German Phonics Game using Speech Synthesis -
A Longitudinal Study about the Effect on Orthography Skills

Kay Berkling¹, Nadine Pflaumer², Rémi Lavalley³

¹Cooperative State University, Karlsruhe, Germany
² Logopraxen, Ettlingen, Germany
³ Inline, Internet Online Dienste, Karlsruhe, Germany

berkling@dgbw-karlsruhe.de, pflaumer@logopraxen.de, remil@singularity.fr

Abstract

Acquisition of orthography is an important problem in German elementary schools. Today, few, if any, schoolbooks can claim to use knowledge of the deep syllable structure of German and its patterns for explicit orthography instruction. To fill this gap, a game described here allows children to explore the complete complexity of the German syllable patterns in analogy to phonics instruction used for English. The game is deployed on iPad and uses the Apple speech synthesis to allow children to listen to all possible letter combinations for legal German syllable structures. Through the explicit teaching of contextual Grapheme-Sound interaction, children are expected to generalize to new words in their writings that are automatically evaluated using speech technology. In the study presented here, 16 children participated in a weekly after-school session of one hour to play the German phonics game. The difference in achievement between the students and their control group were noticeably reduced by the end of the study.

Index Terms: speech synthesis, phonotactics, phonics, education, orthography, serious games

1. Introduction

Reading and Spelling are key skills acquired by children during their first four years of school. According to PISA and IGLU [1] however, a number of schoolchildren are left behind in Germany. PISA (2000-2012) has documented a growing discrepancy between students’ scores. On average, 15-year olds in Germany show comparative results internationally with respect to reading skills. However, Germany has less students at the higher levels of competence, while the number of students in the lower sections remains significant. Additionally, the gender gap has widened [2]. It is generally known that underachievement in reading and spelling acquisition can stem from a lack of a variety of skills, including phonemic awareness, knowledge of grapheme-phoneme correspondences and reading [3, 4, 5].

The driving variable for developing reading skills is orthographic knowledge in the first year and reading practice in subsequent years [6]. Without intervention in the first years, the lack of these skills predict to a large degree the performance at the end of elementary school [7] and further [8]. The degree to which these skills are acquired therefore has a direct impact on students’ scholastic performance across subjects. In order to prevent problematic developments, early reading, spelling and language skills have to be targeted in specific interventions as documented in the US National Early Literacy Panel [9]. Especially reading and spelling interventions administered from Grade 1 to Grade 2 show positive effects [10].

The study presented here is part of a long-term analysis on effective skill sequencing that has not been determined for German yet. Unlike English phonics which has a fairly standard way of sequencing from 1-syllable words with short vowels to 1-syllable words with long vowels and so on, the German sequence is presently not well studied nor apparent in first grade readers [11, 12]. There are several major problems impeding research with respect to orthographic abilities of children. Longitudinal studies are rare due to the enormous human effort of transcribing and annotating the spelling errors. Instead they are either evaluated on small datasets, or on broad categories or focus on children with specific learning disabilities [13, 14, 15, 16, 17, 18, 19]. Furthermore, longitudinal studies that exist often work with standardized tests that have limited retest potential.

Finally, no study proposes a clear scope and sequence for instruction that is then evaluated. For improved interpretation of orthographic error categories, a desirable goal is to determine a logical sequence of learning and analyze committed errors within this progression both at the syntax and orthographic level. With this goal in mind, phonics categories or patterns similar to those used in English phonics teaching [20, 21] for reading and writing have been redefined for German [11]. This study collects data that can be analyzed in a longitudinal manner on these orthographic patterns in order to study acquisition over time by using speech technology to automatically analyze spelling errors on large data sets. In doing so, we can circumvent the usual problem with standardized exams that limits many other studies. In our study we are bound only by the German language itself.

