



HAL
open science

Developing a Reading Tutor: Design and Evaluation of Dedicated Speech Recognition and Synthesis Modules

Jacques Duchateau, Yuk On Kong, Leen Cleuren, Lukas Latacz, Jan Roelens, Abdurrahman Samir, Kris Demuynck, Pol Ghesquière, Werner Verhelst, Hugo Van Hamme

► To cite this version:

Jacques Duchateau, Yuk On Kong, Leen Cleuren, Lukas Latacz, Jan Roelens, et al.. Developing a Reading Tutor: Design and Evaluation of Dedicated Speech Recognition and Synthesis Modules. *Speech Communication*, 2009, 51 (10), pp.985. 10.1016/j.specom.2009.04.010 . hal-00558520

HAL Id: hal-00558520

<https://hal.science/hal-00558520>

Submitted on 22 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Developing a Reading Tutor: Design and Evaluation of Dedicated Speech Recognition and Synthesis Modules

Jacques Duchateau, Yuk On Kong, Leen Cleuren, Lukas Latacz, Jan Roelens, Abdurrahman Samir, Kris Demuynck, Pol Ghesquière, Werner Verhelst, Hugo Van hamme

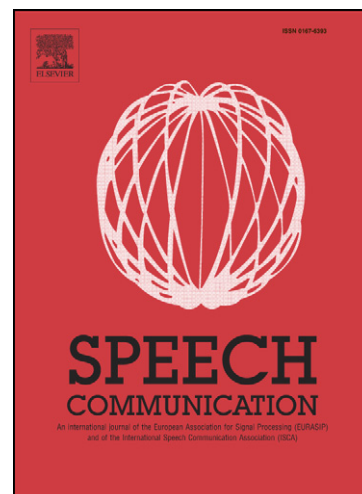
PII: S0167-6393(09)00069-7
DOI: [10.1016/j.specom.2009.04.010](https://doi.org/10.1016/j.specom.2009.04.010)
Reference: SPECOM 1804

To appear in: *Speech Communication*

Received Date: 29 May 2008
Revised Date: 17 April 2009
Accepted Date: 20 April 2009

Please cite this article as: Duchateau, J., Kong, Y.O., Cleuren, L., Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière, P., Verhelst, W., hamme, H.V., Developing a Reading Tutor: Design and Evaluation of Dedicated Speech Recognition and Synthesis Modules, *Speech Communication* (2009), doi: [10.1016/j.specom.2009.04.010](https://doi.org/10.1016/j.specom.2009.04.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Developing a Reading Tutor: Design and Evaluation of Dedicated Speech Recognition and Synthesis Modules

Jacques Duchateau^{a,*}, Yuk On Kong^b, Leen Cleuren^c,
Lukas Latacz^b, Jan Roelens^a, Abdurrahman Samir^a,
Kris Demuyne^a, Pol Ghesquière^c, Werner Verhelst^b,
Hugo Van hamme^a

^a*ESAT department, Katholieke Universiteit Leuven,
Kasteelpark Arenberg 10 PO 2441, B-3001 Heverlee (Leuven), Belgium*

^b*ETRO department, Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussel, Belgium*

^c*Centre for Parenting, Child Welfare and Disabilities, Katholieke Universiteit
Leuven
Vesaliusstraat 2 PO 3765, B-3000 Leuven, Belgium*

Abstract

When a child learns to read, the learning process can be enhanced by significant reading practice with individual support from a tutor. But in reality, the availability of teachers or clinicians is limited, so the additional use of a fully automated reading tutor would be beneficial for the child. This paper discusses our efforts to develop an automated reading tutor for Dutch. First, the dedicated speech recognition and synthesis modules in the reading tutor are described. Then, three diagnostic and remedial reading tutor tools are evaluated in practice and improved based on these evaluations: (1) automatic assessment of a child's reading level, (2) oral feedback to a child at the phoneme, syllable or word level, and (3) tracking where a child is reading, for automated screen advancement or for direct feedback to the child. In general, the presented tools work in a satisfactory way, including for children with known reading disabilities.

Key words: reading tutor, computer-assisted language learning (CALL).

* Corresponding author. Tel: +32 16 321861, Fax: +32 16 321723.

Email address: Jacques.Duchateau@esat.kuleuven.be (Jacques Duchateau).

1 Introduction

2 For a child with reading difficulties, daily reading practice and individualised
3 support is necessary. However, due to practical limitations (e.g. in Flemish
4 regular elementary schools only one teacher is in charge of a group of twenty
5 to thirty students) and federal law restrictions (e.g. only a limited amount
6 of individual reading therapy is covered by the national health care system),
7 it is often difficult to provide the child with sufficient practice and support.
8 Therefore, a fully automated reading tutor, able to track the child's reading
9 progress and to accurately detect reading errors on the one hand, and able
10 to act as a fluent reading model (to read aloud for the child, read along with
11 the child etc.) and to give adequate personalised feedback on the other hand,
12 would be beneficial.

13 This paper discusses our efforts to develop an automated reading tutor for
14 Dutch. The development of such a reading tutor is one of the aims of the
15 SPACE project's educational part (SPACE is the acronym for SPeech ALgorithms
16 for Clinical and Educational applications). The idea is to improve
17 speech technology so that both the speech recogniser and the speech synthe-
18 siser can accomplish the reading tutor tasks described above. More information
19 on the SPACE project can be found on its website¹.

20 In the literature, many systems for CALL (computer-assisted language learn-
21 ing) have been described that incorporate language and speech technology.
22 However, it can be concluded from the proceedings of recent workshops on
23 the topic (e.g. the 2007 SLaTE workshop on Speech and Language Technol-
24 ogy in Education), that a clear majority of these systems focus on second
25 language learning adult users. Some examples of these systems are described
26 in Probst et al. (2002); Neri et al. (2006); Abdou et al. (2006) and D'Mello
27 et al. (2007).

28 Literature on CALL systems for children, including reading tutors, is more
29 scarce, especially papers on systems that have been evaluated in practice on
30 children in general, and on children with reading difficulties in particular.
31 One important reason for this is that the development of CALL systems for
32 children, especially the speech recognition and synthesis modules in it, is a
33 very challenging task due to reading-related and child-related development
34 processes. For example, the awareness of young children of phonetic processes
35 (or linguistic processes in general) is rather limited. Also, due to a child's
36 variable maturation, articulatory competencies differ and due to variable word
37 decoding skills, oral reading of inexperienced readers or readers with reading
38 disabilities can be fraught with oral reading errors. In the past decade, several

¹ SPACE website: <http://www.esat.kuleuven.be/psi/spraak/projects/SPACE>

39 projects on the development of CALL systems were started, as described e.g.
40 in Hagen et al. (2007); Banerjee et al. (2003); Black et al. (2007); Eskenazi
41 and Pelton (2002); Russell et al. (2000) and Adams (2006).

