To link to this article : DOI :10.1007/978-3-319-09761-9_1
URL : http://dx.doi.org/10.1007/978-3-319-09761-9_1

To cite this version : Hämäläinen, Annika and Cho, Hyongsil and Candeias, Sara and Pellegrini, Thomas and Abad, Alberto and Tjalve, Michael and Trancoso, Isabel and Dias, Michael *Automatically Recognising European Portuguese Children's Speech*. (2014) In: International Conference on Computational Processing of Portuguese - PROPOR 2014, 6 October 2014 - 9 October 2014 (São Carlos, Brazil).

Automatically Recognising European Portuguese Children's Speech

Pronunciation Patterns Revealed by an Analysis of ASR Errors

Annika Hämäläinen[1,2], Hyongsil Cho[1,2], Sara Candeias[1,3], Thomas Pellegrini[4], Alberto Abad[5], Michael Tjalve[7], Isabel Trancoso[5,6], and Miguel Sales Dias[1,2]

[1] Microsoft Language Development Center, Lisbon, Portugal
[2] ISCTE - University Institute of Lisbon, Lisbon, Portugal
[3] Instituto de Telecomunicações-pole of Coimbra, Coimbra, Portugal
[4] IRIT - Université Toulouse III - Paul Sabatier, Toulouse, France
[5] INESC-ID Lisboa, Lisbon, Portugal [6] Instituto Superio Técnico, Lisbon, Portugal
[7] Microsoft & University of Washington, Seattle, WA, USA
{t-anhama,t-hych,michael.tjalve,miguel.dias}@microsoft.com,
saracandeias@co.it.pt, pellegri@irit.fr,
{alberto.abad,isabel.trancoso}l2f.inesc-id.pt

Abstract. This paper reports findings from an analysis of errors made by an automatic speech recogniser trained and tested with 3-10-year-old European Portuguese children's speech. We expected and were able to identify frequent pronunciation error patterns in the children's speech. Furthermore, we were able to correlate some of these pronunciation error patterns and automatic speech recognition errors. The findings reported in this paper are of phonetic interest but will also be useful for improving the performance of automatic speech recognisers aimed at children representing the target population of the study.

Keywords: Automatic speech recognition, children's speech, error analysis, European Portuguese, fricatives, pronunciation, vowel formants.

1 Introduction

Speech interfaces have tremendous potential in the education of children. Speech provides a natural modality for child-computer interaction and can, at its best, contribute to a fun, motivating and engaging way of learning [1]. However, it is well known that automatically recognising children's speech is a very challenging task. Recognisers trained on adult speech tend to perform substantially worse when used by children [1-6]. Moreover, word error rates (WERs) on children's speech are usually much higher than those on adult speech, even when using a recogniser trained on children's speech, and they show a gradual decrease as the children get older [1-7].

The difficulty of automatically recognising children's speech can be attributed to it being acoustically and linguistically very different from adult speech [1, 2]. For instance, due to their smaller vocal tracts, the fundamental and formant frequencies of children's speech are higher [1, 2, 7-9]. What is particularly characteristic of children's speech is its higher variability as compared with adult speech, both within and across speakers [1, 2]. This variability is caused by rapid developmental changes in their anatomy, speech production etc., and manifests itself, for example, in speech rate, in the degree of spontaneity, in the frequency of disfluencies, in the values of fundamental and formant frequencies, as well as in pronunciation quality [1, 2, 7-11]. The highly variable values of acoustic parameters converge to adult levels at around 13-15 years of age [9]. Research on age-related pronunciation error patterns, so-called phonological processes or deviations, have also been carried out widely (e.g. [12-14]). Studying and understanding the acoustic and linguistic patterns of children's speech is important for designing and implementing well-functioning speech interfaces for children.

This study focuses on European Portuguese (EP) children's speech in the context of automatic speech recognition (ASR). From the point of view of phonetics, EP has characteristics that make the study of children's speech very interesting. Examples of such characteristics include its high frequency of vowel reduction and consonantal clusters, both within words and across word boundaries [15]. These two characteristics make EP difficult for young speakers to pronounce because their articulatory muscles are not sufficiently developed yet for skilfully articulating all the speech sounds and clusters of speech sounds of the language. In fact, when children attempt to imitate adult speech, they use certain processes to simplify the production of speech sounds. Such simplification may have a negative effect on ASR performance [2].

