THE ROLE OF WRITTEN INPUT IN THE ACQUISITION OF A GERMAN-LIKE
PATTERN OF FINAL DEVOICING BY NATIVE ENGLISH SPEAKERS: EVIDENCE
FROM A LISTENING TASK

by

Cailey N. Lloyd

A Senior Honors Thesis Submitted to the Faculty of
The University of Utah
In Partial Fulfillment of the Requirements for the
Honors Degree in Bachelor of Arts

In

Linguistics

Approved:

Shannon Barrios
Thesis Faculty Supervisor

Scott Jarvis
Chair, Department of Linguistics

Aaron Kaplan
Honors Faculty Advisor

Sylvia D. Torti, PhD
Dean, Honors College

May 2020
Copyright © 2020
All Rights Reserved
ABSTRACT

In German, underlyingly voiced obstruents are devoiced in final position (e.g., the word ‘wheel’ /ʁad/ <rad> is pronounced [ʁat]). Adult, native English-speaking learners of German are known to experience difficulty acquiring phonological processes such as German final devoicing. It has been hypothesized that there is a direct causal relationship between exposure to written input and native English speakers’ ability to learn the final devoicing pattern in German (Hayes-Harb, Brown & Smith, 2018; Young-Scholten 2002). While some literature has reported that written input may interfere with the acquisition of target-like surface voicing (Hayes-Harb, Brown, & Smith., 2018), more recent work has suggested there is a written input trade-off (Barrios & Hayes-Harb, 2020), by which exposure to written input aids learners to establish target-like underlying representations. The studies to date have employed production tasks to investigate learners’ knowledge of German final devoicing. The present study builds on these findings using a listening task to examine how adult, native speakers of English acquire a German-like final devoicing pattern with or without exposure to written forms. Fifty-six participants learned alternating and non-alternating German-like words in both their suffixed and unsuffixed forms via word form-image pairs, in two word learning conditions: Orthography and No Orthography. They were later tested on their knowledge of the words in a word-picture matching task. It was hypothesized that, consistent with previous findings, exposure to orthographic input during word learning would cause adult English-speaking learners to misremember the surface voicing of words containing underlyingly voiced obstruents in their singular forms, but would prove helpful for learning the underlying voicing of words in their plural forms. As expected, participants
generally learned the German-like nonwords, as they scored high on criterion test items, regardless of word learning condition. Moreover, and consistent with past findings, exposure to written input was found to interfere with the acquisition of final devoicing. Specifically, the Orthography Group had more difficulty detecting voicing mismatches for underlyingly voiced words in their singular than plural form. Unexpectedly, the trade-off effect was not robust in the present study. Together, these findings suggest that orthographic input plays a complex role in the acquisition of phonological processes such as German final devoicing.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHODS</td>
<td>9</td>
</tr>
<tr>
<td>RESULTS</td>
<td>14</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>20</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>23</td>
</tr>
</tbody>
</table>
INTRODUCTION

In the phonological acquisition of a second language (L2), learners must acquire a new system of phonological contrasts (used for distinguishing L2 words), as well as any language-specific phonological processes using input from the target language as evidence. Importantly, literate adult learners of a second language often have access to orthographic input from the language’s phonographic writing systems as an additional source of evidence regarding the phonological structure of the target language. To better understand how written input may be utilized by second language instructors, it is essential to understand the role that orthographic input plays in the acquisition of L2 phonology.

