Study on the Text Analysis Technology of Tibetan Text to Speech *

Cai Rang Zhuo Ma, Cai Zhi Jie

Tibetan Intellectual Information Processing Centre of Qinghai Normal University Xining, Qinghai, China

Corresponding author's Email: Cr-zhuoma@163.com

Abstract: Text to speech system(TTS) ,which converts text into speech through computer and then output the speech, is a popular research object in the realm of speech processing in recent years. A successful TTS outputs speech with clear timbre and natural fluency. TTS is composed of Text analysis, prosody control and speech synthesis modules. Text analysis module is in the front end, which gives the linguistics and phonetics text detailed information for the speech synthesis. This paper, based on the Tibetan written, pronunciation characteristics and pronunciation mechanism, gives the basic structure of TTS that is based on large-scale speech corpus and study of standardization, word segmentation and voice tagging technologies for Tibetan TTS.

Keywords: Text to speech; Standardization; Word segmentation; Tagging introduction

1. Introduction

Text to speech (TTS) technology, which is also known as speech synthesis, transfers the text into phonological sequence, then generates speech waveform with speech synthesizer, and finally forms a continuous voice to output. In addition to human computer interaction, TTS technology also has broad application in automatic control, monitoring and control communications systems, office automation, information management systems, intelligent robotics and other fields.

The research for TTS technology was over 200 years at home and abroad, but the practical TTS technology just emerged with the development of computer technology and digital signal processing technology. In the history of TTS technology, the synthesis parameters were mainly used in the early years, but in the late eighties, TTS technology had made new progress, especially PSOLA method proposed (1990), the timbre and naturalness was much better. In recent years,

a new method of TTS-based Corpus TTS technology has caused the concerns of scholars. In the Corpus-based TTS method, since the whole speech unit is derived from original pronunciation as well as the speech corpus including all kinds of context speech units and it is rich enough, you can synthesize statement with high definition and high nature.

There are about 50 million people speaking Tibetan in China. However, most of them live in the remote and inaccessible area, and among them 85% Tibetan can only speak but cannot read and write. Therefore, studying and applying Tibetan TTS technology into education, telecommunications, and finance, will not only improve the life quality of Tibetan but also promote the cultural exchanges and national unity. But the research of TTS technology of Tibetan has just started in recent years, and most of the related work is still in the initial stage. The writing structure and spelling model of Tibetan words are wildly different from any other language in the world, such as English, Japanese and even any minority language in China. According to the written text, the feature and principle of pronunciation of Tibetan, this paper gives the

basic structure of Tibetan TTS system, text analysis and processing, and the design of large-scale speech corpus.

2.   The Structure of Tibetan TTS System

Tibetan is an alphabetic language evolved from the Sanskrit characters which has 30 consonant characters, 4 vowel characters. The basic unit of Tibetan is syllable. There is a set of strict and complete rules of combination and arrangement of characters for Tibetan. Its string of characters is 2-dimensional, written from left to right, however, the characters are not totally linearly arranged, part of compositional initial consonant is presented in a top-down manner. The traditional Tibetan grammar divides the characters into root letters, head letters, under letters, former letters, later letters and backmost letters, as is shown in figure 1.

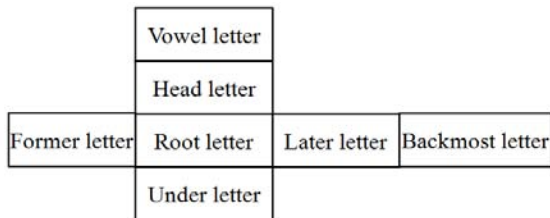|	Vowel letter		
	Head letter		
Former letter	Root letter	Later letter	Backmost letter
	Under letter		

Figure 1 The structure of Tibetan characters

According to the spelling strategy of Tibetan, former letters are put in front of the root letters, head letters are on the top of the root letters, under letters are below the root letters, and backmost letters are behind the later letters. For example: 搬帛罢邦 or 搬舶搬邦.

