Design and Analysis of a Database to Evaluate Children's Reading Aloud Performance

Jorge Proença[1,2], Dirce Celorico[1], Carla Lopes[1,3], Miguel Sales Dias[4,5], Michael Tjalve[6], Andreas Stolcke[7], Sara Candeias[4], and Fernando Perdigão[1,2]

[1] Instituto de Telecomunicações, Coimbra, Portugal
{jproenca,direcelorico,calopes,fp}@co.it.pt
[2] Department of Electrical and Computer Engineering, University of Coimbra, Portugal
[3] Polytechninc Institute of Leiria, Leiria, Portugal
[4] Microsoft Language Development Centre, Lisbon, Portugal
{miguel.dias,t-sacand}@microsoft.com
[5] ISCTE – University Institute of Lisbon, Lisbon, Portugal
[6] Microsoft & University of Washington, Seattle, WA, USA
michael.tjalve@microsoft.com
[7] Microsoft Research, Mountain View, CA, USA
andreas.stolcke@microsoft.com

Abstract. To evaluate the reading performance of children, human assessment is usually involved, where a teacher or tutor has to take time to individually estimate the performance in terms of fluency (speed, accuracy and expression). Automatic estimation of reading ability can be an important alternative or complement to the usual methods, and can improve other applications such as e-learning. Techniques must be developed to analyse audio recordings of read utterances by children and detect the deviations from the intended correct reading i.e. disfluencies. For that goal, a database of 284 European Portuguese children from 6 to 10 years old (1st-4th grades) reading aloud amounting to 20 hours was collected in private and public Portuguese schools. This paper describes the design of the reading tasks as well as the data collection procedure. The presence of different types of disfluencies is analysed as well as reading performance compared to known curricular goals.

Keywords: Reading Aloud Performance, Child Speech, Speech Corpus, Reading Disfluencies.

1 Introduction

The use of automatic speech recognition technologies to analyse reading performance gains prominence as an alternative to any kind of manual or 1-on-1 evaluation. Usually, teachers have to spend a considerable amount of time on the task of manually assessing a child's reading ability. Automatic evaluation of literacy or reading ability (not necessarily of children) is always related to detecting correctly read words, or optionally detecting what kind of mistakes are made. Additionally, there are several systems oriented to improve the literacy of an individual [1,2], ideally denoting and warning about reading errors that occur. Computer-Assisted Language Learning (CALL) is the area of research that focuses on this subject, allowing a self-practice or an oriented training of the language. These systems are most often created for foreign language learning [1], [3], and are therefore targeted at adults or young adults for whom speech technologies are significantly mature. Nevertheless, for children, there are also applications that deal with the improvement of reading aloud performance, such as reading tutors. Some projects aim at creating an automatic reading tutor that follows and analyses a child's reading, such as LISTEN [4], Tball [5], SPACE [6] and FLORA [7]. Most of these applications are helpers, e.g., by highlighting words in a sentence as they are correctly pronounced. The present work is carried out in the scope of the LetsRead project whose overall goal is to have an application that can automatically evaluate the reading aloud performance of European Portuguese (EP) children from 6 to 10 years old (1st-4th grades), and not necessarily provide feedback to them, but to their teachers and tutors.

There are currently no computer assisted applications for EP that automatically evaluate the reading aloud performance of children. Even for other languages, this automatic evaluation is a developing field, and the focus is on reading of isolated words rather than longer sentences [8,9]. To carry out the goals of the LetsRead project, it was necessary to create a large new corpus of EP children's speech with utterances of reading tasks that are rich with common disfluencies that children commit while reading. There are some children's speech databases for EP, such as SPEECON with rich sentences [10]; ChildCAST [11] with picture naming; the Contents for Next Generation (CNG) Corpus targeting interactive games [12] and [13] with child-adult interactions. However, these databases do not exhibit the required samples of disfluent reading speech. Since children's speech has different characteristics from adult speech (such as fundamental frequency, formant frequency variability, vowel duration variability, etc. [14,15]), special care is needed to adapt or create robust acoustic models that target children [16,17].