Research on the role of written input in adult second language acquisition of L2 phonological contrasts has revealed a complex relationship. On the one hand, prior studies have indicated that literate, adult L2 learners can benefit from orthographic input when learning the phonological forms of L2 words that contain novel phonological contrasts known to be difficult for particular groups of L2 learners to perceive (Escudero, Hayes-Harb, & Mitterer, 2008; Escudero, Simon, & Mulak, 2014; Showalter & Hayes-Harb, 2013). For example, using the visual world eye-tracking paradigm, Escudero et al. (2008) demonstrated that exposure to written forms containing familiar graphemes (e.g., graphemes that are used to signal a contrast in both the L1 and L2) can support the acquisition of phonological contrasts. This study consisted of 50 highly proficient Dutch-English bilinguals. Participants were tasked with learning new words containing the /æ/-/ɛ/ contrast, which is known to be difficult for L1 Dutch speakers to distinguish in perception. Participants were divided into two groups, including a control group that was
presented the auditory form of a word and a picture, and the experimental group that was additionally presented with the spelled forms of the words. Using eye-tracking technology, the number of times that a participant looked to the image that had been paired with the target word (involving either /æ/ or /ɛ/) versus the image corresponding to a distractor (containing the other member of the /æ/-/ɛ/ contrast) was measured. If learners failed to represent the difference between words containing /æ/ versus /ɛ/ then words containing /æ/ and /ɛ/ should be confused symmetrically (having heard a word involving /æ/ should produce looks to either the target image or distractor image of both). However, an asymmetric pattern of confusion would suggest that learners distinguish the new words to some extent. At test, participants exposed to words' written forms, but not those who were not, displayed an asymmetric pattern of confusion in which words containing /æ/ were confused more often with words containing /ɛ/ than in the opposite direction. As a result, Escudero et al. (2008) demonstrated that the orthographic input provided to those participants in the experimental group facilitated their ability to encode the difficult contrast. In a separate study, Showalter & Hayes-Harb (2013) demonstrated that orthographic input can be beneficial even when stimuli contain unfamiliar diacritics suggesting that the systematicity of the mapping between grapheme and phoneme may be more important than the familiarity of the grapheme (c.f., Durham et al., 2016).

Another factor that is known to influence the impact of written input on L2 phonology is whether the grapheme-phoneme correspondences (i.e., the mappings between graphemes and phonemes in the L1 and L2) are congruent or incongruent. A grapheme-phoneme correspondence is considered congruent when an L2 phoneme is represented by the same grapheme as it is in the L1 (e.g. the phoneme /p/ is spelled <p>
in both Spanish and English). When grapheme-phoneme correspondences are incongruent, either an L2 grapheme corresponds to different phonemes in the L1 and L2 (e.g. the grapheme <h> is silent in Spanish but is produced [h] in English), or an L2 phoneme is represented by a different grapheme in the L1 and L2. It has been found that written input can interfere with the acquisition of the phonological forms of L2 words, particularly when the grapheme-phoneme correspondences are not one-to-one (Hayes-Harb & Cheng, 2016; Hayes-Harb, Nicol, & Barker, 2010; Showalter, 2018). For example, Hayes-Harb et al. (2010) examined whether learners’ memory of phonological forms of L2 words are influenced by the congruence of grapheme-phoneme correspondences. Participants were assigned to one of three exposure conditions. For one group of participants, grapheme-phoneme correspondences were congruent (e.g., the auditory form [fɾʃə] was spelled <fasha> following English spelling conventions). For another group of participants, grapheme-phoneme correspondences were incongruent (e.g., the auditory form [fɾʃə] was spelled <faza>). A third group of participants was presented with no written forms. The authors later employed an auditory-word picture matching task to examine participants’ memory of the words. At test, participants who were exposed to written forms that were inconsistent with respect to English spelling conventions exhibited interference, suggesting that learners utilize orthographic information when it is available, thereby affecting their phonological representations of the novel words.

This literature has revealed a complex relationship between written input and the acquisition of novel phonological contrasts. However, L2 learners must also acquire novel phonological processes (e.g., liaison in French, spirantization in Spanish, flapping
in English, etc). While a growing body of research has addressed the role of written input in L2 acquisition of nonnative contrasts, there is a lack of equivalent research regarding the role of orthography in the acquisition of L2 phonological processes. In the present study, we investigate the adult second language acquisition of a phonological process known as German final devoicing, with a particular focus on the role that exposure to written input has on its acquisition.

German has a voicing contrast among obstruents which can be observed in the words ‘advice’ /ʁat/ pronounced [ʁɛːt] and ‘wheel’ /ʁəd/ pronounced [ʁəːd]. However, the phonological process of final devoicing results in (near) neutralization of the voicing contrast in word-final position. For example, in alternating words such as ‘wheel’, which undergo the final devoicing process, the underlyingly voiced final consonant /d/ is produced as voiceless /t/ in final position (e.g., the word ‘wheel’ /ʁəd/ <rad> is pronounced [ʁat]). The process does not apply to non-alternating words such as ‘advice’ /ʁat/, pronounced [ʁat], because the consonant /t/ is already voiceless. Though this phonological process of final devoicing in German is evident in the auditory input, its acquisition is nevertheless difficult for native English speakers learning German. It is suspected that this challenge posed to native English speakers arises from the misleading evidence that written input provides. In this sense, it is believed that written input may interfere with the acquisition of final devoicing in German (Barrios & Hayes-Harb, 2020; Hayes-Harb, Brown & Smith, 2018; Young-Scholten, 2002).

