MAXIMUM ENTROPY MOTIVATED GRAPHEME-TO-PHONEME, STRESS AND SYLLABLE BOUNDARY PREDICTION FOR PORTUGUESE TEXT-TO-SPEECH

Maria João Barros[1] & Christian Weiss[2]
[1]ISEL, Polytechnic Institute of Lisbon, Portugal, [2]L2F, Inesc Id, Portugal
mjbarros@cc.isel.ipl.pt, christian.weiss@l2f.inesc-id.pt

ABSTRACT

In this paper we present a framework for grapheme-to-phoneme (G2P) conversion, stress and syllable boundary prediction for European Portuguese (EP) Text-to-Speech (TTS) Systems. For all prediction tasks Maximum-Entropy models were used for classification. Due to the need of expensive work by experts to implement rule based G2P converters there was interest in developing probabilistic models with the Maximum-Entropy approach to solve the previous mentioned symbolic pre-processing within a TTS system. The system presented in this work is a fast and flexible approach which gives good results in each of the prediction tasks, optimal for fast application development in the TTS domain. The data used for training the G2P conversion model is manually labelled from continuous speech with natural vocalic reduction and co-articulation between words effects, common in Portuguese continuous speech. The framework is used for EP but is also usable for Brazilian Portuguese (BP) where minor changes have to be done in the G2P training data whereas stress and syllable models are the same.[1]

1. INTRODUCTION

Portuguese is the sixth most spoken language in the world, with 200 million native speakers. It is the second most spoken Latin language and the third language spoken in Occidental world. It is a Romanic language that is spoken in Angola, Brazil, Cape Verde, Indian Union, Guinea-Bissau, Macau, Mozambique, Portugal, Sao Tome & Principe and Timor Lorosa'e. It is the official language in eight countries and largely used as second language in many others.

Automatically translation of any orthographic word into its phonetic representation is needed. This is mainly known as G2P conversion. The challenge is to convert words in a way that each token sounds natural for each rhythm and not only according to the rules that are followed for paused speech or isolated word reading, this means considering coarticulation between words effects and phonetic reductions of continuous speech, known as Sandhi effects [1], [2].

Rule-based systems are useful to generate standard transcriptions to synthesis applications or to build new corpora for machine learning systems, but these systems are expensive while needing a linguist expert to setup all rules and exceptions needed to produce the results. Probabilistic systems are not that cost intensive and can be setup even without linguist knowledge. They show to be more flexible according to natural sounding synthetic speech of continuous speech, once their statistical models can be trained with data in such manner they were spoken in determined speaking rhythm that is intended to be used in each application, with corresponding allophones from coarticulation effects and phonetic reductions.

There is much work done in the speech processing domain like TTS or Automatic Speech recognition (ASR) for EP and BP. However, there is only few works done in probabilistic motivated G2P conversion, stress prediction and syllabification. But, as it was demonstrated, these are essential tasks for natural language processing with the major aspect on building TTS systems or speech recognition systems.

In this paper we introduce our approach to G2P, stress prediction and syllabification based on the Maximum-Entropy Framework introduced by [3], [4] for Natural Language Processing (NLP).

So far there are two rule based approaches to EP G2P conversion in the literature, one described in [5], [6] and another in [7], [8]. A neural network was introduced by [9] with fairly good results and leading to a CART based G2P conversion which was developed within the DIXI+ framework. A newer approach to G2P conversion was introduced by [10] where a weighted finite state transducer was implemented using the rules of the DIXI+. The disadvantage of generating many thousands of weighted finite-state-machines (WFSTs) resulting from this transformation the authors considered a hybrid approach using a combined knowledge-based and data driven-driven approach. There is also some work made for EP stress and syllable boundary prediction, using rule based approach as in [11] for syllable and [12] for stress prediction, or using neural networks probabilistic approach as in [13] for syllable and [14] for stress prediction.

