University of Copenhagen
Faculty of Humanities

Invented spelling and speech synthesis feedback in a deep orthography

Considerations in designing a speech synthesis feedback

Stine Fuglsang Engmose, lxk494@hum.ku.dk

TASK	FEEDBACK	MAIN FEEDBACK FACILITATOR	STUDY
Invented spelling of isolated words	Modeling feedback	Reading professional	Levin & Aram (2013); Martins & Silva (2003, 2006); Ouellette & Sénéchal (2008); Ouellette, Sénéchal et al. (2012, 2013); Rieben, Ntamakiliro, Gonthier, & Fayol (2005)
		Peer and reading professional	Albuquerque & Alves-Martins (2016)
		Speech synthesis	??

Fig. 1 The idea

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 with both reading professionals, 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.

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 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.

PONY/'pouni/

P	PO	PON	PONY
/p/ /pi:/	/pou/ /pi:ou/	/poun/ /pɑ:n/ (pond)	/pouni/

P	PO	PON	PONE
/p/ /pi:/	/pou/ /pi:ou/	/poun/ /pɑ:n/ (pond)	/poun/

P	PO	POE	
/p/ /pi:/	/pou/ /pi:ou/	/poue/ (poet)	

Fig. 2 The problem with feedback from synthesis during writing

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?

How the speech synthesis read.

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

Type	Text analysis		Waveform synthesis
	Text normalization e.g. = for instance N.Y. = New York UN = United Nations or "read as letter names"	Phonetic analysis Real words: Pronunciation dictionary → e.g. APPLE = /ˈæp.əl/ Unknown words: g2p based on probability of pronunciation in a given width (W) → e.g. PONE = /poun/ or /pouniː/	Generation of waveform Concatenate sequences of units from a recorded database
Unit selection	Inaccessible code Recode abbreviations, numbers, etc. into letter strings	Inaccessible code W as wide as possible	Concatenate units
Diphone	No text normalization code	This level is accessible with regard to g2p strategy: * W can be defined Possible to add LEX strategy: * pronunciation dictionary for real words	Concatenate diphones

Characteristics of feedback	Characteristics of synthesis	Synthesis demands	Type	Operationalization
Draw children's attention to the connection between letters and their most common sound	Read letters using stable g2p conversion rules Letters are converted to the most common sound	1a) Code text normalization so that letter input = letter output 1b) Convert normalized text to internal phonological representation = consequently convert each letter to its most common sound	Unit	Not possible since text analysis level is an inaccessible code a) e.g. can be normalized into "for instance" b) alternating g2p assignment because of deep Danish orthography and as big W strategy as possible
			Diphone	Possible because of access to code at text analysis level a) text normalization unnecessary b) g2p based on W1 strategy
Draw children's attention to the process and result of blending phonemes	Read by blending phonemes during writing Children hear a blend of the entire letter string every time they add a letter to their spelling	2) Code that prescribes when the synthesis is read	Unit	Not possible since this synthesis is read during writing, when the spacebar is pressed Code is not accessible and cannot be modified
			Diphone	Possible because code can be modified based on how frequently it reads e.g. first letter = read, then x msec break, then read again During slow writing, this is perceived as ongoing blending of preceding letter string and the just-added letter
Make children aware of redundant, missing, or odd letters Create an incentive for continued work with spelling	Read with slow speech rate	3) Code the length of each phoneme	Unit	Possible because it is available as an adjustable parameter for the user
			Diphone	Possible to change duration of each phone in code to a set length e.g. 0.5 msec
Make children recognize phonological plausible spellings as acceptable spelling attempts	Read unconventional spelling attempts by blending each phoneme's standard sound	4) Result of 1 and 2	Unit	Not possible because the deep Danish orthography and the g2p strategy "W as big as possible" leads to alternating connections between letters and sounds
			Diphone	Possible because we can code the synthesis to make stable g2p conversions on the basis of an adapted W1 strategy
Make children recognize conventional spellings as acceptable spelling attempts	Read conventionally spelled words in a conventional manner	5) Make transformation into internal phonetic representation by searching a pronunciation dictionary	Unit	Possible because this aligns with the innate purpose of the synthesis – only problem is homographs
			Diphone	Possible to add a LEX strategy to the basic g2p strategy – only problems are homographs and quality of synthesis

Fig. 3 The analysis

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.

Conclusions

By interpreting theoretically established feedback characteristics given in response to invented spelling and transforming these into demands for the synthesis, it became clear that 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?