Listener: a Pronunciation System for Brazilian Portuguese Speakers English L2 Learners

Author: Gustavo Mendonça, MS Student (USP)

gustavoauma@gmail.com

Advisor: Sandra Maria Aluisio, PhD (USP)

sandra@icmc.usp.br

Co-Advisor: Aldebaro Klautau Jr., PhD (UFPA)

aldebaro.klautau@gmail.com

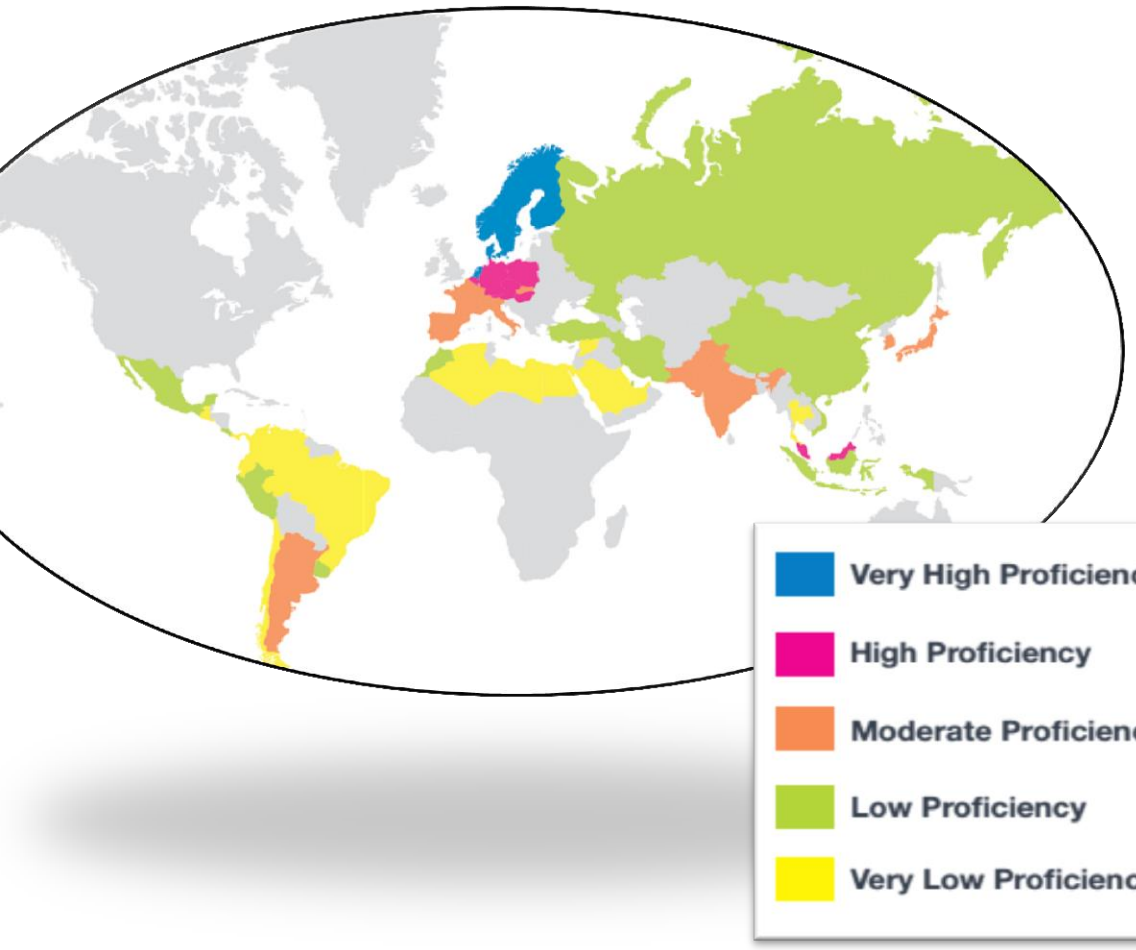
MOTIVATION

Recent surveys have shown that Brazil is among the countries with the lowest knowledge of the English language.