This work was sponsored by Canon Foundation in Europe

This paper is organized as follows. In Section 2 we describe the Maximum-Entropy approach and refer to its difficulties. In Section 3 we explain the training corpus used for the EP models. A summary of G2P EP conversion is made in Section 4, of syllable boundary detection in Section 5 and of stress prediction in Section 6. Section 7 presents the actual results of our trained models and a conclusion is given in Section 8.

2. MAXIMUM-ENTROPY BASED MODEL

As noted above the Maximum Entropy Framework is a well known approach for ambiguities resolution in natural language processing where many problems can be reformulated as a classification problem. The task of such a reformulation is to include a context and to predict a correct class. The objective is to estimate a function $X \rightarrow Y$, which predicts an object $x \in X$ to its class $y \in Y$. Y represents the predefined classes for either each task of our prediction problem.

In the case of G2P conversion each phoneme of the phoneme inventory represents a class. In European Portuguese there are 38 Phonemes (see Table 1) which means that we have 38 classes.

In the field of stress prediction we are dealing with a binary classification where the class is true for stressed syllables and false for non-stressed.

The same binary classification task has to be solved in the domain of syllabification where we have a syllable boundary or not.

X consists of linguistic features where we include the context and the resulting input for the classification is a feature vector containing the object itself which has to be classified as well as the context. The classifier $X \rightarrow Y$ can be seen as a conditional probability model in the sense of

$$C(x) = \arg\max_{y} p(y|x) \qquad (1)$$

where x is the object to be classified and y is the class. Including the context we get a more complex classifier

$$C(x_1, x_2, ..., x_n) = \arg\max_{y_1 \cdots y_n} \prod_{i=1}^{n} p(y_i | x_1 ... x_n, y_1 ... y_{i-1}) \qquad (2)$$

Where $x_1 ... x_n, y_1 ... y_{i-1}$ is the context at the i^{th} decision and y_i is the outcome.

3. THE TRAINING CORPUS

The training corpus for G2P conversion consists of 7352 orthographic words with their phonetic transcription, which although presented to the system as isolated words obey to be a transcription that considers the context of the word, or the coarticulation between words, and include all the phonetic grammar rules. The syllable boundary detection and the stress prediction corpora have respectively 4283 and 4219 phonetic words, the first with their syllable boundary information and the second with their syllable stress classification. These two corpora have a binary classification: in or not in boundary and stressed or non-stressed syllable. These classifications were annotated as *0* for no boundary or no stress; and 1, for boundary or stress.

The G2P conversion is more complex in the sense that instead of predicting binary classes, the system needs to classify 44 different classes, as can be seen in Table 2.

Some classes are a combination of more than one phoneme which is the reason for having 44 classes instead of the 38 corresponding to the number of phonemes for Portuguese, as it was explained before.

4. GRAPHEME-TO-PHONEME CONVERSION

The most important aspect of G2P conversion is the choice of the symbol inventory used for the transcription system. Although International Phonetic Alphabet (IPA) [15] is the most complete and most widely used in transcription systems, as dictionaries, the Computer Readable Phonetic Alphabet (SAMPA) [16] is usually adopted for computational systems. The reason for this is its reduced phonetic set, relating to IPA, which is enough for G2P transcriptions and reduces the systems complexity.

In Portuguese there are 20 phonetic consonants, 14 phonetic vowels and 4 semi-vowels, which are shown in Table 1, opposed to the 18 graphic consonants and 5 graphic vowels.