The nature of surface voicing in German has been a topic of interest for several researchers. For instance, Young-Scholten and her colleagues examined the acquisition of the German final devoicing process (among several other phonological features) for three
adolescents studying abroad for a year in Germany. Their reports mentioned, both, that the acquisition of German final devoicing among these students was incomplete, and that the student who reported the largest amount of time reading in German was less likely to exhibit target-like surface voicing (Young-Scholten, 2002). More recently, a study by Hayes-Harb, Brown, and Smith (2018) provided direct evidence for the causal link between exposure to written input and non-target-like productions of German surface voicing. In an artificial lexicon study, adult, native English speakers were assigned to one of two exposure conditions. During a word learning phase, all participants learned 12 German-like words exhibiting a pattern of surface voicing consistent with German final devoicing in unsuffixed forms via auditory forms paired with pictured meanings. One group of participants additionally saw the words’ written forms. At test, participants completed a picture naming task involving the 12 new words. Later the participants’ productions were coded by an additional group of participants to determine the presence of voiced or voiceless obstruents in the word-final position. It was observed that those participants who were exposed to written input were more likely than those who were not to incorrectly produce underlingly voiced obstruents in alternating words as voiced.

Together, these experiments suggest that exposure to written input may interfere with the acquisition of target-like surface voicing in German. Importantly, the studies reviewed above focused on surface voicing and did not provide direct evidence for a role of orthographic input in the acquisition of a phonological alternation. As Young-Scholten (2002) points out, a learner’s knowledge of an alternation cannot be determined by the non-target-like production of only one of the alternants. In order to demonstrate knowledge of an alternation a learner must identify which words undergo the alternation
(and which do not), as well as extend the process to novel words (Ruben Van de Vijver, 2014).

Barrios and Hayes-Harb (2020) examined whether written input affects the acquisition of German Final Devoicing. Specifically, they asked whether naïve adult learners demonstrate a German-like pattern of final devoicing when they have evidence of the alternation from both suffixed and unsuffixed forms, and whether orthographic input interferes with the acquisition of the alternations when further evidence of the alternation is available in the auditory input.

Experiment 1 included a word learning phase, criterion test, and final test. Participants were randomly assigned to one of two exposure conditions, either the Orthography condition or the No Orthography condition, during which they were exposed to a total of 12 auditory word-picture pairs, with the Orthography group additionally provided the spelled forms of each word. Following a word learning cycle (192 total trials; 12 words x 2 speakers x 2 repetitions x 4 blocks), participants completed a criterion test involving a two-way forced-choice auditory word-image matching test. Each singular image was presented twice: once with the auditory form it was paired with during training (match trial) and once with another auditory form (mismatch trial). In the final test, participants performed a picture naming task where they were presented with each of the singular and plural images twice and were recorded producing the corresponding word in its singular and plural form.

Barrios and Hayes-Harb (2020) reported that the No Orthography group differed in the number of voiceless productions that were made for the underlyingly voiced (alternating) words in their suffixed and unsuffixed forms, whereas the Orthography
group showed no such difference. This finding suggested that the No Orthography group had outperformed their counterparts in learning the final devoicing pattern, and thus, that exposure to written input interfered with the acquisition of the final devoicing pattern by the Orthography group. Moreover, Barrios & Hayes-Harb (2020) observed that the Orthography group incorrectly produced more underlyingly voiced words as voiced in their singular form than the No Orthography group. Interestingly, however, participants in the Orthography group correctly produced more underlyingly voiced obstruents as voiced in the plural than those participants in the No Orthography group, suggesting that a potential “trade-off” effect may exist with respect to exposure to written input. A second experiment replicated these core findings but failed to provide evidence that learners apply the final devoicing process to novel words.