42 This paper is organised as follows. Section 2 gives a description of the synthesis
43 (section 2.1) and recognition (section 2.2) modules. Section 3 presents three
44 reading tutor tools: successively the automatic reading level assessment (sec-
45 tion 3.1), the oral feedback by synthesis (section 3.2), and the reading tracker
46 (section 3.3) are discussed and evaluated. Finally, in section 4 our conclusions
47 and directions for future work are given.

48 2 Description of the synthesis and recognition systems

49 This section discusses the dedicated speech synthesis and recognition modules
50 which are integrated into our reading tutor. The main modules of the reading
51 tutor concern:

- 52 • Management of the reading tasks, to define each task and the possible use
53 of synthesis and recognition in it.
- 54 • Management of the student, to define which student should do what tasks.
- 55 • Management of the recordings, including diagnostic tools for the supervisor.
- 56 • Making of the recordings, which allows the student to do the tasks.

57 The reading tutor was mainly used (1) to record the Chorec database (see
58 section 3.1), which consists of recordings of reading tasks in which the synthesis
59 and recognition modules were *not* used, and (2) to evaluate the use of the
60 synthesis and recognition modules in remedial reading tasks. The remainder
61 of this section details both modules.

62 2.1 *Speech synthesis system*

63 2.1.1 *The synthesiser*

64 Corpus-based concatenative speech synthesis (e.g. Hunt and Black, 1996; Chu
65 et al., 2003; Clark et al., 2007) is the mainstream way to synthesise speech. In
66 such synthesisers, a large speech database is first segmented into small units.
67 To synthesise an input text, the best combination of speech units is selected
68 from the database to match the utterance, based on the sum of weighted
69 cost functions. The selected unit sequence is then concatenated to generate
70 the synthesis. In the SPACE project, the database of the speech synthesiser
71 contains recordings of an early-middle-aged, female, native Flemish speaker.

72 She has a pleasant voice, appropriate for children, who are the target popu-
73 lation of the reading tutor. If necessary, new voices can be created later by
74 recording other speakers, as the synthesiser is data-driven. Alternatively, new
75 voices can be created through voice conversion techniques. Our current speech
76 database contains about 3 hours of recordings. Besides a full Dutch diphone
77 set (which serves as the last resort), the recordings mainly consist of graded
78 Dutch story material, as this suits the application. The database has been seg-
79 mented using the forced aligner described in Demuynck et al. (2004). A unit
80 selection synthesiser has two parts: a language-dependent front-end providing
81 text analysis, and a language-independent back-end providing unit selection.
82 In our case, text analysis is performed by an adapted version of the front-
83 end of the Dutch diphone synthesiser, NeXTeNS (Kerckhoff and Marsi, 2002).
84 Three main changes were implemented:

- 85 • The lexicon is adapted in order to match Flemish pronunciation.
- 86 • The post-lexical rules are not used, so as to match the labeling of the speech
87 database.
- 88 • Besides silences between phrases, silences between words within a phrase can
89 also be predicted. Training is performed on the segmentation and labeling
90 of the speech database, instead of manually labeling the utterances. Silences
91 are predicted using memory-based machine learning techniques (Daelemans
92 and Van den Bosch, 2005). Features used by the memory-based learner
93 include part of speech, word identity, and punctuation before and after
94 words.

95 Actually, our synthesis system is a combination of four different synthesisers
96 under the same framework: (a) a diphone synthesiser, (b) a unit selection
97 synthesiser using uniform units (diphones or monophones), (c) a hierarchical
98 synthesiser, and (d) a novel non-uniform unit synthesiser primarily looking
99 for the longest possible chunks. The front-end is shared by all synthesisers.
100 With the last two synthesisers, we aimed at obtaining higher quality than
101 state-of-the-art, and at capturing the required natural prosody existing in
102 the speech corpus by maximising the mean unit length in syntheses. Both
103 words and full utterances can be synthesised with our synthesis system. The
104 hierarchical synthesiser has been implemented with three levels: phrase, word
105 and diphone. For each input utterance, it first looks for units at the phrase
106 level. If it fails for some part of an utterance or the whole of it, it will back
107 off to the next level and search again for the missing parts. As all Dutch
108 diphones are in the inventory, each utterance can be completed. For the fourth
109 and novel selection algorithm, long non-uniform units are selected from the
110 speech database (Latacz et al., 2007). In brief, these non-uniform units could
111 represent any sequence of diphones so that units as long as syllables, words
112 and beyond could be the target. The best unit sequence is selected according
113 to target and join costs, which can be based on symbolic and/or acoustic
114 features. The join costs measure how smooth the transition between candidate

115 units is. Differences in spectrum, pitch and energy are taken into account. The
116 target costs determine how well a unit matches its symbolic target description.
117 The target is described in terms of linguistic features like syllable structure,
118 lexical stress, and phonemic and syllabic contexts. Synthesising full utterances
119 is more difficult than synthesising single words because phrase breaks could
120 occur between words and the prosody should be more complex and variable.
121 Prosody is not predicted explicitly but is obtained through symbolic features.
122 In addition to the features mentioned above, symbolic features such as syllable
123 accent, part of speech, the position in the utterance and the phrase, and the
124 number of syllables to the next and the previous accent, stress, and phrase
125 breaks are also used for synthesising utterances.

126 As the synthesiser serves partly to model and demonstrate for the child, we
127 strive to achieve high intelligibility. Firstly, we use a speech database contain-
128 ing the same type of material as required for the application, namely stories for
129 children, because the quality of the synthesised speech of a unit selection syn-
130 thesiser is known to be better this way than if generic material is used (Black
131 and Lenzo, 2001). By doing so, the synthesiser is optimised for the application
132 and the target population. Secondly, we use a novel algorithm, hierarchical
133 synthesis, for the synthesiser. As it searches for longer chunks for synthesis,
134 this should raise intelligibility as well. Thirdly, in order to make it easier for
135 the child to understand synthesised speech and to make sure that the child
136 can read along with synthesised speech, the synthesis speech rate must not
137 be too high. The synthesis speech rate can be adjusted online and, if needed,
138 short pauses can be inserted between consecutive words, in order to achieve
139 an even slower speech rate but with lower chance of creating artifacts. The
140 length of these pauses can also be adjusted. For illustration purposes, some
141 audio samples of synthesised words are available on our website².

142 *2.1.2 Phoneme-by-phoneme and syllable-by-syllable synthesis modes*

143 Usually, existing speech synthesisers provide a speaking style corresponding
144 to fluently read text. In contrast, speech therapists or teachers use differ-
145 ent speaking styles when interacting with their patients. Additional speak-
146 ing styles or reading modes increase the effectiveness of the reading tutor, as
147 demonstrated in Heiner et al. (2004). Both phoneme-by-phoneme and syllable-
148 by-syllable speech are needed for modelling and for giving feedback. Two types
149 of syllabified speech are considered, with isolated or lengthened connected syl-
150 lables. The latter is needed for assisting more advanced readers to build up
151 their fluency and reading speed.

