Language*. Heidelberg: Winter.

Stock, E.; Hirschfeld, U. (1996): *Phonothek Deutsch als Fremdsprache. Lehrerhandreichungen*. Leipzig: Langenscheidt.

Strange, W. (Ed.) (1995): *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*. Baltimore: York Press.

Strange, W. (2007): Cross-language phonetic similarity of vowels. Theoretical and methodological issues. In Bohn, O.-S.; Munro, M. J. (Eds.): *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company, 35–55.

Strange, W.; Shafer, V. L. (2008): Speech perception in second language learners. In Edwards, J. G. H.; Zampini, M. L. (Eds.): *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company, 153–191.

Studdert-Kennedy, M. (1981): The emergence of phonetic structure. In *Cognition* (10), 301–306.

Ternes, E. (1978): Der Einfluß der Silbenstruktur auf die lautliche Interferenz. In H. W. Wodarz (Ed.): *Frankfurter Phonetische Beiträge II*. Hamburg: Helmut Buske, 93–119.

Ternes, E. (2012): *Einführung in die Phonologie*. 3rd edition. Darmstadt: Wissenschaftliche Buchgesellschaft.

Thurgood, E. (2001): The phonetic realization of phonologically geminate affricates in Polish: The long and the short of it. In *Speech and Language Technology* (5), 9–19.

Thurgood, E. (2002): The recognition of geminates in ambiguous contexts in Polish. *Speech Prosody*, 11th–13th of April. Aix-en-Provence.

Thurgood, E.; Demenko, G. (2003): Phonetic realizations of Polish geminate affricates. In Solé, M. J.; Recasens, D.; Romero, J. (Eds.): *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Universitat Autònoma de Barcelona, 1895–1898.

Tomaschek, F. (2009): *Eine Perzeptionsuntersuchung des F2-Vokalraumes deutscher und polnischer Muttersprachler anhand der hohen Vokalreihe [i – y – u]*. Master's thesis. Eberhard Karls Universität, Tübingen.

Tomaschek, F.; Truckenbrodt, H.; Hertrich, I. (2011): Processing German vowel quantity: Categorial perception or perceptual magnet effect? In Lee, W.-S.; Zee, E. (Eds.): *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, University of Hong Kong, 2002–2005.

Trubetzkoy, N. S. (1939): Grundzüge der Phonologie. In *Travaux du Cercle Linguistique de Prague* (7). Cited from the 7th edition (1989). Göttingen: Vandenhoeck and Ruprecht.

Tsukada, K.; Birdsong, D.; Bialystok, E.; Mack, M.; Sung, H.; Flege, J. E. (2005): A developmental study of English vowel production and perception by native Korean adults and children. In *Journal of Phonetics* (33), 263–290.

Turk, A. E.; Nakai, S.; Sugahara, M. (2006): Acoustic segment durations in prosodic research: A practical guide. In Sudhoff, S.; Lenertová, D.; Meyer, R.; Pappert, S.; Augurzky, P.; Mleinek, I.; Richter, N.; Schliesser, J. (Eds.): *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 1–26.

Tworek, A. (2012): *Einführung in die deutsch-polnische vergleichende Phonetik*. Dresden: Neisse Verlag.

Weber, A.; Broersma, M.; Aoyagi, M. (2011): Spoken-word recognition in foreign-accented speech by L2 listeners. In *Journal of Phonetics* (39), 479–491.

Weber, A.; Cutler, A. (2004): Lexical competition in non-native spoken-word recognition. In *Journal of Memory and Language* (50), 1–25.

Weirich, M.; Simpson, A. P. (to appear): Impact and interaction of accent realization and speaker sex on vowel length in German. In Leemann, A.; Kolly, M.-J.; Dellwo, V.; Schmid, S. (Eds.): *Trends in Phonetics and Phonology in German-Speaking Europe*. Frankfurt: Peter Lang.

