

# Automatic Discovery of Brazilian Portuguese Letter to Phoneme Conversion Rules through Genetic Programming

Evandro Franzen<sup>1</sup> and Dante Augusto Couto Barone<sup>2</sup>

<sup>1</sup>UNISC – Universidade de Santa Cruz do Sul  
Av. Independência, 2293-Bairro Universitário CEP 96815-900 Santa C. do Sul – RS  
evandro@open.inf.br

<sup>2</sup>UFRGS – Universidade Federal do Rio Grande do Sul. Instituto de Informática  
Av. Bento Gonçalves, 9500 – Campus do Vale – Bloco IV. Bairro Agronomia – Porto Alegre –  
RS – Brasil. CEP 91501-970 Caixa Postal: 15064  
barone@inf.ufrgs.br

**Abstract.** Letter to phoneme conversion is a basic step in Speech Synthesis processes. Traditionally, the activity involves the implementation of rules that define the mapping of letters into sounds. This paper presents results of the application of an evolutionary computation technique (Genetic Programming), in Brazilian Portuguese synthesis, aiming to discover automatically programs implementing specific synthesis rules.

## 1 Introduction

The spoken language is the most used form of communication between humans, being simultaneously powerful and simple. The interaction between men and machines continues to be a hard problem to solve, and the application of Artificial Intelligence techniques to perform these tasks is not straight forward. [5].

Automatic speech processing performed by computers is mainly performed by two different kinds of problems: speech recognition and speech synthesis [5]. The first aims to convert an acoustic signal, captured by a microphone or by a telephone, into a set of intelligible words or phrases [8]. The second one consists in automatic generation of voice waveforms, commonly generated from a written or a stored text [8].

One of the most common approaches for performing speech synthesis is Text To Speech (TTS) technique. In this approach, a text is converted into a set of phonemes which are gathered to produce synthetic "voice" signals [4].

In most of the world's spoken languages, a written text does not correspond to its pronunciation; thus to describe the correct pronunciation, a set of symbolic representations become necessary. Each language possesses some intrinsic characteristics as different phonetic alphabet, set of possible phonemes and its combinations. Each language possess a specific set of phonemes, which can be defined as "elementary" sounds, used as "bricks" to construct any used sound found in speech productions, using that considered language [8]. In many languages there isn't an exact consistency

between phonemes and the corresponding letters (graphemes) that can produce them [11].

The present work is part of the Spoltech Project [2], which aims to create, develop and provide technologies of speech processing: speech recognition and synthesis to Brazilian Portuguese. As synthesis procedure we use concatenation of diphones.

Letter to phoneme conversion is done through described rules in the LTS (letter to sounds) module. One of the major goals of this present work is to provide the Spoltech synthesizer used tool (Festival environment [12]) with additional advanced technologies, based in Genetic Programming, to accomplish the processes that compose the speech synthesis.

## 2 Modeling the Problem Using Genetic Programming

The rules to convert letters into phonemes can be represented through computer programs. If they were implemented by human programmers, they probably would have this form: IF (current letter is x) THEN the phoneme is y ELSE...

In accordance with [6], there are five major steps in preparing the utilization of genetic programming to solve a specific problem: i) determining the set of terminals; ii) determining the set of functions; iii) determining the fitness measure and the fitness cases; iv) determining the parameters and variables for controlling the run, and; v) determining the method of designating a result and the criterion for stopping a run.

The activity of converting letter into phonemes can be summarized as the application of rules on words or letters to discover corresponding phonemes. Being this, the initial step for determining one or more strategies to find solutions through the Genetic Programming is to define if fitness cases will consist of a set of letters or words. In our case, we are using words to have the fitness measured.

The set of fitness cases is defined as a word list, each one composed by a list of letters ((p a t o) (t i p o)). The correct answers for each case is specified as lists of phonemes in the same way ((p a t o) (ts i p o)). The definition of cases and respective answers in a list form was chosen by the easiness found in the LISP language to deal with list processing; however other data structures can be used, not compromising the solution search process.

To evaluate the produced answer for each individual, the Genetic Programming system [3] calculates the raw fitness using the following rules:

- In the case that the produced phoneme is equal to the expected in the specific position, three points are credited to the solution;
- in the case that the phoneme produces differently in a considered position, but represents an expected phoneme in any other word's position, one point is credited to the solution.

Standardized fitness must always indicate better solutions having lesser values, tending to zero. In this problem we start with the biggest expected value for fitness, diminishing it as solutions evolve.

### 3 Experimental Results

The set of rules to convert a letter into a phoneme in Portuguese is extremely wide. Silva [11] describes a set of more than eighty rules that cover the diverse contexts where letters are used.