This paper describes the careful design of the reading tasks as well as the collection procedure of the LetsRead database. Part of the data was verified and annotated manually including tagging of several types of disfluencies. An analysis of reading performance with a comparison to adult speakers is also presented.

2 Reading Tasks

The Portuguese government has defined certain Curricular Goals (CG) with qualitative and quantitative objectives per grade for reading aloud [18]. Some of these objectives include target reading speed of words per minute on different tasks. With the analysis of curricular goals in mind, the reading of sentences and pseudowords was the target material to be collected (described in the following subsections). It was decided not to include reading of isolated words, as the required time for a session could become too long and the child's performance is likely to decrease with extended sessions. The pseudoword reading task provides an objective analysis of reading skills. With sentences, plenty of reading disfluencies can be collected from which the overall reading performance of a child can be evaluated. Each child was presented with a reading task where they were asked to read aloud twenty sentences and ten pseudowords. Forty reading tasks were established (10 per grade) to balance repetition and diversity of the data. At a later stage, these were shortened to 5 tasks per grade, to reinforce repetition. The vocabulary of the set of sentences and pseudowords comprises a total of 2721 words. The distribution of the material for the different grades is described below.

2.1 Sentences: selection and difficulty level

A large set of sentences was extracted from children's tales and school books of the level of the target group (6-10 years old, 1st-4th grades). They are mostly short sentences, although the maximum length is 30 words. Twenty sentences were included in each reading task (for a recording session of one child). The first concern for distributing sentences along the grades was to maintain a good representation of all phones, so that acoustic models with significant quality can be built with the data. The other main concerns to build appropriate reading tasks were to maintain the same average difficulty within a grade (with a rising average difficulty from 1st to 4th grades) and to have sentences of varying difficulty in a task (overlapping distributions of difficulty for the grades). Furthermore, it is necessary to capture all types of reading disfluencies to have examples for training, so the difficulty cannot be too low, although a balance must exist so as not to make the tasks too hard.

A parameter of difficulty was developed to classify sentences according to phonological constraints. Although it would be ideal to also relate a word's difficulty to its age-of-acquisition or familiarity, not all words of the proposed reading tasks were present in available lexical databases such as ESCOLEX [19], and it was not possible to consider such features. The proposed parameter of difficulty is based on the method described in [20] where sentences are evaluated in terms of phonological complexity and variety. All words were split into syllables and a difficulty level was assigned to each syllable, determined from these rules: the length of the syllable; the multiple pronunciation of some graphemes (e.g. <mãe> [mˈɐ̃j̃] and <bem> [bˈɐ̃j̃]); the ambiguous pronunciation of consonant clusters (e.g. <prever> [pɾəvˈeɾ] or <florescer> [fluɾəʃsˈeɾ]) and vocalic encounters (<candeeiro> [kɐ̃diˈɐjɾu] or <veem> [vˈeɐ̃j̃]). Since each syllable has a given minimum difficulty, the length of the sentence also contributes to difficulty.

2.2 Pseudowords creation

Pseudowords (such as <traba> [tɾˈabɐ], <impemba> [ĩpˈẽbɐ] or <culenes> [kulˈɛnəʃ]) represent non-existing or nonsense words which can be used to evaluate morphological and phonemic awareness. A novel method for the creation of pseudowords was developed. Existing tools such as Wuggy [21] take as input existing words and output pseudowords that differ in one or two syllables to the original words. This creates pronounceable words that are similar to existing words (such as <sapado> from

<sapato>). The proposed method creates pseudowords without the starting point of valid words while maintaining full pronounceability. It should create unfamiliar words and the difficulty of reading them should be slightly higher than familiar or existing words. The aim was to create pseudowords of two, three and four syllables. First, the most frequent syllables in each position for words with those number of syllables were extracted from a large lexicon of European Portuguese, CETEMPúblico [22]. Then, words of two or more syllables are created randomly from a set of the most frequent syllables. Words that have syllabic combinations that do not respect pronounceability rules are deleted as are words that exist in the lexicon. The difficulty score for a pseudoword is calculated by the same method described above for sentences. The distribution of the pseudowords along the reading tasks is also similar to sentences, promoting a varying difficulty and a rising average difficulty along the grades.