Oral Vowels and Semi-vowels
6, a, E, e, @, O, o u, i, j, w
Nasal Vowels and Semi-vowels
6~, e~, o~, u~, i~, j~, w~
Fricative Consonants
v, f, z, s, S, Z
Liquid Consonants
L, l, l~
Vibrant Consonants
r, R
Plosive Consonants
b, p, t, k, g, d
Nasal Consonants
m, n, J

Table1. *European Portuguese Phoneme Inventory, (in SAMPA)*

4.1. Portuguese Vowels

From the 5 graphic vowels that exist for Portuguese, <a>, <e>, <i>, <o> and <u>, the <o> with 7 possible transcriptions, /o/, /O/, /u/, /w/, /o~/, /o~j~/, /w~/, and the <e> with 11 possible transcriptions, /E/,

/e/, /@/, /6/, /i/, /j/, /6j/, /e~/, /6~j~/, /6~j~6~j~/ and "&", show to be the most complex ones according to phonetic transcription. There are about 30 grammatical rules for each, only related to isolated words, and even more rules related to the allophones that come from coarticulation effects between words.

The symbol "&" represents a dummy class and appears in those situations of mute graphemes, in our model classified as "&". For instance the word <que> can be transcribed as /k@/ being the <u> mute, or even as /k/ being the /@/ suppressed. The grapheme <u> in both situations and <e> in the last one would be transcribed as "&".

Another problem according to the graphemes phonetic transcription is homonyms, once their transcription gives many times the meaning of homonym words. For example <sede> that can be /sed@/ or /sEd@/, or <acordo> that can be /6kordw/ or /6kOrdw/ depending on the meaning of the word.

The <a> is a grapheme that, although not so complex as the <o> or the <e>, still has 3 phonetic transcriptions: /a/, /6/ and /6~/. The <i> and the <u> represent not only the phonetic vowels <i> and <u> but also the semivowels <j> and <w> respectively and all of these four in their nasal version: <i~>, <j~>, <u~> and <w~>. The <i> can also be transcribed as a <@>, for instance when there is a sequence of two syllables with <i> as in <feminino> that is transcribed as /f@m@ninw/.

4.2. Portuguese Consonants

From the 18 graphic consonants existent for Portuguese there are some particular cases that have to be well considered.

The <h> is a consonant that by itself doesn't have a sonorous transcription as it lost its aspiration with the evolution of the language, but when preceded by one of the consonants <c>, <n> or <l>, gives the consonants combinations <ch>, <nh> and <lh> that have their own particular phonetic transcriptions, /S/, /J/ and /L/ respectively.

The <m> and <n> between a vowel and a consonant or the <m> in the end of a word, don't have a phonetic transcription but they nasalize the precedent vowels.

The most difficult consonants are the <x> which has 4 possible phonetic transcriptions: /s/, /z/, /S/ or /ks/ and the <s> that can be transcribed into /s/, /z/, /S/, or even /Z/ if coarticulation between words is considered.

Also the <l> can be /l/ if followed by a vowel or /l~/ if followed by a consonant or in the end of a word. The <g> can be transcribed as /g/ or /Z/ depending on the context.

There are also cases, although these are not important to the G2P transcription problem, of two graphemes with the same phonetic transcription. For example, both <g> and <j> can be transcribed as /Z/ and both <s> and <c> as /s/.

4.3. Probabilistic Transcription

From the training results we observe the correspondence between Portuguese graphemes and phonemes that is presented in Table 2, where "&" is a dummy as explained in Subsection 4.1.

Grapheme	Phonemes
ú	u~, u
õ	o~
ô	o
ó	O
í	i~, i
ê	6~j~6~j~, e~, 6j, 6~j~, e
é	E, 6, 6~j~, e
ç	s
ã	6, 6~
â	6, 6~
á	a, &
à	a
z	Z, S, z
x	S, ks, z, s
v	v
u	@, u~, &, w, u, w~
t	t
s	Z, S, &, z, s
r	R, &, r
q	k
p	&, p
o	O, o~, @, &, w, u, w~, o
n	J, &, n
m	&, m
l	L, l~, l
k	k
j	Z
i	i~, @, &, j~, j, i
h	&
g	Z, g
f	f
e	i~, E, 6, @, e~, 6~, &, 6j, 6~j~, j~, j, i, e
d	d
c	S, &, s, k
b	b
a	a, 6, 6~, &, 6~w~, o

Table2. *Portuguese Graphemes Phonemic Transcription*

By observation of the table we can see that some graphemes transcriptions have more than one phoneme.