The work described here continues the previously published case study [22] and extends the evaluation to a set of 16 children over a 12 week period comparing their performance over time with themselves as well as with their classmates before and after the study took place. The combination of speech synthesis and game play define the central aspect of the proposed acquisition of orthography in this study. Students practice repeatedly while receiving immediate feedback on their concept of grapheme usage without adult intervention. In contrast, paper, as the usual recipient of student writing, provides no immediate feedback on the accuracy of the spelling. Instead, the game employs on iPad and uses the Apple speech synthesis to allow children to explore the complete complexity of the German syllable patterns in analogy to phonics instruction used for English. The game is deployed on iPad and uses the Apple speech synthesis to allow children to listen to all possible letter combinations for legal German syllable structures. Through the explicit teaching of contextual Grapheme-Sound interaction, children are expected to generalize to new words in their writings that are automatically evaluated using speech technology. In the study presented here, 16 children participated in a weekly after-school session of one hour to play the German phonics game. The difference in achievement between the students and their control group were noticeably reduced by the end of the study.

The rest of this paper is structured as follows. The prerequisites for the study are summarized in Section 2, the experimental set up will be described in Section 3. Section 4 describes the evaluation procedure. Section 5 reports on results followed by a discussion in Section 6 on current and future work needed.
2. Prerequisites

The goal is to elicit and analyse test data in a longitudinal study from children who use a German phonics game in order to see if and how certain skills taught by the game are acquired over time. This section describes both training and testing phase and their dependence on speech synthesis technology.

2.1. Playing - Training

The game used in this study is described in detail in [22]. The following serves as a brief review.

**Phonics** is a method that emphasizes explicit instruction regarding the interplay between graphemes and phonotactics. In English instruction the method adheres to a well-studied scope and sequence. The game uses this method by defining a sequence of syllable patterns from simple to more complex for each level. There are two arguments for starting with the 2-syllable structure. The trochee (two syllables, first one stressed, second one unstressed) is the key to word-segmentation in early language acquisition [23]. Secondly, German orthographic decisions are often based on the 2-syllable form (“Bad” /bat/ has a <d> because of “baden” or “rennt” has two <n> because of “rennen”). For these reasons, the first levels of the German phonics progress consist of consonant and vowel minimal pair changes in simple trochee words. With each level, a further step of complexity is introduced. The first levels are listed in Table 1. Figure 1 shows the highest level reached during the 12 weeks, with complex onset and two forms of usage of grapheme <h> (consonant vs. long vowel marker) in long vowel context.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1: Long Vowel</td>
<td>Magen, gehen, Biene, loben, Bude</td>
</tr>
<tr>
<td>L2: Short Vowel Silent Consonant</td>
<td>Watte, Betten, Spinnen, Robben, Puppe</td>
</tr>
</tbody>
</table>

**English Analogy**

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short vowel</td>
<td>cut, mat, plan</td>
</tr>
<tr>
<td>Long vowel Silent &lt;e&gt;</td>
<td>cute, mate, plane</td>
</tr>
</tbody>
</table>

The player is allowed to change a defined set of graphemes within each position of the syllable appearing in the game as tubes containing letters in movable bubbles. The goal is to build correctly spelled words and submit these for points. Students entering the word “Puppe” (“doll”) in the first level by using the method, where one grapheme represents one phoneme, will write <p><u><e><p><e> and fail. A student who thinks that "Butter", which is pronounced /bʊtaɪ/ is spelled <b><u><e><i><e> will fail. The user that will input <b><i><e><n><e> for "Biene" ("bee") will equally fail. These examples show that the instructional method based on a 1-1 correspondence between grapheme and phoneme does not apply to German. The examples also show the key principles that are taught in the first levels as described next.

**The Alphabetic Principle** As in English, a first step after learning the alphabet is to discover the pronunciation of various consonants in context (“cat” vs. “hat”) and the words’ corresponding semantic change. The ability to manipulate sounds relates to phonemic awareness [24] and by itself as well as in relation to graphemes is an important sub-skill to practice and master. Due to the synthesis system, any combination of letters can be tested by the player. Regular words can be constructed and sounded out: (“beten”; “lesen”; “reden”; “bieten”; “Duden”; “loben”). Equally any word with the pattern CVC<red> can be constructed and sounded out, where C is a single consonant, V denotes any long vowel, and red is part of the unstressed syllable containing an <e>/iː/. Due to the speech synthesis on any combination of letters, vowel length and various functions of grapheme-phoneme connections become apparent. For example, the letter <e> produces a different phoneme in the middle and end of these two-syllable words. While writing on paper does not support this immediate feedback, the game allows each child to proceed at their own speed and receive immediate feedback on their concept of grapheme usage.