In the present study we examine whether the observed “trade-off” effect proves robust to methodological changes. First, research that has explored the role of orthographic input in the acquisition of a phonological process has employed an artificial lexicon study paradigm, including a word-learning phase, a criterion test, and a final test involving picture-naming. While a production task indeed produces data that bears on the research question, it also requires that productions be coded by a second group of individuals (“coders”), and as a result, is both resource and time intensive. The present study investigates the acquisition of German final devoicing using an artificial lexicon study format with a listening task, involving a two-way forced-choice auditory word-image matching test, rather than a production task as the final test. If participants have acquired the final devoicing process, then they should readily detect voicing mismatches at test and d’ scores should be high across conditions. Additionally, the trade-off effect
would be evidenced by the Orthography group experiencing greater difficulty rejecting underlyingly voiced words incorrectly produced as voiced in the singular morphological condition, whereas participants in the No Orthography group should experiencing greater difficulty rejecting underlyingly voiced words incorrectly produced as voiceless in the plural.

Furthermore, studies using the artificial lexicon study format have traditionally adopted an arbitrary threshold of 90% accuracy or better on the criterion test items as evidence that an individual participant has sufficient knowledge of the new words to proceed to the final test. Participants who obtain less than 90% accuracy either complete additional word learning cycles until they reach the criterion threshold (Hayes-Harb et al., 2010) or are excluded from data analysis (Barrios & Hayes-Harb, 2020). Adopting a 90% threshold as an exclusionary criterion may result in the exclusion of many participants. In the present study, we reconsider this arbitrary threshold and ask whether adopting different thresholds (90%, 70%, or no threshold) impacts the core results of the study.
METHODS

Participants

Fifty-seven participants were recruited from the Linguistics Study Pool at the University of Utah and received course credit for their participation (No Orthography = 27, Orthography = 32, Mean age = 21.2 years; range = 18-62 years). All participants were native English speakers with no speech, hearing, or neurological disorders, and had no prior experience with languages that contain the phonological pattern of interest, final devoicing (e.g., Polish, Dutch, German, etc.). The data of two eligible participants were excluded due to experimenter error. Adopting the standard threshold of 90% accuracy or higher on a criterion test resulted in the exclusion of 27 of 57 participants (n=30; No Orthography = 11, Orthography = 19). Adopting a less conservative threshold of 75% accuracy resulted in the exclusion of 5 of 57 participants (n = 52, No Orthography = 22, Orthography = 30).

Stimuli

The stimuli (see Figure 1) consisted of 12 German-like words from Barrios & Hayes-Harb (to appear), all of which were German words or nonwords that were phonotactically legal in German. Half of the words were alternating and contained an underlyingly voiced stop consonant (i.e., /b/, /d/, /g/) in final-position in unsuffixed forms, whereas the other half were non-alternating and contained voiceless stops underlyingly (i.e., /p/, /t/, or /k/).
Each of the nonwords were recorded by two female, native English-speakers with phonetic training in both their singular and “pseudo-plural” forms. We chose to have native English-speakers produce the words because our task required both target-like (e.g. /trob/ [tʁop] ‘fork’) and “incorrectly produced” forms of our German-like words in which the underlyingly voiced obstruents were produced as voiced in a final obstruent.
devoicing context (e.g., [tʰɔb]). Given that knowledge of a phonological process is productive and hard to suppress in native speakers of the language, we reasoned that native English speakers who are accustomed to producing final voicing contrasts would more easily produce these tokens. Moreover, incomplete neutralization is prevalent in the speech of speakers of languages that exhibit the final devoicing process, such as German (Roettger, Winter, Grawunder, Kirby, & Grice, 2014). Having native English speakers produce our stimuli ensured that voicing neutralization is indeed complete in our stimuli. Each of the speakers produced four tokens of each word in its singular and plural form. The second and third tokens from both speakers were extracted via Praat software (Boersma & Weenink, 2018) for use in the experiment.

Procedure

Each participant was positioned in front of a computer screen with headphones in a sound-attenuated booth. DMDX software (Forester & Forester, 2003) was used to control the stimulus presentation and record participants’ responses. The study used an artificial lexicon study design and consisted of three phases described in detail below. Following the experiment, participants completed a language background questionnaire.