Weiss, R. (1974): Relationship of vowel length and quality in the perception of German vowels. In *Linguistics* (123), 59–70.

Werker, J. F.; Tees, R. C. (1984): Phonemic and phonetic factors in adult cross-language speech perception. In *Journal of the Acoustical Society of America* (75), 1866–1878.

White, L.; Melhorn, J. F.; Mattys, S. L. (2010): Segmentation by lexical subtraction in Hungarian speakers of second-language English. In *Quarterly Journal of Experimental Psychology* (63), 544–554.

Wiese, R. (2000): *The Phonology of German*. Oxford: Oxford University Press.

Winter, B. (2011): Pseudoreplication in phonetic research. In Lee, W.-S.; Zee, E. (Eds.): *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, University of Hong Kong, 2137–2140.

Winter, B. (2013): Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf].

Wójtowicz, J. (1981): *Die Polnische Lautlehre*. Warsaw: Panstwowe Wydawnictwo Naukowe.

Wood, S. (1982): *X-ray and model studies of vowel articulation*. Dissertation. Lund University, Working Papers of the Department of Linguistics.

Wängler, H.-H. (1974): *Atlas deutscher Sprachlaute*. 5th edition. Berlin: Akademie-Verlag.

Ylinen, S.; Shestakova, A.; Alku, P.; Huotilainen, M. (2005): The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. In *Language and Speech* (3), 313–338.

Young-Scholten, M. (2002): Orthographic input in L2 phonological development. In Burmeister, P.; Piske, T.; Rohde, A. (Eds.): *An Integrated View of Language Development. Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, 263–279.

Young-Scholten, M. (2004): Prosodic constraints on allophonic distribution in adult L2 acquisition. In *International Journal of Bilingualism* (8), 67–77.

Young-Scholten, M.; Langer, M. (2015): The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. In *Applied Psycholinguistics* (36), 93–114.

Zampini, M. (2008): L2 speech production research. Findings, issues, and advances. In Edwards, J. G. H.; Zampini, M. L. (Eds.): *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company, 219–249.

Appendix

Appendix I: International Phonetic Alphabet

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC)

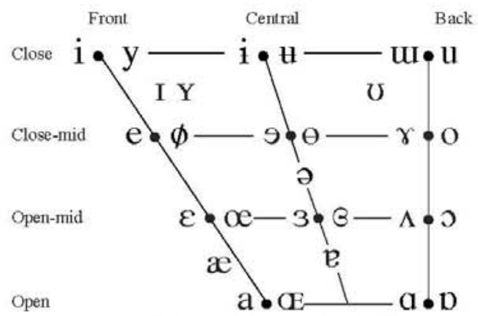
	Bilabial	Labiodental	Dental	Alveolar	Post alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	Examples:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

- ʍ Voiceless labial-velar fricative
- w Voiced labial-velar approximant
- ɥ Voiced labial-palatal approximant
- ʜ Voiceless epiglottal fricative
- ʢ Voiced epiglottal fricative
- ʡ Epiglottal plosive
- ɕ ʑ Alveolo-palatal fricatives
- ɺ Voiced alveolar lateral flap
- ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

Voiceless	n̥ d̥	Breathy voiced	b̤ a̤	Dental	t̪ d̪
Voiced	s̬ t̬	Creaky voiced	b̰ a̰	Apical	t̺ d̺
Aspirated	tʰ dʰ	Linguolabial	t̼ d̼	Laminal	t̻ d̻
More rounded	ɔ̹	Labialized	tʷ dʷ	Nasalized	ẽ
Less rounded	ɔ̜	Palatalized	tʲ dʲ	Nasal release	dⁿ
Advanced	u̟	Velarized	tˠ dˠ	Lateral release	dˡ
Retracted	e̠	Pharyngealized	tˤ dˤ	No audible release	d̚
Centralized	ë	Velarized or pharyngealized	ɫ		
Mid-centralized	e̽	Raised	e̝ (ɹ̝ = voiced alveolar fricative)		
Syllabic	n̩	Lowered	e̞ (β̞ = voiced bilabial approximant)		
Non-syllabic	e̯	Advanced Tongue Root	e̘		
Rhoticity	ɚ a˞	Retracted Tongue Root	e̙		

