The Design and Implementation of HMM-based Dai Speech Synthesis

Jian Yang, Zhan Wang, Xin Yang
Yunnan University

Abstract

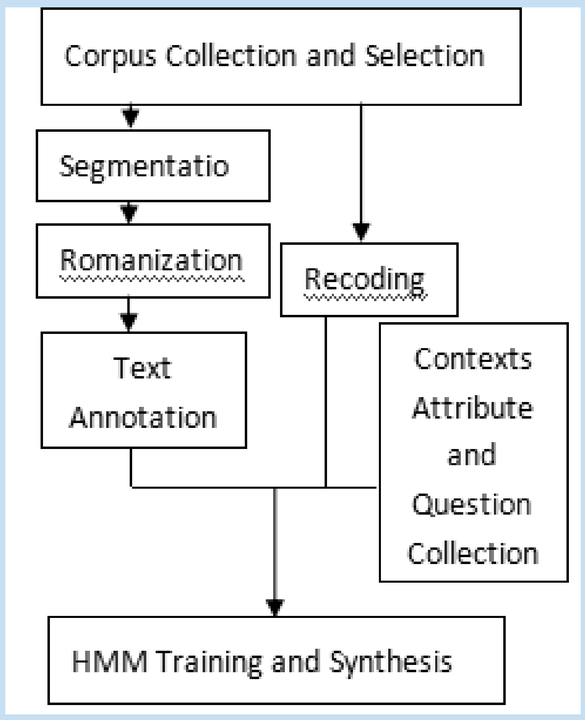
By far there are more than 1.2 million Dai compatriots using Dai language in Yunnan province, researching Dai speech synthesis has great significance in advancing the informationization of Dai.This paper researches the implementation of Dai speech synthesis by taking the HMM speech synthesis framework and STRAIGHT synthesizer into account.

In this paper, collection and selection of Dai text corpus, recording of speech corpus, text normalization, segmentation, Romanization and the implementation of acoustic?model training are described.

Introduction

Dai is a monosyllable, tonal language, there are 91 vowels in Dai, its 42 consonants are divided into high and low two groups, and it has 9 tones. The characteristic of its syllables is C+V+C or C+V, and only -pčň-tčň-kčň-?čň-mčň-n can be the tail consonant

In this paper, the design and implementation of the Dai speech synthesis is based on the basic framework of HMM speech synthesis. Figure 1 shows the basic block diagram of Dai speech synthesis.



Corpus database building

1.Collection and Selection of Text Corpus

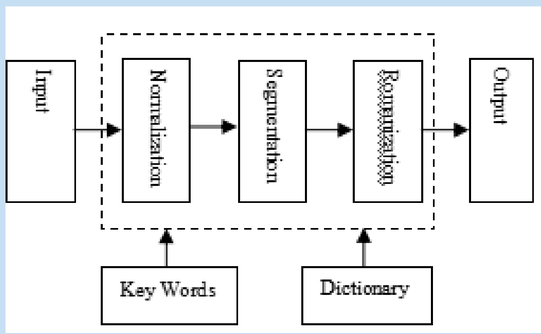
This paper used Teleport Ultra to download and organize the data from the Xishuangbanna news website and built a 60MB Dai text corpus database. As for the collected corpus, we take the maximum of the syllable coverage as the Selection criteria and write a program to select pronunciation corpus,we finally constructed a 6.0MB pronunciation corpus; its full syllable coverage rate is 94.2%, which can cover most of the legal syllables. When the confidence value is 98%, the similarity of text corpus and pronunciation corpus is 0.4.

2.Recording of speech corpus

With the use of digital recording, we finally obtain a text corpus database of 1204 sentences and a pronunciation corpus database of 28.2h, statement number:0001 1204,using CD format to store the original voice we obtain 6 CDs.

