8th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS
Suceava, Romania, May 25 – 27, 2006

SPEECH ANALISYS AND MODELING FOR DEVELOPMENT SPEECH RECOGNITION SYSTEM

Georghe CAPATANA, Nicolae OBJELEAN
State University of Moldova
str. Mateevici 60 MD-2009
Gh_Capatana@yahoo.com, objelean@yahoo.com

Abstract. *This work deals with the series of new experiments to use existing speech tools and speech resources for research the characteristics of East European Spoken Language for building speech interfaces. Was used the open existing speech resources and speech tools.*
Keywords: *Language modeling, Speech Recognition, Signal Processing, Tools.*

Introduction

Language and speech dialogue is the fundamental means of human communication. Today peoples want to use the computers just as watches, phones, TV, therefore interest to the problem speech dialogue between man and computers have increasing. The first simple automatic speech recognition systems already have application in many fields [1]. Speech recognition technology for many languages is already well developed and has found application in many areas such as control systems, automatically call centers in assisting physically handicapped people and etc[2].

The action of an automatic speech recognition system has often been likened to that of a voice "typewriter" which automatically transcribes spoken words into written text. The development of such systems combines techniques from many fields, ranging from pattern recognition and information theory to linguistics and psychology and request using many informational and linguistically resources.

This work deals with the series of new experiments to use existing speech tools and speech resources for research the characteristics of East European Spoken Language and building speech interfaces. For this goal has used the open speech resources and speech tools [3] of CMU Sphinx Speech.

Pronunciation modeling framework

We using methodology developed by Riley and Ljolje of the decision-tree-based pronunciation modeling describing in [4]. We adapt the techniques they developed for read speech to our speech task. The main steps in training and using a pronunciation model for ASR, as shown in the block diagrams of Figure 1, are to:

1. **Obtain a canonical (phonemic) transcription** of some training material. A standard pronouncing dictionary is used for this purpose, with Viterbi alignment when there are multiple pronunciations.
2. **Obtain a surface-form (phonetic) transcription** of the same material.
3. **Align the phonemic and phonetic transcriptions**. A dynamic programming procedure based on phonetic feature distances is used for this purpose.
4. **Estimate a decision-tree pronunciation model**. A decision tree is constructed to predict the surface form of each phoneme by asking questions about its phonemic context.
5. **Perform recognition with this pronunciation model**. A dictionary-based phoneme level recognition network is first obtained from a word lattice generated by an initial recognition pass. The pronunciation model is then used to transformer phoneme sequences in this network to yield a network of surface-form realizations.
6. **Full training set retranscription**. Starting with the canonical transcription of the entire

acoustic training set, the pronunciation model is used to create pronunciation networks represent.

Training data for Language modeling and Speech analysis.

The training and testing data contained two vocabulary alphabetic + digital vocabulary and vocabulary frequently users conversation for traveling and contain about 1300 words. In experiments took part only one-man speaker. The speaker are reading six-time list of sentences that contain words of vocabulary in a natural condition. The first dictionaries contain all letters of Romanian and Russian languages and number 0-9. These dictionary is not very large, but it's very important for building automatic speech recognition systems, because in first it's contain all sounds and phonemes spoken language and in second time this dictionary can be used on many speech technology application. In additionally this vocabulary may be used for investigating the phonological characteristics of language, which very necessary and frequently used for building speech synthesizes systems. The alphabetic + digital vocabulary can be used as etalon also for testing of capacity speech recognition systems which will be developed different firms and companies.

The acoustic and pronunciations vocabulary models was building for words of Romanian and Russian languages and, they can to use for extending vocabulary and removing the language for new tasks. The acoustic models was using for search of the speech sounds – phonemes, and the words pronunciations giving in terms of the phonemes from the phonemes set.

Far the decoder model was used, that includes the knowledge of each the pronunciation of words with a sequence of phones. Acoustics models keep information about sound changes over time and as phonemes are distributed into words. The SphinxTrain was used for training acoustic models in normal environmental condition.

These data was recorded directly of a microphone an 8 kHz sampling rate was used and collected on Speech Database. This database can be used in order to create a training set of acoustical signals for the development automatic speech recognition and for other goals as speakers recognition, speech corpus creation, spectral analyses of speech signals and speech synthesize.

Automatic speech recognition system was build with limited vocabulary in results. This system can work in regimes training - learning new words from new speakers and recognition words in dialogue. Therefore the building system is adaptable for new dictionary and new tasks and can be use data input for in computer by voice.

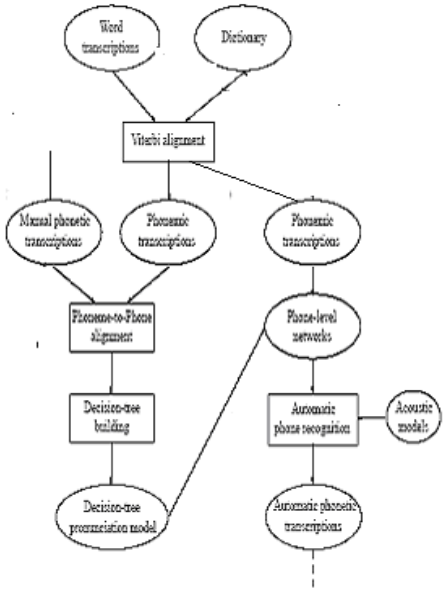

Figure 1. Pronunciation model training.

References

[1] ***(1997) *Spheech Recognition. The future now*! Prentice Hall PTr.

[2] Newell A. at all (1973). *Speech nderstanding System, Final Rapport*. Amstradam .

[3] *The CMU Sphinx Group Open Source Speech Recognition Engines*. www.cs.cmu.edu.

[4] Riley, M. & Ljolje, A. (1995),. *Automatic generation of detailed pronunciation lexicons*. *Automatic Speech and Speaker Recognition : Advanced Topics*, chapter 12, pp. 285–302. Kluwer Academic Press.