Word Learning Phase. During a word learning phase, participants in both word learning conditions heard an auditory form and were presented with an image depicting its “meaning”. The Orthography group was additionally exposed to the written form of the word below the image. There were four exposure blocks separated by a participant-controlled break. Each block consisted of a total of 48 singular and 48 plural exposures (12 words x 2 speakers x 2 tokens).
**Criterion Test.** The criterion test consisted of a two-way forced-choice word-picture matching task. Participants heard an auditory form and were presented with a picture on the computer screen and pressed a button labelled “YES” or “NO” on a keyboard to indicate whether the picture and auditory form matched. The test consisted of 24 trials, half of which were matching (new tokens of familiar auditory forms were paired with the same picture as they were during training, i.e., [tʁɔb] was paired with the fork picture) and the other half were global mismatch trials (new tokens of familiar auditory forms were paired with a different picture than during exposure, i.e., [tʁɔb] was paired with the trumpet picture). As noted earlier, these trials were used to determine whether participants had sufficient knowledge of the words. Data from participants whose accuracy on global mismatch trials was less than 90% (27 of 57 participants who met all other exclusionary criteria) were excluded from analysis.

**Final Test.** Like the criterion test, the final test involved word-picture matching and consisted of 12 match trials and 12 mismatch trials. However, on mismatch trials familiar pictures were paired with new auditory tokens which differed from the correct auditory form in the voicing of the relevant stop consonant. Participants were tasked with indicating whether these picture-auditory word pairings were a correct match via yes/no button press. Given that participants received training on the word tokens, these ‘voicing mismatch’ test trials were used to examine whether participants inferred the correct underlying and surface voicing and could detect minimal mismatches among alternating (i.e., /trob/ ‘fork’ incorrectly pronounced [tʁɔb]) and non-alternating variants (i.e., /krak/ ‘boot’ incorrectly pronounced [kʁag]).
Data Analysis

A d’ score was computed for each participant and for each condition. The d’ score measured participants’ sensitivity to match vs. mismatch trials while factoring out potential response bias (in this case, the propensity of each individual participant to respond ‘match’ or ‘mismatch’ on a given trial). Hits were computed as the number of trials in which the participant responded ‘mismatch’ for different trials divided by the total number of different trials. False alarms were computed as the number of trials in which the participant responded ‘mismatch’ for match trials divided by the total number of match trials. A log-linear correction in which 0.5 was added to all cells of the contingency table without regard to the contents of the cell was applied to correct for extreme proportions (hit rates or false alarm rates of 0 or 1; Hautus, 1995). As a result, the maximum possible d’ score was 2.93. Overall, higher d’ scores indicate greater sensitivity to voicing mismatches on mismatch trials, whereas a d’ score of 0 indicates no sensitivity. If participants were to demonstrate target-like knowledge of the final devoicing process, we would expect to see high d’ scores regardless of underlying voicing and morphological conditions.
RESULTS

Figure 2 displays the mean d’ scores by exposure condition, underlying voicing, and morphological condition when a 90% accuracy threshold was applied for the criterion test. The mean d’ scores were not uniformly high or low but instead varied by condition. Participants in both the No Orthography group and the Orthography group generally demonstrated greater sensitivity to mismatch trials involving underlyingly voiceless (non-alternating) words than underlyingly voiced (alternating) words. Additionally, different patterns of sensitivity can be observed among the underlyingly voiced (alternating) words between the two participant groups depending on morphological condition. The Orthography group outperformed the No Orthography group in detecting voicing mismatches for underlyingly voiced words in their plural form (d’ = 1.02 and 0.4, respectively). The No Orthography group, however, showed greater sensitivity to voicing mismatches for alternating words in their singular form (d’ = 1.35 and 0.6, respectively).

![Figure 2: Mean d’ score by exposure condition, underlying voicing, and morphological condition. The error bars represent one standard error of the mean.](image-url)
Our statistical analyses of the d’ score consisted of a linear mixed-effects model using the *lme4* package (version 1.1-15; Bates, Maechler, Bolker, & Walker, 2015) of R (version 3.3.1). The model was fit using a maximum likelihood technique with the d’ score as the dependent variable. The analysis included contrast coded fixed effects for exposure condition, underlying voicing, and morphological condition in a 2*2*2 factorial design. Random effects were fit using a “maximal” random effects structure (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for participants. Planned comparisons were conducted using simultaneous tests for general linear hypotheses with the *multcomp* package in R (Hothorn, Bretz, & Westfall, 2008). P-values were adjusted using the single-step method and a family-wise error rate protection via Bonferroni correction.

