Abstract

This study investigates the role of orthography in German vowel production by Polish L1 speakers with German as an L2. Eighteen intermediate to advanced Polish L2 German learners and 20 German native speakers were recorded during a picture-naming task in which half of the experimental items were explicitly marked in their orthographic representation for their vowel length (short or long). Duration measurements revealed that explicit orthographic marking helped the Polish L2 German learners produce the short-long contrast more native-like. The analysis of vowel quality further showed that (in)congruencies between L1 and L2 grapheme-to-phoneme correspondences may influence L2 vowel production as well. These findings have important implications for models of L2 speech learning and pronunciation training.

Introduction

In the field of German as a second language (L2), pronunciation research has metaphorically been referred to as the "poor cousin" of other subfields of second language acquisition (SLA) research (Hirschfeld, 2001). However, in recent years, we have witnessed a growing interest in L2 phonetics/phonology research, with special issues and anthologies focusing on L2 speech learning (Archibald & Young-Scholten, 2003; Bohn & Munro, 2007; Trouvain & Gut, 2007; Edwards & Zampini, 2008). In this regard, orthography and its effect on acquiring a second phonology is the cousin of the poor cousin. While pronunciation has had its place in recent works on L2 acquisition, orthography still has not received much attention in (instructed) SLA research (Ellis, 2015; Loewen, 2015; Derwing & Munro, 2015), despite the fact that the way a word is written can profoundly influence the way we perceive and produce it (Bassetti et al., 2015).

Learning the phonology of an L2 is challenging, especially later in life (Flege et al., 2006; Piske, MacKay, & Flege, 2001; Saito, 2015). Yet, it is not impossible to attain high levels of pronunciation abilities in an L2 (Bongaerts, van Summeren, Planken, & Schils, 1997; Moyer, 1999). One of the greatest challenges for Polish L1 speakers with German as an L2 is the German vowel system (Hentschel, 1986; Hirschfeld, 1998; Morciniec, 1990). In contrast to Polish, which has six vowels and no phonological vowel length contrast, German exhibits a relatively high number of 15 vowel phonemes and makes use of a phonological distinction between long/tense and short/lax vowels (Pompino-Marschall, 2009). This leads to a number of minimal pairs that could potentially cause communication problems for L2 learners of German (e.g., *Höhle* /høːlə/ "cave" versus *Hölle* /hœlə/ "hell" or *fühlen* /fyːlən/ "to feel" versus *füllen* /fʏlən/ "to fill"). From an orthographic point of view, the long vowels in *Höhle* and *fühlen* are explicitly marked for their length by so called *Dehnungs-h* ("lengthening *h*"), which is a silent letter and a reliable marker for a preceding vowel to be

long (Eisenberg, 2013). At the same time, the short vowels in *Hölle* and *füllen* are also explicitly marked for their length in that all German vowels which are followed by double consonant letters are short (Ramers, 1999). Since not all short and long vowels are explicitly marked for their length (e.g., *Boden* /boːdən/ "floor" or *Wolke* /vɔlkə/ "cloud")[1], German is an ideal testing ground for the investigation of the effects of orthographic markings on L2 vowel productions.

L2 pronunciation researchers and teachers assume that orthographic cues such as lengthening *h* can help German L2 learners establish different phonetic categories for short and long vowels and hence produce them more native-like (e.g., Dieling, 1983; Dieling & Hirschfeld, 2007). However, experimental evidence to test this assumption is still missing. Furthermore, German and Polish use the same graphemes to represent similar vowel phonemes. Escudero et al. (2014) have shown that the "congruency" between L1 and L2 grapheme-to-phoneme (G-P) correspondences plays an important role in L2 sound perception and can both help and hinder L2 word learning. The influence of (in)congruencies between L1 and L2 G-P mappings might also be relevant for L2 vowel production. Polish L2 German learners have been reported to produce German vowel quality either too tense or too lax, depending on the L2 vowel category and its similarity to the closest L1 sound (Morciniec, 1990; Müller, 2003). Some L2 German researchers have specifically commented on the production of German /eː/ and reported that this vowel is often diphthongized by Polish L2 German learners (Dieling, 1992; Hirschfeld, 1998; Müller, 2003), based on perceptual impressions. There is still a lack of experimental data to support these observations concerning vowel quality productions of Polish L2 German learners.

The present study set out to investigate whether orthographic cues in the German writing system really are used by L2 learners when learning to produce L2 sounds. Since both length (duration cues) and quality (spectral cues) distinctions are central to German vowel

acquisition, our research questions concern the production of both vowel length and quality. For the two dimensions, orthography is assumed to have an effect in different ways. As regards vowel length, we are specifically investigating the influence of L2 orthographic length markers (lengthening *h* and double consonant letters). With regard to vowel quality, we are more generally investigating how Polish L2 German learners produce German vowel quality, and subsequently discuss how their quality productions may be related to L1 and L2 G-P (in)congruencies. Our two main research questions are therefore:

1) Can explicit orthographic length markers in German help Polish L2 German learners produce German vowel length distinctions more native-like?

2) How are German vowel qualities produced by Polish L2 German learners and can L1 and L2 G-P (in)congruencies help explain possible deviations from native speaker productions?

By choosing Polish as L1 and German as L2, this study makes an important contribution to a number of studies which have already investigated orthographic effects in (L1 or L2) English (e.g., Bassetti, 2006; Bassetti and Atkinson, 2015; Dornbusch, 2012; Escudero, Hayes-Harb, & Mitterer, 2008; Mathieu, 2014; Simon, Della Chambless, & Alves, 2010).

L2 phonological acquisition and the role of orthography

In the field of L2 sound acquisition, two very influential models are the *Speech Learning Model* (SLM, Flege, 1995) and the *Perceptual Assimilation Model* (for L2) (PAM(-L2)) (Best, 1995; Best & Tyler, 2007). At the heart of both models lies the idea that difficulties in the perception of L2 sounds can be predicted based on the similarity/dissimilarity between the closest L1 and L2 sounds. If L1 and L2 sounds are

perceived as similar, learners are unlikely to differentiate them well, i.e., establish new L2 sound categories and produce them in a native-like manner. On the other hand, if sounds are sufficiently different, learners are likely to learn new sound categories well. Both the SLM and the PAM(-L2) focus on the perceptual side of the learning process, even though the SLM also proposes that accurate production depends on accurate perception. Interestingly, none of the models discusses the potential influence of orthography in any detail, even though it seems to be a well-acknowledged factor in the field of pronunciation teaching (e.g., Dieling 1992; Sobkowiak 2001) as well as in recent L2 speech research (Bassetti et al., 2015). The authors of PAM(-L2) do point out that, for example, English learners of French tend to equate French [ʁ] (a voiceless uvular fricative) with the phonetically very different English liquid [ɹ]. While this is difficult to be captured by a model that is based on perceptual similarity alone, this effect could be explained by the fact that the French and the English phoneme /r/ are represented by the same grapheme in the two orthographic systems. Yet, the authors of the most prominent L2 speech learning models do not elaborate on this or other orthography-related issues. Recent research into the role of orthography in L2 speech learning suggests that the inclusion of this factor may be a necessary step (e.g., Bassetti, Escudero, & Hayes-Harb, 2015; Erdener and Burnham, 2005; Escudero et al., 2008; Escudero et al., 2014; Mathieu, 2014; Showalter & Hayes-Harb, 2015; Young-Scholten, 2004).

Perception studies

A paradigm that is often used in perception studies is that of novel word learning, as it allows researchers to manipulate the orthographic input given to the participants (e.g., Escudero et al., 2008; Escudero et al., 2014; Mathieu, 2014; Showalter & Hayes-Harb, 2015; Simon et al., 2010). In these studies, participants learn new words paired with pictures of objects and, for some participants, the written form of the word as well. For example, in an

eye-tracking experiment by Escudero et al. (2008), 50 Dutch participants learned 20 English nonce words. Ten of these novel words were the critical test items which differed in the /ɛ/-/æ/ contrast (e.g., <tenzer> versus <tandik>), which is difficult for Dutch learners. 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 showed an asymmetric pattern where items containing [ɛ] were fixated more than words containing [æ]. Importantly, this was only the case for the group which had been learning the nonce words along with their spelling. The authors interpreted this finding as evidence that orthographic information is used to establish lexical-phonological representations of novel L2 words: In Dutch, only the letter <e> matches a front central vowel, while <a> corresponds to a back vowel, hence making the perception of the English front vowel [æ] while exposed to <a> unlikely. The study showed that L2 learners are influenced by their L1 orthographic system when processing the written and auditory cues of their L2.