In Education First's *English Proficiency Index* 2012 [1], for instance, Brazil ranked 46th out of 54 countries.

Brazil achieved a similar position in GlobalEnglish's *Business English Index 2013* [2], being ranked as 71nd across 77 positions. The country was classified among those with very low proficiency, such as El Salvador and Kuwait.

• English Proficiency Index 2012



Proficiência muito baixa

Classificação	País	EF EPI
39	Chile	48.41
40	Venezuela	47.50
41	El Salvador	47.31
42	Síria	47.22
43	Equador	47.19
44	Argélia	47.13
45	Kuwait	47.01
46	Brasil	46.86
47	Guatemala	46.66
48	Egito	45.92
49	Emirados Árabes Unidos	45.53
50	Colômbia	45.07
51	Panamá	44.68
52	Arábia Saudita	44.60
53	Tailândia	44.36
54	Líbia	42.53

Figure 1. Education First's *English Proficiency Index* (2012) worldmap.

- Business English Index 2013

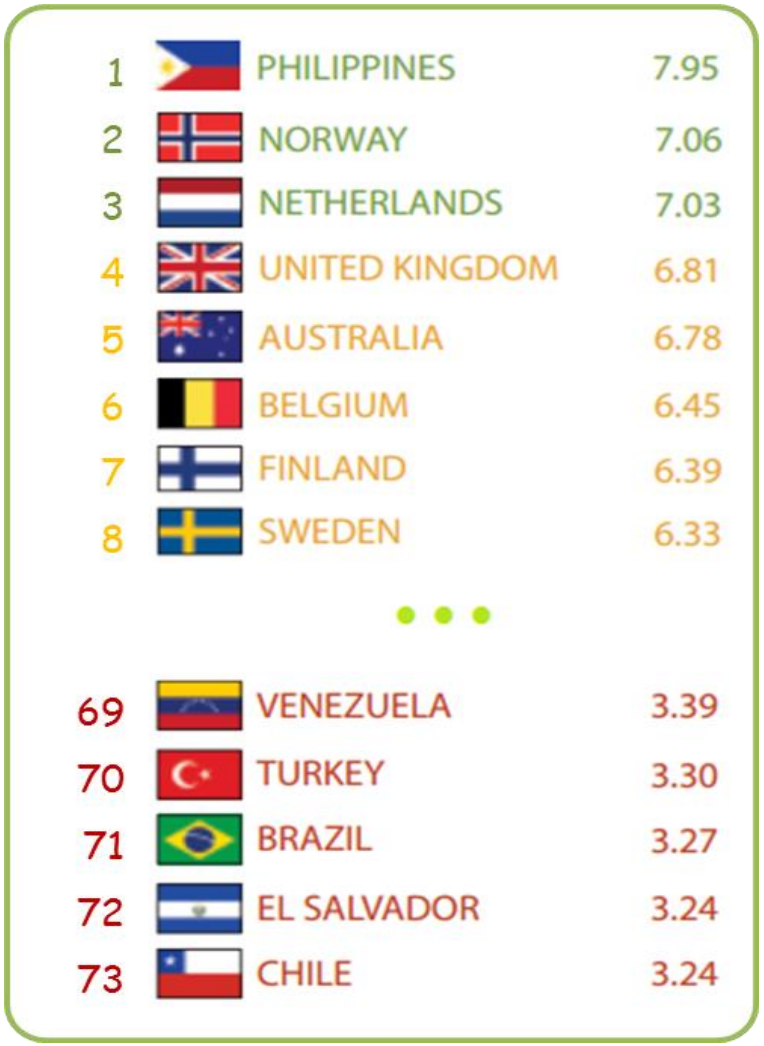


Figure 2. *Global English's Business English Index* (2013) partial ranking .