3 Data Collection

The corpus of children reading aloud was collected at 2 private and 9 public schools in urban centres and periphery areas of the central Coimbra region with children that attend primary school, aged 6 to 10 years old. A specific application was developed in which the sentences are displayed in a large font size on a computer screen simultaneously with the start of recording. This means that there is no practice time to influence performance. A screenshot of the application can be seen in Fig. 1 as well as an example of the recording environment. The recordings were performed in school classrooms chosen for their low reverberation and noise. The children were asked to read aloud a set of 20 sentences and 10 individual pseudowords. A lapel Lavalier microphone (Shure WL93) was used as the main recording device, accompanied by a standard table top PC microphone as backup (Plantronics Audio 10). The background noise could not always be completely controlled but was mostly low, also because the main recording microphone did a good job at filtering out background noise.

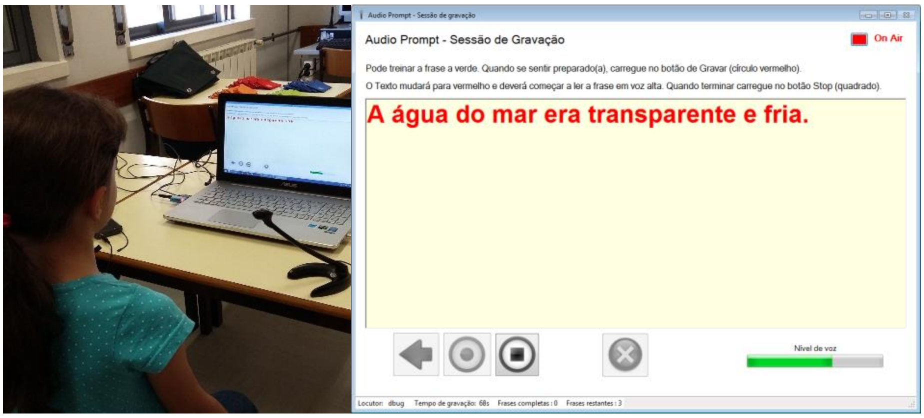

Fig. 1. Example of the recording environment (left) and software (right).

4 Corpus and Disfluencies

The collected database consists of around 20 hours of recorded speech from 284 children, 147 female and 137 male, distributed from the 1st to the 4th grade with 68, 88, 76 and 52 children, respectively. A set of 104 children's speech utterances of pseudowords and sentences has been fully and manually annotated and these children (46 male and 58 female) are equally distributed among the 4 grades (26 per grade). Speech from an additional 100 children was annotated only for the pseudoword reading task. The annotated data amounts to approximately 6 hours and 15 minutes of speech. The annotated speech exhibits a large variety of disfluencies that represent the most common types of errors in reading aloud by children. Based on [23], the rules for the annotation and labelling procedure were defined and several types of disfluency were identified:

- PRE – False starts that are followed by the attempted correction (pre-corrections, multiple can occur). Example: for prompt "*grande espanto*" [gɾɐ̃də iʃpɐ̃tu], utterance is "*grande **espa** espanto*" [gɾɐ̃də ˈ**iʃpɐ** iʃpˈɐ̃tu].
- SUB – Substitution or severe mispronunciation of a word. Example: for prompt "*voava em largos círculos*" [vuˈavɐ ɐ̃j̃ lˈaɾguʃ sˈiɾkuluʃ], utterance is "*voava em **lares sicos***" [vuˈavɐ ɐ̃j̃ **lˈaɾəʃ sˈikuʃ**].
- PHO – Small mispronunciation of a word, usually with a change in one phone or a phone extension (EXT, marked with the symbol [:]). Example: for prompt "*A Lena chegou a casa, da escola*" [ɐ lˈenɐ ʃəgˈo ɐ kˈazɐ dɐ iʃkˈɔlɐ], utterance is "*A Lena **chegou** a casa, da **escola***" [ɐ lˈenɐ **ʃə:gˈo** ɐ kˈazɐ dɐ **ɛʃkˈɔlɐ**].
- REP – Repetition of a word (multiple repetitions may occur). Example: for prompt "*Ele já me deu*" [ˈelə ʒˈa mə dˈew], utterance is "*Ele, **ele** já me deu*" [ˈelə **elə** ʒˈa mə dˈew].
- INS – An inserted word that is not part of the original sentence. Example: for prompt "*mas também dizem*" [mɐʃ tɐ̃bɐ̃j̃ dˈizɐ̃j̃], utterance is "*mas também **me** dizem*" [mɐʃ tɐ̃bɐ̃j̃ **mə** dˈizɐ̃j̃].
- DEL – The word was not pronounced (deletion). Example: for prompt "*onde morava **uma** velha*" [ˈõdə muɾˈavɐ ˈ**umɐ** vˈɛʎɐ], utterance is "*onde morava velha*" [ˈõdə muɾˈavɐ vˈɛʎɐ].
- CUT – The word is cut, usually in the initial or final syllable, but not corrected later. Example: for prompt "*dá água ao papagaio*" [dˈa ˈagwɐ aw pɐpɐgˈaju], utterance is "*dá água ao **papaga***" [dˈa ˈagwɐ aw **pɐpɐgˈa**].
- PAU (…) – Intra-word pause, when a word is pronounced syllable by syllable and silence occurs in between. The symbol [...] can also appear in other disfluency events denoting a pause. Example: for prompt "*formosa e bonitinha*" [fuɾmˈɔzɐ i bunitˈiɲɐ], utterance is "*formosa e **boni...tinha***" [fuɾmˈɔzɐ i **buni...tˈiɲɐ**].

Silence and noise events such as breathing, labial and background noise were also annotated. Extensions and intra-word pauses may occur simultaneously with other disfluencies and are marked with [:] and […] inside phonetic transcriptions. The number of occurrences for each type of disfluency and their percentage of total uttered words in the database are presented in Table 1 for each of the 4 grades.

Table 1. Distribution of disfluency types in sentences for each of the four grades and in pseudowords (number of events and % of total uttered words).

Tags	Sentences					Pseudowords
	1st grade	2nd grade	3rd grade	4th grade	Total	Total
PRE	295 (7.4%)	278 (5.7%)	281 (4.4%)	302 (4.1%)	1156 (5.1%)	318 (15.6%)
SUB	182 (4.6%)	149 (3.1%)	215 (3.4%)	208 (2.8%)	754 (3.3%)	263 (12.9%)
PHO	214 (5.4%)	169 (3.5%)	203 (3.2%)	143 (1.9%)	729 (3.2%)	476 (23.3%)
REP	122 (3.1%)	89 (1.8%)	129 (2.0%)	161 (2.2%)	501 (2.2%)	4 (0.2%)
INS	30 (0.8%)	42 (0.9%)	42 (0.7%)	65 (0.88%)	179 (0.8%)	20 (1.0%)
DEL	5 (0.1%)	14 (0.3%)	16 (0.3%)	50 (0.68%)	85 (0.4%)	3 (0.2%)
CUT	11 (0.3%)	15 (0.3%)	29 (0.5%)	27 (0.37%)	82 (0.4%)	2 (0.1%)
EXT :	256 (6.5%)	145 (3.0%)	212 (3.3%)	73 (1.0%)	686 (3.0%)	431 (22.7%)
PAU…	179 (4.5%)	126 (2.6%)	102 (1.6%)	65 (0.9%)	472 (2.1%)	251 (13.1%)