Text Analysis and Processing

The flow chart of text analysis is shown in figure 1:



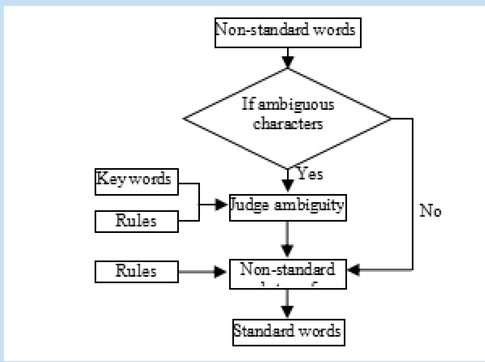


Figure 1: figure 1:Text analysis process (left);figure 2:The process of normalization (right)

1.Text Normalization

In general, the process of normalization is divided into four steps: recognize non-standard words, judge ambiguity, eliminate ambiguity, and transform non-standard words into standard word. The flow chart is shown in figure 2:

2.Words Segmentation

MMSEG is a Chinese word segmentation method based on dictionary,in order to use MMSEG to segment Dai sentence, we propose a mapping between Dai and Chinese, then segment Chinese text by using the tool of MMSEG, lastly mapping the Chinese to Dai language.

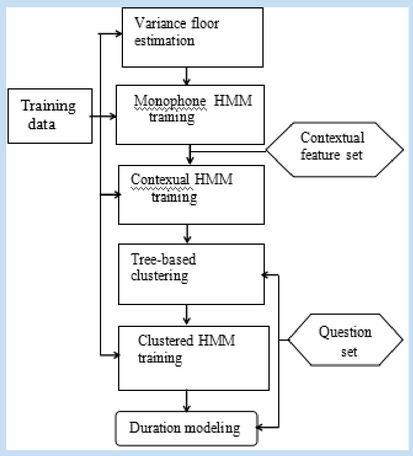
3.Romanization

In the Dai dictionary, many Dai vocabularies are come from Chinese vocabularies, these kinds of words are called Chinese Loanwords. The pronunciation of Chinese loanwords, some are in accordance with Chinese pronunciation, so we need to increase some Chinese phoneme in the Romanization scheme.

Acoustic Model Training

Before training,we use context triphone models to describe the context dependent models, and describe the context attribute through pronunciation information and prosodic hierarchy information.

In this project, we take vowels and consonants as the model unit, use SP to model the silence section. The frame shift of extracting training speech data is 5ms, the final spectrum and fundamental frequency feature not only contains the static parameters, but also contains one order and two order difference parameters. As for spectrum and fundamental frequency characteristics, we use 5-state left-to-right HMMs, and Gauss distribution is used to represent the length of the phoneme. In the training process of the context dependent statistical model, we use the decision tree to solve the problem of sparse data.The training process of the whole acoustic model is shown in Figure:



Experimental results and analysis

An example for original and synthesized speech spectrum of a sentences within the training set is shown in Figure4, 5.

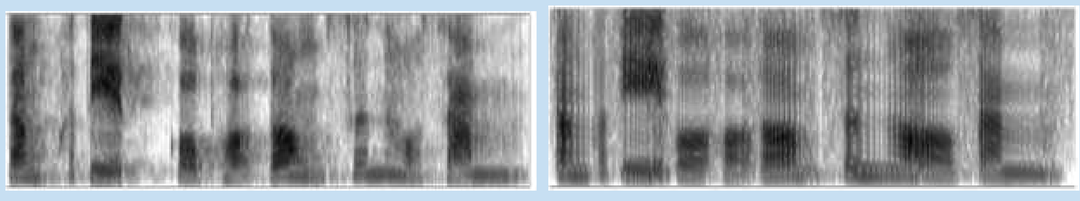


Figure 2: figure 4:Original spectrum (left);figure 5:Synthetic spectrum (right)

We can see, the resonance peaks of the two sentences are basically the same, but the prosodic information and fundamental frequency information are not consistent.

This paper researches the implementation of Dai speech synthesis by taking the HMM speech synthesis framework and STRAIGHT synthesizer into account. We focus on the process of the whole speech synthesis framework, mainly including: corpus building, text analysis and acoustic model training, then under the Cygwin platform do HMM training and synthesis by using HTS synthesis tools.After manual evaluation, the Intelligibility degree of synthesized speech can reach to 100%, but the natural degree of synthesized speech needs to be improved. Our future word will be modify and adjust the text annotation file, make the label text more accurate, and in order to produce a higher quality voice, we will process the boundary noise of wave file.