PROPOSAL

This project aims at developing a tool for improving these indexes. The goal is to build up an ASR-based Pronunciation System for Brazilian Portuguese (BP) speakers English L2 learners. The Pronunciation System proposed herein, called *Listener*, will be able to provide online feedback regarding the pronunciation of the user.

GAP

Similar tools are available for other languages, such as japanese [3], french [4], spanish [5] and dutch [6], however, for BP, there is still a gap to be explored.

RESEARCH QUESTION

The research hypothesis states that it is possible to build up an efficient Pronunciation System through:

(i) an error classification that takes into account phonetic and phonological transfer from L1 to L2;
(ii) an acoustic model that contains speech data from both native speakers and English L2 learners;
(iii) a pronunciation dictionary which includes the transcription of the mispronunciations that learners are likely to make;
(iv) and a language model befitting the syntax of the learner.

A Brief Review of Automatic Speech Recognition through Hidden Markov Models (HMM)

- Noisy-channel Metaphor [7]: the recognition system tries to estimate, for a language $\mathcal{L}$, given a certain acoustic input O, what is the most likely sentence $\hat{W}$ out of all sentences W :

ESTIMATED BY: AM LM

$$\hat{W} = \underset{W \in \mathcal{L}}{argmax}\, P(W|O) \quad \therefore \quad \hat{W} = \underset{W \in \mathcal{L}}{argmax}\, P(O|W)\, P(W)$$