SUPRASEGMENTALS

- ˈ Primary stress
- ˌ Secondary stress ˌfoʊnəˈtɪʃən
- ː Long eː
- ˑ Half-long eˑ
- ˘ Extra-short ĕ
- | Minor (foot) group
- ‖ Major (intonation) group
- . Syllable break ɹi.ækt
- ‿ Linking (absence of a break)

TONES AND WORD ACCENTS

LEVEL		CONTOUR	
e̋ or ˥	Extra high	ě or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˦˥	High rising
è ˨	Low	e᷅ ˩˨	Low rising
ȅ ˩	Extra low	e᷈ ˧˦˧	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

Appendix II: Polish test words

Vowel	Word	Meaning	Context
/a/	**papa**	*tar paper*	bilabial
/a/	**tata**	*daddy*	postdental
/a/	**skakać**	*to jump*	velar
/ɛ/	**mebel**	*furniture*	bilabial
/ɛ/	**teza**	*thesis*	postdental
/ɛ/	**gekon**	*gecko*	velar
/ɨ/	**pypeć**	*blotch/mole*	bilabial
/ɨ/	**tytan**	*titan*	postdental
/ɨ/	**chyba**	*maybe*	(velar)
/i/	**biba**	*party*	bilabial
/i/	**sinus**	*sinus*	postdental
/i/	**kikut**	*stump*	(velar)
/u/	**puma**	*puma*	bilabial
/u/	**tutaj**	*here*	postdental
/u/	**kukać**	*to cuckoo*	velar
/ɔ/	**popyt**	*demand*	bilabial
/ɔ/	**soda**	*soda*	postdental
/ɔ/	**kogut**	*cock*	velar
/a/	**niania**	*nanny*	palatal
/a/	**dziadzio**	*grandpa*	palatal
/a/	**ziajać**	*to pant*	palatal
/ɛ/	**pieniacz**	*choleric person*	palatal
/ɛ/	**dzieciak**	*child*	palatal
/ɛ/	**sieci**	*nets*	palatal
	Geminate experiment		
/ɛ/	**Grecy**	*Greeks (nom.)*	singleton
/ɛ/	**greccy**	*Greek (adj.)*	geminate
/ɛ/	**leki**	*medicines*	singleton
/ɛ/	**lekki**	*light (weight)*	geminate
/a/	**pana**	*man (acc.)*	singleton
/a/	**panna**	*maiden*	geminate
/u/	**uczę**	*teach (1. pres.)*	singleton
/u/	**uczczę**	*celebrate (1. fut.)*	geminate
/u/	**buda**	*doghouse*	singleton
/u/	**Budda**	*Buddah*	geminate
/i/	**rodziny**	*families*	singleton
/i/	**rodzinny**	*family (adj.)*	geminate
/a/	**saki**	*bags*	singleton
/a/	**ssaki**	*mammals*	geminate
/ɛ/	**lecie**	*summer (loc.)*	singleton
/ɛ/	**lećcie**	*fly (2. pl. imp.)*	*geminate*

Appendix III: Segmentation criteria for vowels

Consonantal contexts:

- **Plosives:**
  - Preceding vowels: first positive zero crossing of the first periodic waveform
  - Following vowels: last positive zero crossing of the periodic waveform where F2 is still visible
- **Fricatives:**
  - Preceding vowels: first positive zero crossing of the first periodic waveform
  - Following vowels: last positive zero crossing of the periodic waveform where F2 is still visible and frication noise is not yet prevalent
- **Nasals:**
  - Preceding vowels: point of change in spectrogram when vowel formants become more visible combined with changes in waveform
  - Following vowels: point of change in spectrogram when vowel formants fade out combined with changes in waveform
- **Lateral [l]:**
  - Preceding vowels: point of change in spectrogram when vowel formants become more visible combined with changes in waveform
  - Following vowels: point of change in spectrogram when vowel formants fade out and/or change, combined with change in waveform; furthermore, in difficult cases, careful comparison of visual segmentation with auditory impression
- **R-Allophones:**
  - Preceding vowels: first positive zero crossing of the first periodic waveform (following [ʁ]); first positive zero crossing of the first continuous periodic waveform combined with the point of change of formants into a continuous movement ([r])
  - Following vowels: last positive zero crossing of the periodic waveform where F2 energy is still visible and frication noise is not yet prevalent (for [ʁ]); last positive zero crossing of the periodic waveform where F2 energy is still visible (for [ɾ])

Appendix IV: German and Polish participants

a) Polish participants

Subject	Sex	Class	Age	Place of birth	Language mother	Language father	Other languages	Dialect/lang. at home	In G > 1 month	Problems (hear./sp.)?
P1	female	4th	19	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P2	female	4th	19	Warsaw	Polish	Polish	Engl. (7), Jap. (2)	no	no	no
P3	female	4th	19	Warsaw	Polish	Polish	English (7), Russ. (5)	no	3 months	no
P4	male	4th	19	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P5	female	4th	19	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P6	female	4th	19	Warsaw	Polish	Polish	Engl. (12), Russ. (2)	no	no	no
P7	male	3rd	18	Warsaw	Polish	Bulgarian	Engl. (11)	no	no	no
P8	female	4th	19	Czestochowa	Polish	Polish	Engl. (9), Russ. (3)	no	no	no
P9	female	4th	18	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P10	female	4th	19	Warsaw	Polish	Polish	Engl. (9)	no	3 months	no
P11	female	4th	19	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P12	female	3rd	18	Warsaw	Polish	Polish	Engl. (5)	no	no	no
P13	female	3rd	18	Otwock	Polish	Polish	Engl. (9)	no	no	no
P14	female	3rd	17	Warsaw	Polish	Polish	Engl. (11)	no	no	no
P15	female	3rd	18	Warsaw	Polish	Polish	Engl. (11)	no	no	no
P16	male	3rd	18	Warsaw	Polish	Polish	Engl. (12)	no	no	no
P17	male	3rd	18	Warsaw	Polish	Polish	Engl. (13)	no	no	no
P18	female	4th	19	Warsaw	Polish	Polish	Engl. (13)	no	no	no
P19	female	3rd	18	Warsaw	Polish	Polish	Engl. (13)	no	no	no
P20	female	4th	19	Legnica	Polish	Polish	Engl. (9)	no	no	no
P21	female	3rd	17	Warsaw	Polish	Polish	Engl. (13)	no	7 years	no
P22	female	4th	18	Sofia	Bulgarian	Polish	Engl. (12) Bulg. (native)	Bulg.	no	no