Previous work on EP children's speech includes several linguistic research projects focused on children's language, especially on language acquisition [16-21]. Studies have also been carried out to identify common age-related phonological processes in EP children’s speech [15, 22-23]. To the best of our knowledge, no other studies on the characteristics of EP children's speech have been published in the context of ASR.

In this paper, we report findings from a detailed analysis of errors made by an automatic speech recogniser trained and tested with 3-10-year-old EP children's speech. The goal of the study was to identify pronunciation patterns in children's speech that might be important from the point of view of ASR performance. The results of the study will allow us to understand the mechanisms of EP children’s pronunciation and to find ways of improving the accuracy of ASR systems aimed at them and, hence, to improve their experience with speech-enabled applications.

2 Methodology

To reach our goal, we analysed EP children's speech with specific reference to a speech recogniser built for a multimodal educational game aimed at 3-10-year-old Portuguese children [24]. The recogniser was trained and tested with speech extracted from a corpus of EP children's speech, which was specifically collected for this purpose. When carrying out the analysis, we focused on utterances that had not been recognised correctly, as well as on utterances that had been recognised correctly but with a low confidence score. This chapter describes the speech material, the automatic speech recogniser, and the methodology used in our study. The results of our analysis are reported in Section 3.

2.1 Speech Material

We used speech extracted from the CNG Corpus of European Portuguese Children's Speech [24]. The corpus contains four types of utterances recorded from children aged 3-10: phonetically rich sentences, musical notes (e.g. *dó*), isolated cardinals (e.g. *44*), and sequences of cardinals (e.g. *28, 29, 30, 31*). The children were divided into two groups when developing the corpus: 3-6-year-olds and 7-10-year-olds. The prompts for both the cardinals and the cardinal sequences were designed to be easier in the case of the 3-6-year-olds, who were also asked to produce fewer prompts. Depending on their age and reading skills, the children either read the prompts, or repeated them after a recording supervisor. The corpus comes with manually verified transcriptions, as well as annotations for filled pauses, noises, and incomplete, mispronounced and unintelligible words. Table 1 presents the main statistics of the training and test data used in this study.

Table 1. The main statistics of the speech material

	Training	Test
#Speakers	432	52
#Word types	605	521
Ages 3-6	*557*	*319*
Ages 7-10	*585*	*494*
#Word tokens	102,537	12,029
Ages 3-6	*9553*	*1148*
Ages 7-10	*92,984*	*10,881*
hh:mm:ss	17:42:22	02:05:34
Ages 3-6	*02:30:24*	*00:18:31*
Ages 7-10	*15:11:58*	*01:47:03*

2.2 Automatic Speech Recognition

In [24], several different Hidden Markov Model (HMM) -based speech recognisers for EP children's speech were trained and tested. Table 1 summarises the datasets used for training and testing the recognisers. The best-performing recogniser, which we are also using in this study, was a cross-word triphone recogniser trained using a standard acoustic model training procedure with decision tree state tying (see e.g. [25]). Thirty-eight phone labels were used for training the triphones, which have 14 Gaussian mixtures per state. The recogniser also comprises a silence model, a hesitation model and a noise model; the last two were trained utilising the annotations for filled pauses and noises that are available in the corpus. The recogniser was specifically trained for a multimodal educational game, which was developed in the

Contents for Next Generation Networks (CNG) project and expects isolated cardinals, sequences of cardinals and musical notes as speech input [24]. Therefore, [24] used constrained grammars for language modelling purposes: a list grammar for the musical notes, and structure grammars for the isolated cardinals and the cardinal sequences. The grammar for the isolated cardinals allowed cardinals from 0 to 999, whereas the grammar for the cardinal sequences allowed sequences of 2-4 cardinals ranging from 0 to 999; the grammars corresponded both to the recorded data and to the expected speech input. During the experimentation phase, [24] recognised the phonetically rich sentences using a list grammar consisting of the phonetically rich sentences recorded for the corpus; the CNG game itself does not use this type of speech input.

