L2 SOUND PERCEPTION: DOES ORTHOGRAPHY MATTER?

Katharina Nimz[a] *& Ghada Khattab*[b]

[a]IDEALAB (Universities of Potsdam, Newcastle, Trento, Groning and Macquarie University)
[b]Newcastle University
`katharina.nimz@uni-potsdam.de; ghada.khattab@newcastle.ac.uk`

ABSTRACT

In recent years, there has been growing interest in how L2 learners' perceptual abilities relate to their lexical representation of foreign words, and in how orthography may play a role in this. In this study we address both questions using two perception experiments that were concerned with the discrimination and representation of German long vowels by Polish learners of L2 German and a native speaker control group. While the first experiment tested phonetic discrimination abilities using nonsense words, the second experiment was a judgement task that was designed to tap into the participants' lexical representations. Half of the test items in the judgement task were real words containing vowels that were explicitly marked for length in their orthography, while the remaining items were not explicitly marked. The findings of the two experiments are dissociated; interestingly, orthography did not seem to be the driving factor.

Keywords: L2 vowel perception, vowel length, orthography, Polish, L2 German

1. INTRODUCTION

Many studies investigating L2 perception have been concerned with the discrimination and identification of foreign sounds while the nature of lexical representation involving these segments has only recently attracted more attention [4]. Some researchers have put forward the idea that, besides perceptual abilities, other factors such as ancillary knowledge in the form of for example orthographic information may play a role in phonological representation of novel L2 words [6].

Most research which has discussed the possible influence of the L1 and L2 orthographic system on the acquisition of L2 sounds is concerned with English as an L2. However, English is regarded as a language with relatively deep orthography [8], which makes it less suitable for investigating the role of orthography in L2 perception than a language with a shallower orthography like German.

Both German and English have vowel pairs which exhibit qualitative as well as quantitative contrasts (e.g. /ɪ/-/iː/). Unlike most varieties of English, however, vowel length is more prominent in German due to for example the /a/-/aː/ pair, where length is considered the primary feature [13]. Additionally, vowel length in German may be signalled orthographically by the so called *lengthening h*, namely the letter <h> following a long vowel, e.g. <Sahne> [zaːnə] ("cream"). It has been hypothesized – yet not experimentally tested – that this may aid German-as-a-Foreign-Language (GFL) learners in building phonological representations of vowel length when they do not make use of this feature in their native language [11].

Polish learners of L2 German have been reported to have problems producing a distinction between German long and short vowels: long vowels are generally substituted by their short counterparts [10]. In contrast to German, Polish does not differentiate between long and short vowels which, according to the feature hypothesis by McAllister et al. [9], means that both the production and the perception of German long vowels will be deficient. On the other hand, some researchers have argued that duration cues such as vowel length are always easy to access whether listeners have had specific linguistic experience with them or not [2].

This study explored the perception of German vowels by Polish L2 learners in order to address both the issue of general length perception abilities by L2 learners and the influence of orthography therein. Two experiments were carried out: a discrimination task with manipulated nonsense words which differed either in length, vowel quality, or both, and a judgement task with real German words that were either marked or unmarked for vowel length in their orthography.

2. DISCRIMINATION TASK

2.1. Participants

Participants for both the discrimination and judgement task were recruited at a Polish high school in Warsaw, Poland (experimental group) and at a German high school in Dortmund, Germany (control group). The Polish institution was a school with special emphasis on German as a Foreign

Language (GFL) and all Polish participants had received at least two years of intensive GFL instruction (in their first year about 18h per week, and in the following years about 10h of German per week).

20 Polish GFL learners (4 males, average age 18.5 (SD=.6)) and 20 German native speakers (6 males, average age 17.9, SD=1.1)) participated in the discrimination task.