² See <http://www.etro.vub.ac.be/Research/DSSP/Demo/SpeechCommunication1.htm>

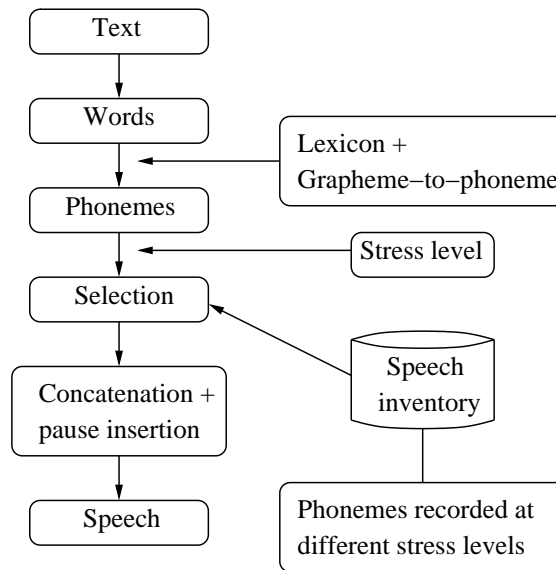


Fig. 1. Overview of phoneme-by-phoneme mode synthesis (Latacz et al., 2006)

152 *Phoneme-by-phoneme mode.*

153 In most current speech synthesisers, orthographic spelling mode, or letter-by-
 154 letter spelling, is implemented. A phoneme-by-phoneme mode which allows
 155 stressing/emphasis is, however, more useful for training people with read-
 156 ing disabilities. In this mode, each single word is pronounced phoneme by
 157 phoneme. If necessary, a phoneme can be pronounced with stress or emphasis
 158 to attract the attention of the child to this particular phoneme. Note that
 159 stress here does not mean lexical stress.

160 An overview of the phoneme-by-phoneme mode synthesis is given in figure 1.
 161 Phonemes of the same stress level were recorded with similar pitches, speech
 162 rates and loudness levels to obtain natural-sounding synthesised speech. There
 163 are two different stress levels to choose from, as we recorded each Dutch
 164 phoneme from the speaker in two versions: with stress and without stress.
 165 No signal processing is involved.

166 The input text is processed into a list of words. Words are converted into a
 167 stream of phonemes using a lexicon and grapheme-to-phoneme conversion. The
 168 appropriate phonemes with their associated stress-levels are then selected from
 169 a speech inventory and concatenated with a silence in between. Utterances
 170 can then be spelled phoneme by phoneme. As expected, the quality of the
 171 synthesised speech of this mode is found to be very high. The domain could
 172 be extended in order to synthesise feedback like *not /h a t/ but /h a d/*, which
 173 is straightforward to implement.

174 *Syllable-by-syllable mode.*

175 The purpose of this reading mode is to synthesise speech as either isolated
 176 or lengthened connected syllables. To synthesise speech in isolated syllables,

177 silences are inserted between neighbouring syllables by analysing the input
178 text. The output synthesis is then syllabified. As for synthesising lengthened
179 connected syllables, pre-synthesised speech is stretched (Latacz et al., 2006).

180 2.2 *Speech recognition system*

181 The task of the speech recognition system in a reading tutor seems to be
182 easy as the words or sentences that should be read, are known. But the child
183 will not read everything correctly, and it is difficult to predict what the child
184 will say exactly: he/she may even give comments that are not related to the
185 reading task. Therefore, a recogniser set-up was adopted that is very similar
186 to a recognition system which we use for large-vocabulary continuous speech.
187 This section describes the acoustic modelling, the search engine(s) and the
188 *language* models involved in the system.

189 The acoustic modelling is based on a 22-hour read speech database in Dutch,
190 which is different from the databases we used for the experiments below in
191 section 3. It contains recordings of continuous sentences read or spoken flu-
192 ently by children aged between 5 and 11 years. Cross-word context-dependent
193 acoustic models were estimated with 1400 tied Hidden Markov Model (HMM)
194 states and 16000 tied Gaussian distributions in total. A straightforward sig-
195 nal processing scheme based on the Mel-spectrum was adhered to, including
196 cepstral mean subtraction, discriminant analysis and Vocal Tract Length Nor-
197 malisation (VTLN). As we are interested in tracking where a child is read-
198 ing, we developed a VTLN system that does not introduce latency in the
199 pre-processing. In the estimation of the warping factor in the VTLN system,
200 latency (or the use of an initial recording for the estimation) could be avoided
201 by using an adaptive warping factor estimation. This way, the only latency
202 in our pre-processing is due to taking first and second order time derivatives,
203 which results in a latency of 3 frames (30 msec).

204 For the decoding of the speech, a recognition system architecture with two
205 layers was adopted, as depicted in figure 2. This architecture can also be used
206 for other recognition tasks. It was, for instance, applied successfully in large-
207 vocabulary continuous speech recognition. This recognition system is discussed
208 in detail in Demuyne et al. (2006).

209 In the first layer, a task-independent phoneme recogniser generates a phoneme
210 lattice. The phoneme recogniser is based on the acoustic modelling detailed
211 above, and on a trigram phoneme sequence model estimated from a Dutch
212 database with correctly read sentences.

213 In the second layer, the task-dependent information is modelled. As the sen-
214 tence that should be read is known in a reading tutor, we opted for a finite-

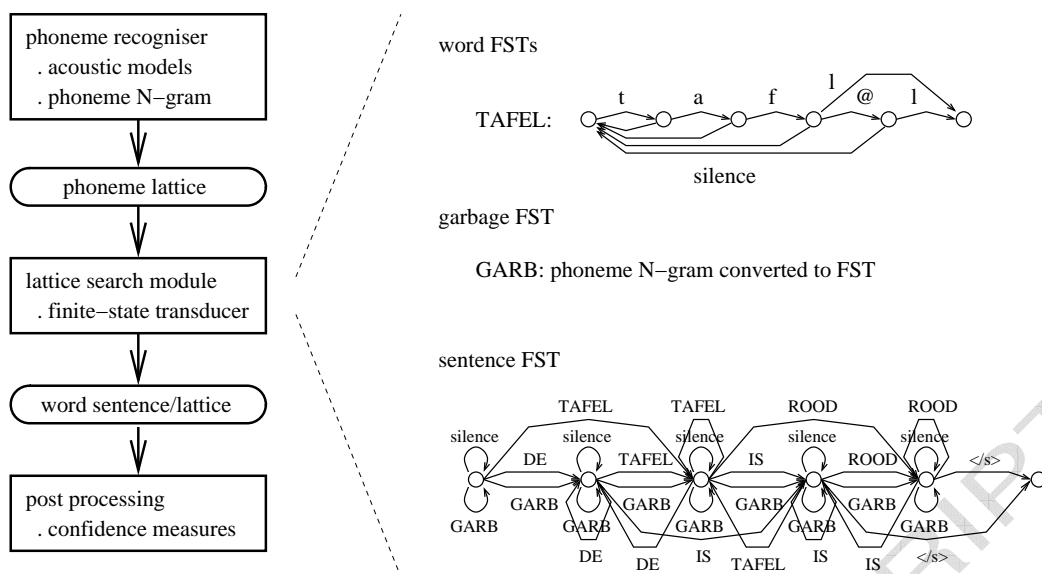


Fig. 2. Recognition system architecture with 2 layers (Duchateau et al., 2006): a phoneme recogniser and an FST-based word recogniser

215 state transducer (FST) to produce a detailed model for the speech to be
 216 expected from the child. The FST used is a composed FST: the word FSTs
 217 (top right in figure 2) and the garbage FST (middle right) are inserted at the
 218 right places in the sentence FST (bottom right). The Dutch sentence in the
 219 example is *De tafel is rood* [The table is red]. The search engine in the second
 220 layer turns the phoneme lattice into a word level recognition result. This result
 221 may be a lattice that can be used in further processing. In the experiments in
 222 this paper, the recognition results in the best path through the sentence FST,
 223 as this allows detection of reading errors. More details on this recognition sys-
 224 tem architecture and its advantages, and on the VTLN system in the acoustic
 225 modelling can be found in Duchateau et al. (2006).