Table 2. WERs (%) with a 95% confidence interval for all, for 3-6-year-old, and for 7-10-year-old speakers in the evaluation test set

	Full Test Set	Ages 3-6	Ages 7-10
Children's ASR	10.0 ± 0.5	27.1 ± 2.6	8.2 ± 0.5

Table 3. The WERs (%) of the children's speech recogniser per utterance type

	Full Test Set	Ages 3-6	Ages 7-10
Phonetically rich	10.4	25.6	6.6
Musical notes	4.2	13.3	2.2
Isolated cardinals	6.3	27.4	3.9
Sequences of cardinals	10.6	33.3	9.7
Overall (excl. phon. rich)	9.8	29.3	8.7

Table 4. The number of word substitution, insertion and deletion errors made by the children's speech recogniser, excluding the phonetically rich sentences

	Full Test Set	Ages 3-6	Ages 7-10
Substitutions	345	60	285
Insertions	198	15	183
Deletions	303	60	243

Table 2 summarises the speech recognition results obtained with the children's speech recogniser. Similar to other studies [3-5, 7], the WERs were considerably higher in the case of the younger children.

Table 3 lists the WERs of the children’s speech recogniser for each of the recorded utterance types. It also includes the overall WERs without phonetically rich sentences, which represent a prompt type that is not applicable to the CNG game. Table 4 presents the corresponding number of substitution, insertion and deletion errors made by the children’s speech recogniser; the higher number of errors in the case of the 7-10-year-olds reflects the larger amount of test data in their case. The results in Table 3 make it clear that the recognition performance of 3-6-year-olds leaves much to be desired. While the recognition performance of the different types of prompts also leaves room for improvement in the case of 7-10-year-olds, it may already be acceptable for the CNG game – in particular in the case of musical notes and isolated cardinals.

2.3 Auditory Analysis

We analysed the word substitution, insertion and deletion errors made by the children's speech recogniser on the set of test utterances excluding the phonetically rich sentences (see Section 2.2 and Table 4). In total, we analysed 87 errors made in the case of the 3-6-year-olds and 39 errors made in the case of the 7-10-year-olds. In some cases, the recogniser did not output any words for the whole utterance. A preliminary analysis of the utterances with recognition errors suggested that the word substitution errors were the most interesting errors for a thorough auditory phonetic analysis, so we focussed on those types of errors in particular. To get a better overall picture of the pronunciation patterns that might be important from the point of view of ASR performance, we also analysed utterances that had been recognised correctly but with a low confidence score (51 utterances from the 3-6-year-olds and 51 utterances from the 7-10-year-olds).

Two qualified phoneticians, one an expert in Portuguese phonetics and another an expert in general auditory phonetics, carefully listened to all the test utterances that had been misrecognised by the children's speech recogniser. They transcribed the children's phonetic realisations of the misrecognised words using SAMPA (Speech Assessment Methods Phonetic Alphabet; [27]), compared their transcriptions with the standard transcriptions of the words in question, and categorised the differences between the two. The results of this analysis are reported in Section 3.

2.4 Acoustic Analysis of Vowel Formants

The auditory analysis carried out by the phoneticians suggested that vowels are usually pronounced correctly by the children in the corpus and do not play a role when it comes to ASR performance. However, for a more complete analysis of EP children's speech, we computed the average formant values for the 3-6-year-old and the 7-10-year-old children. As there was no obvious correlation between the ASR errors and the realisation of the vowels, we did not limit this analysis to the utterances with ASR errors but extended it to all the vowels in the phonetically rich sentences of the corpus (1848 and 7077 phonetically rich sentences recorded from 3-6-year-olds and 7-10-year-olds, respectively).

To be able to compute the average formant values for vowels, we obtained phoneme-level segmentations by carrying out a forced alignment of the phonetically rich sentences using an in-house (adult) speech recogniser [28]. We used context-independent acoustic models for the forced alignment, as they are considered more suitable for linguistically motivated research than context-dependent models (e.g. [29]). We extracted the formant values, filtered out aberrant values, and drew the vowel charts using the Praat software [30]. To define the threshold values for filtering, we used the average formant values for EP adult females [31] as a reference (cf. Section 2.2). Formant values that were 400 Hz below or above the reference values were considered as artefacts and were discarded. After filtering, we were left with a set of 5,100 and a set of 24,100 vowels for computing the average F1/F2 values for the 3-6-year-olds and the 7-10-year-olds, respectively. Figure 1 illustrates the F1/F2 values for the nine oral vowels of EP, showing the expected shift in formant frequencies. The F1/F2 chart is discussed in more detail in Section 3.2.