2.2. Experimental items

Three German vowel pairs were chosen to test whether Polish GFL learners have problems perceiving length differences in the German vowel pairs /eː/-/ɛ/, /oː/-/ɔ/, and /aː/-/a/. Since most long and short vowels not only differ in their length but also in their quality, the test items were manipulated similar to Sendlmeier [13]: Using PRAAT [3], a prototypical long vowel spoken by a female German native speaker in a bilabial consonantal context was shortened to the average length of its corresponding short counterpart, while a prototypical short vowel was lengthened to the average length of its corresponding long counterpart. Pairs were then matched for three conditions: (1) condition "proto": non-manipulated long, tense vowel vs. non-manipulated short, lax vowel, e.g. [boːp] vs. [bɔp], (2) condition "length": non-words were matched in a way that pairs only differed in their length, e.g. [boːp] vs. [bop] or [bɔp] vs. [bɔːp], (3) condition "quality": non-words were matched so that pairs only differed in their vowel quality, e.g. [bop] vs. [bɔp] or [boːp] vs. [bɔːp].

2.3. Experimental design

The discrimination experiment was administered through PRAAT. Each vowel pair was judged 8 times in each condition for being "same" or "different", plus the same amount of filler pairs that were exactly the same, and a control condition with 12 pairs that were clearly different. In all, each subject rated 156 nonsense word pairs, which were presented randomly for each subject with an ISI of 0ms. The experiment lasted about 10 min.

2.4. Results

Of interest to the analysis were those trials which were "different" (4320 data points). As evident in Figure 1, results for the low central vowel pair ("a-pair") clearly differed from those for the mid vowel pairs ("e- and o-pairs"). For this reason, two separate generalized linear mixed models were fit to the binomial accuracy data in R [12] with language, condition, and their interaction as fixed factors and

participants as a random factor. Pairwise comparisons (Tukey) of all factor levels of the model were then run, which revealed that group differences in the a-pairs were driven by the "length" (p<.001) and "proto" (p<.001) condition. In the mid-vowel pairs significant group differences were only present in the "length" condition (p<.001).

Figure 1: Correct responses for the three vowel groups by condition and language group (N=20 per language group; error bars show 2 SE)
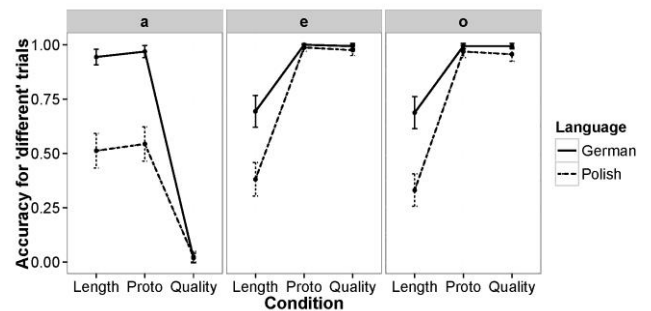
2.5. Discussion

The results of the discrimination task show clearly that Polish GFL learners have difficulties perceiving pure length differences between vowels, a result which is in line with McAllister et al. [9]. Quality differences in mid vowels (e- and o-pairs) on the other hand can be differentiated in a native-like manner by Polish GFL learners. Similarly, prototypical e- and o-pairs are discriminated well, as these represent the same quality difference as those items in the "quality" condition.

Polish native speakers' difficulty with length discrimination becomes crucial for the a-pair: Since prototypical /aː/ and /a/ only differ in length, Polish native speakers do not easily perceive the differences in the "proto" condition, just as they have difficulty perceiving length differences in the manipulated "length" conditions. Neither German native speakers nor Polish GFL learners perceive differences in the "quality" condition of the a-pair, as a lengthened short /a/ sounds like a long /aː/ and a shortened /aː/ sounds like a short /a/ – to any language group.

Because of the results in the discrimination task, Polish GFL learners were expected to have comparable problems when judging real word items that were manipulated in the same way as the nonsense word pairs. That is, it was expected that the length feature of long vowels is less likely to be represented in the learners' lexicons than vowel quality features. The next experiment tested whether the orthographic marking of the test items may play a role in the participants' performance.

3. JUDGEMENT TASK

3.1. Participants

The participants of the judgement task (which was administered on another day prior to the discrimination task) were the same as those in the discrimination task, with the exception of two female Polish native speakers. These two participants did not know at least 75% of the test words used, which was established as a cut-off point before the analysis.