226 Currently, we use this architecture with 2 layers for the detection of reading
 227 errors only. In order to provide low-latency feedback when tracking where a
 228 child is reading, a set-up with one layer was preferred to the system with two
 229 layers presented above. It is possible to implement feedback for tracking in the
 230 system with two layers, but the intermediate representation with phoneme arcs
 231 in the lattice will produce additional latency. The set-up with one layer is based
 232 on the same search engine and the same acoustic models as the first layer in the
 233 set-up with two layers. The difference is in the language model: the FST is used
 234 instead of the trigram phoneme sequence model. Every 150 msec, the search
 235 engine infers the word in the sentence with the highest probability (for being
 236 spoken) at that point in time from the different hypotheses in the search, and
 237 sends this information to the reading tutor. This feedback frequency (once
 238 every 150 msec) seemed to be fine in practice. It can be decreased, but a
 239 high frequency (e.g. once per frame) will slow down the recognition system
 240 noticeably. It should be noted that in the reading tutor, there is additional,

241 not negligible latency due to data transfer (audio, feedback) between modules,
242 screen generation, handling and plotting (the screens with the words to be
243 read), and even the refresh rate of the computer screen.

244 **3 Evaluation of the diagnostic and remedial tools**

245 In this section, we discuss three diagnostic and remedial tools: (1) the auto-
246 matic assessment of a child's reading level, (2) the use of synthesis for oral
247 feedback, and (3) the tracking of the child's reading for automated screen
248 advancement and for visual feedback.

249 *3.1 Automatic assessment of children's reading level*

250 Early identification of children with reading disabilities in primary school is
251 a major concern as their overall academic development depends heavily on
252 it. Automatic reading level assessment may help in this task as a form of
253 screening, so that speech therapists can have time to provide both a more
254 detailed assessment and adequate intervention for children with (according to
255 the automatic assessment) a low reading level.

256 *3.1.1 Experimental set-up*

257 In Duchateau et al. (2007), we proposed a baseline automatic assessment sys-
258 tem. This assessment is based on a reading test with 40 isolated words. The
259 score that expresses the reading performance of a child is defined as the total
260 time needed to read the 40 words divided by the number of correctly read
261 words. Since timing is known from the recording, we only need to know, for
262 every word, whether it was read correctly, to deduce a score for an individual.
263 Therefore, a human score can be determined from a manual annotation, while
264 an automatic score is based on a speech recognition system. Based on these
265 scores, each of the children in a particular grade is classified into one of 5
266 performance groups³. For more background on this scoring and classification
267 method, used in schools in Flanders, the reader is referred to Duchateau et al.
268 (2007).

³ In the Netherlands and Flanders, CITO (Central Institute for Test Development) introduced the use of 5 performance groups of unequal sizes: best performing 25%, above average performing 25%, below average performing 25%, far below average performing 15%, and worst performing 10%.

269 The experiments are based on the Chorec database (described in Cleuren
 270 et al., 2008), which contains reading sessions (of both isolated words and
 271 stories) from 400 children, including children with known reading disabilities.
 272 We selected the 3 real word reading tasks for the experiments: one task with
 273 40 1-syllable words, one with 40 2-syllable words, and one with 40 3- and 4-
 274 syllable words. Previous experiments (see Duchateau et al., 2007) showed that
 275 the use of tasks with (non-existent) pseudowords resulted in worse agreements
 276 for classification. Each real word reading task was read once by each of the
 277 children, however some recordings are lacking, for instance because the task
 278 was too difficult for that child. For grades 2, 3 and 4, the Chorec database
 279 contains about 75 recordings per real word reading task. For grade 1, this
 280 number is 55, 32 and 9 only for the three tasks respectively. Therefore, we will
 281 not report classification results for the 3- and 4-syllable word task for grade 1.

282 To automatically decide if a word was read correctly, we used the recognition
 283 system with 2 layers described in section 2.2. Since, by definition, a word is
 284 read correctly when the child reads it correctly at his/her final attempt, we
 285 can easily tag a word as read correctly or not by inspecting the end of the
 286 best path through the sentence FST (see figure 2), which is the result of the
 287 recogniser. If the best path ends in the word FST, the word is tagged as read
 288 correctly, if it ends in the garbage FST, the word is tagged as read incorrectly.

289 3.1.2 *Experimental results*

290 *Improvements over the baseline.*

291 In the baseline, a word recognised as being read correctly contributes *one* in
 292 the score calculation. But the recogniser is not always correct, so we improved
 293 on that by contributing only the probability that the word is really read cor-
 294 rectly given that it was recognised as correct. This task-dependent probability
 295 is estimated from the Chorec data. Similarly, for words recognised as read
 296 wrongly, we contribute the probability for a false alarm given that the word is
 297 recognised as wrong (rather than contributing *zero* as in the baseline). In the
 298 result tables, we call this system *general probabilities*.

299 Furthermore, we found that these probabilities depend on the width of the
 300 phoneme lattice that is generated by the first pass of the recogniser (more
 301 precisely, we use the average phoneme lattice width over the final attempt
 302 by the child for that word). This dependency is shown in figure 3 for the 1-
 303 syllable task read by the grade 3 children. This is not surprising as this infor-
 304 mation source is also used in typical confidence measures in large-vocabulary
 305 recognisers, either explicitly as in Duchateau et al. (2002) or implicitly e.g.
 306 by using posterior probabilities in lattices. In practice, we divided the range
 307 of the phoneme lattice width into 7 bins and estimated task-dependent (not
 308 word-dependent) probabilities for each bin. We call this system *lattice-width-*

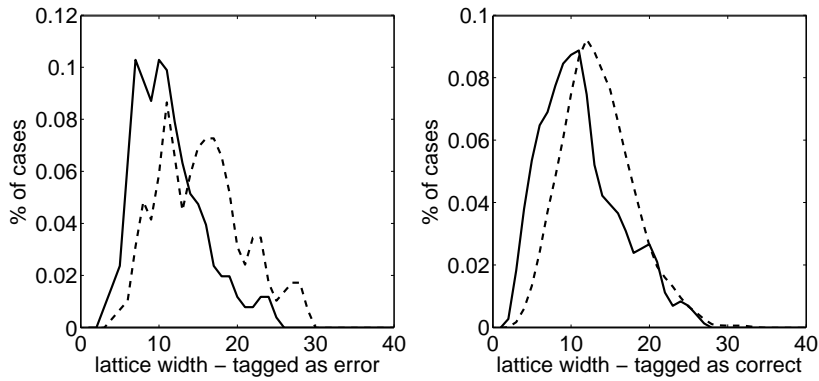


Fig. 3. Distribution of words annotated in Chorec as wrong (solid line) and correct (dashed line) as a function of the lattice width. The figure to the left is for words tagged as wrong by the recogniser, the one to the right is for words tagged as correct.