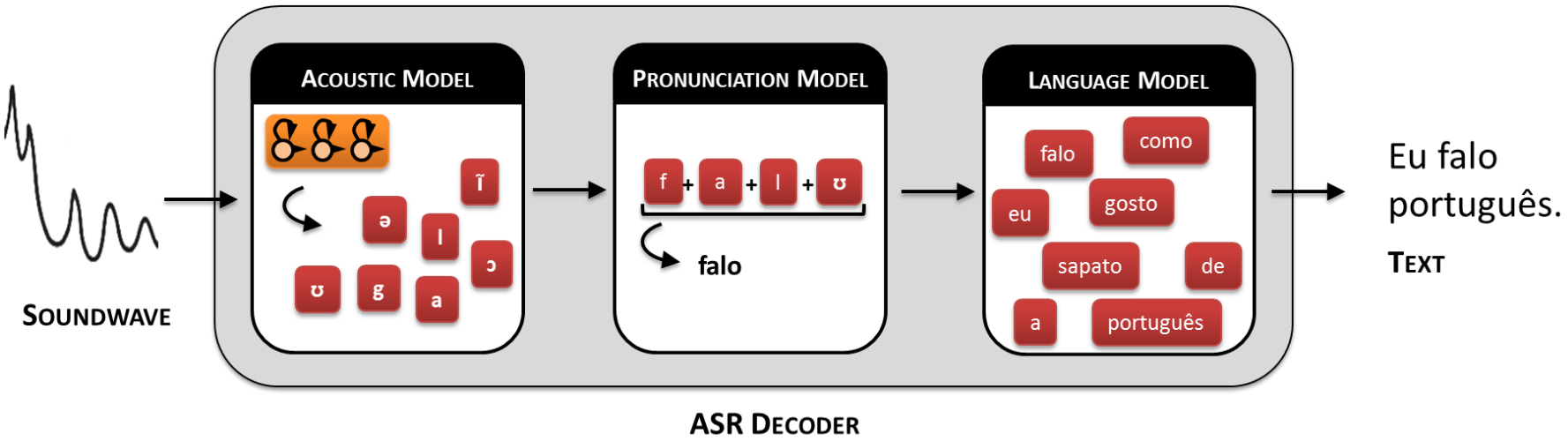


Figure 3. *Architecture of an HMM Automatic Speech Recognition System.*

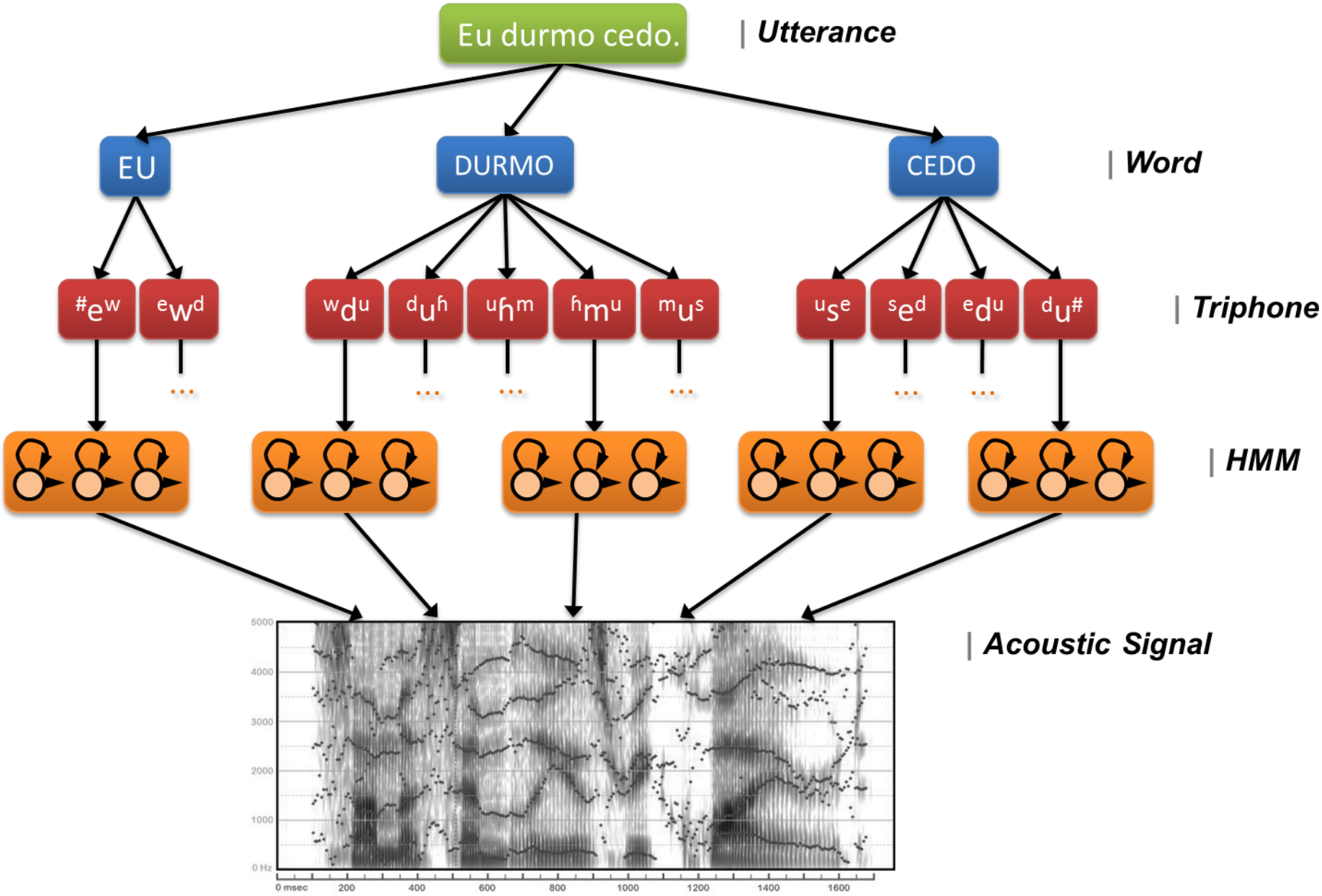


Figure 4. *Architecture of an HMM Automatic Speech Recognition System.*

METHOD

- The General American (GA) will be considered the standard accent.