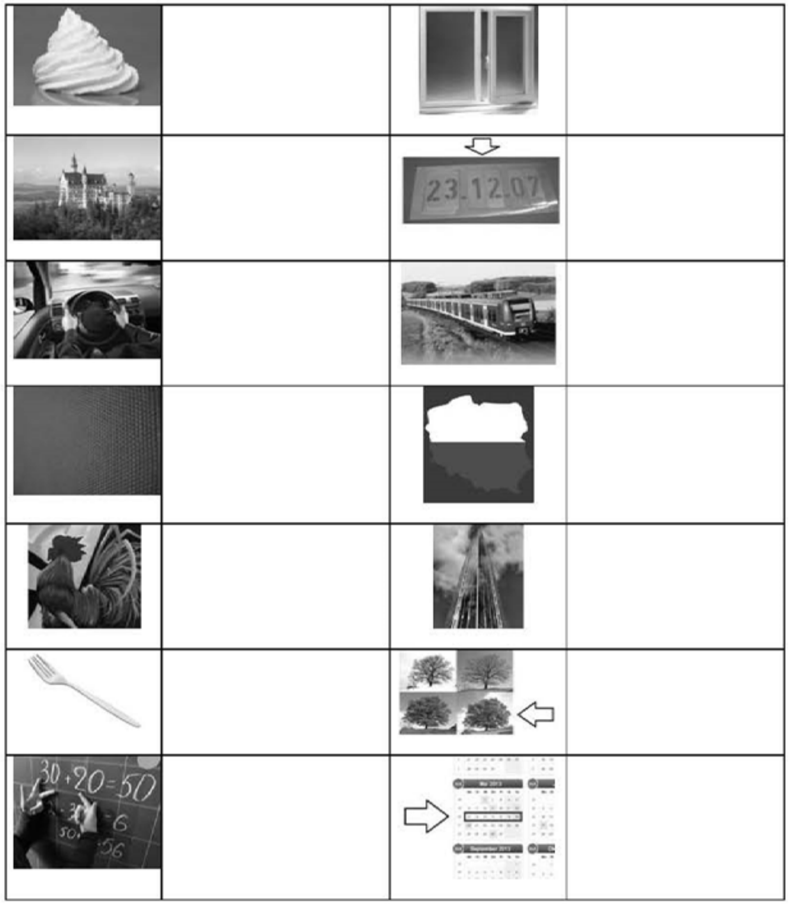
b) German participants

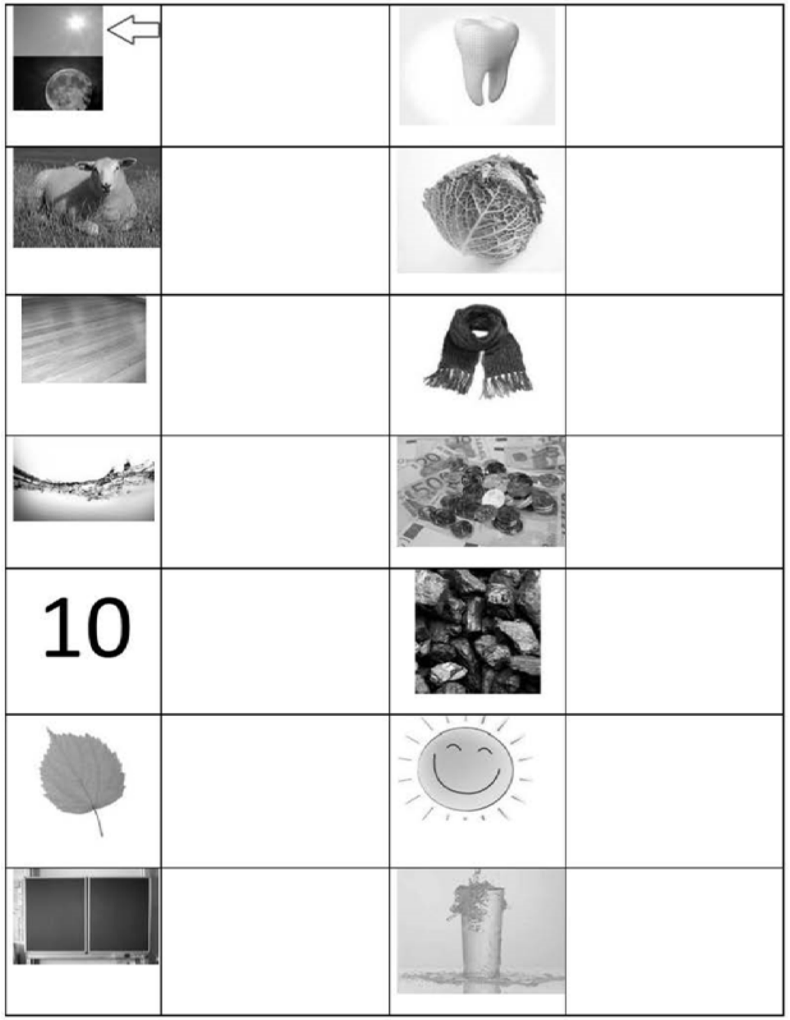
Subject	Sex	Grade	Age	Place of birth	Dialect home?	Other languages (learning in years)	Problems (hear./sp.?	Discr. task
G1	female	12	17	Dortmund	no	Engl. (10), Sp. (3), Kata. (3), Por. (5), Fr. (2)	no	yes
G2	female	12	18	Dortmund	no	Engl. (11), Span. (3)	no	yes
G3	female	12	18	Dortmund	no	Engl. (10), Fren. (7)	no	yes
G4	female	12	18	Dortmund	no	Engl. (8), Span. (3)	no	yes
G5	female	12	17	Dortmund	no	Engl. (9), Fren. (7)	no	no
G6	female	12	17	Herdecke	no	Engl. (9 + 1 year in USA), Fren. (4)	no	yes
G7	male	12	18	Dortmund	no	Engl. (10)	dyslexic	yes
G8	male	12	18	Hamm	no	Engl. (11), Fren. (8)	no	yes
G9	female	12	17	Dortmund	no	Engl. (10), Fren. (5)	no	yes
G10	male	12	17	Dortmund	no	Engl. (11), Fren. (6)	no	yes
G11	female	12	18	Herdecke	no	Engl. (9), Fren. (7), Span. (3)	no	yes