These phonemes sequences were considered as individual classes, what justifies having 44 classes instead the 38 correspondent to the number of EP phonemes.

Examples of these situations or others that can be not so clear are: <tem>, where the <e> is transcribed as /6~j~/ while the <m> is mute: /t6~j~/; <têm>, where the

<ê> is transcribed as /6~j~6~j~/ while the <m> is mute: /t6~j~6~j~/; <mandam>, where the second <a> is transcribed as /6~w~/: /m6~d6~w~/; <excelente>, where the first <e> is transcribed as /6j/: /6jSs@le~t@/; <além>, where <é> is transcribed as /6~j~/ while the <m> is mute: /al6~j~/; <óptimo>, where the <p> is mute: /Otimu/; situations with double consonants being transcribed as one phoneme: <fosse> that is transcribed as /fos@/ or <torre> that is transcribed as /toR@/; or consonants followed by <h> that lead to a particular phoneme: <acho> that is transcribed as /aSw/, <ilha> that is transcribed as /iL6/, <ninho> that is transcribed as /niJu/.

5. SYLLABLE BOUNDARY DETECTION

Syllable is what denominates the part of the word which is pronounced in only one voice emission and can be classified in non-stressed, pos-stressed, pre-stressed or stressed.

Portuguese syllabic segmentation follows a set of few grammatical rules, from which 6 are related to the vowels, 6 to the consonants and 3 to others cases which has to be considered [17].

Although syllable boundary detection is not a difficult task to achieve using rule-based systems, the probabilistic approach is very simple because it is a problem of binary classification: in boundary or not in boundary of syllable, and it allows considering Sandhi effects, which are very prominent in the EP language [1], [2].

6. STRESS PREDICTION

In the Portuguese grammar the words can be classified according to the stressed syllable position in the word as oxytone, paroxytone or proparoxytone, depending if the stressed syllable is the last syllable, the one before the last or the third from the end of the word, respectively.

Many times in Portuguese the stress can be given by graphic accents, but these don't always exist in words. The stress prediction, as the syllable boundary detection, is also a problem of binary classification: stressed or non-stressed syllable.

The accents rules for Portuguese can be found in the Inter-institutional style guide from the European Union Publications Office [18].

7. RESULTS

The results for the three tasks are presented by giving the Log likelihood and the performance of the system.

The G2P transcription's model results are shown in Table 3, the syllable boundary detection's model results in Table 4 and the stress prediction's model ones in Table 5.

The performance can be seen as the number of correct classified elements divided by the number of overall elements multiplied by one hundred.

	Result
Log likelihood	-12329.53
Performance	88.94 %

Table3. *Prediction results of G2P transcription*

	Result
Log likelihood	-1200.10
Performance	97.64 %

Table4. *Prediction results of syllable boundary detection*

	Result
Log likelihood	-5465.59
Performance	85.57 %

Table5. *Prediction results of stress prediction*

Besides these results another test was made to the G2P converter once this module is more complex for classification than the other two, having 44 classes or possible classifications while the other two have a binary classification.

The corpus for this test has 550 words comprising 3430 phonemes. All the phonemes from the EP inventory were covered, with different number of occurrences.

The test consisted in comparing the system results to the test corpus with the entries from the training corpus, giving the average number of phoneme errors taking into account three kind of errors: when the system replaced the phoneme by another (substitution); when the system gave a result to a phoneme that in the manually transcription is missed (insertion); and when the system gave nothing and the manually transcription gives a phoneme (deletion). This kind of measure is known in the speech recognition domain by word error rate (WER).

The results, given by the number of correct classified phonemes divided by the number of overall phonemes multiplied by one hundred, are shown in Table 6 and Table 7, using the confusion matrix method [19], where the rows are the actual classes and the columns the predicted classes.

As explained in Subsection 4.3, some classes are a combination of phonemes in order to represent all possible graphemes transcriptions.