309 *dependent probabilities*. Note that in order to calculate the score of a particular
 310 child, all of the above probabilities are estimated on the data from the other
 311 children only, to avoid a bias.

312 The results of the reading level assessment experiments are given in table 1.
 313 For each grade, for the three tasks and for the three classifiers, the agreement
 314 (linearly weighted Cohen's Kappa⁴) with the reference human classification is
 315 given, and also the percentage of correctly classified children. We can see that
 316 the agreements are typically over 0.8 (known as *almost perfect*). Moreover,
 317 wrongly classified children can be put into a neighbouring performance group
 318 in all cases. Given the results in the table, we can conclude that there is a
 319 significant improvement from the baseline to the systems with probabilities.
 320 The systems with probabilities perform equally on average. The value of the
 321 added information in the system with lattice-width-dependent probabilities
 322 seems to be too small to improve the classification.

323 *Human-human vs. human-machine agreements*.

324 One third of the Chorec database is annotated a second and third time, by
 325 different annotators (the aim being the assessment of inter-annotator agree-
 326 ments). Annotators may disagree if a word has been read correctly because
 327 some decisions are disputable and because annotations are never error-free.
 328 Therefore, based on the additional annotations, other human classifications
 329 can be made. Comparing these with the (in fact arbitrarily chosen) reference
 330 human classification results in human-human agreements on the classification
 331 task.

332 In table 2, the average agreement (and the percentage correct) between the
 333 2 new human classifiers and the reference are compared with the agreements
 334 between the automatic classifiers and the reference. As classification of the

⁴ See e.g. <http://faculty.vassar.edu/lowry/kappaexp.html>

Grade 1	1-syllable	2-syllable	3+4-syll.
baseline	0.79/73%	0.72/63%	- / -
general probs.	0.81/76%	0.77/69%	- / -
lattice-width-dep.	0.84/80%	0.70/63%	- / -
Grade 2	1-syllable	2-syllable	3+4-syll.
baseline	0.82/74%	0.90/85%	0.85/77%
general probs.	0.83/76%	0.91/88%	0.89/84%
lattice-width-dep.	0.86/80%	0.94/91%	0.85/79%
Grade 3	1-syllable	2-syllable	3+4-syll.
baseline	0.85/79%	0.74/70%	0.81/75%
general probs.	0.93/92%	0.89/86%	0.89/86%
lattice-width-dep.	0.93/90%	0.85/82%	0.89/86%
Grade 4	1-syllable	2-syllable	3+4-syll.
baseline	0.85/80%	0.87/82%	0.80/72%
general probs.	0.88/84%	0.95/93%	0.89/84%
lattice-width-dep.	0.86/81%	0.93/90%	0.92/88%

Table 1

Evaluation of the proposed automatic classifiers for the different grades and reading tasks. Agreement (Kappa value) and percentage of correctly classified children given.

	1-syllable	2-syllable	3+4-syll.
human (average)	0.94/92%	0.92/88%	0.93/90%
baseline	0.91/88%	0.90/88%	0.88/84%
general probs.	0.94/91%	0.90/85%	0.90/86%
lattice-width-dep.	0.95/92%	0.92/88%	0.86/80%

Table 2

Comparison of human-human and human-machine agreements. Kappa value and percentage of correctly classified children given.

335 children per grade based on only one third of the recordings would lead to
 336 results with a high variance, we decided to classify all children from grades 1
 337 to 4 together. From the table we can conclude that the human-machine agree-
 338 ments obtained equal the human-human agreements. Only in the case of the
 339 3- and 4-syllable word task, the human-machine agreement is slightly, but
 340 insignificantly, worse.

341 3.2 *Synthesis for oral feedback*

342 One of the reading tutor tasks is inspired by the strategy of a human reading
343 tutor (a teacher or a clinician) to provide feedback and assistance while a
344 child is reading aloud and struggles or makes a mistake. Feedback has been
345 shown to be of great importance for the development of reading for both
346 children with and without reading difficulties (e.g. McCoy and Pany, 1986;
347 Pany and McCoy, 1988; Perkins, 1988; Spaai et al., 1991; Wise and Olson,
348 1998). Since a beginning or struggling reader who makes a reading mistake
349 has great difficulties correcting himself or herself, feedback should be provided
350 by someone else. Therefore, the role of an intervening human tutor has always
351 been indispensable for children with reading difficulties. As MacArthur et al.
352 (2001) discussed, a computerised reading tutor could also be well suited to
353 provide feedback and decoding assistance to a struggling reader by means of
354 digitised (recorded) or synthesised speech, and the ability to simultaneously
355 highlight words or word parts.

356 In the SPACE project, synthesised speech is used to provide three different
357 kinds of oral corrective feedback of which the effectiveness has been proven in
358 the literature: whole-word-feedback, syllable-by-syllable feedback, and phoneme-
359 by-phoneme feedback (e.g. McCoy and Pany, 1986; Perkins, 1988; Spaai et al.,
360 1991; Wise and Olson, 1998). With respect to the first type of feedback, the
361 computerised reading tutor supplies the child with the whole word whenever
362 an error is made against that word (e.g. *mother*). The second and third types
363 of feedback concern the provision of segmentation cues by segmenting the er-
364 roneously read word into its constituent syllables (e.g. *mo-ther*) or phonemes
365 (e.g. *m-o-th-e-r*). Immediately following the feedback presented by the read-
366 ing tutor, the child is asked to read or reread the word. If the word is again
367 not read without any reading error, feedback is given again until the child
368 succeeds in decoding the word correctly (with a maximum of three feedback
369 prompts for any particular word). During feedback, the word of interest is
370 highlighted in the text. After the feedback has been given, the next word to
371 read is underlined in order to indicate to the child where to continue his/her
372 reading.

373 Additionally, the SPACE project has made an effort to build a synthesiser that
374 is able to emphasise particular phonemes in phoneme-by-phoneme feedback.
375 Inspiration for this feature comes from reading therapy experience that shows
376 a tendency to emphasise the wrongly read phoneme in a word. For example, if
377 a child has to read the word *wood* but reads the word *mood* instead, the human
378 tutor tends to overemphasise the phoneme /w/ in the segmental (phoneme-
379 by-phoneme) feedback presented to the child.