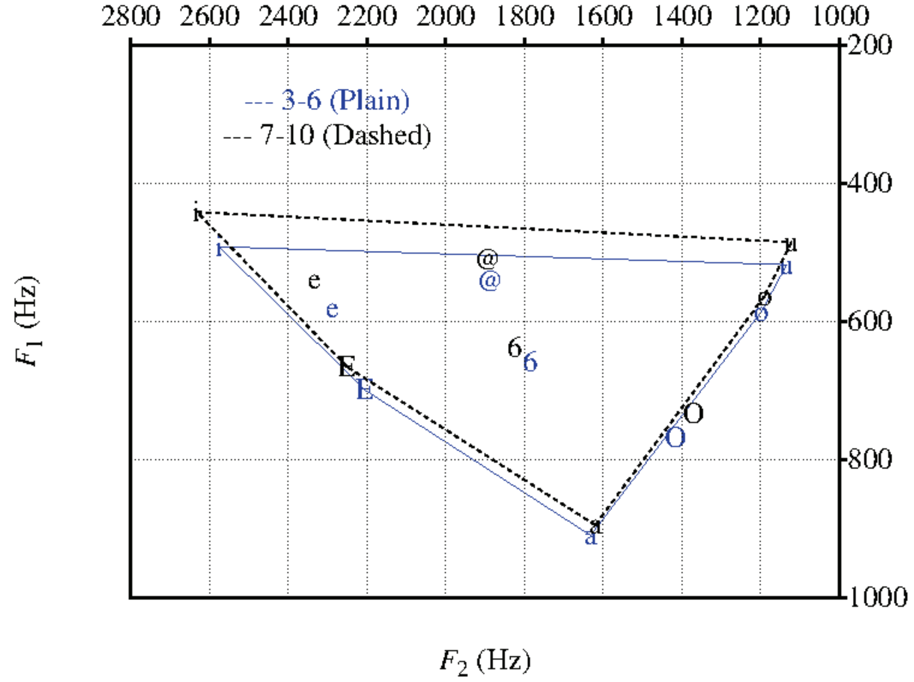

Fig. 1. F1/F2 chart for 3-6-year-olds and 7-10-year-olds

3 Pronunciation Patterns of European Portuguese Children's Speech

This section describes the findings from the auditory analysis (see Section 2.3) and the analysis of vowel formants (see Section 2.4). Before describing any pronunciations in this section, we must clarify an aspect concerning the phonetic/phonological binomial (see [32, 33]): we have here adopted a phonetic representation of sound patterns because it is closer to the physical reality of language than a phonological representation.

3.1 Consonants

ASR errors were often related to the reduction or truncation of consonant clusters, especially in the case of liquids. Previous studies [22, 23] have shown a high occurrence of consonant cluster reductions in European Portuguese children's speech.

We observed the same phenomenon in our data. For example, the word *três* ('three') was often pronounced as [t”eS] instead of the standard pronunciation [tr”eS]. This mispronunciation accounted for 10% of the misrecognitions and 10% of the correct recognition results with a low confidence score that we analysed. Considering the fact that children acquire the ability to accurately produce liquid consonants, such as [l] and [r], at the latter stage of their language acquisition process (at around 4 or 5 years old of age), this finding is not surprising.

The word *um* ([”u~]; 'one') was sometimes incorrectly recognised as the word *onze* ([”o~z@]; 'eleven'). We hypothesise that these ASR errors were related to background noise in the recordings or to the audible breathing of the speakers right after the production of the word *um*, which might have led the recogniser to output the word *onze* whose pronunciation includes the alveolar fricative [z].

As for fricative consonants, the substitution of the phones [s] and [z] with their palatal equivalents [S] and [Z] was common in the case of the 3-6-year-olds. Examples of such substitutions include:

- *sete* ('seven'): [s”Et@] → [S”Et@]
- *cinco* ('five'): [s”i~ku] → [S”i~ku]
- *dezasseis* ('sixteen'): [d@z6s”6jS] → [dZ6S”6jS]
- *dezassete* ('seventeen'): [d@z6s”Et@] → [dZ6S”Et@]
- *dezoito* ('eighteen'): [d@z”Ojtu] → [dZ”Ojtu]

There was some evidence suggesting that these phone substitutions might be correlated with ASR errors but the analysis did not provide conclusive results yet. Therefore, we must look into these phone substitutions again if and when more data becomes available in the future.