Some interesting phenomena can be observed, such as 1st grade children being the ones that exhibit more intra-word pauses and extensions (due to slower reading), and 4th grade children having more insertions and deletions (due to faster reading). Furthermore, the defined false start type (PRE) is the most common disfluency for sentences, whereas in pseudowords mispronunciations are more common since there are fewer attempts to correct unknown words. Unexpectedly, children did not use filled pauses when trying to read aloud as teen and adults do in spontaneous speech [24], and instead pause with silence while thinking about how to read.

5 Reading Performance

5.1 Reading Speed

With annotated data, a simple analysis of the reading performance of each individual child can be done. A common metric is to evaluate reading speed considering only correctly read words, which is defined as Words Correct Per Minute (WCPM) [25]. The average values of WCPM per grade of 80 children of our corpus at the end of school year are shown in Table 2, side-by-side with the target curricular goals. A large inter-grade overlap of the distributions is observed, showing a variability in reading performance of different children, although the average does increase per grade. Fig. 2 displays this behaviour with a boxplot of the distributions of WCPM, showing one clear outlier for the third grade. On data of adults and elderly speakers reading [26,27], average words per minute are 130.3±17.8 and 118.6±21.7 respectively (on average, there are lower reading speeds for people above 60 years). Comparing these values to the observed child performance, there may still be expected improvement from 4th grade children, although some perform as well as adults. For sentence reading, the difference from average WCPM to curricular goals increases in absolute terms along the grades, and these lower WCPM values may be explained by the difficulty of the reading tasks. It can be concluded that the suggested increase of difficulty along the grades could be too steep to directly evaluate CG as intended, and, for overall reading ability evaluation, this difficulty needs to be taken into account. For pseudowords, although there are no CG for the third and fourth grades, average WCPM values are significantly lower than CG, suggesting that the created pseudowords (based on joining common syllables and not on existing words) are of high difficulty.

Table 2. Per grade Mean and Standard Deviation of measured Words Correct per Minute (WCPM), Curricular Goals (CG) of WCPM and relative difference of WCPM to CG, for sentences and pseudowords reading tasks.

	Words in Sentences			**Pseudowords**		
Grade	***WCPM***	***CG***	***WCPM-CG***	***WCPM***	***CG***	***WCPM-CG***
1st	59.7±18.1	55	+8.5%	18.8±8.0	25	-24.8%
2nd	85.2±22.9	90	-5.3%	26.7±8.4	35	-23.7%
3rd	97.1±23.5	110	-11.7%	26.1±6.5	-	
4th	110.4±22.7	125	-16.7%	34.9±9.6	-	

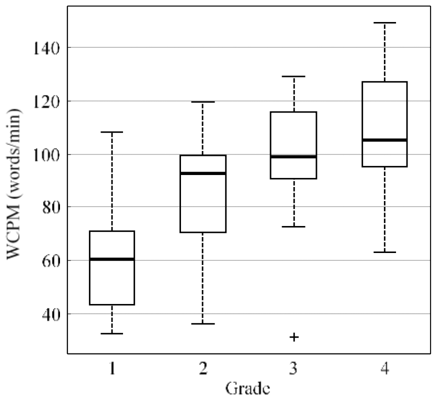

Fig. 2. Median and quartiles boxplots of Words Correct per Minute (WCPM) for sentence reading tasks for each of the 4 grades.

The defined curricular goals can be a starting point to appraise a child's reading ability. However, these do not take into account factors such as task difficulty or type of disfluencies and other ways to qualify reading performance should be considered. One possibility is to gather the opinion of experts and teachers, asking them to quantitatively rate the reading aloud performance of children (by listening to recordings of reading tasks). Their subjective opinion will be based on the several aspects of reading (speed, fluency, number of mispronunciations, etc.), and if these parameters can be quantified, there can be a correlation between the human score and a weighted average of the parameters. This is the premise of building an overall reading ability score that is well correlated with the opinion of expert evaluators [8], [9].