Escudero et al. (2014) further investigated the role of the L1 orthographic system in relation to the orthographic system of the L2. Using a word-picture matching task, Spanish learners of Dutch were taught pseudowords in an auditory-only and an auditory-orthographic condition. Some of these new words formed perceptually difficult contrasts, such as /pʏx/ and /pyx/, and subgroups of these difficult contrasts were classified as either orthographically congruent or incongruent; for example, the /ʏ/-/y/ contrast was classified as “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/ and, with that, the same grapheme <u>. 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. The authors interpreted this finding as evidence for the reinforcement of congruent G-P correspondences during L2 language learning, while incongruent correspondences persistently interfered with the acquisition of an L2 phonology.

Another strand of research is concerned with the transparency and familiarity of orthographic systems and how this might influence L2 phonological and lexical representations (e.g., Dornbusch, 2012; Mathieu, 2014, Showalter & Hayes-Harb, 2015). In an auditory lexical decision task looking at the orthographic consistency effect in L1 and L2 speakers, English native speakers and advanced L2 English learners with Danish or German as their L1 judged English real and nonce words for their lexical status (Dornbusch, 2012). 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 were orthographically inconsistent (for example, /iːp/ can be spelled <eap> or <eep>). Dornbusch (2012) found an orthographic consistency effect 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 more slowly on auditorily presented inconsistent items than the other participants. This finding both supposes the existence of a link between orthography and auditory L2 lexical processing, and highlights the importance of the concept of orthographic depth: German participants coming from a transparent orthographic background are more affected by spelling-to-sound inconsistencies than Danish participants, whose L1 shows a high degree of complexity in the mapping between graphemes and phonemes and is therefore classified as a more opaque orthographic writing system. This research shows that effects caused by the L2 spelling (i.e., the orthographic consistency effect) originate in the characteristics of the L1 orthographic system.

Taken together, perception studies show that orthographic input may be beneficial in establishing lexical-phonological representations when G-P correspondences between the

native and target language are congruent and languages are similar in orthographic depth. This makes German and Polish an interesting case to study, as both languages use similar graphemes and have relatively transparent orthographic systems.

Production studies

One of the first studies to investigate orthographic effects in L2 production was conducted by Erdener and Burnham (2005). The authors presented 32 native Turkish speakers (transparent L1 orthography) and 32 native Australian English speakers (opaque L1 orthography) with Spanish (transparent orthography) and Irish (opaque orthography) nonword stimuli. Participants were asked to repeat the nonwords upon presentation and their productions were recorded and scored for phoneme errors. 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, the Turkish participants' performance was significantly worse than that of the Australian participants, while the performance of the Australian participants 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 correspond to the L2 phonological system in a straightforward way. This conclusion concurs with results from other perception studies (e.g., Dornbusch, 2012; Simon et al., 2010).

In another study on L2 pronunciation errors, Young-Scholten (2004) collected monthly production data from three American students over the course of a year which they spent at a German secondary school. The researcher auditorily analyzed and transcribed pronunciation errors concerning, for example, German final devoicing (e.g., [kɪnd] for <Kind> "child" instead of [kɪnt]). The results showed that the subjects had persistent

problems with the neutralization of the German voiced final stops. Young-Scholten hypothesized that orthographic input over the course of L2 learning was 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. Although the author did not explain the results with reference to “spelling pronunciation”, one could argue that producing *[kind] for <Kind> is exactly that, as the term is used for the pronunciation of a word according to its spelling (Neuman, 2013). Interestingly, “spelling pronunciation” was not used in any of the research articles presented here, while it is a well-acknowledged phenomenon in the field of pronunciation teaching. For example, Polish L2 English learners are found to produce sounds that would normally be silent, e.g., the <e> in the past-tense suffix <-ed> (Sobkowiak, 2001).

Another series of experiments was concerned with the effects of spelling on the pronunciation of known words in experienced Italian learners of L2 English (Bassetti and Atkinson, 2015). In one of these experiments, the authors collected production data on the past tense marker <ed> by means of a verb paradigm-production task. 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/ could be explained by the fact that <d> represents /d/ both in Italian and English (with only a few exceptions).

The authors further investigated the production of vowel duration as a function of orthographic marking through vowel digraphs in seven English word pairs where the two words contained the same target long vowel, spelled with either a double letter grapheme or not (e.g., <seen> versus <scene>). Productions were elicited by means of a reading-aloud task, with target words placed in a carrier phrase and repeated three times. The authors found that vowels spelled with digraphs were on average 237 ms (SD = 58 ms) long, while vowels

spelled with single letters were 29 ms shorter (208 ms, SD = 47 ms). 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 they may not be as stark in the lexical-phonological representations of learners. 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 experiments conducted by Bassetti and Atkinson (2015) can easily be interpreted as evidence for spelling pronunciation in L2 learners who apply their L1 G-P correspondences incorrectly, as in *[wɔlk] for <walk> or in *[askɛd] for <asked>. 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 direct G-P correspondence between a digraph and a long vowel in Italian. The marking of length in English must therefore be interpreted by means of a more abstract rule ("double vowel letter means long vowel") and cannot be explained by spelling pronunciation based on L1 orthographic rules alone. Unfortunately, English orthography is not very consistent in this regard. 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* /fuːd/.

Taken together, experimental studies on orthography and L2 perception and production point to the need of investigating further how *real words* and their phonological forms are learned by L2 speakers. 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, it is 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 relatively transparent orthographies, with German being an interesting example for marking vowel length in its orthography.

The German and Polish vowel system

For Polish learners of German, the German vocalic system is considered one of the most difficult phonological aspects to learn (Morciniec, 1990). While Polish is a consonant-rich language with over 30 consonant phonemes (Jassem, 2003), German contrasts a relatively large number of vowels. Most researchers agree that there are 15 contrastive vowels in German, excluding the diphthongs /aɪ, aʊ, ɔɪ/ and the two German schwa sounds /ə/ and /ɐ/ (Hall, 2011; Pompino-Marschall, 2009; Ternes, 2012). Eight of these 15 vowel phonemes are considered long, as they are on average twice as long as their short counterparts (Antoniadis and Strube, 1984). Polish, on the other hand, is described as a language with only six (short) vowel phonemes without any vocalic length contrasts (Gussmann, 2007; Hentschel, 1986; Sadowska, 2012).

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. In what follows, vowel length and vowel quality differences between the languages will be reported separately, so as to mirror the key aspects of the research questions above: research question 1 is concerned with vowel length differences between the speakers groups; research question 2 focusses on vowel quality differences.

Vowel length

L2 German researchers have stressed that long vowels are specifically difficult for Polish learners, who tend to substitute them with short vowels (Dieling, 1992; Morciniec, 1990; Slembek, 1995; Müller, 2003). Since Polish does not differentiate between short and long vowels, there are only few studies which have investigated Polish vowel length experimentally and reported the specific vowel durations. Yet, Polish does differentiate short and long consonants (Thurgood, 2001; Thurgood & Demenko, 2003; Rojczyk & Porzuczek, 2014). One strand of research investigated whether Polish vowels are shorter/longer before singleton versus geminate consonants (Rojczyk & Porzuczek, 2014; Nimz, 2016). Rojczyk and Porzuczek (2014) investigated vowel and consonant durations in the words *pana* [pana] ("gentleman") and *panna* [panːa] ("maiden") spoken in a carrier phrase. They found that vowels before the singleton consonant were 73 ms long, while they measured 85 ms before geminates. Nimz (2016) investigated five different Polish minimal pairs spoken in isolation and in a carrier phase. In contrast to Rojczyk and Porzuczek (2014), Nimz (2016) found that vowels before geminates were 13% shorter. On average, vowels in this study were 81.7 ms long before singletons and 92.1 ms before geminate consonants. When statistically controlling for how the geminates were realized (single versus double articulation), this difference did not reach significance. What is more, Slowiaczek and Szymanska (1989) tested whether Polish native speakers might 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 13% difference mentioned above), subjects did not attend to the difference in vowel durations and did not perform significantly above chance in identifying underlying voiced items. Even though final consonants are arguably phonologically different from geminate versus singleton consonants, these studies on Polish vowel duration suggest that this cue does not play a role in Polish phonology.

German long vowels are reported to be on average twice as long as their short counterparts. Antoniadis and Strube (1984) measured German vowel durations spoken in three different consonantal contexts in the form [CVCə] spoken in a carrier phrase. Their long vowels measured on average 154 ms and their short vowels measured about 72 ms. The comparison of the Polish and German duration data suggests that Polish vowels are comparable in their duration to the German short vowels, possibly slightly longer.