380 To evaluate the reading tutor's speech synthesis abilities, an intervention study

381 was carried out in a group of ten Flemish elementary school children with
 382 reading difficulties. During a period of four weeks, each child received twenty
 383 minutes of computerised reading intervention on each school day. In each read-
 384 ing session, the child read instructional-level story material that was presented
 385 paragraph by paragraph on a touch screen connected to a laptop computer.
 386 Whenever the child experienced difficulties in decoding a particular word, help
 387 could be asked for by touching that word on the screen. If the child did not
 388 ask for help but did not succeed in reading the word correctly, feedback was
 389 automatically given. Although no automatic speech recognition was used yet
 390 in the presented study, the children were told that the computer could really
 391 listen to them reading aloud. In reality, a human listener, *hidden* behind the
 392 large touch screen, controlled the feedback supply in case of reading errors.

393 From this study, based on the different opinions of the three examiners in-
 394 volved, and based on the lack of reaction with respect to this matter from the
 395 children that exercised with the reading tutor, we conclude that the speech
 396 synthesis quality was good for all three of the feedback types: feedback sounded
 397 natural, and was clearly audible and well understandable. Children got quickly
 398 used to the reading tutor's voice and were very understanding with respect
 399 to the reading tutor's occasional mistakes (e.g. giving feedback although the
 400 word was decoded correctly) that were introduced by the human listener to
 401 emulate a computer's behaviour.

402 However, with respect to the phoneme-by-phoneme feedback mode, some con-
 403 siderations for future improvements should be taken into account, for in-
 404 stance concerning the progressive and regressive assimilation processes in
 405 Dutch. These processes imply that a phoneme takes over the characteristics of
 406 the preceding (progressive assimilation) or following (regressive assimilation)
 407 phoneme in a word or a sentence. To obtain correctly pronounced whole-word
 408 and syllable-by-syllable feedback, it is important that these processes are ac-
 409 counted for by the speech synthesiser. But when synthesising phoneme-by-
 410 phoneme feedback, these processes should be neglected and phonemes should
 411 be produced as if they were standing alone. For example: in the Dutch word
 412 *zakdoek* [handkerchief], the phoneme /k/ before the /d/ becomes voiced (/zAg-
 413 duk/). Nevertheless, when synthesising phoneme-by-phoneme feedback for this
 414 word, that phoneme needs to remain voiceless (/z/-/A/-/k/-/d/-/u/-/k/).

415 The situation is different with respect to the more specific phenomenon of *final*
 416 *devoicing*, where a voiced phoneme in a word becomes voiceless at the end of
 417 the word when followed by a pause. When synthesising phoneme-by-phoneme
 418 feedback for such a word, the phoneme at the end of the word needs to remain
 419 voiceless. For example, in the Dutch word *hond* [dog], the voiced phoneme
 420 /d/ becomes voiceless because it is placed at the end of the word: *hond* is
 421 pronounced as /hOnt/, not as /hOnd/. The phoneme-by-phoneme feedback
 422 should be /h/-/O/-/n/-/t/, keeping the final devoiced consonant, not altering

423 it in the way described in the previous paragraph.

424 Another valuable conclusion inferred from the intervention study, is that em-
425 phasis on the wrongly decoded phoneme does not seem to be feasible in every
426 word under any circumstance. Although a more detailed observational study
427 is recommended to detect the optimal conditions for specific phoneme empha-
428 sis, some general conclusions can already be made. The first conclusion is that
429 specific phoneme emphasis does not feel natural if the child makes multiple
430 errors within one word. Only for words in which the child makes an error on
431 only one phoneme, a human tutor would consider to emphasise that phoneme
432 when giving feedback to the child. Another observation is that emphasising a
433 particular phoneme seems to be most effective for simple, monosyllabic words
434 (e.g. when the child reads *wood* for *mood*, or *mad* for *map*). For example, if the
435 child makes an error on the word *handkerchief*, it is not very helpful for the
436 child to emphasise the one phoneme that has not been decoded correctly. How-
437 ever, for such long words, specific phoneme feedback seems to sound natural
438 again when the error falls on the first phoneme of the word.

439 3.3 *Tracking where the child is reading*

440 In an automated reading tutor, it is also important to track where the child
441 is reading so that the progress through the reading task is known. As ex-
442 plained in section 2.2, the recognition system therefore provides the reading
443 tutor with the current position (the most probable position according to the
444 hypotheses in the search engine) in the FST that models the sentence being
445 read. This tracker can be used for diagnostic and for remedial purposes. On
446 the one hand, the tracker will be used for advancing automatically from one
447 screen to the next in reading tasks that are presented on consecutive screens.
448 On the other hand, the tracker can be used for generating feedback to the
449 child, for instance, by highlighting the word he/she should read. This section
450 presents our practical tests on children with known reading disabilities using
451 the reading tutor supported by automatic tracking, and the improvements to
452 our system based on the tests.

453 3.3.1 *Automated screen advancement*

454 To evaluate the automated screen advancement, 8 reading disabled children
455 were asked to read a word reading exercise (7 exercises in total) or a story
456 (2 in total). The selected reading tasks were *difficult* compared to the child's
457 reading level. The word reading tasks consist of 40 single words presented one
458 by one on the screen, the stories are presented on 4 or 5 consecutive screens.
459 To prevent the children from getting confused when the tracker was making

460 a mistake, they were told that even a computer can make mistakes, and that
461 they did not need to pay attention to the computer's possible mishearings.
462 In total, the examiner had to intervene twice by reading the word herself,
463 in 2 word reading exercises read by 2 different children. In 3 out of the 7
464 word reading exercises, the child needed to repeat one word (out of the 40
465 words) to get the tracker to advance to the next word. In 2 other word reading
466 exercises, the child needed to repeat 2 respectively 3 words. For 9 words (from
467 6 different word reading exercises), the system advanced too fast so that the
468 child did not get enough time to try to read that particular word. Automatic
469 screen advancement in the two stories went perfectly. Thus, from this first
470 qualitative evaluation, it became clear that some robustness had to be added
471 to the system. As a result, four timing parameters were introduced to define
472 the behaviour of the automatic screen advancement during a reading exercise:

- 473 ● Minimal time of screen appearance: avoids that the system advances before
474 anyone can read the screen.
- 475 ● Maximal time of screen appearance: forces the system to advance, so that
476 the system will not get stuck just because the tracker does.
- 477 ● Amount of time the tracker feedback indicates the child is ready before the
478 tutor really advances to the next screen, typically this is about 1 second.
479 Basically, this means that the tutor advances when the child is silent for
480 this amount of time: in this case, the child is supposed not to have another
481 attempt.
- 482 ● Amount of time the system is waiting between 2 screens, typically half a
483 second. This avoids that the child, out of surprise that the tutor advances
484 to the next screen, still says something that belongs to the previous screen
485 but that will erroneously be interpreted by the tracker as part of the next.

486 A second qualitative evaluation of this automatic screen advancement option
487 was carried out by getting another 22 reading disabled children read one (and
488 in 2 cases: 2 respectively 3) of the relatively difficult word reading exercises (16
489 in total) or stories (9 in total). Again, they were asked not to pay attention to
490 the computer's possible mishearings. This time, for the word reading exercises,
491 the examiner did not have to intervene once, and no repetition of words was
492 needed in order to get the tracker proceed to the next word. The tracker moved
493 too fast for only one word in one word reading exercise. Automatic screen
494 advancement for all 9 stories went flawlessly. In summary, no real difficulties
495 were encountered. Thus, by using proper values for the timing parameters
496 mentioned above, the overall system for advancing to the next screen seems to
497 work satisfactorily for diagnostic purposes when used in daily school practice.