- The engine *Julius* [8] is used as the basis of the recognizer.

- Ten mispronunciations were selected, according to manuals and tutorials in english pronunciation teaching for native brazilian portuguese speakers [9, 10, 11].

Table 1. Selected mispronunciations for the Pronunciation System.

SELECTED MISPRONUNCIATIONS	
NO.	**DESCRIPTION**
1	Epenthesis of [i, ɪ]
2	Vocalisation of [l, ɬ] in coda position
3	Deletion of [m, n] in coda position + Nasalisation of previous vowel
4	Deletion [ŋ] in word final position + Nasalisation of previous vowel + Insertion of [g]
5	Realisation of [ɹ] as [h, ɦ, x, ɣ, ɾ, r]
6	Realisation of [θ] as [f, t, s, d], and of [ð] as [d]
7	Lack of aspiration in [p^h, t^h, k^h] in stressed syllables
8	Simplification of 3rd person singular present tense verbal forms: [s, z, ɪz] endings are realised as /S/
9	Simplification of regular simples past and past participle forms: [t, d, ɪd] endings are realized as [ɪd, ed]
10	Shortening of long vowels [iː,ɑː, ɜ˞, ɔː, uː]

- The Acoustic Model (AM) will be built up, in a pooled fashion, based on two speech corpora, one containing data of native English speakers: *TIMIT Acoustic-Phonetic Continuous Speech Corpus* [12], and another of English L2 learners: *COBAI - Corpus Oral Brasileiro de Aprendizes de Inglês* [13].