Table 1
*Mixed effects model analysis examining the effects of exposure condition, underlying voicing, and morphological condition on d’ score when an accuracy threshold of 90% was adopted as criterion (n=30).*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>1.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Exposure Condition</td>
<td>-0.04</td>
<td>0.17</td>
</tr>
<tr>
<td>Underlying Voicing</td>
<td>0.71</td>
<td>0.09</td>
</tr>
<tr>
<td>Morphological Condition</td>
<td>0.41</td>
<td>0.09</td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing</td>
<td>0.24</td>
<td>0.18</td>
</tr>
<tr>
<td>Exposure Condition x Morphological Condition</td>
<td>-0.60</td>
<td>0.18</td>
</tr>
<tr>
<td>Underlying Voicing x Morphological Condition</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing x</td>
<td>1.19</td>
<td>0.35</td>
</tr>
<tr>
<td>Morphological Condition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* All factors were contrast coding as follows: Exposure Condition (-0.5 = NoOrthography, 0.5 = Orthography), Underlying Voicing (-0.5 = voiced, 0.5 = voiceless), Morphological Condition (-0.5 = plural, 0.5 = singular). Model Formula: dprime ~ Exposure Condition * Underlying Voicing * Morphological Condition + (1 | Part_ID).

*|t| > 2.0, indicating a significant effect (Gelman & Hill, 2007)
The model output is summarized in Table 1. The results indicate that participants demonstrated greater sensitivity to voicing mismatches for underlyingly voiceless than underlyingly voiced words (main effect of underlying voicing) and for unsuffixed (singular) than suffixed (plural) forms (main effect of morphological condition). These main effects were qualified by a two-way interaction of exposure condition and morphological condition, as well as a three-way interaction of exposure condition, underlying voicing, and morphological condition, suggesting that d’ scores varied as a function of exposure condition, underlying voicing, and morphological condition. No other main effects or interactions were found to be significant.

Pairwise comparisons of the two exposure groups for underlyingly voiced words in their singular form was marginal (estimate = 0.75, SE = 0.32, z = 2.33, p = 0.07). No difference was observed between the groups for underlyingly voiced words in their plural form (estimate = -0.63, SE = 0.32, z = -1.95, p = 0.18), underlyingly voiceless words in their singular form (estimate = -0.27, SE = 0.32, z = -0.84, p = 0.85) or underlyingly voiceless words in their plural form (estimate = -0.29, SE = 0.32, z = -0.91, p = 0.82).

Given that the traditional 90% criterion threshold resulted in the exclusion of a large number of participants (i.e., 27 of 57), we also analyzed the data using the less conservative criterion threshold of 75%, as well as no threshold at all. The adoption of a 75% accuracy threshold resulted in the exclusion of only 5 participants (3 from the No Orthography group and 2 from the Orthography group, n=52). The data under this adjusted threshold is displayed in Figure 3, again showing the mean d’ score by exposure condition, underlying voicing, and morphological condition. What is noteworthy is that, while adopting a less stringent threshold resulted in lower mean d’ scores across
conditions relative to the 90% criterion, the overall qualitative pattern observed in the data is remarkably similar.

Figure 3: Mean d’ score by exposure condition, underlying voicing, and morphological condition. Error bars represent one standard error of the mean.

Our statistical analysis of the data under the 75% criterion threshold was conducted as described above. Table 2 summarizes the model output. Given the similarity between the data under the two thresholds, it is not surprising that we again observed a main effect of underlying voicing and morphological condition, as well as the two-way interaction of exposure condition and morphological condition, and the three-way interaction of exposure condition, underlying voicing, and morphological condition.