498 3.3.2 *Highlighting the word to be read*

499 The feedback provided by the tracker can also be used to visually highlight a
 500 word (or words) on the screen. For 2 stories (read by 2 different children), the
 501 reading tutor highlighted the word the tracker feedback indicates as the word
 502 the child is reading. However, this form of feedback appeared to be useless
 503 for reading intervention because the child knows what he/she tries to read at
 504 that moment, so the feedback is not needed (in case the tracker is correct)
 505 or confusing (in case the tracker is wrong). Therefore, the word which the
 506 child *should* be reading, is highlighted. This is implemented by highlighting
 507 the word indicated by the tracker's feedback, except in two cases:

- 508 ● On the one hand, if the tracker's feedback points to a word in an earlier
 509 part of the sentence (because the child repeats some words, or because the
 510 tracker makes an error), the highlighted word does not follow, it can only
 511 stick to the same word or advance. Besides maintaining the meaning of the
 512 highlighted word (namely that this is the word that *should* be read), this
 513 also avoids the situation when the possibly jumpy behaviour of the raw
 514 tracker feedback is shown directly on the screen.
- 515 ● On the other hand, robustness is added to the system by allowing the teacher
 516 to set a maximum reading speed (in characters per second). If the tracker
 517 erroneously skips several words, the highlighted word will follow only slowly,
 518 so that the tracker is able to correct its error.

519 When providing the altered highlighting feedback to 8 other reading disabled
 520 children (9 stories in total), another problem became immediately clear: when
 521 a child is silent between two words (while decoding the next word), the tracker
 522 indicated the previous word. Because in this case, the next word should be
 523 highlighted, the tracker's behaviour needed to be changed. This idea was sup-
 524 ported by a third evaluation, in which another 10 reading disabled children
 525 were asked to read relatively difficult stories supported by highlighting feed-
 526 back. However, overall, the current feedback system tends to proceed to the
 527 next word quite quickly, even if the previous word has not been pronounced
 528 correctly or completely. This behaviour is satisfactory for slow, both inte-
 529 rior/silent and exterior/aloud decoding readers⁵ who need to learn to read
 530 faster. For fast and erroneous readers, however, highlighting seems to be of
 531 little use as a remedial tool. It may be more useful in this case to indicate
 532 words read incorrectly *after* reading: these readers need to be made aware of
 533 their errors and need to be forced to read at a slower pace.

⁵ At the word recognition level, the slow/accurate-fast/inaccurate dichotomy has been associated with the indirect versus direct word approach (Coltheart, 1978).

534 4 Conclusions and future work

535 In this paper, we presented several diagnostic and remedial reading tutor tools
536 that are based on speech technology: tools for synthesis in phoneme or syl-
537 lable mode, for tracking where the child is reading, and for assessment of
538 the child's reading level. In general, these tools seem to work satisfactorily
539 when evaluated on children, including children with reading disabilities. As
540 for the assessment, which was based on a reading task with only 40 words, we
541 found that a child can be assessed in 5 performance groups equally well by a
542 computer than by a human. The outcome of the extensive evaluation of the
543 remedial tools was dual. On the one hand, as described in detail in this paper,
544 the evaluation resulted in several improvements of the tutor and the speech
545 technology involved, as to make the overall system more robust. On the other
546 hand, the evaluation proved that the presented tools work in practice, in a
547 real school environment and for the target users, children.

548 Future work on the reading tutor includes the development of an automatic
549 quality measure for the synthesised speech, which may, for instance, help the
550 teacher to prepare the exercises in the tutor. Also, the current automatic de-
551 tection of reading errors needs to be improved as it is insufficient for adequate
552 word-by-word feedback to the child.

553 Furthermore, one of the planned features of the SPACE reading tutor is syn-
554 chronised reading, the ability to read along with the child. The reading tutor
555 plays the role of a speech therapist, which sets the reading pace. While read-
556 ing, the reading speed is maintained, unless the feedback from the recogniser
557 indicates that the child has difficulties to follow the tempo. Then the reading
558 tutor slows down, or falls back to the syllable mode for synthesis.

559 5 Acknowledgements

560 The research in this paper was supported by IWT project SPACE (sbo/040102):
561 SPEech Algorithms for Clinical and Educational applications, home page:
562 <http://www.esat.kuleuven.be/psi/spraak/projects/SPACE>. We would like to
563 thank the participating schools, teachers and pupils for their collaboration in
564 this project.

565 References

566 Abdou, S. M., Hamid, S. E., Rashwan, M., Samir, A., Abd-Elhamid, O.,
567 Shahin, M., Nazih, W., Sep. 2006. Computer aided pronunciation learning

- 568 system using speech recognition techniques. In: Proceedings of Interspeech
569 2006 - ICSLP, 9th International Conference on Spoken Language Process-
570 ing. Pittsburgh, PA, U.S.A., pp. 849–852.
- 571 Adams, M. J., 2006. The promise of automatic speech recognition for fostering
572 literacy growth in children and adults. In: McKenna, M. C., Labbo, L. D.,
573 Kieffer, R. D., Reinking, D. (Eds.), *International Handbook of Literacy and*
574 *Technology*. Vol. 2. Lawrence Erlbaum Associates, Mahwah, NJ, U.S.A. and
575 London, U.K., pp. 109–128.
- 576 Banerjee, S., Beck, J. E., Mostow, J., Sep. 2003. Evaluating the effect of
577 predicting oral reading miscues. In: Proceedings of Interspeech 2003 - Eu-
578 rospeech, 8th European Conference on Speech Communication and Tech-
579 nology. Geneva, Switzerland, pp. 3165–3168.
- 580 Black, A. W., Lenzo, K. A., Aug. 2001. Optimal data selection for unit se-
581 lection synthesis. In: Proceedings of the 4th ISCA Workshop on Speech
582 Synthesis. Perthshire, Scotland, pp. 63–67.
- 583 Black, M., Tepperman, J., Lee, S., Price, P., Narayanan, S., Aug. 2007. Auto-
584 matic detection and classification of disfluent reading miscues in young chil-
585 dren’s speech for the purpose of assessment. In: Proceedings of Interspeech
586 2007 - Eurospeech, 10th European Conference on Speech Communication
587 and Technology. Antwerp, Belgium, pp. 206–209.
- 588 Chu, M., Peng, H., Zhao, Y., Niu, Z., Chang, E., Apr. 2003. Microsoft Mulan
589 - a bilingual TTS system. In: Proceedings of the IEEE International Con-
590 ference on Acoustics, Speech, and Signal Processing. Vol. I. Hong Kong,
591 China, pp. 264–267.
- 592 Clark, R., Richmond, K., King, S., Apr. 2007. Multisyn: Open-domain unit
593 selection for the Festival speech synthesis system. *Speech Communication*
594 49 (4), 317–330.
- 595 Cleuren, L., Duchateau, J., Ghesquière, P., Van hamme, H., May 2008. Chil-
596 dren’s oral reading corpus (CHOREC): Description and assessment of an-
597 notator agreement. In: Proceedings of the 6th International Conference on
598 Language Resources and Evaluation. Marrakech, Morocco.
- 599 Coltheart, M., 1978. Lexical access in simple reading tasks. In: Underwood,
600 G. (Ed.), *Strategies in Information Processing*. Academic Press, New York,
601 NY, U.S.A., pp. 151–216.
- 602 Daelemans, W., Van den Bosch, A., 2005. *Memory-Based Language Process-*
603 *ing*. Cambridge University Press, Cambridge, U.K.
- 604 Demuyne, K., Laureys, T., Wambacq, P., Van Compernelle, D., May 2004.
605 Automatic phonemic labeling and segmentation of spoken Dutch. In: Pro-
606 ceedings of the 4th International Conference on Language Resources and
607 Evaluation. Vol. I. Lisbon, Portugal, pp. 61–64.
- 608 Demuyne, K., Van Compernelle, D., Van hamme, H., Sep. 2006. Robust
609 phone lattice decoding. In: Proceedings of Interspeech 2006 - ICSLP, 9th
610 International Conference on Spoken Language Processing. Pittsburgh, PA,
611 U.S.A., pp. 1622–1625.
- 612 D’Mello, S. K., King, B., Stolarski, M., Chipman, P., Graesser, A., Oct. 2007.

