Invented spelling and speech synthesis feedback in a deep orthography

Considerations in designing a speech synthesis feedback

Stine Fuglsang Engmose, lk494@hum.ku.dk

This poses a potential problem for the pre-reader or beginning reader seeking to master the alphabetical principle. The child must learn the most salient letter-to-sound correspondences and recognize them when segmenting a word. Developing this ability is unlikely to be supported by a synthesis that reads with alternating letter-to-sound conversions (see example in Fig. 2). A synthesis that consistently converts letters into sound using the salient letter-to-sound correspondences is more likely to support acquisition of the alphabetical principle.

The problem

Speech synthesis is usually designed to read aloud in as natural-sounding a way as possible. In a deep orthography, like that of Danish, the synthesis must thus assign different sounds to the same letter depending on e.g. position in the word, adjacent letters, morphological structure, or word-specific pronunciations. This is how the speech synthesis read. Two Danish speech syntheses were selected for comparison in this analysis: one commercially available unit-selection speech synthesis (IntonHonds) and one diphone synthesis developed and available for research purposes (Hennrichsen, 2004). The two synthesis transform letters into sounds using the same basic processes, text analysis, and waveform synthesis (Jønsten & Martin, 2014).

How the speech synthesis read.

Two Danish speech syntheses were selected for comparison in this analysis: one commercially available unit-selection speech synthesis (IntonHonds) and one diphone synthesis developed and available for research purposes (Hennrichsen, 2004). The two speech syntheses transform letters into sounds using the same basic processes, text analysis, and waveform synthesis (Jønsten & Martin, 2014).

Table 1. Synthesis demands and operationalization

<table>
<thead>
<tr>
<th>Characteristics of feedback</th>
<th>Characteristics of synthesis</th>
<th>Synthesis demands</th>
<th>Type</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draw children's attention to the process and result of blending phonemes</td>
<td>Letters are converted to the most common sound</td>
<td>1a) Code text normalization so that letter input = letter output</td>
<td>Diphone</td>
<td>Possible because access to code at text analysis level</td>
</tr>
<tr>
<td>Draw children's attention to the connection between letters and their most common sound</td>
<td>Letters are converted to the most common sound</td>
<td>1b) Alternating g2p assignment because of deep Danish orthography and big W strategy as possible</td>
<td>Diphone</td>
<td>Possible because code can be modified based on how frequently it reads</td>
</tr>
</tbody>
</table>

Background

Results from the intervention studies (see Fig. 1) shows, in different degrees, positive influence on early literacy skill from letting kindergarten or pre-school children, with limited or none reading ability, write with invented spelling and supportive corrective feedback. Positive effect has been shown when feedback is given by both the teacher and peers and professional, in combination, facilitating corrective feedback.

It seems obvious that providing tailored feedback on invented spelling is a task that is time consuming and requires teacher knowledge. If a synthetic voice could support or replace feedback from the teacher then it would be a potential beneficial tool in a kindergarten classroom context, where time and teacher's knowledge about literacy instruction are limited.

Question

On the basis of the theoretical established characteristics…

...in which of the selected speech syntheses, diphone or unit-selection, can feedback in response to invented spelling be operationalized?

Analysis

The feedback aims to heighten the quality of children's invented spelling. Feedback characteristics must thus support acquisition of knowledge of common letter-to-sound correspondences and develop ability to segment words into phonemes. These feedback characteristics are interpreted into speech synthesis characteristics, which translate into a list of demands for the speech synthesis. The final column describes whether and how these demands can be operationalized in the two selected speech syntheses.

Table 2. Characteristics of feedback

<table>
<thead>
<tr>
<th>Type</th>
<th>Text analysis</th>
<th>Phonetic analysis</th>
<th>Waveform synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit selection</td>
<td>Inaccessible code</td>
<td>Real words: Pronunciation dictionary = e.g. APPLE = /æp.əl/</td>
<td>Concatenate sequences of units from a recorded database</td>
</tr>
<tr>
<td>Diphone</td>
<td>No text normalization code</td>
<td>This level is accessible with regard to g2p strategy. Pronunciation dictionary for real words</td>
<td>Concatenate diphones</td>
</tr>
</tbody>
</table>

Conclusions

By interpreting theoretically established feedback characteristics given in response to invented spelling and transforming these demands into the synthesis, it became clear how it was possible to operationalize these demands in only one of the two analyzed speech syntheses. The commercially available unit selection speech synthesis did not meet the theoretically desired characteristics of corrective feedback because of ongoing alteration in connections between letter and sound. This is due to a combination of:

1. The synthesis' innate purpose of reading aloud in as natural-sounding a way as possible.
2. The deep Danish orthography.
3. Inaccessible text analysis level. No possibility of adapting code that prescribes how text normalization and phonetic analysis are conducted.

The diphone synthesis, developed for research purposes, met the demands since it was possible to adjust the code at the text normalization and phonetic analysis levels. These adjustments are necessary if a speech synthesis developed for a deep orthography like Danish is to assign sound to letters in a stable way.

It is not the degree of sophistication of the technology per se – diphone is the less advanced of the two syntheses – that prescribes its relevance. It is necessary to evaluate whether the established criteria can be operationalized by a technology before using it for one's research.

Perspectives

Results from an ongoing intervention study will reveal whether the feedback on invented spellings, provided by the synthesis, is as effective as corrective feedback from a reading professional in stimulating kindergarten children’s early literacy skills. New questions arise as to whether invented spelling with carefully designed speech synthesis feedback can develop early literacy skills in a deep orthography. Do the synthesis characteristics need to be the same as in a shallow orthography? Which characteristics of the speech synthesis are most salient? Do word regularity and difficulty affect learning? Are the children’s skills to start from important for the learning outcome? And could feedback from a speech synthesis support mastery of conventional spelling in a deep orthography?