# DEVELOPING A VOICED INFORMATION RETRIEVAL SYSTEM FOR THE PORTUGUESE LANGUAGE CAPABLE TO HANDLE BOTH BRAZILIAN AND PORTUGUESE SPOKEN VERSIONS

*M.N. Souza*, E.J. Caprini*, C.G. Machado*, M.V. Ludolf*, L.P. Calôba*, J.M. Seixas*, F.G. Resende*, S.L. Netto*, D.R. Freitas**, J.P. Teixeira**, C. Espain**, V. Pera**, F. Moreira***

* Universidade Federal do Rio de Janeiro, COPPE/EE

** Universidade do Porto, FEUP, CEFAT

## ABSTRACT

The two versions of the Portuguese language, the one spoken in Portugal and the one spoken in Brazil differ both phonetically and lexically. The paper reports on a menu driven voiced information retrieval system, BomdePapo, which can handle both versions of the language. The system was developed in Delphi during a research project joining teams from the Universidade do Porto, Portugal, and the Universidade Federal do Rio de Janeiro, Brazil. Information retrieval is processed through a tree menu. The system, BomdePapo, is currently being tested on an application for accessing information on sports, culture and politics news. The paper also reports on the main phonetic differences between the two versions.

## 1. INTRODUCTION

The common language is a safe ground for work to be done by joint teams from both countries. The two teams now submitting this paper were joint by a research program ruled by scientific national institutions in both countries, CAPES in Brazil and ICCTI in Portugal.

The program, started 1996, saw some changes along its three years. From the goal of building an information retrieval system to be used over the telephone line across the Atlantic, it turned to simpler and more effective objectives, also information retrieval systems working with simpler databases, both in organization and size, over the telephone or not.

The Brazilian team has been producing work in cryptanalysis of speech signals [1, 2], speaker verification [7] and auditory based time frequency distributions applied to speech signals [9]. The Portuguese team worked in algorithm acceleration [3, 4, 6] and voice control of computers [5, 8].

Another important issue was making the system usable by both communities. This is still not entirely done but an effort was made in studying phonetic conversion rules that allow one of the TTS to emulate the other version of the language with reasonable success. There are also experiences to be made concerning the recognition ratio changes when these rules are applied to the spoken digits.

The system BomdePapo is shown in Fig.1. Section 2 describes the system functioning and Section 3 details some of the investigation going on around the system.

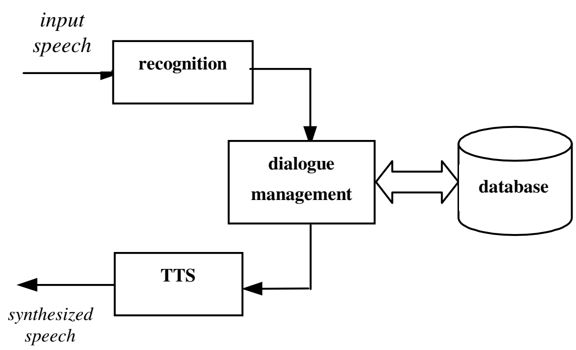


*Fig.1 BomdePapo*

## 2. SYSTEM

### 2.1 Speech Processing

Speech was first pre-emphasized using a filter of the form $(1\text{-}0.97z^{-1})$. The signal was then segmented applying a 23 ms Hamming window once every 9 ms. Mel-frequency cepstral coefficients and a normalized energy term were generated. For each frame, a set of 22 mel-scaled triangular bandpass filters was applied to the short-time power spectrum obtained by a 512-point FFT. Using a DCT, 12 MFCC were computed and then liftered. First-order weighted derivatives were used, resulting in a 26-dimensional feature vector.

## 2.2 Text-to-Speech Converters

In the context of PAPO two text-to-speech converters (TTS) were used. The brazilian TTS system was specifically designed and developped and is based on a time-domain diphone concatenation system [10]. The portuguese TTS is the MULTIVOX-Porto system that was developped in 1996/1997 in the scope of another project [11]. It is a formant coded speech synthesiser.

Both systems handle prosodic features at a basic level.

## 2.3 Phonetic Transcription Rules

A careful study of the grapheme-to-phoneme conversion rules in the two versions of the portuguese language was done, resulting in a set of notes regarding the identification of important differences between the two language versions. First of all the basic phonemes lists were compared resulting in the identification of the common and the specific phonemes sub-sets. Secondly the phonological grapheme-to-phoneme conversion rules were listed and compared. A list of phonetic transcription rules was produced. This will allow conversion from one TTS to the other. This will also be the starting point for building a module that will detect the version the user speaks, by phoneme frequency determination, and trigger the correct TTS. The following list presents some examples (SAMPA):

2.3.1. Non-tonic vowels are not reduced or little reduced in Brazil:

Ex: p[a]rtir, m[o]rar, l[e]var;

but in Portugal these are generally quite reduced

Ex: p[6]rtir, m[u]rar ,l[]var;

2.3.2. /t/ and /d/ suffer palatalization when ocurring before tonic or atonic /I/ and postonic /e/ in Brazil:

Ex: [tS]io, [dZ]irector, ba[tS]i

But in Portugal no palatalization occurs in these cases:

Ex: [t]io, [d]irector, ba[t]i

2.3.3. Supression or velarization of the /r/ in end position, in Brazil

Ex: senh[o], faz[e], am[a]

In Portugal the simple /r/ is kept when in end position:

Ex: senh[oR], faz[eR], am[aR]

2.3.4. In Brazil a semi-vocalization of the final /l/ ending a syllable and a word is usual:

Ex: anim[aw], Brasi[u], sa[u]tar

Differently the /l/ is velarized at the end of a syllable or a word, in Portugal:

Ex: anima[l], Brasi[l], sa[l]tar

2.3.5. In Brazil a non- palatalization of the sibilants at the end of syllable or word is common, except for the Rio de Janeiro speaking style:

Ex: me[z]mo, menino[s]

But in Portugal this palatalization of the final sibilants of syllables or words practically doesn't occur.

