A set of Phonological Rules for Mexican Spanish

Esmeralda Uraga,
Luis Alberto Pineda Cortés

In this paper a set of phonological rules for grapheme to phones conversion for Mexican Spanish is presented. These rules permit the automatic generation of phonetic transcriptions and phonetic pronunciation models out of text corpora. A phonetic alphabet called Mexbet, which is based on Worldbet, including representations for all Mexican phones is proposed. The phonological rules cover most phonetic types of contexts of Mexican Spanish and map orthographic representations into sequences of Mexbet symbols. The rules have been used for the creation of speech corpora for training acoustic-phonetic models using neural networks, hidden Markov models and hybrid approaches. The speech recognition results and a simple application are also presented**.**

1 EXTENDED ABSTRACT

An essential component of a speech recognition system is the acoustic model. Acoustic models have been created using Neural Networks, Hidden Markov Models or through hybrid approaches. All these models are produced out of a speech corpus by automatic training techniques in which speech signal segments, represented through a number of parameters, are correlated with their corresponding linguistic units, like phonemes or phones. For the creation of reliable and robust models large amounts of data are required. Traditionally phonetic units associated to speech segments are labeled by hand by human experts; however, this is an expensive and time consuming task. In particular, there are no available labeled corpora for Mexican Spanish with the size and quality that the task demands, and the human and material resources to create such corpora by hand are not forthcoming. For this reason it is highly desirable to obtain speech data out of text, in the same way that human speakers are able to tell which sounds correspond to sequences of characters when reading. A facility to interpret text and produce the corresponding phonetic units is useful to train the acoustic-phonetic models, diminishing considerably the human effort invested in the task. The rules to map textual representation (graphemes) to phonetic symbols underlying such facility will be referred to as *phonological rules*. The main of purpose of this paper is to present a set of phonological rules for Mexican Spanish, facilitating greatly the creation of acoustic models for this language.

With these rules, for instance, human subjects can read a predefined text corpus to obtain the speech data, while the phonetic representation can be obtained directly through the phonological rules, providing the information required to train the acoustic-phonetic model.

The phonological rules are also useful for the creation of phonetic dictionaries, as standard dictionary inputs can be processed with these rules to produce the pronunciation models of words in a systematic fashion. This is helpful because in speech recognition acoustic models recover the most likely phonetic units which correspond to the input speech and, to recover the words pronounced by the speaker in most systems, the output of the acoustic model (i.e., the sequence of phonetic symbols) must be matched against the entries of a phonetic dictionary (e.g., a list of words with the corresponding phonetic pronunciation models), in which all words that can be recognized are represented.

It is also important to consider that human pronunciations in spontaneous, conversational, speech is much more variable than in careful reading where the pronunciations of words is more likely to adhere to the standard pronunciation. Also, most phonemes can have allophonic variations depending on the different ways of speaking of communities (i.e., regional accents) and contextual inter and intra word factors (i.e., in Spanish the phoneme /d/ has two allophones, the interdental and the dental-palatal, which can occur freely in any context). Most speech recognition systems, however, rely on pronouncing dictionaries which contain few or none alternative pronunciation models for most words and this limitation is a significant cause for the relatively poor performance of recognition systems for large vocabulary conversational speech recognition (lvcsr) tasks. Interestingly enough, as allophones can be known through empirical investigation, once the standard pronunciation model of a word is known the set of possible allophonic variants of the phonemes contained in the word can be obtained by a second set of rules, and robust pronunciation models can be produced in an automatic fashion. Phonological rules have also been used for a variety of purposes including pronunciation modeling in speech recognition, phonetic dictionaries generation, word or name lookups for database searches and speech synthesis.

In this project the phonological rules were applied to labeling the training corpus. Temporal alignment of labels was performed manually. The rules were also applied to the recognition vocabulary in order to create the required pronunciation models. The speech corpus was used to train acoustic-phonetic models using two approaches based on neural networks (NN) and hidden Markov models (HMM). Then, forced alignment (automatic generation of a phonetic transcription using a previously trained acoustic model) was performed with the best acoustic model based on neural networks. The acoustic-phonetic models were also retrained with a phonetic transcription of the training data obtained using a hybrid NN-HMM-model. Recognition using this acoustic model and the pronun-

ciation models resulted in an increment of 7% in the recognition performance at sentence level, which is partly attributable to some changes in the acoustic training procedure (using a hybrid model), and partly to improved training transcriptions and the pronunciation models (using the phonological rules).

Finally, we created a robust spoken language system that integrates a human-machine interface with speech recognition, spoken language generation, spatial deictic events manager, input and output text, and a database for information retrieval in the call phone domain.

Esmeralda Uraga is a member of the Academic Staff of the Department of Computer Science at the Institute for Applied Mathematics and Systems (IIMAS) of the National University of México (UNAM), AP. 20-726 ADMON. No. 20. Del. Alvaro Obregon 01000 Mexico, D.F She can be reached at euraga@leibniz.iimas.unam.mx, see htpp://leibniz.iimas.unam.mx/~euraga.

Luis Pineda is a Titular Investigator and Head of Computer Science at the Institute for Applied Mathematics and Systems (IIMAS) of the National University of México (UNAM), AP. 20-726 ADMON. No. 20. Del. Alvaro Obregon 01000 Mexico, D.F. He is the author of more than 30 papers in computer graphics, intelligent CAD systems, knowledge representation, diagrammatic reasoning and computational linguistics. National Investigator (SNI) since 1993. He can be reached at luis@leibniz.iimas.unam.mx, see htpp://leibniz.iimas.unam.mx/~luis.

This work was done under partial support of Conacyt/NSF bilateral program for the development of computer science in a collaboration with the University of Rochester and ITESM, Campus Morelos, Conacyt grant 400316-5-C092A. We express special thanks to Dr. Enrique Sucar and Prof. James Allen for valuable discussions.