When analysing the pronunciation of plosives, we observed the velar consonant /k/ often being substituted with an alveolar stop in words like *quinze* ('fifteen'; [k”i~z@] → [t”i~z@]). This pronunciation pattern has also been reported in the literature [22] as one of the most common pronunciation patterns in EP children's speech. Interestingly, this phone substitution, which crosses phonological categories, did not seem to have any major impact on ASR performance.

We also found a devoicing deviation for the alveolar fricative [z] in words like *zero* and *doze*:

- *zero* ('zero'): [z”Eru] → [s”Eru] or [z”Eru] → [S”Eru]
- *doze* ('twelve'): [d”oz@] → [d”os@]

To further analyse devoicing deviations in EP children’s speech, we will carry out an acoustic analysis of VOT (Voice Onset Time) in future research, and analyse whether or not there is a correlation with ASR performance.

3.2 Vowels

We could not identify any word substitution errors caused by deviations in the pronunciation of vowels. As for word deletion errors, one specific word caught our attention: the word *e* ('and') was often deleted by the recogniser in the case of cardinals between 22 and 99. Although monosyllabic function words are known to be a common source of ASR errors, these errors also seemed to correlate with a pronunciation pattern that we could observe in the children's speech. In Portuguese, the orthographic form of these cardinals includes *e* between the tens and the units (e.g. *vinte e cinco* (‘twenty-five’)). However, there are two alternative ways of pronouncing these cardinals: one with the *e* (pronounced as an unstressed [i]) and another without. The speakers in the corpus often merged the pronunciation of *e* into the final vowel of the previous word. This phenomenon, which is typical of EP continuous speech also in the case of adult speakers, gives rise to a change in the syllable structure of the syntagm, which seemed to cause the children's speech recogniser to make a number of word deletion errors. Examples of this phenomenon include, for instance:

- *vinte e cinco* ('twenty-five'): [v”i~t@ i s”i~ku] → [v”i~t i s”i~ku]
- *cinquenta e quatro* ('fifty-four'):
  [si~k”we~t6 i k”watru] → [s”i~kwe~t i k”watru]

The vowel formants F1 and F2 (see Figure 1) showed age-related tendencies that did not seem to correlate with ASR errors. Although the vowel triangles of the 3-6-years-olds are very similar to those of the 7-10-year-olds, the triangle of the 3-6-year-olds has higher F1 values, mainly for close and mid-close vowels. This slight increase in F1 values could be expected as the "closer" articulation of the 3-6-year-olds is related to their vocal tracts being smaller than those of the 7-10-year-olds. The centralization of the front vowels [i], [e] and [E] is reinforced by the total absence of lip rounding, showing that children become more skilled in their ability to control the articulators with age. This is a view shared by many experts in child language acquisition [12, 14].

3.3 Other Characteristics of EP Children's Speech

We also observed other linguistic events, such as truncated words and repetitions (e.g. [k”wa k”watru] for *qua- quatro* ('fo- four')), especially in the case of the 3-6-years-olds. We expected to observe these events, well-known as hesitations or disfluencies, as they are a characteristic of read speech [34]. However, similarly to [10], they did not have an impact on ASR performance.

Compared with adult speech corpora, some children in this study uttered words with a reduced duration and/or a quiet voice. We believe that there is a psychological explanation for this: especially the younger children often reacted to the recording situation with shyness [24]. The words with a short duration and/or a low volume - in particular monosyllabic words with a simple syllable structure, accounted for a large part of the word deletion errors made by the recogniser. Examples of words that were frequently deleted include, for instance, *e* ('and'; [”i]), *um* ('one'; [”u~]), and *sim* ('yes'; [s”i~]).

4 Conclusions and Discussion

This paper reported our findings from a detailed analysis of errors made by an automatic speech recogniser trained and tested with 3-10-year-old European Portuguese children's speech. The goal of the study was to identify pronunciation patterns in children's speech that might be important from the point of view of ASR performance. The analysis confirmed the general tendencies in EP children's pronunciation that have been described by others but it also provided us with valuable information on the pronunciation patterns that actually have an impact on ASR performance. Using the findings from the analysis, we intend to derive pronunciation rules for adding relevant pronunciation variants into a pronunciation lexicon used by the children's speech recogniser. Such an approach has previously led to significant decreases in WER when automatically recognising preschool children's speech [35].