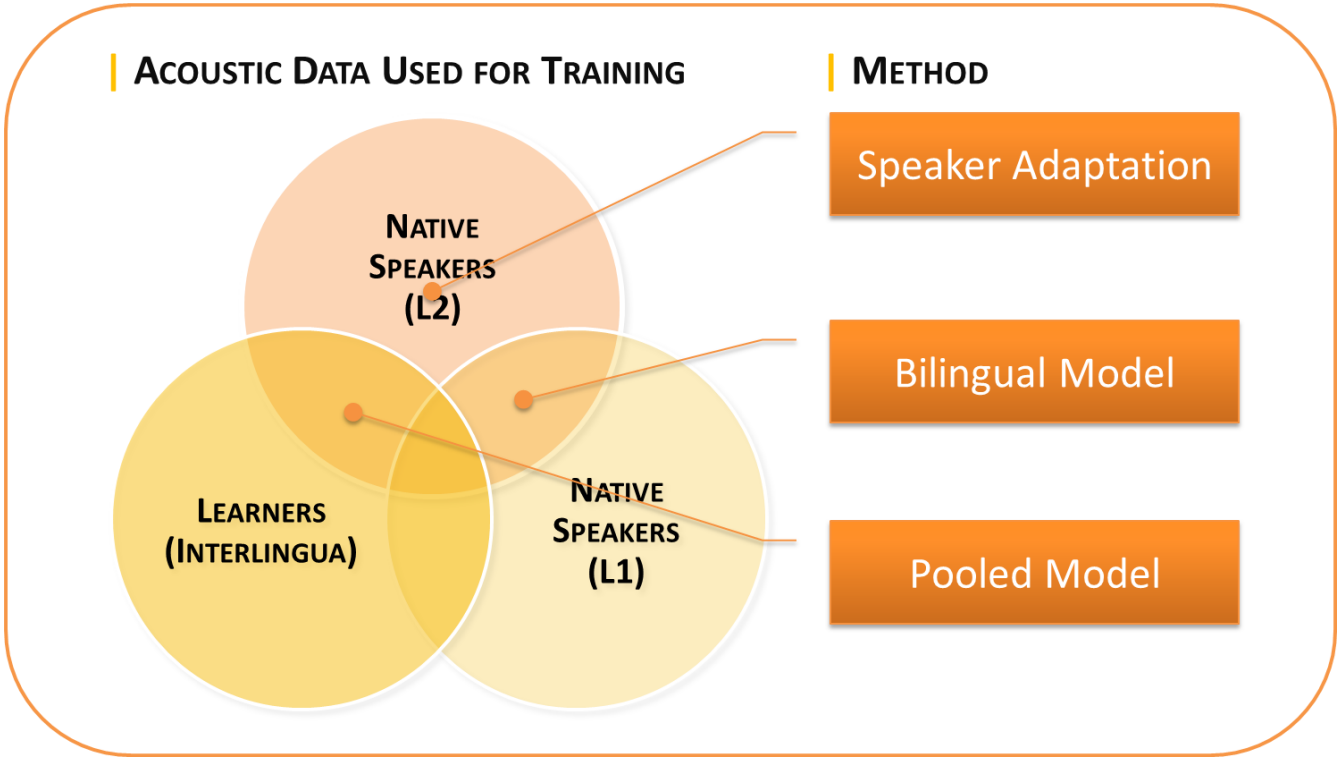


Figure 5. Methods of adapting the Acoustic Model (AM) to non-native speakers.

- *VoxForge Speech Corpus*‘ [14] word list will serve as the basis of the Pronunciation Model (MP).
- Mispronunciations of the learners will be added to the dictionary, manually and also automatically by machine learning algorithms, such as *Transformation-Based Learning* (Brill, 1995).
- The Language Model (LM) will be compiled over 99,508 articles from Simple English Wikipedia [15].

EVALUATION

Word Error Rate (WER), Character Error Rate (CER) and confusion matrices will be the measurements used to evaluate the performance of the recognizer. These metrics will be applied to both the corpora used through a ten-fold stratified cross-validation technique.

REFERENCES

[1] Education First (2012) English Proficiency Index 2012. Retrieved from: http://www.ef.com.br/epi/downloads. May, 2013.

[2] GlobalEnglish (2012) The Business English Index 2012 Report: Analyzing the Trends of Global Readiness for Effective 21st-Century Communication. Brisbane: GlobalEnglish.

[3] Tsubota, Y., Dantsuji, M., & Kawahara, T. (2004). An English pronunciation learning system for Japanese students based on diagnosis of critical pronunciation errors. ReCALL, 16, 173-188.

[4] Genevalogic. (2006). SpeedLingua - Language Learning Accelerator. Retrieved from: http://www.speedlingua.com/ Speedlingua. Sept, 2013.

[5] Reis, J., & Hazan, V. (2011). Speechant: a vowel notation system to teach English pronunciation. ELTJournal, 66/2, pp. 156-165.

[6] Strik, H., Doremalen, J., & Cucchiarini, C. (2008). A CALL system for practicing speaking . Proceedings of the XIIIth Int. CALL Conference, (pp. 123-125). Antwerp.

[7] Jurafsky, D.; Martin, J. (2009) Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. London, Pearson Education Ltd.

[8] Lee, A.; Kawahara, T. (2009) Recent Development of Open-Source Speech Recognition Engine Julius. In: Proceedings APSIPA ASC 2009 - Asia-Pacific Signal and Information Processing Association, pp. 131-137.

[9] Godoy, S., Gontow, C., & Marcelino, M. (2006). English Pronunciation for Brazilians. Porto Alegre: Disal.

[10] Zimmer, A., Silveira, R., & Alves, U. (2009). Pronunciation Instruction for Brazilians: Bringing Theory and Practice Together. Newcastle: Cambridge Scholars.

[11] Cristófaro-Silva, T. (2012). Pronúncia do Inglês para Falantes do Português Brasileiro. São Paulo: Contexto.

[12] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren (1990) DARPA, TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM. National Institute of Standards and Technology.

[13] COBAI

[14] Voxforge

[15] Simple English Wikipedia