**Silent Consonant** In analogy with the silent <e> in English, German has the silent consonant. In the middle of a trochee, a single consonantal sound following a short consonant is doubled at the grapheme level. This regular pattern is used to shorten the vowel preceding the consonant. An example of this is the word “remmen” where the double consonant <nn> is used to shorten the vowel belonging to the stressed grapheme <e> to an <l>. There are few minimal pairs like “Hütte” (hut) vs. “Hütte” (hat); were adding the silent consonant changes the meaning of the word. Correct writing of this orthographic pattern is not mastered by all students even by Grade 8, especially within compound words and high-frequency words [11].

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;i:e&gt; for /iː/</td>
<td>This category refers to the default German grapheme &lt;i:e&gt; marking the long vowel /iː/. This spelling pattern is not mastered by all students by Grade 8. The game does not allow these misspellings in the first levels and later, when given the choice of grapheme, the dictionary will not accept misspellings. The child will fail to gain points, thereby receiving immediate feedback on their misconception. Exceptions to regular patterns are not taught at the lower levels of the game.</td>
</tr>
</tbody>
</table>

**Reduction Syllable containing /a/** The synthesis clearly shows that any ending in their constructed word is not stressed and the constructed words are always read out by the synthesiser containing the shwa /a/. This regular feedback emphasizes the phonotactics for the grapheme <e> in long and short form and

![Figure 1: Screen shot of game at a high level. List of words collected can be seen along with progress bar and stars received. Clicking on "Lesen" will sound out the letter sequence in the window: "blühen".](image-url)
in stressed and unstressed position.

2.2. Writing - Testing

Based on an automatic tool for spelling error annotation [25] large amounts of text can be processed automatically. This tool builds on the assumption that the child will use graphemes based on the pronunciation of the word, which is misleading because the phoneme-grapheme correspondence is complex (for example, <sch> but not <sch>). The misspelled and the correctly spelled words are first converted to their pronunciation using rule based synthesis of MARY speech synthesis system [26]. The phonemes are then aligned based on their acoustic similarity. In a next step the graphemes are segmentated according to their pronunciation. The joint alignment of graphemes, phonemes, syllable boundary, stress patterns and morpheme boundaries (obtained from [27]) allows for a detailed automatic spelling error classification. In texts, each of the spelling errors are marked in two ways, the number of times the pattern occurs and the number of times there was a misspelling in the pattern.

In accordance with the German orthographic patterns emphasised in the game, the following error categories are analyzed.

Silent Consonant: The spelling error for the omission of the silent consonant is denoted by KV. Closely associated with this is the overgeneralization of the rule. An example is the word for “round” with correct writing of “rund” and misspelled as “runnt” in analogy with “rennt” (he runs). Erroneously doubling the silent consonant after short vowels is denoted by KV Hyp. The automatic spelling tagger marks both the number of times the pattern occurs overall (Base) and the number of times the pattern occurs and a mistake is committed (Error) for both of these categories. Equation 1 defines the reported result for this category.

$$\frac{Error[KV]}{Base[KV]} + \frac{Error[KV Hyp]}{Base[KV Hyp]}$$

(1)

For both of these categories. Equation 2 defines the reported result for this category.

$$\frac{Error[ie]}{Base[ie]} + \frac{Error[ie]}{Base[ie]}$$

(2)

Reduction Syllable containing - This category refers to misspellings in the reduction syllable. So few occurrences of spelling errors in this category appeared in the data that the category is not covered in detail in this publication.