-	ks	S	d	Z	k	g	t	J	v	s	b	&	z	r	l~	L	f	n	m	l	p	R
ks	87	10	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-
S	1	96	-	-	-	-	-	-	-	2	-	-	1	-	-	-	-	-	-	-	-	-
d	-	-	99	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	4	-	93	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-
k	-	-	-	-	97	-	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-
g	-	-	-	2	-	98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
t	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
J	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
v	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-
s	1	4	-	-	-	-	-	-	-	93	-	1	1	-	-	-	-	-	-	-	-	-
b	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-
z	-	3	-	-	-	-	-	-	-	3	-	-	94	-	-	-	-	-	-	-	-	-
r	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-
l~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	83	-	-	-	-	17	-	-
L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-
f	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-
n	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	98	-	-	-	-
m	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-
l	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	98	-	-
p	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	99	-
R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100

Table 6. *Confusion matrix for consonants (in percent.)*

-	e~	a	E	j	j~	u	6~j~	e	6~j~6~j~	u~	&	i~	w	w~	@	i	6	O	o~	6~	6j	o
e~	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
a	-	83	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-
E	-	-	78	1	-	-	-	8	-	-	4	-	-	-	5	4	-	-	-	-	-	-
j	-	-	-	77	2	-	-	4	-	-	6	-	-	-	-	11	-	-	-	-	-	-
j~	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
u	-	-	-	-	-	63	-	-	-	-	12	-	17	-	4	-	-	-	-	-	-	4
6~j~	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
e	-	-	-	-	-	-	-	66	-	-	27	-	-	-	7	-	-	-	-	-	-	-
6~j~6~j~	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-	-
u~	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-	-
&	-	-	1	-	-	1	-	1	-	-	90	-	3	-	2	2	-	-	-	-	-	-
i~	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-	-
w	-	-	-	-	-	18	-	-	-	-	17	-	59	-	-	-	-	1	-	-	-	5
w~	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-	-	-	-	-	-
@	-	-	-	-	-	-	-	4	-	-	40	-	12	-	36	4	-	-	-	-	-	4
i	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	97	-	-	-	-	-	-
6	-	8	-	-	-	-	-	-	-	-	3	-	-	-	-	-	89	-	-	-	-	-
O	-	-	-	-	-	3	-	-	-	-	3	-	5	-	-	-	-	75	-	-	-	14
o~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-
6~	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-	-
6j	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	-
o	-	-	-	-	-	3	-	-	-	-	2	-	2	-	-	-	-	13	-	-	-	80

Table 7. *Confusion matrix for vowels (in percentage)*

8. CONCLUSION

The vowels show to be more difficult than the consonants due to the fact of most of them having several possible transcriptions. But there are also very complex consonants as the <x> and the <s> that even the rules don't cover all of their possible cases.

As one can see in Table 6 the /ks/, /S/, /Z/, /s/ and /z/ are the most difficult cases. It is convenient to refer that although the /l/ and the /l~/ were sometimes confused with each other we don't consider these as difficult cases, because this is not perceptually relevant and even some systems don't distinguish between them.

Regarding to the vowels, from the analysis of Table 7, we see that some of the phoneme substitutions are not even errors because have no perceptual significance or are both acceptable transcriptions. For example, substituting an /i/ by an /j/ and vice versa, or an /u/ by an /w/ and vice versa is completely acceptable and it is even acceptable that these are substituted by /@/ or "&", because the system considers the phonemic transcription. This means that sometimes there are more than one possible choice.

Suppressing the /@/ means substituting the /@/ by "&", what is also reasonable and the only thing that is affected is the rhythm of the speech, not the meaning. If we attend to these considerations we can consider that the /u/ has 96% of correct values, the /w/ has 94%, the /j/ has 88% and the /@/ has 88%.

The system represents a first approach to the given tasks and can be improved by adding more data in the training corpora, but attending to the considerations made about the results in the confusion matrices the system shows to be a reliable solution and the probabilistic Maximum-Entropy Framework shows to be a good and simple approach.