3.2. Experimental items

The judgement task investigated the perception of the same (long) vowels used in the discrimination task; however, this time the vowels appeared in real German words which had to be judged for their correctness ("correct" vs. "incorrect"). Similar to the manipulation in the discrimination task, long vowels (spoken by a female German native speaker) were shortened to the average length of their short counterparts in the same consonantal context. For example, the vowel in the test word /bo:dən/ ("floor") was shortened to the length of the short vowel in the nonsense word /bɔdən/ (condition called "length", as *length* in this item is incorrect). At the same time, the vowel in the nonsense word /bɔdən/ was lengthened to the length of the vowel in the real word, yielding another test condition named "quality" (as *quality* in this item is incorrect). Furthermore, a correct prototypical rendition of the test item was included called "proto", as well as a word, which was incorrect in both length and quality (condition "both"). To sum up, the word <Boden> appeared as [bo:dən] ("proto"), [bodən] ("length"), [bɔ:dən] ("quality"), and [bɔdən] ("both") in the judgement task.

24 test words were chosen of which one third contained long /e:/, one third long /o:/, and one third long /a:/. Half of all the test words were words that were marked in their length through so called *lengthening h* ("Dehnungs-h"), yielding the experimental variable "orthographic marking". For example, the words <Mehl> /me:l/ ("flour"), <Sohn> /zo:n/ ("son"), and <Sahne> /za:nə/ ("cream") were items within the condition "orthographically marked", while the words <Weg> /ve:k/ ("path"), <Boden> /bo:dən/ ("floor"), and <Gabel> /ga:bəl/ ("fork") were instances of orthographically unmarked items. 24 filler items were added, which appeared in the experiment both as correct or completely incorrect items, in that the vowel in the filler item <Bett> /bɛt/ ("bed") was either presented in its correct rendition or with a vowel that was very different from the correct one, for example /bɔt/ instead of /bɛt/.

3.3. Experimental design

The judgement task was administered through the software DMDX, which allowed for collecting both accuracy and reaction time (RT) data. Participants were presented visually with the pictures they had been familiarized with in a prior production task and, crucially, received no orthographic input. Upon seeing the picture, they were presented auditorily with the corresponding test word which appeared either in the condition "proto", "both", "length", or "quality" over the course of the experiment. Participants were asked to judge as fast as possible via button press on a gamepad whether the word they just heard was a correct or incorrect rendition of the image they just saw. The experiment was administered in four blocks, in which each item appeared once. Presentation order of blocks and items was randomized for each participant. In all, 192 items had to be judged; the experiment lasted about 20 min.

3.4. Results

Two different aspects were of interest in the analysis of the judgement task data: 1) Would perceptual difficulties in length differentiation (as found in the discrimination task) be reflected in a task that tapped into learners' lexical representations of words containing long vowels? 2) Does orthographic marking of the test words play a role in how learners (or even native speakers) judge incorrect, manipulated items?

3.4.1. Length versus quality versus both

Of special interest were those conditions which should have been identified as "incorrect", because they were either non-target-like in their length, their quality, or both. Figure 2 shows the accuracy results for all words (excluding the "quality" condition in words containing the vowel /a:/, due to the results of the discrimination task).

A linear mixed model was run to predict the binomial accuracy data (2224 data points, excluding items which were not known to participants during the production task) by the factors language and condition, with participants and words as random factors. The factors were contrast coded in a way that allowed for testing the overall effect of language as well as direct comparisons of the factor levels within each group. There was a clear main effect of language in that German participants are significantly better than the Polish participants in

identifying manipulated items as incorrect. While there were no significant differences between the conditions within the German group, comparisons of the conditions within the Polish group revealed a significant difference between the "length" and the "quality" condition, in that learners were significantly better at judging items which were too short as "incorrect" than judging items that were wrong in their vowel quality (p=.03).