The main rules that can be considered in our Genetic Programming system are: i) direct relation between letter and phones, searching the special occurrence of letters "t" and "d" before "i"; ii) letter "c" when used before letters "e" or "i" is represented by phoneme [s], in the other cases by [k] phoneme; iii) letter "s" represents phoneme [z] in two situations, when it occurs between two vowels or when it is used before voiced consonants ("d", "b", "m", "g", "n"); iv) the use of "ss" results in the production of only one phoneme [s]; v) letter "z" correspond to phoneme [s] when there is a vowel at the end of the word. In the cases where it's followed by a vowel, the phoneme corresponds to the letter itself.

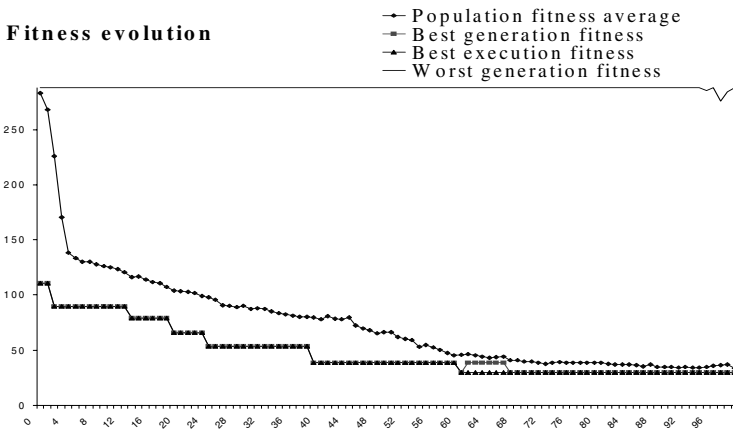


Fig. 1. Graphic of fitness evolution

The standard fitness of the best individual was found at generation 60, corresponding to the value 30. Optimized solutions tend to lower values. A "perfect" solution corresponds to 0, since in our modelling we have used as fitness measure the difference between the expected value of the pronunciation of a string of words (correct one by definition) and the obtained fitness value through the application of rules described above. In Fig. 1, we show the fitness evolution to one of the tested cases in the work.

### 4 Final Conclusions

This article has presented a technique for the automatic discovery of programs and also has showed the results of its application in the automatic obtention of rules to convert a letter into a phoneme to Brazilian Portuguese. The realized experiments

have demonstrated that it is possible to construct systems able to discover rules through a supervised learning technique using Genetic Programming.

The research was developed in the context of the SPOLTECH project (International cooperation between Brazil (UFRGS) and USA (University of Colorado), adding a specific tool for Portuguese phonetic rules and using Genetic Programming approach. The easiness offered by the representation of the technique which uses common instructions in a programming language offers a flexibility and easiness for the approach of different problems in speech synthesis.

One of the major difficulties found in the results analysis consists in describing properly the implemented activity by each solution. This comes directly from the increase of complexity of the solutions and also the number of instructions that constitute the solution individuals. Another important issue is the definition of a set of fitness cases which can represent properly the rules to be discovered, without compromising other contexts where some letters can belong to.

## References

1. Banzhaf, Wolfgang; Nordin, Peter; Keller Robert; Francone, F.D. Genetic Programming An Introduction. On the Automatic Evolution of Computer Programs and Its Applications. San Francisco: Morgan Kaufmann, 1998. 470 p.
2. Spoltech. Advancing Human Language Technology in Brazil and the United States Through Collaborative Research on Portuguese Spoken Language Systems. In: PROTECC, 4., 2001, Rio de Janeiro. Projects Evaluation Workshop: international cooperation: proceedings. Brasília: CNPw, 2001. p. 118–142.
3. Branko Soucek; Iris Group. Dynamic, Genetic and Chaotic Programming. New York: John Wiley & Sons, 1992.
4. Dutoit, Thierry. An Introduction to Text-To-Speech Synthesis. Dordrecht: Kluwer Academic, 1996. 280 p.
5. Hausser, Roland. Foundations of Computational Linguistics. Man-Machine Communication in Natural Language. Berlin: Springer-Verlag, 1999. 534 p.
6. Koza, John R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge: The MIT Press, 1992. 819 p.
7. Koza, John R. Genetic Programming II: Automatic Discovery of Reusable Programs. Cambridge: The MIT Press, 1998. 746 p.
8. Lemmetty, Sami. Review of speech synthesis technology. Disponível em: <<http://www.acoustics.hut.fi/~slemmet/dippa/chap1.htm>>. Acesso em 20 dez. 2000.
9. Mitchell, Melanie. An Introduction to Genetic Algorithms. Cambridge: The MIT Press, 1992. 205 p.
10. Silva, T.C. Fonética e Fonologia do Português, roteiro de estudos e guia de exercícios. São Paulo: Contexto, 1999. 254 p.
11. Silva, Miriam B. Ensaios. Leitura, Ortografia e Fonologia. São Paulo: Ática, 1993. 110 p.
12. Festival. The Festival Speech System. System Documentation. Disponível em: <[http://www.cstr.ed.ac.uk/projects/festival/manual/festival\\_toc.html](http://www.cstr.ed.ac.uk/projects/festival/manual/festival_toc.html)>. Acesso em 06 fev 2003.