3. Experimental Setup

An average elementary school in the state of Baden-Württemberg in southern Germany was approached with the idea of asking students to voluntarily join an experiment regarding the use of an iPad game to improve their orthography. The authors expected that around 10 kids would participate once a week after school. Instead, 16 2nd graders from three parallel classrooms enrolled. Two sections were opened, with eight children each, attending Tuesdays or Thursdays from 12:15 to 1:00 pm once a week. The protocol was similar each time. After 30 minutes of playing the game, the kids spent 15 minutes writing a text. The school has adapted the book “ABC der Tiere” [28] for teaching the first two years in elementary school. This is a syllable-based method. It explicitly teaches the function of the silent consonant by pronouncing it. It is less explicit on teaching shw in the reduction syllable or explaining the function of grapheme <ie>.

3.1. Playing - Training

For the first several weeks the levels of the game opened up progressively as children collected enough words for the presented pattern. More complex graphemes and combinations opened up that matched the patterns defined in Table 1. After all levels were opened, children were able to choose their own levels to play at. The game was usually played in pairs of two or alone according to the wishes of the children. The game required no explanation. Five correct words submitted gain one star and five stars accomplish a level. Failures were not penalized. Once the children discovered the use of the morpheme endings (last tube) to create new words quickly, they blindly tested all endings for quick points. A system of “lives” had to be introduced, where players had three mistakes (hearts) before failing the entire level of collecting 25 words, obtaining one star for every 5 words.

3.2. Writing - Testing

The goal was to observe the major spelling error categories that are addressed by the first levels of the game in order to see if their spelling changes over time. During the writing phase the kids are presented with pictures that elicit enhanced output with respect to the error categories under observation. Spelling errors are not corrected and no feedback is given on the writing. Figure 2 shows one writing sample, containing words like “Ziege” (goat), “liest” (reads), “liegt” (lying - misspelled), “gießt” (watering - misspelled) and “Bett” (bed) from the desired spelling patterns.

![Example of elicited text in week 10.](image)

Figure 2: Example of elicited text in week 10.
4. Evaluation

The success of the project was evaluated in three ways. Children’s progress was measured each week. All collected texts were transcribed digitally and annotated automatically with the error tagger. The progression of spelling ability is compared to the students’ peer group, a larger data base from surrounding schools and to themselves over time.

A pre- and post test was administered within the three classroom settings to all children in the study group within their peer group. The tests contained the words listed in Table 2.

Table 2: Wordlist for Pre- and Post-test. Words are elicited via pictures or dictation*.

<table>
<thead>
<tr>
<th>Category</th>
<th>Wordlist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest (23 items)</strong></td>
<td></td>
</tr>
<tr>
<td>Long Vowel (LV)</td>
<td>Lupe, Hose, Besen, Nadel</td>
</tr>
<tr>
<td>LV Challenges</td>
<td>Schuhe*, Sahne*, Beule*</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Koffer, Tunnel, Sonne</td>
</tr>
<tr>
<td>SC Challenges</td>
<td>Teller, Wippe, Butter</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Schnecke*, Katze*</td>
</tr>
<tr>
<td>SC Challenges</td>
<td>Zunge*, Töpfle*</td>
</tr>
<tr>
<td>Short Vowel</td>
<td>rund*, Murmel*</td>
</tr>
<tr>
<td>Short &lt;i&gt;</td>
<td>Kiste, Spinne*</td>
</tr>
<tr>
<td>Long &lt;ie&gt;</td>
<td>Biene, sieben*</td>
</tr>
<tr>
<td><strong>Post-test (45 items)</strong></td>
<td></td>
</tr>
<tr>
<td>Long Vowel (LV)</td>
<td>Lupe, Hose, Besen, Nadel, Rose, Feder</td>
</tr>
<tr>
<td>LV Challenges</td>
<td>Schuh*, Sahne*, Beule*</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Koffer, Tunnel, Sonne</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Wippe, Butter, Hammer, Roller</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Tanne, Tonne, Wasser, Kanne</td>
</tr>
<tr>
<td>SC Challenges</td>
<td>Sessel, Ritter</td>
</tr>
<tr>
<td>Silent Consonant (SC)</td>
<td>Zunge*, Töpfle*</td>
</tr>
<tr>
<td>SC Challenges</td>
<td>rund*, Murmel*, Wolke, Pinsel</td>
</tr>
<tr>
<td>Short Vowel</td>
<td>Kiste, Spinne*</td>
</tr>
<tr>
<td>Short &lt;i&gt;</td>
<td>Biene, sieben*</td>
</tr>
<tr>
<td>Long &lt;ie&gt;</td>
<td>Dieb, Brief, lieben*</td>
</tr>
</tbody>
</table>

Results are compared to two subgroups of the Karlsruhe Spontaneous Text Corpus [29] that were evaluated on spontaneously written texts with less density of test pattern material when compared to the pre- and posttest material.