- 613 The effects of speech recognition errors on learner's contributions, knowl-
614 edge, emotions, and interaction experience. In: Proceedings of the SLaTE
615 Workshop on Speech and Language Technology in Education. Farmington,
616 PA, U.S.A., pp. 49–52.
- 617 Duchateau, J., Cleuren, L., Van hamme, H., Ghesquière, P., Aug. 2007. Auto-
618 matic assessment of children's reading level. In: Proceedings of Interspeech
619 2007 - Eurospeech, 10th European Conference on Speech Communication
620 and Technology. Antwerp, Belgium, pp. 1210–1213.
- 621 Duchateau, J., Demuynck, K., Wambacq, P., May 2002. Confidence scoring
622 based on backward language models. In: Proceedings of the IEEE Inter-
623 national Conference on Acoustics, Speech, and Signal Processing. Vol. I.
624 Orlando, FL, U.S.A., pp. 221–224.
- 625 Duchateau, J., Wigham, M., Demuynck, K., Van hamme, H., May 2006. A
626 flexible recogniser architecture in a reading tutor for children. In: Proceed-
627 ings of the ITRW on Speech Recognition and Intrinsic Variation. Toulouse,
628 France, pp. 59–64.
- 629 Eskenazi, M., Pelton, G., Sep. 2002. Pinpointing pronunciation errors in chil-
630 dren's speech: Examining the role of the speech recognizer. In: Proceedings
631 of the ITRW on Pronunciation Modeling and Lexicon Adaptation for Spo-
632 ken Language Technology. Estes Park, CO, U.S.A., pp. 48–52.
- 633 Hagen, A., Pellom, B., Cole, R., Dec. 2007. Highly accurate children's speech
634 recognition for interactive reading tutors using subword units. *Speech Com-
635 munication* 49 (12), 861–873.
- 636 Heiner, C., Beck, J., Mostow, J., Jun. 2004. Improving the help selection policy
637 in a reading tutor that listens. In: Proceedings of the InSTIL/ICALL Sym-
638 posium on NLP and Speech Technologies in Advanced Language Learning
639 Systems. Venice, Italy, pp. 195–198.
- 640 Hunt, A., Black, A., May 1996. Unit selection in a concatenative speech syn-
641 thesis system using a large speech database. In: Proceedings of the IEEE
642 International Conference on Acoustics, Speech, and Signal Processing. At-
643 lanta, GA, U.S.A., pp. 373–376.
- 644 Kerkhoff, J., Marsi, E., Nov. 2002. Nextens: a new open source text-to-speech
645 system for Dutch. In: Proceedings of the 13th meeting of Computational
646 Linguistics in the Netherlands. Groningen, The Netherlands.
- 647 Latacz, L., Kong, Y. O., Mattheyses, W., Verhelst, W., Nov. 2006. Novel
648 text-to-speech reading modes for educational applications. In: Proceedings
649 of the ProRISC/IEEE Benelux Workshop on Circuits, Systems and Signal
650 Processing. Veldhoven, The Netherlands.
- 651 Latacz, L., Kong, Y. O., Verhelst, W., Aug. 2007. Unit selection synthesis using
652 long non-uniform units and phonemic identity matching. In: Proceedings of
653 the 6th ISCA Workshop on Speech Synthesis. Bonn, Germany, pp. 270–275.
- 654 MacArthur, C., Ferretti, R., Okolo, C., Cavalier, A., 2001. Technology applica-
655 tions for students with literacy problems: A critical review. *The Elementary
656 School Journal* 101 (3), 273–301.
- 657 McCoy, K., Pany, D., 1986. Summary and analysis of oral reading corrective

- 658 feedback research. *The Reading Teacher* 39 (6), 548–554.
- 659 Neri, A., Cucchiarini, C., Strik, H., Sep. 2006. ASR-based corrective feedback
660 on pronunciation: Does it really work? In: *Proceedings of Interspeech 2006 -*
661 *ICSLP, 9th International Conference on Spoken Language Processing*. Pitts-
662 burgh, PA, U.S.A., pp. 1982–1985.
- 663 Pany, D., McCoy, K., 1988. Effects of corrective feedback on word accuracy
664 and reading comprehension of readers with learning disabilities. *Journal of*
665 *Learning Disabilities* 21 (9), 546–550.
- 666 Perkins, V., 1988. Feedback effects on oral reading errors of children with
667 learning disabilities. *Journal of Learning Disabilities* 21 (4), 244–248.
- 668 Probst, K., Ke, Y., Eskenazi, M., Jul. 2002. Enhancing foreign language tutors
669 - in search of the golden speaker. *Speech Communication* 37 (3-4), 161–173.
- 670 Russell, M., Series, R. W., Wallace, J. L., Brown, C., Skilling, A., Apr. 2000.
671 The STAR system: an interactive pronunciation tutor for young children.
672 *Computer Speech and Language* 14 (2), 161–175.
- 673 Spaai, G., Ellermann, H., Reitsma, P., 1991. Effects of segmented and whole-
674 word sound feedback on learning to read single words. *Journal of Educa-*
675 *tional Research* 84 (4), 204–213.
- 676 Wise, B., Olson, R., 1998. Studies of computer-aided remediation for reading
677 disabilities. In: Hulme, C., Joshi, R. (Eds.), *Reading and Spelling: Devel-*
678 *opment and Disorders*. Lawrence Erlbaum Associates, Mahwah, NJ, U.S.A.
679 and London, U.K., pp. 473–487.