Figure 2: Correct responses to the manipulated "incorrect" conditions by language group (N(Ger)=20, N(Pol)=18; error bars show 2 SE)
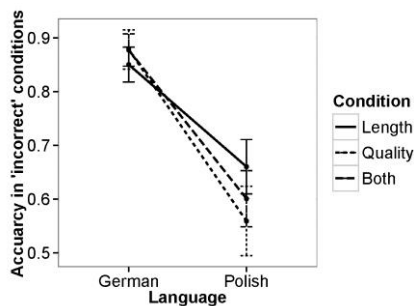

Apparently, difficulties with length/ease with quality perception in a discrimination task do not reflect themselves in a task that taps into the learners' lexical representations. It was then tested whether orthography might play a role in this, i.e. whether learners may have profited from the fact that half of the words were explicitly marked by *lengthening h*.

3.4.2. Orthographically marked vs. unmarked

As Figure 3 clearly shows, orthographic marking of the test items had no effect on the accuracy of the participants' performance in the length condition. Neither did the inspection of the reaction times for the correct items show any influence of this factor.

Figure 3: Effect of orthographic marking on accuracy (left) and RTs (right) (N(Ger)=20 German, N(Pol)=18; error bars show 2 SE)
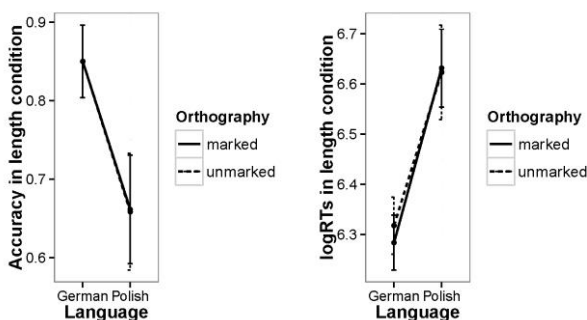
Mixed models with language, orthography, and their interaction as fixed factors and participants and words as random factors confirmed that orthographic marking does not significantly influence the performance of GFL learners (p=.98 (accuracy model) and p=.56 (RT model)).

3.5. Discussion

The accuracy data for the three "incorrect" conditions show an interesting dissociation with the results of the discrimination task: While Polish GFL learners have obvious difficulties in discriminating length differences (and are much better in discerning vowel quality differences), they perform best in the "length" condition of a judgement task involving real words. In previous research it has been suggested that orthography may help built more accurate lexical representations [6, 11]. However, statistical analyses of both accuracy and reaction time data show clearly that this is not the case for this specific example of explicit marking of German long vowels. Similarly, Escudero [7] did not find that orthography helps in novel spoken-word learning when pairs were perceptually difficult.

An alternative way of explanation may lie in the Polish participants' language learning background. 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; however, only one participant knew that vowels differ in their quality as well. Therefore, a different kind of ancillary knowledge than orthography, namely metalinguistic awareness, may play an important part in forming L2 lexical representations and with that more target-like pronunciations [5].

4. CONCLUSION

This study showed that research in L2 phonetics and phonology profits from task diversity in order to better appreciate when L2 learners might exhibit awareness of phonetic contrasts which do not exist in their L1. The results of the two perception experiments show an interesting dissociation as far as perception of phonetic length, as signalled by durational contrasts, and phonological length, as signalled by lexical contrasts, is concerned. Recent research in the field has focussed on orthography as a possible explanatory variable [1]. However, in the case of German vowel length perception by Polish GFL learners, explicit orthographic marking is unlikely to play a role. Further research into other factors in L2 sound acquisition, such as for example metalinguistic knowledge, is necessary.

5. REFERENCES

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

[2]   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*. Timonium, MD: York Press, 279-304.

[3]   Boersma, P., Weenink, D. 2014. *Praat: doing phonetics by computer* [Computer program].

[4]   Darcy, I., Daidone, D., Kojima, C. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *ML,* 8, 372–420.

[5]   Derwing, T. M., Munro, M. J., Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*. 48, 393–410.

[6]   Escudero, P., Hayes-Harb, R., Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics.* 36, 345–360.

[7]   Escudero, P. 2015. Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*. 36, 7–22.

[8]   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.

[9]   McAllister, R., Flege, E. J., 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.

[10] Morciniec, N. 1990: *Das Lautsystem des Deutschen und des Polnischen. Heidelberg*: Julius Groos.

[11] Nimz, K. 2011. Vowel perception and production of late Turkish learners of L2 German. *Proc. 17th ICPhS* Hong Kong, 1494–1497.

[12] R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna.

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