Due to the nature of the corpus and the methodology used in this study, the analysis reported in this paper has its limitations. Because the types of utterances recorded are not fully representative of everyday language use, the findings of the study are difficult to generalise. Furthermore, the data in the children's speech corpus is read or repeated speech and, as such, not fully representative of the speech input expected in the multimodal educational game that the children's speech recogniser was built for. Therefore, future studies will have to focus on collecting speech data with a wider variety of utterance types to ensure the diversity of the data from the phonetic and phonological point of view. In addition to that, the setting of future recordings will need to be reviewed to make sure that the recorded data is more representative of the type of speech that is of interest to us (spontaneous speech instead of read or repeated speech). The best option would be to collect more speech data by recording children's verbal interaction with the multimodal educational game itself.

Acknowledgements. Microsoft Language Development Center carried this work out in the scope of the following projects: (1) QREN 5329 Fala Global, co-funded by Microsoft and the European Structural Funds for Portugal (FEDER) through POR Lisboa (Regional Operational Programme of Lisbon), as part of the National Strategic Reference Framework (QREN), the national program of incentives for Portuguese businesses and industry; (2) QREN 7943 CNG – Contents for Next Generation Networks, co-funded by Microsoft and FEDER through COMPETE (Operational Program for Competitiveness Factors), as part of QREN. Sara Candeias is supported by the FCT grant SFRH/BPD/36584/2007 and is involved in the LetsRead: Automatic Assessment of Reading Ability of Children project of Instituto de Telecomunicações, co-funded by FCT. This work has also been partially been supported by FCT grant PEst-OE/EEI/LA0021/2013.

References

1. Gerosa, M., Giuliani, D., Narayanan, S., Potamianos, A.: A Review of ASR Technologies for Children's Speech. In: Workshop on Child, Computer and Interaction, Cambridge, MA (2009)

2. Russell, M., D’Arcy, S.: Challenges for Computer Recognition of Children’s Speech. In: Workshop on Speech and Language Technology in Education, Farmington, PA (2007)
3. Potamianos, A., Narayanan, S.: Robust Recognition of Children’s Speech. IEEE Speech Audio Process 11(6), 603–615 (2003)
4. Wilpon, J.G., Jacobsen, C.N.: A Study of Speech Recognition for Children and Elderly. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA, pp. 349–352 (1996)
5. Elenius, D., Blomberg, M.: Adaptation and Normalization Experiments in Speech Recognition for 4 to 8 Year Old Children. In: Interspeech, Lisbon (2005)
6. Gerosa, M., Giuliani, D., Brugnara, F.: Speaker Adaptive Acoustic Modeling with Mixture of Adult and Children’s Speech. In: Interspeech, Lisbon (2005)
7. Gerosa, M., Giuliani, D., Brugnara, F.: Acoustic Variability and Automatic Recognition of Children’s Speech. Speech Commun. 49(10-11), 847–860 (2007)
8. Huber, J.E., Stathopoulos, E.T., Curione, G.M., Ash, T.A., Johnson, K.: Formants of Children, Women and Men: The Effects of Vocal Intensity Variation. J. Acoust. Soc. Am. 106(3), 1532–1542 (1999)
9. Lee, S., Potamianos, A., Narayanan, S.: Acoustics of Children’s Speech: Developmental Changes of Temporal and Spectral Parameters. J. Acoust. Soc. Am. 10, 1455–1468 (1999)
10. Narayanan, S., Potamianos, A.: Creating Conversational Interfaces for Children. IEEE Speech Audio Process. 10(2), 65–78 (2002)
11. Eguchi, S., Hirsh, I.J.: Development of Speech Sounds in Children. Acta Otolaryngol. Suppl. 257, 1–51 (1969)
12. Bowen, C.: Children’s Speech Sound Disorders. Wiley-Blackwell, Oxford (2009)
13. Grunwell, P.: Clinical Phonology, 2nd edn. Wiliams & Wilkins, Baltimore (1987)
14. Miccio, A.W., Scarpino, S.E.: Phonological Analysis, Phonological Processes. In: Ball, M.J., Perkins, M.R., Muller, N., Howard, S. (eds.) The Handbook of Clinical Linguistics. Wiley-Blackwell, Malden (2008)
15. Candeias, S., Perdigão, F.: Syllable Structure in Dysfunctional Portuguese Children Speech. Clinical Linguistics & Phonetics 24(11), 883–889 (2010)
16. Freitas, M.J.: Acquisition in European Portuguese: Resources and Linguistic Results. Project funded by FCT: PTDC/LIN/68024/2006, Centro de Linguística da Universidade de Lisboa (CLUL) (2006)
17. Vigário, M.: Development of Prosodic Structure and Intonation (DEPE). Project funded by FCT: PTDC/CLELIN/108722/2008, Centro de Linguística da Universidade de Lisboa (CLUL) (2008)
18. Costa, J.: Syntactic Dependencies from 3 to 10. Project funded by FCT: PTDC/CLELIN/099802/2008, Centro de Linguística da Universidade Nova de Lisboa (CLUNL) (2008)
19. Freitas, M.J., Gonçalves, A., Duarte, I.: Avaliação da Consciência Linguística: Aspectos fonológicos e sintácticos do Português. Ed. Colibri, Lisbon (2011)
20. Faria, M.I.H.: Reading Comprehension. Word, Sentence and Text processing. Project funded by FCT: PTDC/LIN/67854/2006, Centro de Linguística da Universidade (2006)
21. Frota, S., Correia, S., Severino, C., Cruz, M., Vigário, M., Cortês, S.: PLEX5 A Production Lexicon of Child Speech for European Portuguese / Um léxico infantil para o Português Europeu. Laboratório de Fonética CLUL/FLUL, Lisbon (2012)
22. Guerreiro, H., Frota, S.: Os processos fonológicos na fala da criança de cinco anos: tipologia e frequência, vol. 3. Instituto de Ciências da Saúde, UCP (2010)