An important study which addresses vowel length production (in L2 Swedish) is McAllister et al. (2002). Their *Feature Hypothesis* states that, if vowel length is not used to signal a phonological contrast in the native language, L2 learners will have difficulties producing contrasts based on this feature. For Polish L2 German learners, this would mean that they would not produce length differences between German short and long vowels. However, as has been laid out in Section 2, orthographic marking could potentially help L2 learners produce this contrast more native-like.

Vowel quality

German long vowels are usually tense, while short vowels are lax. Two exceptions are the vowel pairs /a/-/aː/ and /ɛ/-/ɛː/, but the functional load of /ɛː/ is questionable as many German speakers realise it as [eː] instead of [ɛː] (Ternes, 2012). Traditionally, phonetically similar short and long vowels have been grouped into pairs, e.g., “a-pair” for /aː/ and /a/ (Sendlmeier, 1985), which is also reflected in the same graphemes for these vowels. In this study, the German vowel pairs /a/-/aː/, /eː/-/ɛ/, and /oː/-/ɔ/ are investigated for reasons described below. The closest Polish vowels to these German pairs are /a/, /ɛ/, and /ɔ/. Since it has been stressed by a number of researchers that a simple comparison of phonetic symbols is insufficient for determining phonetic similarity (Bohn, 2002; Flege, 1997; Strange, 2007), we draw on acoustic data from a study by Nimz (2016), in which she investigated the production

of German and Polish vowels as spoken by native speakers of each language. While the study was a small-scale study with only seven Polish males (between 17–18 years of age) from the East of Poland and four German males (all 18 years old) from the West of Germany, these are the most recent comparative acoustic data on the two vowel systems in question. Furthermore, the Polish data collection for the vowels /ɛ/ and /a/ differentiated between palatal (in Figure 1 marked with "+") and non-palatal consonantal contexts, as Jassem (2003) predicted that these vowels would be considerably fronted in palatal contexts. Figure 1 shows all Polish vowels in comparison to the six German vowels investigated in this study.

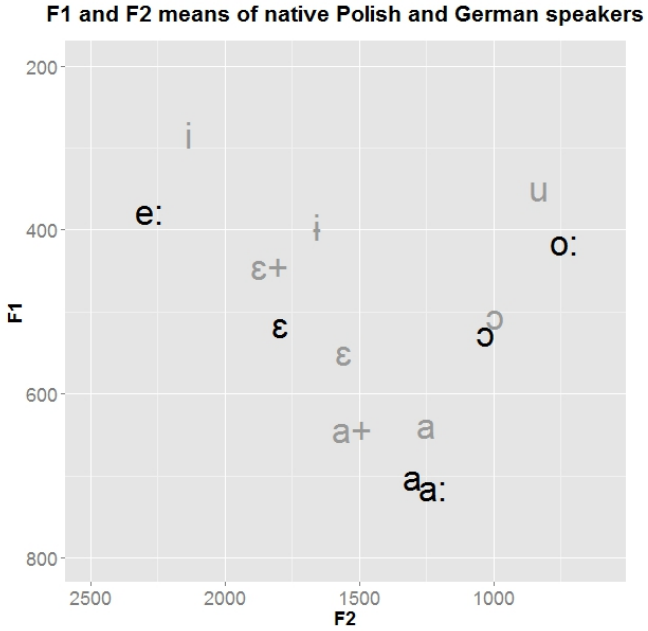

Figure 1. Average Polish (grey) and German (black) vowel qualities as spoken by native speakers; Polish vowels in palatal context are marked by "+" (N(Polish) = 7, N(German) = 4)

The vowel quality data corroborate Jassem's (2003) predictions and show that both /ɛ/ and /a/ are more front and, in the case of /ɛ/, higher in palatal contexts. Just looking at the non-palatal vowels, we see that Polish /a/ is produced slightly higher than German /a(ː)/ and

Polish /ɛ/ is produced slightly lower and further back than German /ɛ/. German and Polish /ɔ/ realizations are almost identical in their vowel qualities. With regards to the German long vowels /oː/ and /eː/, we see the largest quality differences in relation to the closest Polish vowels: German /oː/ is closest in its quality to Polish /u/ (not /ɔ/) and German /eː/ is closest in its quality to Polish /i/ (not /ɛ/).

These acoustic data are supported by perceptual data presented in Hentschel (1986). In this early study on the perception of German vowels by speakers with Polish as an L1, all German vowels were presented auditorily to naïve Polish listeners, who were asked to label these German vowels as one of the six Polish vowel categories (or as "foreign", if they found them to be too different). The acoustic similarities established above match his perceptual results in that the vowels that seem to be almost identical or close acoustically were clearly mapped onto one native category (i.e., German /ɔ/ to Polish /ɔ/, German /ɛ/ to Polish /ɛ/ and German /a(ː)/ to Polish /a/). At the same time, /eː/ and /oː/ were mapped onto two Polish categories: German /eː/ was perceived as Polish /i/ 74% of the time and as Polish /ɨ/ 23% of the time. German /oː/ was perceived as Polish /u/ 73% of the time and as Polish /ɔ/ 26% of the time. In an additional modified identification task, Hentschel (1986) further investigated whether Polish listeners would perceive the German long vowels as diphthongs. 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, a longer version of a Polish vowel, or a diphthongized version of a Polish vowel with either /i/ or /u/ as the second element. Even though his instructions might have influenced the answers of the participants[2], a surprisingly high percentage of the long vowels in question were judged to be diphthongs: the vowel /eː/ was perceived as a diphthong with /i/ as the second element 42% of the time, the vowel /oː/ was perceived as a diphthong with /u/ as the second element 41% of the time. Interestingly,

/aː/ was never perceived as a diphthong, which allowed Hentschel to formulate rules for diphthong perception regarding vowel length and tongue height.

From the contrastive analysis above, we would expect the German vowels /a(ː)/, /ɔ/, and /ɛ/ to be produced with native-like vowel qualities by Polish L2 German learners. Both the perceptual assimilation patterns as well as the acoustic data suggest that the German and Polish vowels are so similar that if Polish learners produce the closest Polish category, their vowel qualities will match those of German native speakers. Furthermore, the G-P correspondences of the two languages are congruent with respect to these vowels (see below).

With regards to /oː/ and /eː/, the analysis above would predict that Polish L2 German learners would produce deviant vowel qualities in comparison to German native speakers. The vowel /oː/ is acoustically most similar to Polish /u/ and predominantly perceived as such. Sometimes it is also perceived as /ɔ/ (26% of the time in Hentschel's study). The vowel /eː/ on the other hand is most similar to Polish /i/, both acoustically and perceptually. In some cases, it might also be perceived as /ɨ/ (23% of the time).

Furthermore, these two vowels are often perceived as diphthongs by naïve Polish listeners, and /eː/ (but not /oː/) has been reported to be produced as [ei] by a number of L2 German researchers (Dieling, 1992; Hirschfeld, 1998; Müller, 2003). However, up to this point, there does not exist any acoustic evidence for this claim. Finally, when comparing the L1 and L2 orthographic systems and the respective G-P correspondences of the vowels /eː/ and /oː/, they behave very differently from the other categories (see below).

Orthographic representation

The orthographic representation of the 15 German vowel phonemes is covered by 8 vowel graphemes. Table 1 shows the German G-P correspondences according to Eisenberg (2013).

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

Table 1. German G-P correspondences on the basis of Eisenberg (2013)

With the exception of /iː/ and /ɛː/, it becomes apparent that most vowels of a short-long pair are mapped onto the same grapheme, for example /aː/ and /a/ are both mapped onto <a>. Still, there are additional orthographic markings to differentiate between the long and shorts vowels of a pair. For example, the doubling of vowels is used as a means to explicitly mark long vowels (e.g., *Boot* /boːt/ “boat”). 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 which shows that less than 1% of all long vowels are marked by means of double vowel letters. On the other hand, /eː/, /oː/, and /aː/ are represented with lengthening *h* about 12% of the time, which is why these vowels were chosen for the present study.

Lengthening *h* is a silent letter which reliably marks its preceding vowel as long. It precedes the sonorant sounds [r, l, m, n]; this is a necessary but not a sufficient condition in that there are words with long vowels written as <Bohne> (“bean”) or <Sohn> (“son”), but also words without the lengthening *h* such as <Monat> (“month”) or <Ton> (“sound”). This characteristic made possible the experimental approach of the study, in which the productions of real words were investigated without having to manipulate the test items, as was the case for most previous studies.