G12	female	12	20	Oelde	no	Engl. (10), Fren. (2), Est. (2,5 + 1 y. Est.)	stammering (past)	yes
G13	female	11	16	Dortmund	no	Engl. (8), Fren. (5)	no	yes
G14	female	12	17	Dortmund	no	Engl. (10), Fren. (4), Span. (2), Chin. (2)	no	yes
G15	male	12	18	Aachen	no	Engl. (11), Fren. (4), Span. (3)	no	yes
G16	female	12	17	Dortmund	no	Engl. (7), Fren. (5)	no	yes
G17	female	12	17	Dortmund	no	Engl. (9), Fren. (7), Chin. (5)	no	no
G18	female	11	18	Herdecke	no	Engl. (8 + 1 y. USA), Fren. (5), Span. (2)	no	yes
G19	male	12	18	Herdecke	no	Engl. (9), Fren. (3)	lisping (past)	yes
G20	female	12	17	Herne	no	Engl. (10), Fren. (7), Span. (2)	no	yes
G21	female	12	19	Dortmund	no	Engl. (11), Lat. (4 + 1 y. Lat.), Span. (3)	no	yes
G22	female	finished	18	Oberhausen	no	Engl. (9), Fren. (2)	no	yes
G23	male	finished	21	Bottrop	no	Engl. (11), Span. (2)	no	yes