23. Almeida, L., Costa, T., Freitas, M.J.: Estas portas e janelas: O caso das sibilantes na aquisição do português europeu. In: Conferência XXV Encontro Nacional da Associação Portuguesa de Linguística, Porto (2010)
24. Hämäläinen, A., Miguel Pinto, F., Rodrigues, S., Júdice, A., Morgado Silva, S., Calado, A., Sales Dias, M.: A Multimodal Educational Game for 3-10-year-old Children: Collecting and Automatically Recognising European Portuguese Children's Speech. In: Workshop on Speech and Language Technology in Education, Grenoble (2013)
25. Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.: The HTK Book (for HTK Version 3.2.1). Cambridge University, Cambridge (2002)
26. Microsoft Speech Platform Runtime (Version 11), `http://www.microsoft.com/en-us/download/details.aspx?id=27225` (accessed March 25, 2013)
27. Wells, J.C.: Portuguese (1997), `http://www.phon.ucl.ac.uk/home/sampa/portug.htm`
28. Meinedo, H., Abad, A., Pellegrini, T., Neto, J., Trancoso, I.: The L2F Broadcast News Speech Recognition System. In: FALA, Vigo, pp. 93–96 (2010)
29. Vieru, B., Boula de Mareüil, P., Adda-Decker, M.: Characterisation and Identification of Non-Native French Accents. Speech Commun. 53(3), 292–310 (2011)
30. Boersma, P.: Praat, a System for Doing Phonetics by Computer. Glot International 5(9/10), 341–345 (2001)
31. Pellegrini, T., Hämäläinen, A., Boula de Mareüil, P., Tjalve, M., Trancoso, I., Candeias, S., Sales Dias, M., Braga, D.: A Corpus-Based Study of Elderly and Young Speakers of European Portuguese: Acoustic Correlates and Their Impact on Speech Recognition Performance. Interspeech, Lyon (2013)
32. Mateus, M.H., d'Andrade, E.: The Phonology of Portuguese. Oxford University Press, Oxford (2000)
33. Barbosa, J.M.: Introdução ao Estudo da Fonologia e Morfologia do Português. Almedina, Coimbra (1994)
34. Veiga, A., Celorico, D., Proença, J., Candeias, S., Perdigão, F.: Prosodic and Phonetic Features for Speaking Styles Classification and Detection. In: Toledano, D.T., Ortega, A., Teixeira, A., Gonzalez-Rodriguez, J., Hernandez-Gomez, L., San-Segundo, R., Ramos, D. (eds.) IberSPEECH 2012. CCIS, vol. 328, pp. 89–98. Springer, Heidelberg (2012)
35. Cincarek, T., Shindo, I., Toda, T., Saruwatari, H., Shikano, K.: Development of Preschool Children Subsystem for ASR and Q&A in a Real-Environment Speech-Oriented Guidance Task. In: Interspeech, Antwerp (2007)