All six Polish oral vowels are symbolized by one corresponding grapheme, with the exception of /u/, which may be represented in the Polish orthography as <u> or <ó>[3] (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

With few exceptions, G-P correspondences in Polish are very consistent, which is why it is generally classified as a language with transparent orthography (Kaminska, 2003). German may be classified as slightly less transparent than Polish in that speakers/readers cannot rely on G-P correspondences alone, as German makes use of additional ways of marking vowel length, such as silent lengthening *h*. While the function of this marking is specific to German, it may help L2 learners in acquiring some German words more native-like (research question 1).

When comparing the German and Polish graphemes and their respective corresponding phonemes, we see that the German graphemes relevant to this study (<a> for /a/-/aː/, <e> for /eː/-/ɛ/, <o> for /oː/-/ɔ/) find corresponding counterparts in Polish, where these graphemes correspond to Polish /a/, /ɛ/, and /ɔ/. However, it has to be kept in mind that the German vowels /eː/ and /oː/, unlike the other German vowels, are each mapped onto two different Polish categories: German /eː/ is perceived as Polish /i/ or /ɨ/ (which in turn correspond to the Polish graphemes <i> and <y>), while German /oː/ is perceived as either Polish /ɔ/ or /u/ (which in turn correspond to the Polish graphemes <o> and <u>) . These incongruencies could lead to potential problems in vowel quality productions, as Escudero et al. (2014) have already shown effects of (in)congruencies in L2 perception (research question 2).

Methodology

Two groups of German and Polish participants were recorded during a picture-naming task and their vowel productions were analyzed acoustically both for their vowel length (duration in ms) and vowel quality (F1 and F2).

Participants

The participants were recruited at a Polish high school in the East of Poland (experimental group) and at a German high school in the West of Germany (control group). In both regions, Standard Polish and Standard German are spoken and in an additional questionnaire, none of the participants reported to speak a specific dialect. The Polish school put special emphasis on German as a Foreign Language (GFL), which was important for the study in two respects: first, participants had to be advanced enough to be able to name the experimental items. Second, it was important to recruit participants with an extensive amount

of written input in order to be able to make claims about the role of orthography in the acquisition of a second phonology.

At the Polish high school, students from the so-called bilingual branch 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. In a preparative interview with a German teacher of the school, it was stated that students also receive a few hours of phonetic instruction; however, in the questionnaires administered after the experiment hardly any of the students mentioned phonetic training (2 out of 22). Most likely the phonetic instruction was not very extensive. The German lessons at the Polish school 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 (CEFR)), which, if they pass, classifies them as intermediate speakers. This qualifies them to join the bilingual branch of the school. During the remaining three 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, 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 CEFR 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 bilingual branch (during the first half of their school year). 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). Twenty-two Polish students took part in the experiment, of which two were excluded from further analysis because they did not know at least 75% of the test words used, which was

established as the cutoff point before the acoustic analysis was carried out. Two further participants were excluded 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, and subject P22 spoke Bulgarian as a native language. All remaining Polish participants were native speakers of Polish and late foreign language learners of German, that is, none of them had received any intensive GFL teaching (more than 6 hours per week) before they entered high school at around the age of 15. All of them spoke English as their first foreign language. All of them spoke Polish (and no other language) at home. None of them reported any hearing or learning problems. The average age of the remaining Polish participants was 18.5 years (SD = 0.6); four of them were male.

Twenty-one German native speakers were recruited at a German high school as a control group. One student reported to be dyslexic so this participant was subsequently excluded. None of them reported any other hearing or learning problems. All of them reported to speak German at home and had learnt English as their first foreign language at school. The average age of the 20 German participants was 17.9 years (SD = 1.1); five of them were male.

Experimental items and conditions

The primary prerequisite for the experiment was picturability, as a simple reading task would not allow for the differentiation between orthographic effects in reading and actual phonological representation. Secondly, words had to be familiar to the majority of the participants. Because it was assumed that a simple frequency measure might not reflect the actual use of words in a foreign language classroom, German teachers at the Polish school where the experiment was to be conducted rated the experimental items on a scale from 1 (“Very unlikely that the students know this word”) to 7 (“Very likely that all students know

this word") depending on how familiar they thought their students were with the test items. Only items which reached an average familiarity of at least 5 were included in the final items list.

The vowels investigated in this study were the long vowels /eː/, /aː/, /oː/, and their short counterparts /ɛ/, /a/, /ɔ/. For each vowel in the experiment, eight test items were chosen, of which four were explicitly marked for their length (see Appendix). The marked long vowels were all written with lengthening *h*; the marked short vowels were all followed by double consonant letters. Half of all test items were one-syllable words and the other half were two-syllable words. The difference in orthographic marking (*marked* versus *unmarked*) constituted one of the main experimental factors of the study, together with the length of the vowel (*short* versus *long*). In all, there were 48 test items (6 vowels x 8 words).

Procedure

Pictures of all the test items were presented on an Acer Timeline laptop using PowerPoint slides. The same order was used for each participant and no filler items were used. Subjects were instructed to name the pictured items as they appeared on the screen. There was no previous practice session, and the productions were not timed. Once the target word was named, the German experimenter, who did not speak Polish, presented the next picture. If a different word than the target was produced, the 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, in order to prompt an authentic production even if the picture could not be named. If a participant still did not know the respective word, the experimenter would produce the word and the participant would repeat it. These repetitions (8% of the whole dataset) were not included in the final analysis. In the second production

cycle of the experiment, the same pictures were presented in a different random order. 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.

After the main test phase, a post-test was administered which addressed the spelling skills of the Polish participants. Since it was crucial that the participants knew the orthographic representation of the words (in order to make reliable claims about the influence of orthography), they were ask to write down each word next to the pictures they had just seen in the oral production task. The post-test was administered after the main experimental task in order not to prime the participants in any possible way. Those items which were spelled incorrectly (e.g., *<Schaff> for <Schaf> /ʃaːf/ “sheep”) were excluded from the analysis (14 words in the whole data set). Finally, all participants were given questionnaires which collected additional information concerning the participants’ detailed language (learning) background as well as small test concerned with the orthographic rules for the marking of short and long vowels.

Acoustic analysis

The productions of 38 participants were analyzed using PRAAT (Boersma & Weenink, 2014). Target vowels were manually labeled, with vowel onset being defined as the first positive zero crossing of the first periodic waveform, and vowel offset being the last positive zero crossing before the following consonant. Additionally, when the surrounding consonants were voiced obstruents or sonorants, vowel on- and offsets were labeled by taking into account formant energy as well as changes in the waveform. Figure 2 exemplifies the segmentation of the vowel /eː/ in the test word *Zehn* /tseːn/ (“ten”). The beginning of the vowel is marked at the point where the periodic waveform begins, while vowel offset is

marked where formant energy fades out, combined with a lower amplitude in the waveform. Sixty-one vowels were labeled as missing data due to background noise, incorrect utterances, unnatural (exaggerated) productions, unreliable segmentation, or interruptions.

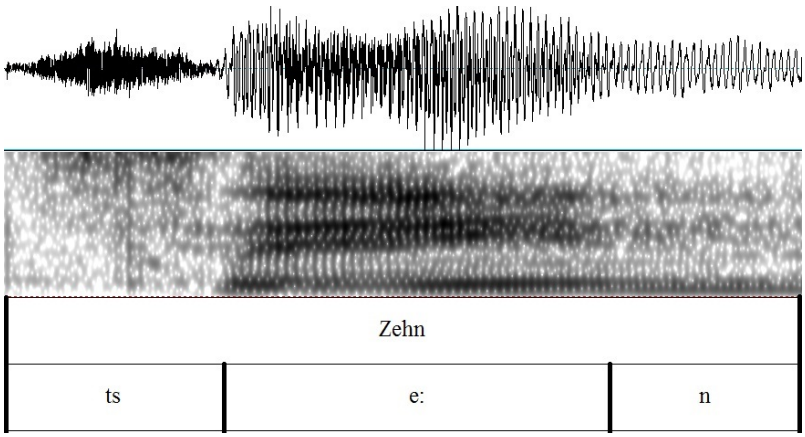

Figure 2. Segmentation of the vowel /e:/ in the test word *Zehn*, produced by a Polish learner

In order to check the reliability of the first author's acoustic measurements, a student assistant was hired to segment about 20% of the same production data. The assistant was given the same segmentation criteria the first author used in segmenting the items. 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 indicates acceptable reliability (Anderson-Hsieh et al., 1992).