Pairwise comparisons of the exposure groups revealed a significant difference for underlyingly voiced words in their singular form (estimate = 0.81, SE = 0.24, z = 3.30, p<0.005), but not for underlyingly voiced words in their plural form (estimate = -0.47, SE = 0.24, z = -1.92, p = 0.18), or underlyingly voiceless words in their singular form (estimate = -0.12, SE = 0.24, z = -0.51, p = 0.97) or their plural form (estimate = -0.05, SE = 0.24, z = -0.22, p = 1.00).
Table 2

Mixed effects model analysis examining the effects of exposure condition, underlying voicing, and morphological condition on d’ score when an accuracy threshold of 75% was adopted as criterion (n=52).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.12</td>
<td>0.09</td>
<td>12.23</td>
<td>0.56</td>
</tr>
<tr>
<td>Exposure Condition</td>
<td>-0.04</td>
<td>0.18</td>
<td>-0.22</td>
<td></td>
</tr>
<tr>
<td>Underlying Voicing</td>
<td>0.75</td>
<td>0.09</td>
<td>7.95</td>
<td></td>
</tr>
<tr>
<td>Morphological Condition</td>
<td>0.43</td>
<td>0.09</td>
<td>4.61</td>
<td></td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing</td>
<td>0.26</td>
<td>0.19</td>
<td>1.38</td>
<td></td>
</tr>
<tr>
<td>Exposure Condition x Morphological Condition</td>
<td>-0.67</td>
<td>0.19</td>
<td>-3.59</td>
<td></td>
</tr>
<tr>
<td>Underlying Voicing x Morphological Condition</td>
<td>0.22</td>
<td>0.19</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing x</td>
<td>1.20</td>
<td>0.37</td>
<td>3.21</td>
<td></td>
</tr>
</tbody>
</table>

*Note. All factors were contrast coding as follows: Exposure Condition (-0.5 = No Orthography, 0.5 = Orthography), Underlying Voicing (-0.5 = voiced, 0.5 = voiceless), Morphological Condition (-0.5 = plural, 0.5 = singular). Model Formula: \( d' \) ~ Exposure Condition * Underlying Voicing * Morphological Condition + (1 | Part_ID). *|t| > 2.0, indicating a significant effect (Gelman & Hill, 2007)

We additionally conducted an analysis of our data when no criterion threshold was adopted. Figure 4 shows the results for all 57 participants who met the inclusionary criterion. Again, the data looks remarkably similar to the data presented under the 90% and 75% criterion threshold, though d’ scores for each condition were a bit lower.

Figure 4: Mean d’ score by exposure condition, underlying voicing, and morphological condition. Error bars represent one standard error of the mean.
Statistical analyses were conducted as before and are summarized in Table 3 below. Again, we observed a main effect of underlying voicing, and of morphological condition. The two-way interaction of exposure condition and morphological condition and the three-way interaction of underlying voicing, exposure condition and morphological condition were again significant. Moreover, the two exposure groups differed from one another on the underlying voiced words in their singular form (estimate = 0.76, SE = 0.23, z = 3.28, p<.005), but not their plural form (estimate = -0.43, SE = 0.23, z = -1.86, p = 0.20), or for underlyingly voiceless words in their singular form (estimate = -0.08, SE = 0.23, z = -0.35, p = 1.00) or plural form (estimate = -0.08, SE = 0.23, z = -0.34, p = 1.00).

Table 3
Mixed effects model analysis examining the effects of exposure condition, underlying voicing, and morphological condition on d’ score when no criterion threshold is used (n=57).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>Estimate 1.28</td>
<td>SE 0.11</td>
</tr>
<tr>
<td>Exposure Condition</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>Underlying Voicing</td>
<td>0.87</td>
<td>0.14</td>
</tr>
<tr>
<td>Morphological Condition</td>
<td>0.49</td>
<td>0.14</td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>Exposure Condition x Morphological Condition</td>
<td>-0.68</td>
<td>0.27</td>
</tr>
<tr>
<td>Underlying Voicing x Morphological Condition</td>
<td>0.45</td>
<td>0.27</td>
</tr>
<tr>
<td>Exposure Condition x Underlying Voicing x Morphological Condition</td>
<td>1.39</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note. All factors were contrast coding as follows: Exposure Condition (-0.5 = NoOrthography, 0.5 = Orthography), Underlying Voicing (-0.5 = voiced, 0.5 = voiceless), Morphological Condition (-0.5 = plural, 0.5 = singular). Model Formula: dprime ~ Exposure Condition * Underlying Voicing * Morphological Condition + (1 | Part_ID).

