Learning L2 pronunciation with a text-to-speech synthesizer

Walcir Cardoso

Abstract. The study explored the pedagogical use of Text-To-Speech (TTS) synthesizers by comparing two groups of students learning the pronunciations associated with Regular Past Tense (RPT) marking in English (e.g. talk[t], clean[d], want[id]). While one group received TTS-based instruction, the other was taught RPT with teacher assistance. Our findings indicate that both groups behaved similarly regarding the production of RPT, and that the target allomorphs are acquired in the following order: id>d>t.

Keywords: text-to-speech synthesis, L2 pronunciation.

1. Introduction

TTS are applications that convert written text into spoken voice output. Although the literature on its pedagogical applications in L2 education is still scarce, the few available studies indicate that the technology has potential for the teaching/learning of L2 pronunciation (Liakin, Cardoso, & Liakina, 2017). Research has also shown that the technology is ready for use in the L2 classroom, as its output is comparable to that of humans on measures such as intelligibility, comprehensibility, and morphophonemic accuracy (Cardoso, Smith, & Garcia-Fuentes, 2015).

We hypothesize that TTS can contribute to the learning of L2 pronunciation because it can enhance the target input in both quantity and quality via features such as speed control (slow/fast speech), multimodal visual presentation (through orthography and the highlighting of words being read/synthesized), word/phrase repetition, and voice selection (e.g. male or female voices for high phonetic
variability, as recommended by Logan, Lively, & Pisoni, 1991). These features are shown in Figure 1.

Figure 1. Input enhancement via TTS: features (reproduced with permission from NaturalSoft)

To examine the pedagogical suitability of TTS, this study focused on the acquisition of the allomorphy found in RPT marking in English (i.e. talk[t], clean[d] and want[id]). As suggested in the literature (Collins et al., 2009), the acquisition of RPT and related morphophonemics is difficult because -ed forms are not readily available to L2 learners (an issue of input ‘quantity’) and they occur in hard-to-perceive phonetic contexts (an issue of input ‘quality’). By allowing learners to manipulate the quantity and quality of exposure to the forms being learned, TTS can address the issues of availability and accessibility reported in Collins et al. (2009). This study was guided by the following question:

- Which type of instruction is more effective in the learning of simple past -ed allomorphy: TTS-based instruction or one led by a language teacher?

2. Method

2.1. Participants and experimental groups

Eighteen English-as-a-foreign-language students (proficiency: intermediate; age: 13-22; gender: seven females, 11 males) were stratified among two groups: while the TTS Group was asked to complete learning activities using the TTS application in Figure 1 (thus manipulating the L2 input in quantity and quality), the Non-
TTS Group received the same treatment with the assistance of an English teacher. Over a four-week period, all participants engaged in incidental learning of RPT allomorphy via listening activities that focused on the pronunciation of the target forms (e.g. listen and fill-in the gap, and sound categorization).

2.2. Instruments

The data were collected through two oral tasks consisting of 70 ed-final verbs and 30 distractors: Word Reading (id=10; d=23; t=12) and Guided Task (id=3; d=14; t=8) in which participants were asked to answer questions about the past holidays of fictional characters (see Figure 2 for an illustration).

Figure 2. Guided oral task

```
Last July, Lena took a four-week vacation in Japan. What did she do?

<table>
<thead>
<tr>
<th>Activity</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrive safely?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>work during her vacation?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>call her boss every day?</td>
<td>✓ (week)</td>
<td></td>
</tr>
<tr>
<td>stay at a hotel?</td>
<td>✓ (with friends)</td>
<td></td>
</tr>
<tr>
<td>wash her clothes every day?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>clean her bedroom?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>dress informally?</td>
<td>✓ (formally)</td>
<td></td>
</tr>
<tr>
<td>need guides to travel in Japan?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>love the weather?</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>travel with her friends?</td>
<td>✓ (boyfriend)</td>
<td></td>
</tr>
</tbody>
</table>
```

2.3. Research design and procedures

The study followed a pretest-posttest design. Participants were pretested using the two instruments described above and tested within one week of completion of the learning activities described above. The posttest was followed by a 20-minute oral interview in which participants were asked about their pedagogical experience. Figure 3 illustrates the research design and procedures adopted.

The data were coded by two research assistants (1=correct, 0=incorrect) and checked for consistency by a third rater (95.56% rater consistency). The interview
data were audio-recorded, transcribed, and categorized into pertinent themes (e.g. morphophonemic awareness and perceived pedagogical benefits).