Statistical analysis

The following graphs and statistical models are based on all available data points excluding the repetitions, misspelled items, and other missing data due to reasons mentioned in the methodology section. For the German native speaker group, 1902 data points served the vowel length analysis, which was operationalized by vowel duration in milliseconds from

vowel on- to offset. For the Polish L2 learner group, 1384 data points could be used to investigate vowel length production.

The investigation of vowel quality was operationalized by measuring the first (F1) and second formant (F2) by means of a PRAAT script[6] at vowel midpoint, which is common practice for describing vowel qualities in the field (e.g., Bohn & Flege, 1992; Darcy & Krüger, 2012; Escudero, Simon, & Mitterer, 2012; Flege, Bohn, & Jang, 1997). 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 investigate whether /eː/ really is diphthongized, as was claimed before (Dieling, 1992; Hirschfeld, 1998; Müller, 2003). Since speaker sex is a confounding factor in formant measurements, as female formants tend to be higher than those of male speakers (Simpson & Ericsdotter, 2007), only the data of the female participants served the vowel quality analysis (14 Polish and 15 German female speakers). Items for which either F1 and/or F2 could not be measured reliably were further excluded from the analysis. In the end, 2477 data points served the vowel quality analysis.

All analyses were carried out using the R statistical environment (R Core Team, 2017) and fitting series of linear mixed models (LMMs) to the data by means of the R-package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). All p-values were obtained by likelihood ratio tests of the final model with the effect in question against the model without the effect in question (Winter, 2013). The results will be reported separately for vowel length (research question 1) and vowel quality (research question 2).

Results

Vowel length

Figure 3 shows the average durations of German short and long vowels in explicitly marked (presence of lengthening *h* or double consonant letters) and unmarked (no explicit orthographic marking) conditions as produced by German native speakers. The error bars show +/- 2 SE of the aggregated duration data per subject in each condition (i.e., long-marked, etc.).

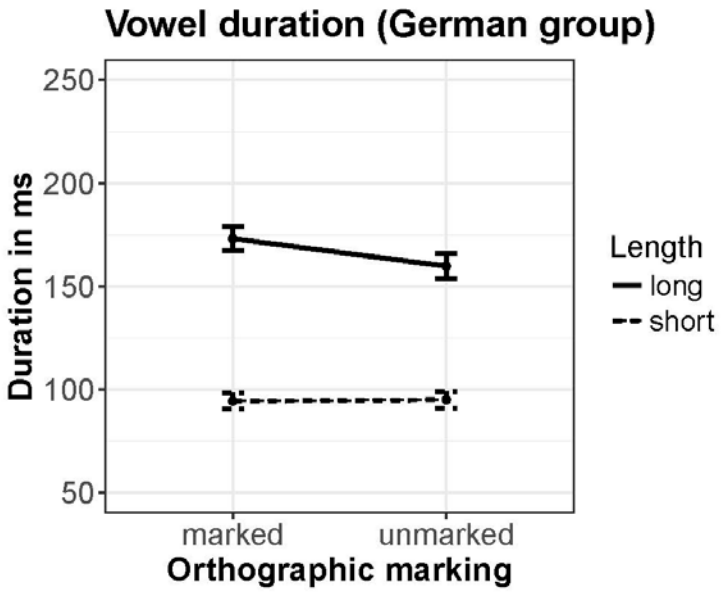

Figure 3. Duration (in ms) of German short and long vowels for orthographically marked and unmarked vowels produced by German native speakers

From the visual inspection it becomes clear that German native speakers produce long vowels with significantly longer durations than short vowels. German short vowels are on average 94.7 ms long, while German long vowels are about 76% longer and measure 166.4 ms on average. With that, vowels in this study are slightly longer than those in Antoniadis and Strube (1984), possibly because words in our study were spoken in isolation and not in a carrier phrase. It seems that long vowels which are explicitly marked in their orthographic representation are produced slightly longer than unmarked ones. On average, marked long

vowels are produced 13.3 ms longer than unmarked long vowels, while marked short vowels are produced only 0.6 ms shorter than their unmarked counterparts.

In order to test whether this interaction is statistically significant, an LMM was fit to the duration data with random intercepts for participants and items (Cunnings, 2012). By-participant random slopes for the effect of LENGTH were further added, as they improved model fit. Random slopes account for the fact that participants and items may vary with regards to how sensitive they are to the experimental variables. It is crucial to add random slopes as they decrease Type I error rates, i.e., rejecting the null hypothesis when it is true (Barr, Levy, Scheepers, & Tily, 2013; Winter, 2013). Besides the main fixed factors of interest, i.e., LENGTH and MARKING, it was decided to further add the variables RUN (first or second production cycle) and VOICING of the following consonant as additional control factors. Instead of averaging over the first and second production cycle, we included RUN in order to control for pseudoreplication and at the same time be able to keep all available data points (Winter, 2011). VOICING was included in order to account for the fact that the consonantal environment could not be completely balanced in our experimental design, even though we know that the voicing of the following consonant can affect the duration of the preceding vowel (Chen, 1970). Model comparisons by means of likelihood ratio tests revealed that only the factors LENGTH ($\chi^2(1) = 76.11$, $p < 0.001$) and RUN ($\chi^2(1) = 132.41$, $p < 0.001$) improved model fit significantly. The effects of MARKING ($\chi^2(1) = 1.19$, $p = 0.28$) and VOICING ($\chi^2(1) = 0.01$, $p = 0.94$) did not turn out to be significant; neither did the interaction between LENGTH and MARKING ($\chi^2(1) = 1.49$, $p = 0.22$).[4]

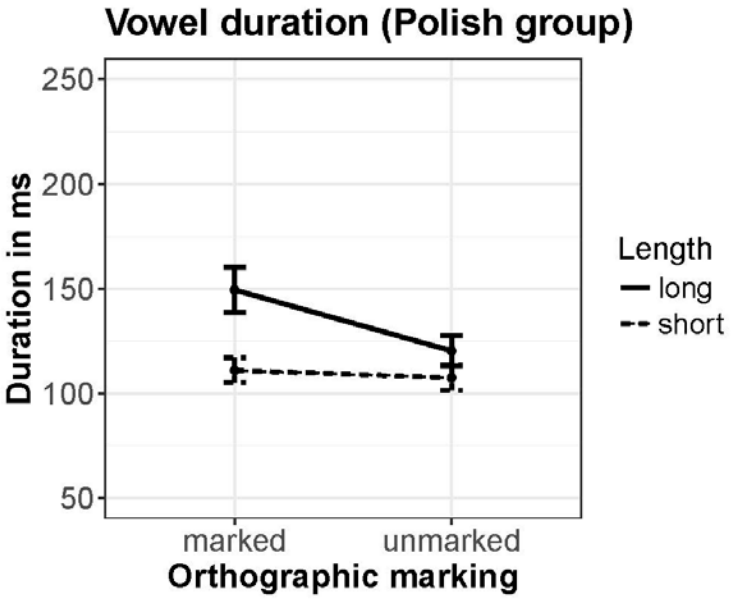

Figure 4. Duration (in ms) of German short and long vowels for orthographically marked and unmarked vowels produced by Polish native speakers

Figure 4 shows the average vowel durations of the Polish L2 German leaners. In contrast to the German native speakers, marking seems to have a strong effect on the production of vowel length: While unmarked short vowels are on average 107 ms ± 3.0 ms (SE) long, marked short vowels measure on average 110 ms ± 3.0 ms (SE). Long vowels, on the other hand, measure only 120 ms ± 3.0 ms (SE) when they are not explicitly marked, but 149.3 ms ± 5.3 ms when marked by lengthening *h* in their orthographic representation. This means that in the unmarked condition, long vowels are only 12% (13 ms) longer than the respective short vowels. In the marked condition, this difference is three times as large (36%). Again, a LMM was fit to the duration data with random intercepts for participants and items. Model comparisons with the same fixed factors as in the German group revealed that LENGTH ($\chi^2(1) = 17.31$, $p < 0.001$), MARKING ($\chi^2(1) = 9.57$, $p = 0.002$, and RUN ($\chi^2(1) =$

111.70, $p < 0.001$) all improved model fit significantly. The control variable VOICING again did not improve model fit ($\chi^2(1) = 0.32$, $p = 0.56$). In contrast to the German group, the interaction between LENGTH and MARKING was significant in the Polish group ($\chi^2(1) = 7.11$, $p = 0.01$).[5]

Vowel quality

Figure 5 summaries the mean formant frequencies of the six German vowels of interest as produced by Polish L2 German learners and German native speakers.