Appendix V: Writing task including familiarity ratings

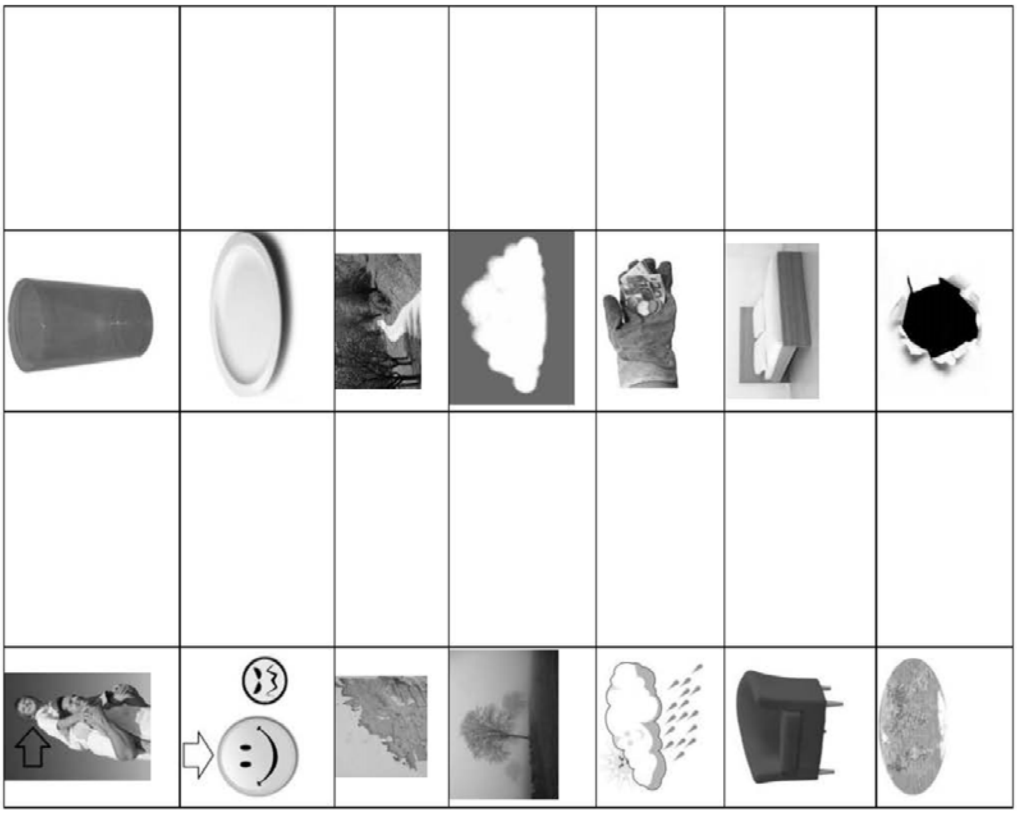
Schreibe die Wörter und sage von 1 – 7 wie gut du die Wörter generell kennst (1 = kenne ich nicht, 7 = kenne ich gut)!

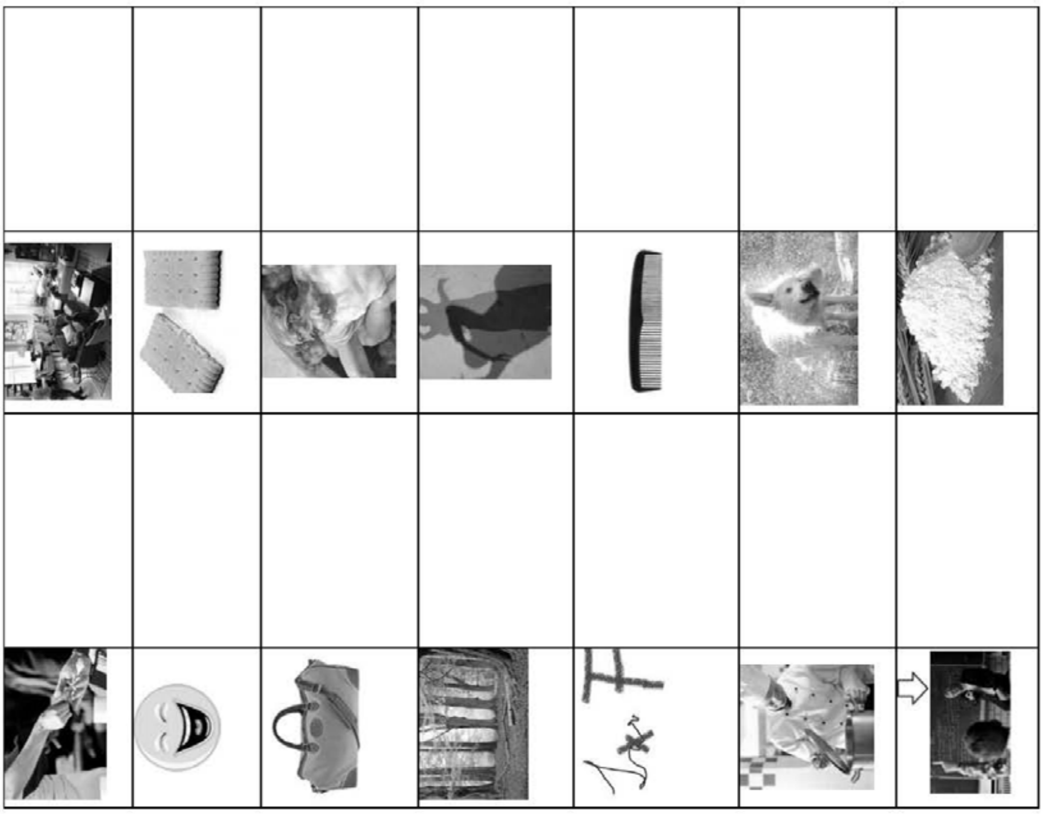
Versuchsperson Nr.: ______ Namensabkürzung: __________