5.2 Pseudoword Performance

To further analyze children's performance on the task of reading individual pseudowords, the additional annotation of 100 children is considered, where they read 10 individual pseudowords each. This task differs substantially from sentence reading as morphological and phonemic awareness are the factors that influence a good performance on reading unknown words. Several interesting metrics can be extracted here, which will contribute to the overall reading performance. First, the reaction time of starting to read the word (the time between the start of recording and first try of uttering the word) reflects how fast the child is confident on reading the entire word or the first syllable, especially for first graders. However, it is not considered if the word is read correctly or not, and there are children with fast reaction times who do make several mistakes. Still, the average reaction time decreases along the grades as observed in Table 3 and Fig. 3, with only a small increase from third to fourth grades.

Table 3. Mean and standard deviation per grade of pseudoword reading reaction times (in seconds), number of uttered words with any kind of disfluency event (including extensions and intra-word pauses) and number of incorrect words.

	1st grade	**2nd grade**	**3rd grade**	**4th grade**
Reaction Time (s)	1.65±0.83	1.35±0.43	1.14±0.23	1.19±0.35
Number of disfluent words (out of 10)	6.54±2.89	3.23±2.32	2.96±1.87	2.70±2.24
Number of incorrect words (out of 10)	4.29±2.33	2.31±2.06	2.19±1.57	2.17±1.92

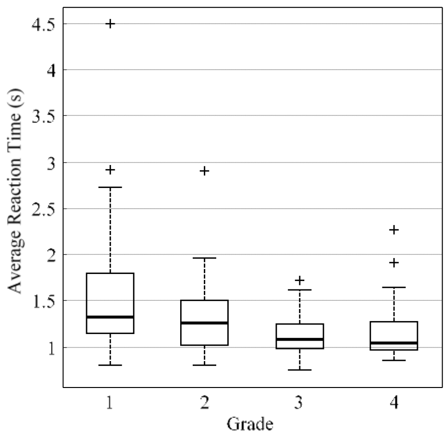

Fig. 3. Median and quartiles boxplots of average Reaction Times for the pseudoword reading task for each of the 4 grades.

Also in Table 3, the number of words that had any disfluency event is quantified. For the first grade, the average of 6.5 disfluent words out of 10 is much higher than other grades. Still, this measure is not identical to incorrect words (also presented in Table 3), since only phone extensions or intra-word pauses may occur.

6 Conclusions

A large corpus of children reading aloud was collected and analysed in terms of disfluencies and reading performance in both sentence and pseudoword reading tasks. Various types of disfluent events were observed with the most prominent being false starts and mispronunciations. Curiously, hesitations with filled pauses were not identified. The analysed reading speed metrics fall averagely close to curricular goals per grade, although the difficulty of the tasks given to children is apparently higher for third and fourth grades than what may be expected for curricular goals evaluation. To be able to give an overall score of reading performance based on reading speed and types of disfluencies committed, a difficulty metric has to be considered as a parameter.

The main goal of the LetsRead project is to automatically analyse the overall performance of children on reading aloud tasks. For that goal, great care and effort must be taken to be able to automatically detect disfluency events and analyse disfluent reading, which are the next steps to be explored. Other necessary efforts include the optimization of the difficulty metric given to sentences and pseudowords, and obtaining the opinion of teachers on the overall reading performance of children of the LetsRead dataset, so the influence of all the parameters that contribute to reading ability can be adjusted.