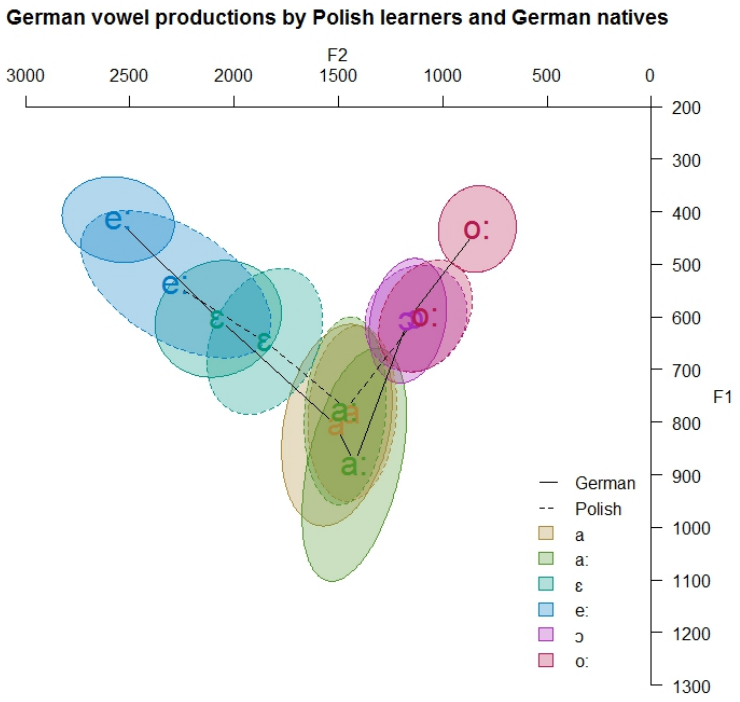

Figure 5. Mean formant frequencies for the German vowels /a, aː, ɛ, eː, ɔ, oː/ produced by Polish L2 German learners and German native speakers. Ellipses correspond to a confidence level of ± 1 standard deviation from the bivariate mean.

The graphical display shows German /oː/ to be produced by Polish L2 German learners with vowel qualities identical to /ɔ/. With that, the Polish group does not differentiate German /oː/ and /ɔ/ quality-wise, and only /ɔ/ is produced native-like by the Polish speakers. To test this statistically, separate LMMs were fit to the F1 and F2 data for all vowel categories with LANGUAGE (L1 Polish or L1 German) as fixed factor, random intercepts for participants and items, and by-item random slopes for the effect of language, if these improved model fit. Table 3 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 ($\chi^2(1) = .10$)	p < 0.001 *** ($\chi2(1) = 11.65$)	p = 0.53 ($\chi2(1) = .40$)	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 ($\chi2(1) = .91$)	p = 0.14 ($\chi2(1) = 2.16$)	p = 0.002 ** ($\chi2(1) = 9.80$)	p < 0.001 *** ($\chi2(1) = 13.14$)

Table 3. p-values of likelihood ratio tests for the factor language on F1 and F2 for all vowels of interest. Significant results are marked with asterisks.

While F1 and F2 of the vowel /oː/ are significantly different, the values for /ɔ/ do not differ significantly. The Polish and German productions of German /a/ also overlap almost entirely and do not differ significantly in either F1 or F2. As far as /aː/ is concerned, the Polish native speakers produce it significantly higher than the German native speakers.

German /ɛ/ is produced further back and slightly more open by Polish L2 German learners than by German native speakers. Only the F2 values differ significantly though. While the German group produced a clear difference between /eː/ and /ɛ/, the productions by

the Polish group span over German /eː/ and /ɛ/. Both F1 and F2 for this vowel category differ significantly between the groups. The large span of the Polish ellipsis can be related to formant movements within this vowel category.

In order to investigate this further, F1 and F2 measurements at 25% and 75% of each vowel production were plotted for /eː/. Figure 6 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.

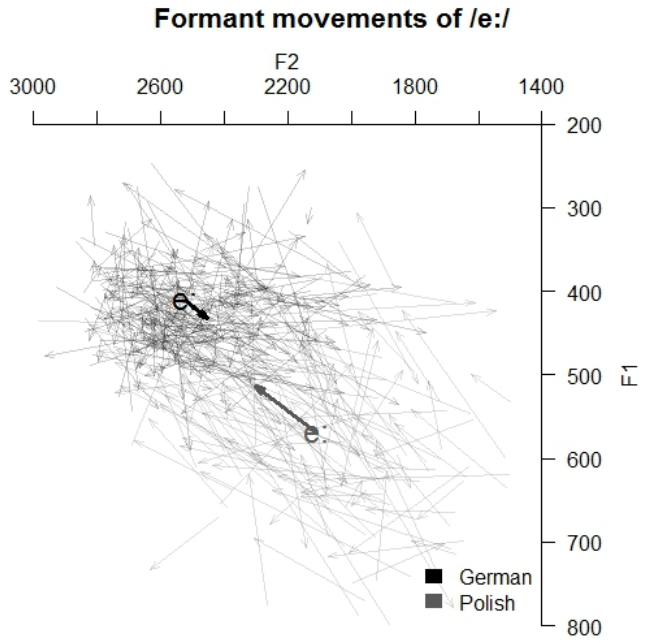

Figure 6. Formant movements for the German vowel /eː/ in Polish GFL learners and German native speakers. Overall group means at 25% and 75% are represented in bold.

While there is obvious variation between the speakers, two important aspects still become visible in the above graph. First, Polish speakers show a much greater formant movement for /eː/. This is also exemplified in the spectrogram in Figure 2 above, where we see clear formant movements in the test word *Zehn* as produced by a Polish L2 German

learner. Second, the direction of the movement is diametrically opposed to the German native 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: $p < .001$ ($\chi2(1) = 52.16$); German group: $p = 0.002$ ($\chi2(1) = 10.06$)). The changes in F2 are only significant in the Polish group ($p < .001$ ($\chi2(1) = 59.88$); German group: $p = .08$ ($\chi2(1) = 3.13$)). In her study on L2 learners' productions of English vowels, Steinlen (2009) had used a benchmark of 10% to characterize significant formant movement. With this benchmark, the average German production of /eː/ is negligible. The average Polish production of German /eː/, however, shows F1 and F2 movements that are very close to the 10% benchmark (F1: 9.8%; F2: 9.1%). At first sight, the fact that Polish L2 German learners tend to show significant formant movements (only) for German /eː/ is puzzling. For example for /oː/, spectral changes were as small as 1.2% (F1 decrease) and 3.8% (F2 increase). In the following discussion, we explain this differences between the vowel categories with reference to the L1 and L2 G-P (in)congruencies between the German and Polish orthographic systems.

Discussion

Research question 1 concerned the effect of explicit orthographic markers on vowel length production and whether these could help Polish L2 German learners produce German vowel length distinctions more native-like. For the German native speakers, we found that long vowels in our data set were on average 76% longer than short vowels. Interestingly, long vowels written with lengthening *h* were produced 13.3 ms longer than unmarked vowels by the German native speakers; however, this difference was not significant. In the Polish group, on the other hand, the interaction between LENGTH and MARKING was significant: As can be seen in Figure 5, error bars overlap in the unmarked condition, and unmarked long vowels are on average only 12% longer than unmarked short vowels. In the marked condition,

however, long vowels are produced 36% longer than the short vowels, which is a difference that is still not as large as in the native speaker group, but substantially more native-like than the difference in the unmarked condition. This difference was driven by the orthographic marking of the long vowels, i.e., lengthening *h*. Short vowels were produced with very similar vowel durations in all groups, independent of the orthographic marking. This could be due to the fact that Polish vowels are comparable in their duration to the German short vowels, hence, explicit orthographic marking of short vowels is not a necessary cue for Polish L2 German learners.

Even though lengthening *h* seems to help Polish L2 German learners produce German long vowels longer, it would need to be investigated further whether the difference is in fact enough for German native speakers to perceive them as target-like. Nooteboom and Doodeman (1980) show that Dutch speakers are able to notice vowel duration changes of only 6% in a forced-choice identification task. Yet, depending on the native phonological system, this high sensitivity for durational contrasts may vary (Altmann, Berger, & Braun, 2012; McAllister et al., 2002). Polish L2 German learners, for example, perform at chance level when they discriminate German non-sense word pairs (e.g., /baːp/-/bap/) which differ only in duration and where the long vowels are about twice as long as their short counterparts (Nimz & Khattab, 2015). German native speakers, on the other hand, might be very sensitive to vocalic durational differences; for example, they were found to perform slightly above chance in an experiment on incomplete neutralization of voiced final stops where stimuli differed (on average) by only 8 ms, along with other cues such as loudness (Röttger, Winter, & Grawunder, 2011).