Ex: me[Z]mo, menino[S]

2.3.6. In Brazil there is an introduction of [i] between two consonants that usually don't make up a group in Brazilian Portuguese:

Ex: cap[i]tura, ab[i]surdo,p[i]neu

However, in Portugal there is a conservation of the consonant group

Ex: captura, absurdo, pneu.

Having implemented the appropriate mappings in the MULTIVOX-Porto system the produced speech could easily be recognised as a brazilian speech by the listeners that informally tested the system. This very important preliminary result inspired a whole task that is under way of execution regarding the implementation of the lacking phonemes and rules in each of the two systems in order to make them practically capable of speaking the two versions of the portuguese. This work will later be extended to the prosodic aspects as well.

### 2.4 Recognition

A single discrete HMM recognizer was also designed for isolated digits. Ten 5-states left-to-right models for each version were trained over a 256 points codebook. The Portuguese models were trained with a database collected at the Universidade de Coimbra, Telefala, at 8 kHz, and the Brazilian models with a database collected at UFRJ, at 11 kHz. For training the Portuguese models 378 utterances per digit were used, while for the Brazilian models 114 were used.

### 2.5 Dialogue

Initially the user is asked to be silent for 1.5 seconds. This is needed to give the system time to evaluate the level of background noise.

Then a first question is posed by the TTS with three options: *Brazil, Portugal and quit*. The user must answer 1, 2 or 3.

The recognition module decodes this answer and the dialogue manager accordingly moves into the next question.

Let's for instance imagine the user was interested in football in Brazil. He would have then answered 1 to the first question. Suppose the recognition system understood it well. Next question would then be: *What are you interested in about Brazil: sports (1), culture (2), politics (3), previous question (4)?* To which the answer would be 1.

Suppose now the system understood 2. The user would have been asked the same question concerning Portugal. He knows then the system misunderstood him and answers 4 (*previous question*) to the second question. This gives him a second chance to answer the first question.

The desired information will be given at one end of the tree.

## 3. CONCLUSIONS

This BomdePapo we have developed is just a first prototype of a dialogue system involving the two versions of the language. Much remains to be done.

First of all, testing the performance of the system trained with a database collected in one country when users belong to the other. The evaluation parameter might be the rate of success in retrieving a piece of information from the end of the information tree.

Then it would be interesting to use the phonetic information that was collected to trigger the correct TTS and the correct recognizer, using the correct database. This is in fact a problem of accent identification.

## REFERENCES


[1] Apolinário Jr., J. A., Mendonça, P. S. R., Calôba, L. P., "Criptoanalise de Sinais de Voz Cifrados por TSP Usando Redes Neurais e Mean Field Annealing", XIV Simpósio Brasileiro de Telecomunicações, pp 515-520, Curitiba, 1996.

[2] Apolinário Jr., J.A., Mendonça, P.S.R., Chaves, R. D., Calôba, L.P., "Cryptanalysis of Speech Signals Ciphered by TSP Using Annealed Hopfield Neural Networks and Genetic Algorithms", Proc. 39th Midwest Symp. on Circuits and Systems, pp 821-824, Ames, 1996.

[3] Araújo, A.J., Pera, V., Espain, C., Matos, J.S., "A Vector Quantizer Architecture for an Automatic Speech Recognizer.", XIII Conference on Design of Circuits and Integrated Systems, Madrid, 1998.

[4] Araújo, A.J., Pera, V., Souza, M.N., "A Vector Quantizer Accelerator for an ASR Application", ICSLP'98, Sidney, 1998

[5] Boticario, L., Sá, L., Pinho, J., Espain, C., Freitas, D., "Controlo de um computador utilizando algoritmos de reconhecimento de voz", I Conferência Nacional de Telecomunicações, Aveiro, 1997.

[6] Pera, V., Espain, C., "Aceleração do algoritmo de backpropagation", III Congresso Brasileiro de Redes Neurais, Florianopolis, 1997.

[7] Pinto, H.L.C.P., Pinto, R.G.C.P., Calôba, L.P., "A Speaker Verification Method Using LPC Singularities Location", Proc. IEEE Intern. Telecommunic. Symp., Acapulco, pp 39-43, 1996.

[8] Santos, D., Costa, J., Vasconcelos, R., Rodrigues, S., Freitas D., Espain, C., "COMVOZ - DSP Speech Control of a PC-Based Virtual Instrument System", First

European DSP education and research conference, Paris, 1996.

[9] Souza, M.N., Calôba, L.P., "A Comparison Between Fourier and Biological Auditory Based Time Frequency Distributions Applied to Speech Signals", Proc. 39th Midwest Symp. on Circuits and Systems, pp 807-810, Ames, 1996.

[10] Souza, M.N., Calôba, L.P., Seixas, J.M., Machado, C.G., Ludolf, M.V., "Sintetizador de Voz do Projeto "Processador Automático de Português", XII Congresso Brasileiro de Automática, Uberlândia, 1998.

[11] Teixeira, J.P., Freitas, D., Gouveia, P., Olaszy, G., Németh, G., "Multivox - Conversor Texto-Fala para o Português", III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada, Porto Alegre, 1998.