Acknowledgements. This work was supported in part by Fundação para a Ciência e Tecnologia under the projects UID/EEA/50008/2013 (pluriannual funding in the scope of the LETSREAD project), and Marie Curie Action IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP). Jorge Proença is supported by the SFRH/BD/97204/2013 FCT Grant. We would like to thank *João de Deus*, *Bissaya Barreto* and *EBI de Pereira* school associations and CASPAE parent's association for collaborating in the database collection.

References

1. S. M. Abdou, S. E. Hamid, M. Rashwan, A. Samir, O. Abdel-Hamid, M. Shahin, and W. Nazih, "Computer aided pronunciation learning system using speech recognition techniques," in INTERSPEECH, 2006.
2. K. Probst, Y. Ke, and M. Eskenazi, "Enhancing foreign language tutors – In search of the golden speaker," Speech Communication, vol. 37, no. 3–4, pp. 161–173, Jul. 2002.
3. T. Cincarek, R. Gruhn, C. Hacker, E. Nöth, and S. Nakamura, "Automatic pronunciation scoring of words and sentences independent from the non-native's first language," Computer Speech & Language, vol. 23, no. 1, pp. 65–88, Jan. 2009.

4. J. Mostow, S. F. Roth, A. G. Hauptmann, and M. Kane, “A Prototype Reading Coach That Listens,” in Proceedings of the Twelfth National Conference on Artificial Intelligence (Vol. 1), Menlo Park, CA, USA, 1994, pp. 785–792.
5. M. Black, J. Tepperman, S. Lee, P. Price, and S. Narayanan, “Automatic detection and classification of disfluent reading miscues in young children’s speech for the purpose of assessment,” presented at the Proc. of Interspeech, 2007, pp. 206–209.
6. J. Duchateau, Y. O. Kong, L. Cleuren, L. Latacz, J. Roelens, A. Samir, K. Demuynck, P. Ghesquière, W. Verhelst, and H. V. hamme, “Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules,” Speech Communication, vol. 51, no. 10, pp. 985–994, Oct. 2009.
7. D. Bolaños, R. A. Cole, W. Ward, E. Borts, and E. Svirsky, “FLORA: Fluent Oral Reading Assessment of Children’s Speech,” ACM Trans. Speech Lang. Process., vol. 7, no. 4, pp. 16:1–16:19, Aug. 2011.
8. M. P. Black, J. Tepperman, and S. S. Narayanan, “Automatic Prediction of Children’s Reading Ability for High-Level Literacy Assessment,” Trans. Audio, Speech and Lang. Proc., vol. 19, no. 4, pp. 1015–1028, May 2011.
9. J. Duchateau, L. Cleuren, H. V. hamme, and P. Ghesquière, “Automatic assessment of children’s reading level.,” in Proc. Interspeech, Antwerp, Belgium, 2007, pp. 1210–1213.
10. ELRA, “ELRA - ELRA-S0180 : Portuguese Speecon database.” [Online]. Available: http://catalog.elra.info/product_info.php?products_id=798. [Accessed: 06-May-2015].
11. C. Lopes, A. Veiga, and F. Perdigão, “A European Portuguese Children Speech Database for Computer Aided Speech Therapy,” in Computational Processing of the Portuguese Language, H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão, Eds. Springer Berlin Heidelberg, 2012, pp. 368–374.
12. A. Hämäläinen, S. Rodrigues, A. Júdice, S. M. Silva, A. Calado, F. M. Pinto, and M. S. Dias, “The CNG Corpus of European Portuguese Children’s Speech,” in Text, Speech, and Dialogue, I. Habernal and V. Matoušek, Eds. Springer Berlin Heidelberg, 2013, pp. 544–551.
13. A. L. Santos, M. Généreux, A. Cardoso, C. Agostinho, and S. Abalada, “A Corpus of European Portuguese Child and Child-directed Speech,” in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14), Reykjavik, Iceland, 2014.
14. A. Hämäläinen, H. Cho, S. Candeias, T. Pellegrini, A. Abad, M. Tjalve, I. Trancoso, and M. Dias, “Automatically Recognising European Portuguese Children’s Speech: Pronunciation Patterns Revealed by an Analysis of ASR Errors,” in Proc. International Conf. on Computational Processing of Portuguese - PROPOR, São Paulo, Brazil, 2014, pp. 1–11.
15. S. Lee, A. Potamianos, and S. Narayanan, “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” The Journal of the Acoustical Society of America, vol. 105, no. 3, pp. 1455–1468, 1999.
16. A. Hämäläinen, S. Candeias, H. Cho, H. Meinedo, A. Abad, T. Pellegrini, M. Tjalve, I. Trancoso, and M. S. Dias, “Correlating ASR Errors with Developmental Changes in Speech Production: A Study of 3-10-Year-Old European Portuguese Children’s Speech,” in Proc. WOCCI 2014 – Workshop on Child Computer Interaction, Singapore, 2014, pp. 7–11.
17. A. Potamianos and S. Narayanan, “Robust recognition of children’s speech,” IEEE Transactions on Speech and Audio Processing, vol. 11, no. 6, pp. 603–616, Nov. 2003.
18. H. C. Buescu, J. Morais, M. R. Rocha, and V. F. Magalhães, “Programa e Metas Curriculares de Portugês do Ensino Básico,” Ministério da Educação e Ciência, May 2015.