Furthermore, vowel length correlates with vowel tenseness and for high and mid vowels, vowel quality seems to be even more important for vowel identification in German native speakers than vowel length (Sendlmeier, 1981; Nimz, 2015). For this reason, the

investigation of vowel quality productions by Polish L2 German learners becomes essential. Hence, the second research question was concerned with how German vowel qualities are produced by Polish L2 German learners and, further, how L1 and L2 G-P (in)congruencies might play a role in possible deviations from native speaker productions.

As concerns the productions of the German short vowels /ɔ/ and /a/, Polish L2 German learners produced the L2 vowel qualities native-like. As was shown earlier in the discussion of Hentschel's (1986) perceptual data, these two German categories are clearly mapped onto the respective Polish categories and are very similar acoustically (see Figure 1). German /ɛ/ and /aː/ show a little less overlap, in that German /aː/ is produced slightly higher and German /ɛ/ slightly further back by the learners. This, too, can be explained well with reference to the earlier acoustic comparison of the two vowel systems, as non-palatal Polish /ɛ/ is produced further back and slightly lower than German /ɛ/, and non-palatal Polish /a/ is produced slightly higher than German /aː/. Despite these differences, we see a large overlap between the German and Polish productions of German /ɛ/ and /aː/, which is why it can be assumed that the Polish learners' vowel quality productions will most likely be perceived as the respective German qualities. Further research with native German listeners will have to test this assumption empirically. Additionally, it would be of interest to combine vowel length and vowel quality perception. The duration of the German short vowels does not seem to be very problematic for Polish L2 German learners, as the average Polish short vowel durations are very similar to the German ones; however, long German /aː/, for example, will most likely be produced too short by Polish L2 German learners, unless it is explicitly marked for its length orthographically. Nimz and Khattab (2015) found that German native speakers are not able to distinguish /aː/-/a/ when length differences are neutralized, which stresses the fact that, especially for /aː/, Polish learners need to pay special attention to the length contrast. As was mentioned in the description of the participants above, it is unlikely

that the Polish participants were aware of this (only 2 out of 22 participants mentioned phonetic instruction in the questionnaire). We assume that, with more explicit phonetic instruction, Polish L2 German learners are able to produce vowel length differences more native-like (see also Saito, 2013). Additionally, this instruction needs to comprise information about the L2 orthographic system as well. For example, only seven out of all Polish participants could (explicitly) name the function of lengthening *h*. We might have seen an even stronger effect if phonetic/orthographic instruction had been more extensive. Furthermore, none of the Polish participants mentioned more "implicit" orthographic markers, such as the fact that if a vowel is followed by only one consonant in the writing, it is likely that it is long. As discussed in the first footnote, such implicit marking still finds little attention in publications on German pronunciation training (but see, for example, Hirschfeld et al., 2007).

The quality analysis of German /oː/ and /eː/ showed that these two vowels are produced very differently by Polish L2 German learners in comparison to German native speakers. Unlike the other vowel categories, /oː/ and /eː/ are perceptually mapped onto two different Polish categories (Hentschel, 1986). As described in the section on the German and Polish vowel systems, German /eː/ is perceived as Polish /i/ or /ɨ/, and German /oː/ is perceived as Polish /u/ or /o/. Assuming a close relationship between L2 sound perception and production (Flege, 1995), it is not very surprising that the Polish productions of these vowels differ significantly from those of German native speakers. However, the way in which the two vowel qualities deviate is more intriguing, as their specific productions can mainly be explained with reference to the L1 and L2 orthographic systems of the learners.

With regards to /oː/, it would have been more likely that Polish L2 German learners produce it with the quality of Polish /u/ instead of /ɔ/, as in 73% of the cases, naïve Polish listeners map German /oː/ onto Polish /u/ (Hentschel, 1986). We suggest that orthography can

explain this rather unexpected production, as the two languages use the same grapheme <o> for two very different vowel qualities: While in German the grapheme <o> is used to represent /oː/, Polish speakers associate <o> with /ɔ/ in their L1. At the same time, German /oː/ is at least sometimes (26% of the time) perceived as /ɔ/ by naïve Polish listeners, which is why the misleading graphical representation of the German vowel (in this case) is partially supported by the L2 perceptual system. In other words, the grapheme <o> might suggest to a Polish learner that the vowel in question is /ɔ/ (based on the L1 orthographic system) and since it is partially mapped perceptually to Polish /ɔ/ as well, the learner produces /ɔ/, which however deviates greatly from native speaker productions.

As has been reported earlier, the vowel /eː/ is indeed diphthongized by Polish L2 German learners. It is unlikely that perception (alone) can explain the production of /eː/. As was presented above, Hentschel (1986) investigated the perception of German long vowels as diphthongs. He found that both /eː/ and /oː/ are perceived as sequences of vowels in over 40% of the cases. Yet, only /eː/ is *produced* as a diphthong. Again, the G-P correspondences of the L1 and L2 can help explain this finding. In the case of /eː/ (in contrast to /oː/), orthographic and perceptual facts do not support each other. The grapheme <e> suggests Polish /ɛ/ according to the L1 orthographic system of the learners, but /eː/ is never perceptually assimilated to this sound. Thus, the diphthong may be an attempt by the learners to combine orthographic and perceptual influences: they incorporate the orthographic and perceptual interferences by starting out with the Polish spelling production of <e>, but satisfy their auditory perception by moving towards the quality of a higher vowel. Rafat and Stevenson (2018) refer to similar effects in English-speaking learners of Spanish as an orthographic McGurk(-like) effect. In their study, English L2 Spanish learners produced non-target-like sound combinations such as [*lj] for Spanish words like <pollo> (“chicken”), which is

interpreted as an auditory-orthographic fusion effect, as <ll> would be pronounced [l] in English but [j] in Spanish.

Current models of L2 speech learning do not yet account for such effects, even though previous research has already shown for other languages (with mainly L2 English) that orthographic effects in L2 perception and production are ubiquitous and should receive more attention in the field of L2 speech (Bassetti et al., 2015; Colantoni, Steele, & Escudero, 2016). As we demonstrated for Polish L2 German learners, explicit orthographic markers such as lengthening *h* can help produce a phonological contrast more native-like. While this is only true for long vowels which are explicitly marked, it is clear that orthography plays a role in the production of L2 sounds. Additionally, in the analysis of vowel quality productions, the learners' productions of German /oː/ and /eː/ can again be explained by drawing on orthographic effects. We argue that (in-)congruencies between the L1 and L2 G-P correspondences of the two languages investigated are responsible for the deviant L2 productions. Consequently, orthography can both help and hinder L2 sound production. Flege (2016) briefly discusses the influence of orthography by making reference to Italians who need to learn English /ɪ/, a vowel not found in their L1 Italian. The closest Italian vowel is written with <i>, while in English ("alas"), /ɪ/ is written with <i> and /i/ is written with <ee>, etc. Still, Flege (2016) concludes that "orthography creates problems that have nothing to do with speech learning". While we agree that speech learning fundamentally depends on speech perception (Best & Tyler, 2007; Flege, 1995), we would still draw the conclusion that in Flege's example above (similar to our Polish vowel quality data) orthography plays a crucial role–at least when it comes to phonological representations at the lexical level. In our study, we investigated the production of vowels in real words, hence the participants drew on their lexical representations to produce the words in question. A few studies have shown that it is possible that the development of a lexical contrast can precede reliable phonetic category

formation (Darcy et al, 2012; Darcy et al., 2013) and it may be possible that orthography and explicit instruction might help learners to bootstrap the contrast lexically (Darcy et al., 2013: 416–417). It is still unknown exactly how knowledge of contrastive spellings of vowels (e.g., <Höhle> [hø:lə] "cave" versus <Hölle> [hølə] "hell") would translate into knowledge of contrastive phonological content of words if it is not through perceptual learning alone; "more research is needed to determine the role of orthographic knowledge in L2 phonological development" (Hayes-Harb & Masuda, 2008: 28).

Bassetti et al. (2015) attribute the disregard of orthographic effects to the fact that language teaching has been dominated by the communicative approach as well as to the general primacy of spoken over written language in linguistics (see also Derwing, 1992). Yet, in the field of (German) L2 teaching, a new strand of research has emerged which stresses the fact that explicit orthographic instruction can promote metalinguistic knowledge and, with that, help acquire an L2 (phonological) system more successfully (e.g., Bredel, 2013; Röber, 2012).