S. Rückseite! ☺





Appendix VI: Orthography task including metalinguistic questions

Universität Potsdam — Katharina Nimz

Fragebogen für VersuchsteilnehmerInnen II

(Nach allen Experimenten auszufüllen)

Versuchsperson Nr.: ______ **Namensabkürzung:** ___________

Wie findest du die deutsche Sprache auf einer Skala von 1 (sehr schlecht) – 10 (sehr gut) – sei ehrlich ;-): _____

Wie findest du den Klang (Klang = dźwięk) der deutschen Sprache auf einer Skala von 1 (sehr schlecht) – 10 (sehr gut)? _____

Wie findest du es, wenn jemand einen polnischen Akzent hat, wenn er Deutsch spricht, auf einer Skala von 1 (sehr schlecht) – 10 (sehr gut): _____

Wie gern würdest du wie ein deutscher Muttersprachler sprechen, auf einer Skala von 1 (überhaupt nicht gern) – 10 (sehr gern): _____

Tust du etwas Spezielles, um eine bessere Aussprache (Aussprache = *wymowa*) zu bekommen (z. B. spezieller Phonetikunterricht)?

____

____

Weißt du, dass es im Deutschen lange und kurze Vokale (Vokal = samogłoska) gibt? JA [ ] NEIN [ ]

Weißt du, dass "kurze" und "lange" Vokale sich nicht nur in ihrer Länge unterscheiden? Falls ja, wie unterscheiden sie sich noch?

____

____

Universität Potsdam — Katharina Nimz

Kleine Aufgabe:

Was denkst du: Welche der unterstrichenen Vokale sind kurz oder lang? Warum denkst du das?

(Es ist kein Problem, wenn du keine Antwort auf "warum" weißt!) ;-)

Satz 1: Pochel ging mit Tossel, Dahn und Beg zum Kend.

Pochel: KURZ [ ] LANG [ ] – Warum? ________________

Tossel: KURZ [ ] LANG [ ] – Warum? ________________

Dahn: KURZ [ ] LANG [ ] – Warum? ________________

Beg: KURZ [ ] LANG [ ] – Warum? ________________

Kend: KURZ [ ] LANG [ ] – Warum? ________________

Satz 2: Pahler sah Gusch, Dull und Falde den Teper.

Pahler: KURZ [ ] LANG [ ] – Warum? ________________

Gusch: KURZ [ ] LANG [ ] – Warum? ________________

Dull: KURZ [ ] LANG [ ] – Warum? ________________

Falde: KURZ [ ] LANG [ ] – Warum? ________________

Teper: KURZ [ ] LANG [ ] – Warum? ________________

Kommentar zu dieser Aufgabe? ________________

____