19. A. P. Soares, J. C. Medeiros, A. Simões, J. Machado, A. Costa, Á. Iriarte, J. J. de Almeida, A. P. Pinheiro, and M. Comesaña, "ESCOLEX: a grade-level lexical database from European Portuguese elementary to middle school textbooks," Behav Res Methods, vol. 46, no. 1, pp. 240–253, Mar. 2014.
20. G. Mendonça, S. Candeias, F. Perdigao, C. Shulby, R. Toniazzo, A. Klautau, and S. Aluisio, "A method for the extraction of phonetically-rich triphone sentences," in Proc. of the International Telecommunications Symposium (ITS), São Paulo, Brazil, 2014, pp. 1–5.
21. E. Keuleers and M. Brysbaert, "Wuggy: a multilingual pseudoword generator," Behav Res Methods, vol. 42, no. 3, pp. 627–633, Aug. 2010.
22. P. Rocha and D. Santos, "CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa," presented at the PROPOR, 2000, pp. 131–140.
23. S. Candeias, D. Celorico, J. Proença, A. Veiga, and F. Perdigão, "HESITA(tions) in Portuguese: a database," in ISCA, Interspeech satellite workshop on Disfluency in Spontaneous Speech - DiSS, KTH Royal Institute of Technology, Stockholm, Sweden, 2013, pp. 13–16.
24. A. Veiga, D. Celorico, J. Proença, S. Candeias, and F. Perdigão, "Prosodic and Phonetic Features for Speaking Styles Classification and Detection," in Advances in Speech and Language Technologies for Iberian Languages, D. T. Toledano, A. O. Giménez, A. Teixeira, J. G. Rodríguez, L. H. Gómez, R. S. S. Hernández, and D. R. Castro, Eds. Springer Berlin Heidelberg, 2012, pp. 89–98.
25. J. Hasbrouck and G. A. Tindal, "Oral reading fluency norms: A valuable assessment tool for reading teachers," The Reading Teacher, vol. 59, no. 7, pp. 636–644, 2006.
26. T. Pellegrini, A. Hämäläinen, P. B. de Mareüil, M. Tjalve, I. Trancoso, S. Candeias, M. S. Dias, and D. Braga, "A corpus-based study of elderly and young speakers of European Portuguese: acoustic correlates and their impact on speech recognition performance.," in INTERSPEECH, 2013, pp. 852–856.
27. A. Hämäläinen, J. Avelar, S. Rodrigues, M. S. Dias, A. Kolesiński, T. Fegyó, G. Németh, P. Csobánka, K. Lan, and D. Hewson, "The EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 2014.