Conclusion

This is the first study to have acoustically investigated the production of German vowels by Polish L2 German learners with regard to the role of orthography in both vowel length and vowel quality production. Our results show that lengthening *h* helps Polish learners produce German long vowels significantly longer and, with that, more native-like. At the same time, deviations in vowel quality productions may be attributed to (in)congruencies between the L1 and L2 G-P correspondences. While the effects of orthography both help and hinder native-like L2 vowel production and are located at different levels of orthographic structure, it is important to recognize the influential role orthography can play in L2

phonological acquisition. This acknowledgement is vital for the advancement of both theoretical models and pronunciation training.

Notes

[1] It could be argued that a word such as <Boden> is implicitly marked for vowel length because <o> is followed by one consonant only. In German, stressed open syllables must always contain a long vowel (Eisenberg, 2013). Hence, <Boden> has to be pronounced ['boː.dən] because the first syllable is stressed and open, since <d> belongs to the second syllable. Similarly, <Wolke> is implicitly marked because the first stressed syllable is closed: ['vɔl.kə]. Yet, this "implicit" marking finds little mention in publications for German pronunciation training (e.g., Dieling, 1992; Neumann, 1981; but see Hirschfeld et al., 2007).

[2] By giving the options "long vowel" or "diphthongized vowel", Hentschel implied 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.

[3] 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).

[4] R-syntax of the final model for the German duration data: model.german = lmer(Duration ~ Length + Run + (1+Length|ID) + (1|Word), data=german, REML=FALSE).

[5] R-syntax of the final model for the Polish duration data: model.polish = lmer(Duration ~ Length*Marking + Run + (1+Length|ID) + (1|Word), data=polish, REML=FALSE).

[6] 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 /oː/: 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.

[7] Attention was paid to exclude German-Polish cognates. However, choices for the one-syllable, unmarked /eː/-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". Statistical analyses showed that the inclusion of *Keks* did not change the results.

References

Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, 51, 1219–1223.

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.

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

Archibald, J. & Young-Scholten, M. (2003). The second language segment revisited. *Second Language Research, 19*, 163–167.

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

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

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

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

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

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer (Version 5.0.25) [Computer software]. Retrieved from http://www.praat.org.

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

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

Bohn, O.-S. & Munro, M. J. (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. *Studies in Second Language Acquisition*, *19*, 447–465.

Bredel, U. (2013). Silben und Füße im Deutschen und Türkischen. Orthographieerwerb des Deutschen durch türkischsprachige Lerner/innen. In A. Deppermann (Ed.), *Das Deutsch der Migranten*. (pp. 368–390). Berlin: De Gruyter.

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

Colantoni, L., Steele, J., & Escudero Neyra, P. R. (2016): *Second Language Speech. Theory and Practice* (Reprint). Cambridge: Cambridge University Press.

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

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

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. *Second Language Research*, *28*, 5–40.

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

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

Derwing, B. L. (1992). Orthographic aspects of linguistic competence. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The Linguistics of Literacy* (pp. 193–210). Amsterdam: Benjamins.

Derwing, T. M.; Munro, M. J. (2015). *Pronunciation Fundamentals. Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: Benjamins.

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. Fremdsprachenunterricht in Theorie und Praxis.* Berlin, New York: Langenscheidt.

Dieling, H., & Hirschfeld, U. (2007). *Phonetik lehren und lernen. Fernstudienprojekt zur Fort- und Weiterbildung im Bereich Germanistik und Deutsch als Fremdsprache Teilbereich Deutsch als Fremdsprache: Fernstudieneinheit 21*. Berlin: Langenscheidt.

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

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

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

Ellis, R. (2015). *Understanding Second Language Acquisition* (2nd edition). Oxford: Oxford University Press.

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

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *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. *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. *Bilingualism: Language and Cognition, 17*, 384–395.

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

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

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

Flege, J. E. (2016). The role of phonetic category formation in second language speech acquisition. Talk at *New Sounds*, 10-12 June, Aarhus.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics, 34*, 153–175.

Grainger, J., & Ferrand, L. (1994). Phonology and Orthography in Visual Word Recognition: Effects of Masked Homophone Primes. *Journal of Memory and Language, 33*, 218–233.

Grainger, J., & Ferrand, L. (1996). Masked Orthographic and Phonological Priming in Visual Word Recognition and Naming: Cross-Task Comparisons. *Journal of Memory and Language, 35*, 623–647.

Gussmann, E. (2007). *The Phonology of Polish. The phonology of the world's languages.* Oxford: Oxford University Press.

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

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

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

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

Hirschfeld, U. (2001). Vermittlung der Phonetik. In G. Helbig, L. Götze, G. Henrici, & H.-J. Krumm (Eds.), *Deutsch als Fremdsprache. Ein internationales Handbuch* (pp. 872–879). Berlin: De Gruyter.

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

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

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

Loewen, S. (2015): *Introduction to Instructed Second Language Acquisition*. New York: Routledge.

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

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

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. *Journal of Phonetics*, *30*, 229–258.

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

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81–108.

Müller, U. (2003). Polnisch. In U. Hirschfeld (Ed.), *Phonetik International. Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache. CD-ROM* (4th edition, pp. 1–17). Waldsteinberg: Popp.

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

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

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.

Nimz, K. (2015). The Perception of Vowel Quality and Quantity by Turkish Learners of German as a Foreign Language. In E. Raimy & C. E. Cairns (Eds.), *The Segment in Phonetics and Phonology* (pp. 253–266). Hoboken: John Wiley.

Nimz, K. (2016). *Sound perception and production in a foreign language: Does orthography matter?* Potsdam Cognitive Science Series: Vol. 9. Potsdam: Potsdam Universitätsverlag.

Nimz, K. & Khattab, G. (2015). L2 sound perception: Does orthography matter? In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow.

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

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

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

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rafat, Y. & Stevenson, R. A. (2018): Auditory-orthographic integration at the onset of L2 speech acquisition. In *Language and Speech*, 1–25.

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

Röber, C. (2012). Die Orthografie als Lehrmeisterin beim Spracherwerb: Die Bedeutung der Rechtschreibung für die Veranschaulichung der Strukturen des Deutschen im DaZ-Unterricht. *Deutsch als Zweitsprache, 2*, 4–49.

Röttger, T. B., Winter, B., & Grawunder, S. (2011). The robustness of incomplete neutralization in German. In W.-S. Lee & E. Zee, E. (Eds.): *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, University of Hong Kong, 1722–1725.

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

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

Saito, K. (2013). Reexaming effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition*, *35*, 1–29.

Saito, K. (2015). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition*, *37*, 713–743.

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

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

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

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

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

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.

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

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

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

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. *Speech and Language Technology*, *5*, 9-19.

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

Trouvain, J. & Gut, U. (2007). *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin: De Gruyter.

Turk, A. E., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, et al. (Eds.), *Methods in Empirical Prosody Research* (pp. 1–26). Berlin: De Gruyter.

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

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

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].

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

Appendix

	Marked	**Unmarked**
/a/	Wasser (“water”)	lachen (“to laugh”)
	Schatten (“shadow”)	Tasche (“bag”)
	Kamm (“comb”)	Wald (“forest”)
	nass (“wet”)	Wand (“wall”)
/a:/	Sahne (“cream”)	Gabel (“fork”)
	fahren (“to drive”)	Tafel (“blackboard”)

	Zahn (“tooth”)	Tag (“day”)
	Bahn (“train”)	Schaf (“sheep”)
/ɔ/	Sommer (“summer”)	Woche (“week”)
	Sonne (“sun”)	Wolke (“cloud”)
	voll (“full”)	Loch (“hole”)
	Gott (“god”)	Koch (“cook”)
/o:/	Kohle (“coal”)	Boden (“floor”)
	wohnen (“to live”)	Monat (“month”)
	Sohn (“son”)	rot (“red”)
	Lohn (“salary”)	hoch (“high”)
/ɛ/	Wetter (“weather”)	Becher (“cup”)
	Sessel (“armchair”)	Fenster (“window”)
	Bett (“bed”)	Geld (“money”)
	nett (“nice”)	Welt (“world”)
/e:/	Fehler (“mistake”)	Nebel (“fog”)
	Lehrer (“teacher”)	geben (“to give”)
	Zehn (“ten”)	Weg (“way”)
	Mehl (“flour”)	Keks (“cookie”)

Table 4. Experimental items for the production task[7]