*|t|> 2.0, indicating a significant effect (Gelman & Hill, 2007)
DISCUSSION

In the present study we set out to investigate the role of orthographic input in the adult acquisition of an L2 phonological process, German final devoicing. In particular, we asked whether the interfering effect of orthographic input and the trade-off effect are robust to methodological changes. Following Barrios and Hayes-Harb (2020), the present study employed an artificial lexicon study format, with the modification of a listening task with a two-way forced choice picture matching test rather than a production task. We additionally explored the effects of adjusting the criterion threshold that has been employed in artificial lexicon studies. We observed that participants generally performed well on global mismatch items regardless of word learning condition, suggesting that they had sufficient knowledge of the novel words. Of interest are the varied results found for participants’ performance on “voicing” mismatch trials. With respect to the interfering effect of written input, we found that the Orthography group was less sensitive to voicing mismatches for underlyingly voiced (alternating) words in their singular form than the No Orthography group, replicating findings reported by Hayes-Harb et al., (2018) and Barrios and Hayes-Harb (2020) using a picture-naming task. Unexpectedly, however, we failed to provide support for the trade-off effect observed by Barrios & Hayes-Harb (2020). Namely, while the d’ scores of the No Orthography group differed numerically from the Orthography group with respect to underlyingly voiced words in their plural form, this difference failed to reach significance.

We additionally investigated the potential implications for the generalizability of our results obtained from the experiment when a 90% accuracy criterion threshold is adopted. With respect to this, we observed that many of the participants were unable to
pass the criterion test with an accuracy of 90% or better. Adopting this threshold resulted in the exclusion of data from 27 of 57 participants (47% of the participants run on the protocol who met all other inclusionary criteria), which suggests that maintaining a stringent threshold can potentially impose limitations on study resources. The finding that adjustments to the criterion threshold only had the effect of modulating the d' scores up and down across the board but had no qualitative effect on our data suggests that, at least for the listening data that was collected in this study, a conservative criterion threshold is not required to ensure the quality of the data. Instead, adopting a stringent threshold only had the effect of reducing the power to observe effects in the data by reducing the total number of participants.

The present study is not without limitations which may be responsible for the lack of significant trade-off effect. Importantly, this may include the number of participants and size of the trade-off effect. Though the goal of the present study was to obtain 40 participants (20 participants per word learning condition), inclusionary criteria reduced this number to 26 participants. Other factors may include the task used to elicit the effect or even confounding variables posed by the stimuli. Participants may have developed a strategy for learning the tokens. For example, differences between speaker productions may have been large enough for participants to differentiate the tokens. Whatever strategy participants may have used, the experimental design likely did not account for all the possibilities.

In conclusion, it is known that adult, native English-speakers have difficulty acquiring the phonological process of final devoicing in L2 German. To adequately understand L2 phonology in literate, adult L2 learners, it is necessary to consider both
auditory and orthographic input during L2 acquisition. The finding that exposure to orthographic input during word learning appears to interfere with the acquisition of a German-like pattern of final devoicing suggests that orthographic input may not be a reliable resource during L2 acquisition. Orthographic input appears to affect L2 learners’ memory of L2 phonology, which may be beneficial for inferring the phonological forms of words in some cases but adverse in others. L2 instructors will need to take such findings into account when they are designing course content and study materials for their students. To avoid the adverse effects that orthographic input may pose, L2 instructors should moderate use of orthographic input in the classroom as well as inform students of strategies that may help them differentiate phonological forms when there are incongruencies between their L1 and L2 (Hayes-Harb et al., 2018).

It may be fruitful for future research to design studies that involve both phonological contrasts as well as phonological processes to address the research questions presented here. In these studies, it is necessary to determine if there is indeed a trade-off effect for orthographic input so that L2 instruction can be adjusted accordingly for the needs of L2 learners.
REFERENCES


doi:10.1002/bimj.200810425

doi:10.1177/0023830918761489


doi:10.1017/S0142716414000411


Name of Candidate: Cailey N. Lloyd

Birth date: March 31, 1996

Birth place: Fairbanks, Alaska

Address: 5446 SOUTH 3400 WEST

Roy, Utah 84067