Figure 3. Research design and procedures

3. Results

Due to space limitations, only the final stage of the statistical analysis is reported in detail. Mixed between-within subjects analysis of variance (ANOVA), followed by pairwise comparisons, were conducted with RPT allomorphy (-id, -d, -t) as dependent variable, and Time (pretest, posttest), Group (TTS, Non-TTS), and Task (Word Reading, Guided Task) as independent variables. For both Word Reading and Guided Task, there was no effect for Time or Group, indicating no significant improvement: the RPT allomorphs remained the same from pretest to posttest, across the two groups and tasks. There was, however, a difference for RPT production in the two tasks on the pretests and posttests, attesting that the three allomorphs behaved significantly differently; while -id was unproblematic, -d was relatively less difficult to produce than -t: id>d>t (where > indicates “easier to produce than” or “acquired before”). Finally, additional analysis revealed no significant improvement in Word Reading. Due to the uneven distribution of RPT forms, the results are provided in mean percentages in Table 1.

Table 1. RTP allomorphs by group (type of instruction), time, and task

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest (%)</th>
<th>Posttest (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word Reading</td>
<td>Guided Task</td>
</tr>
<tr>
<td></td>
<td>-id</td>
<td>-d</td>
</tr>
<tr>
<td>TTS</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>Non-TTS</td>
<td>97.8</td>
<td>13.3</td>
</tr>
</tbody>
</table>

Table 1.
Because the two groups were statistically similar and no improvement was observed in Word Reading, follow-up Within-Subjects ANOVAs involving Guided Task and the two groups combined were conducted. This revealed an interaction between Time and RTP allomorphy ($F(1,17)=15.609$, $p=0.001$ (alpha level set at .003)), suggesting a performance improvement for -d (shaded in Table 1), but not for -t on the posttest.

To summarize, the TTS and Non-TTS groups behaved similarly regarding the production of RPT: while they both improved in producing [d], there was no significant improvement in their production of -id (participants scored at the ceiling level on the pretest) and [t].

4. Discussion and concluding remarks

The study examined the effects of two types of instruction in the learning of RPT allomorphy: TTS-based instruction and one led by a language teacher. The results indicate that the TTS and Non-TTS groups behaved similarly regarding the production of RPT. An interesting pattern emerged in the analysis: an improvement of allomorph -d by both groups, suggesting a development sequence in the acquisition of RPT allomorphy. While -ed is acquired first, due to L1 influence (see Cardoso, 2011), the order -d>-t can be explained by the fact that syllabic coda -d is more sonorous than -t (languages favor sonorous codas; Cardoso, 2011).

The lack of major improvements observed can be attributed to at least two factors. Firstly, it is possible that an exclusive focus on perception via the proposed TTS-based listening activities (under the ‘perception precedes production’ hypothesis – Flege, 1995) was not appropriate to lead learners to transfer the newly-acquired knowledge into production. If these activities had been complemented with speaking activities (e.g. via speech recognition software or even silent practice), learners would have had an opportunity to practice and possibly transfer this novel perceptual knowledge into production. That many learners became aware of RPT allomorphy became evident in the analysis of their interview data, particularly among participants in the TTS Group. When asked what they had learned in the study, most participants reported that they had become aware of the differences in RPT pronunciation: “if I were to [listen to these stories] again, I wouldn’t pronounce as much this “ed” at the end”; “I never knew that the {ed} sometimes sounds strong, sometimes weak”. Secondly, it is possible that the four hours allotted for listening practice was not enough, particularly in an experiment in which
Learning L2 pronunciation with a text-to-speech synthesizer

pronunciation was learned implicitly (see Saito & Lyster, 2012, for the effects of explicit pronunciation teaching).

An important pedagogical implication of these findings is that TTS has the potential to increase and improve the target language input (preferably accompanied by explicit instruction and oral practice), develop better grapheme-to-phoneme associations (there is no clear relationship between orthographic -ed and its allomorphs -id, -d and -t in English), and address some of the time constraints that affect many L2 learning environments (TTS has the potential to foster autonomous and anytime-anywhere learning). Using TTS as an out-of-class pedagogical tool could then increase in-class time so that teachers and students could focus on other important tasks such as providing feedback and engaging in real-life communicative activities.

References