9. REFERENCES

[1] D. Braga, D. Freitas, H. Ferreira: Processamento Linguístico Aplicado à Síntese da Fala. In: 3º Congresso Luso-Moçambicano de Engenharia, Maputo, Mozambique (Aug. 2003).

[2] R. Amaral, P. Carvalho, D. Caseiro, I. Trancoso, L. Oliveira: Anotação fonética automática de corpora de fala transcritos ortograficamente. In: PROPOR'99 - IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada, Evora, Portugal (Sep. 1999).

[3] A. Berger, S.A. Della Pietra, V.J. Della Pietra: A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics 22(1) (1996).

[4] A. Ratnarparkhi: Maximum Entropy Models for Natural Language Ambiguity Resolution. PhD Dissertation, University of Pennsylvania (1998).

[5] L. Oliveira, M. Viana, I. Trancoso: A Rule-Based Text-to-Speech System for Portuguese. In: Proc. ICASSP'92, San Francisco, USA (Mar. 1992).

[6] F. Barbosa, G. Pinto, F. G. Resende, C. A. Gonçalves, R. Monserrat, M. C. Rosa: Grapheme-Phone Transcription Algorithm for a Brazilian Portuguese TTS. In: Computational Processing of the Portuguese Language, 6th International Workshop (PROPOR 2003), Springer Verlag, Faro, Portugal (Jun. 2003) 23-30.

[7] J. P. Teixeira, D. R. Freitas, P. D. Gouveia, G. Olaszy, G. Németh: MULTIVOX – Conversor Texto Fala Para Português. In: III Encontro Para o Processamento Computacional da Língua Portuguesa Escrita e Falada, Porto Alegre, Brasil (Nov. 1998.

[8] J. P. Teixeira: Conversor Texto-Fala para o Português – Desenvolvimentos / Ferramentas. In: 1as jornadas do CEFAT, Bragança, Portugal (1998).

[9] Trancoso, M. Viana, F. Silva, G. Marques, L. Oliveira: Rule-Based vs. Neural Network Based Approaches to Letter-to-Phone Conversion for Portuguese Common and Proper Names. In: Proc. ICSLP'94, Yokohama, Japan (Sep. 1994).

[10] D. Caseiro, I. Trancoso, L. Oliveira, C. Viana: Grapheme-to-Phone using Finite-State Transducers. In: IEEE Workshop on Speech Synthesis Santa Monica, California (Sep. 2002).

[11] P. D. Gouveia, J. P. Teixeira, D. R. Freitas: Divisão Silábica Automática do Texto Escrito e Falado. In: V PROPOR – Processamento Computacional da Língua Portuguesa Escrita e Falada, S. Paulo, Brazil (Nov. 2000).

[12] J. P. Teixeira: A Prosody Model to TTS Systems. PhD dissertation, Faculty of Engineering of University of Porto, Portugal (2004).

[13] H. Meinedo, J. Neto, L. Almeida: Syllable onset detection applied to the Portuguese language. In: EUROSPEECH'99 - 6th European Conference on Speech Communication and Technology, Budapest, Hungary (Sep. 1999).

[14] C. J. Teixeira, I. M. Trancoso, A. J. Serralheiro: Accent Identification. In: ICSLP'96- 4th Int. Conf. on Spoken Language Processing, Philadelphia, USA (Oct. 1996).

[15] The International Phonetic Association, (http://www2.arts.gla.ac.uk/IPA/ipa.html)

[16] SAMPA-Speech Assessment Methods Phonetic Alphabet, (http://www.phon.ucl.ac.uk/home/sampa)

[17] "Código de Redacção Interinstitucional", (http://publications.eu.int/code/pt)

[18] http://publications.eu.int/code/pt/pt-4100100pt.htm#i137

[19] http://dms.irb.hr/tutorial/tut_mod_eval_1.php