5. Results

This section reports on the data collected and the results regarding the spelling errors over time and compared to peers and surrounding schools.

5.1. Data Collection

Not all children attended all sessions. Table 3 lists the number of texts that were available for each week. Week 9 was collected before and after a one week vacation. The data is then transcribed. The number of (unique) words elicited for each of the pictures presented to the children is shown in Figure 3.

5.2. Results Compared to Peers

For the group of kids taking part in the weekly sessions, words that were practiced more than once in their texts were excluded from the final test. Removing between 2-5 words for each child depending on which words they have practiced more than once during the last 12 weeks, made virtually no difference in their performance. This supports the theory that patterns were generalized to new words from the training sessions. Two children were outliers due to known learning deficiencies. While they also made progress, a detailed discussion about their performance is beyond the scope of this paper. No similar outliers existed in the peer group. They are removed from the average results depicted in Figure 4, including confidence interval, showing a significant improvement (CI 95%; p=0.0046) in the “ie+i” error category in the group under study. The error category “KV+KVHyp” was not quite statistically significant in improvement (CI 95%; p=0.090). Even though the numbers are not quite significant across the groups due to the small sample size, there was a noticeable growth in quality and length of children’s texts, and their confidence in explaining the spelling patterns on the qualitative side. Final results showed no significant difference in performance when compared to their peers at the end of the study.

5.3. Progression of Performance

Figure 5 depicts the average progression of performance for the group of 16 kids in the study. However, these average values are only an indicator of the group performance. In reality, each child has a very unique learning curve. To discuss these is beyond the scope of this paper. While this is a well studied phenomenon, it is remarkable to detect the apparent impact of the vacation around week 9 on the average result of the children. While there is large variation during the weeks, a trend towards improvement is discernible. Week 3 seems to be an outlier; it also represents the first week of spontaneous writing for the children, which might have impacted their performance in text length and quality.

5.4. Results Compared to Other Schools

Figure 5 also depicts the performance of results from this project compared to data sets from surrounding schools denoted with KK in the Graph. A subset of texts were chosen from dates closest to the study during the school year are November (47 texts) and June (40 texts). The reported scores are based on spontaneously written texts from second graders that are less dense in test items than the pre- and post tests administered to the peer group. They are more comparable to the spontaneous texts of the students during the study, keeping in mind that the pictures for the study contained a larger number of items with
patterns under observation. The comparison across data sets is therefore limited but serves as an indicator, showing that the kids in the study group of week 11 seem to outperform the KK group from June on spontaneously written text at the end of March.

6. Discussion and Future Work

In this study, we showed that children persisted in taking part in voluntary after-school activity including playing a German phonics game followed by a 15-minute writing session. While there are only 16 children in this study, it can be said that significant progress was made and that the children were motivated to play each week. The data shows that most of the kids that volunteered for the first session had on average more difficulties with spelling than their peers. The study did not include randomly chosen kids from three classrooms but rather selected kids whose parents thought the kids would benefit from extra support. Despite the promising results, the data set is clearly too small to make strong significant statements about results that generalize; larger studies are needed and planned. A closer look at pre-vacation fatigue is necessary. More levels will be added to the game to move on to more complex word structures. There are some indicators that the performance of children using the syllable method of instruction during school is slightly above the performance of surrounding schools that mostly use the method of 1-1 grapheme phoneme correspondence ("Lesen durch Schreiben"). This effect should be studied in more detail in future work.

7. Acknowledgements

We would like to thank the developer Alexei Copylove for donating his time to programming the application.
8. References