Although Tibetan is an alphabetic writing character, its structure is more complicated than English, since it can be spelled vertically and horizontally. Besides, there are blanks between words in English, but the Tibetan words are adjoined together without any spaces. Therefore, the research of Tibetan TTS needs word segmentation just like Chinese. In addition, the spell order of Tibetan is different from other languages, as is shown in figure 2.

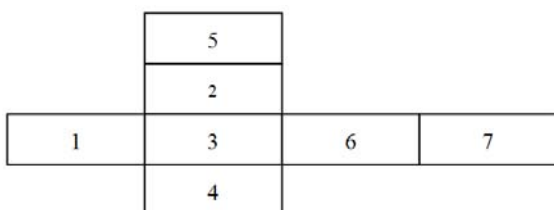|	5		
	2		
1	3	6	7
	4		

Figure 2 The spell order of Tibetan characters

Considering Tibetan is different from other languages in language type, character structure and syntactic structure, the basic structure of Tibetan TTS is designed specifically as is shown in figure 3.
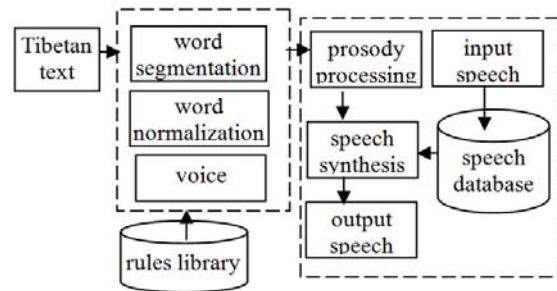

Figure 3 The procedure of Tibetan TTS

As is shown above, according to the rules in semantic, lexical and speech levels and content of the text, we first transfer the text into the sequence of rhyme, and generate the voice waves with respect to the sequence. The first step is related to linguistic processing and the second step requires advanced TTS technology thus to synthesize voice stream with high quality.

3.   Text Analysis and Processing

Text analysis and processing is the vital part of TTS system, which aims to enable the computer to understand the input text, to provide essential information for TTS. As for Tibetan, text analysis includes word segmentation, transformation from special symbols and numbers to proper tags, and analyzing the length of pronunciation of syllables, and the level and severity of a standstill position and length of pause.

As is shown in the figure 4. Through the processing module, the characters string is converted into argument code string with the feature of pronunciation. There are 3 steps in the text analysis and processing: text normalization, text segment and voice tagging.
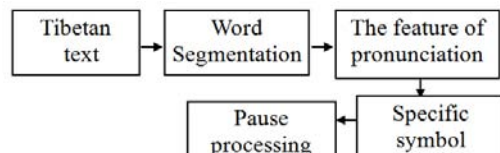

Figure 4 The processing analyzes flow of Tibetan text

3.1   Text normalization

The text in reality contains a lot of non-standard Tibetan characters. These words cannot be found in

dictionaries, and their pronunciation would not be obtained from the common rules. By analyzing the context of the text to implement text normalization and by recognizing the non-Tibetan characters to transfer standard ones, we can present the examples in table 1:

གྲོང་ཁྱེར་ཀ་གེ་ལོའི་སྒྲོག་འཕྲེན་རྩུ་�d་ཡིས 2002 ལོའི་ཟླ 5 ཚེས 30 ནས་

བཟུང་ཚེ་གཏོང་སྒྲོག་སྦྱད་པའི་ཆེད་ལྟ་སྟེ། ཉིན་རེའི་དུས་ཚོད 8:00-22:00 བར་

ཚེ་སྒྲོག་རིན་གོང་ཡིན་ལ་སྒྲོག་བཀོལ་ན 1kwh སྒོར 0.56 ཡིན། ཚེ་སྒྲོག་རིན་གོང་

ནི་གཏོང་སྒྲོག་རིན་གོང་ལས 50%ཡི་མཐོ།

Table 1 The comparison between original and normalized text

No.	Original text	Normalized text
1	2002 ལོའི་ཟླ 5 ཚེས 30	ཉེར་སྟོང་གཉིས་ལོའི་ཟླ་ལྔ་པའི་ཚེས་གསུམ་བཅུ་ཉིན
2	-	...ནས་...བར
3	8:00-22:00	ལྟ་དྲུའི་དུས་ཚོད་བཅུད་ནས་ཕྱི་དྲུའི་དུས་ཚོད་བཅུ་བར
4	0.56	ཤིག་དང་གྱངས་ལྔ་དྲུག
5	1kwh	སྒོང་ཨ་རྒྱ་ཚོད་རེར
6	50%	བརྒྱ་ཆའི་ལྔ་བཅུ

The normalized text as follows:

གྲོང་ཁྱེར་ཀ་གེ་ལོའི་སྒྲོག་འཕྲེན་རྩུ་d་ཡིས་ཉེར་སྟོང་གཉིས་ལོའི་ཟླ་ལྔ་པའི་

ཚེས་གསུམ་བཅུ་ཉིན་ནས་བཟུང་ཚེ་གཏོང་སྒྲོག་སྦྱད་པའི་ཆེད་ལྟ་སྟེ། ཉིན་

རེའི་ལྟ་དྲུའི་དུས་ཚོད་བཅུད་ནས་ཕྱི་དྲུའི་དུས་ཚོད་བཅུ་བར་ཚེ་སྒྲོག་རིན་

གོང་ཡིན་ལ་སྒྲོག་བཀོལ་ན་སྒོང་ཨ་རྒྱ་ཚོད་རེར་སྒོར་ཤིག་དང་གྱངས་ལྔ་

དྲུག་ཡིན། ཚེ་སྒྲོག་རིན་གོང་ནི་གཏོང་སྒྲོག་རིན་གོང་ལས་བརྒྱ་ཆའི་ལྔ་བཅུ་

ཡི་མཐོ།

Text normalization is the vital and essential step in the text processing. After the processing of the model, the characters string can be transferred to argument code string with the feature of pronunciation, thus to provide the succession module.

3.2 Segmentation

The Tibetan words are adjoined together without spaces between words, therefore, the research of Tibetan TTS needs first segmentation just like Chinese. Through word segmentation, semantic analysis, speech unit splitting and tagging, the TTS finally synthesizes natural,

accurate voice. The word segmentation of Tibetan text needs three steps: splitting text into sentences, splitting sentence into chunks, splitting chunks into words. Its processing procedure is shown in Figure 5.
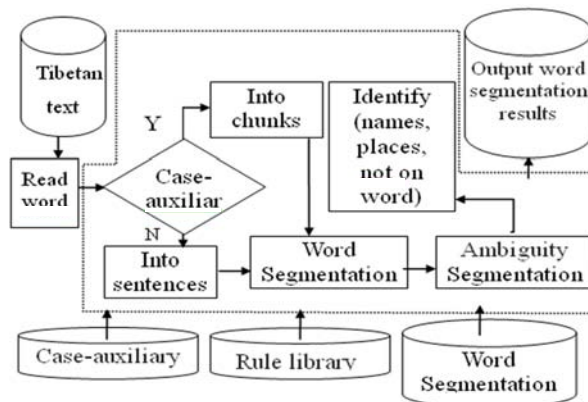

Figure 5 Word segmentation model

During the procedure of word segmentation, we first recognize some words with distinct features, such as name entities and special case particles, then split the original character sequence to smaller ones according to these words, and segment inside the small chunk. We take the method of maximum match to segment the chunk. The procedure of word segmentation (words are separated with "/": For example:

གྲོང་ཁྱེར་ཀ་གེ་ལོའི་སྒྲོག་འཕྲེན་རྩུ་d་ཡིས /ཉེར་སྟོང་གཉིས་ལོའི་ཟླ་ལྔ་

པའི་ཚེས་གསུམ་བཅུ་ཉིན་ནས་བཟུང་ཚེ་གཏོང་སྒྲོག་སྦྱད་པའི་ཆེད་ལྟ་སྟེ//

ཉིན་རེའི་ལྟ་དྲུའི་དུས་ཚོད་བཅུད་ནས་ཕྱི་དྲུའི་དུས་ཚོད་བཅུ་བར་ཚེ་སྒྲོག

རིན་གོང་ཡིན་ལ་སྒྲོག་བཀོལ་ན་སྒོང་ལ་རྒྱ་ཚོད་རེར་སྒོར་ཤིག་དང་གྱངས

རྒྱུང་ལྔ་དྲུག་ཡིན// ཚེ་སྒྲོག་རིན་གོང་ནི་གཏོང་སྒྲོག་རིན་གོང་ལས་བརྒྱ་ཆའི

ལྔ་བཅུ/ཡི/མཐོ།

Step1: Into chunks or sentences

We get seven chunks by Case-auxiliary words ཡིས , �q and three terminators (|), here present the details in table 2:

In this step, the first case-auxiliary word ཡིས splits the first sentence into three chunks, this sentence is ended by the first terminator (|), and the chunks are from No.1 to No.3; the second sentence split by the second terminator (|), and the chunk is No.4; the third case-auxiliary word ཨི splits the last sentence into three chunks, this sentence is also ended by terminator (|), and the chunks are from No.5 to No.7.

Table 2 The text chunks

No.	The text chunks
1	གྲོང་ཁྱེར་ག་གེ་མོའི་སྒྲོག་འཕྲིན་ཅུའུ
2	ཡིས
3	ཉེར་སྟོང་ག་ཉིས་རྣ་ལུ་བའི་ཚོས་གཟུགས་བཅུ་ཉིན་ནས་བཟུང་ཅེ་ག་གོང་སྒྲོག་སྦྱད་པའི་ཆོད་ལྷ་སྟེ།
4	ཉིན་རེའི་སྐུ་རྡོ་དུས་ཚོད་བཅུད་ནས་ཕྱི་དོའི་དུས་ཚོད་བཅུ་པར་ཆེ་སྒྲོག་རིན་གོང་ཡིན་ལ་སྒྲོག་བཀོའན་སྟོང་ལ་ཆུ་ཚོད་རེར་སྒྲོར་ཕྱེག་དང་གཅས་ཆུ་ལྷ་དྲུག་ཡིན།
5	ཅེ་སྒྲོག་རིན་གོང་ནི་ག་གོང་སྒྲོག་རིན་གོང་ལས་བཅུ་ཆའི་ལྷ་བཅུ
6	ཝི
7	ཟཧོ

Step2: Word segmentation

In this step, we use word segmentation algorithm as follows.

```
function seach1(s:string):integer;
  var
  f1:textfile;st:string[4];
  begin
  assignfile(f1,'.\continous word.txt');
          //special case-auxiliary file
      reset(f1);
    //special case-auxiliaries are:གེ་ཞི་ཇི་ཡི་གིས་ཀྱིས་ཀྱི་ལས་ཡིས
      result:=0;
    while not eof(f1) do
  begin
      readln(f1,st);
      if s=st then
  begin
        result:=1;
      exit;
      end;
      end;
end;
procedure fc();              //splitting
a:sz;m,mds:integer;
begin
    mds:=0;
    while (not eof(inputfile)) do
  begin
      m:=0; k:=0;
    mds:=mds+1;
    repeat
    begin
      m:=m+1;
      a[m]:='';
      for i:=1 to 4 do
      begin
```

```
read(inputfile,ch1);
k:=k+length(trim(ch1));
if (ch1<>'ཿ') and (ch1<>'ༀ') then
    a[m]:=a[m]+ch1;
else
    begin
      if ch1='ༀ' then
      begin
        m:=m+1; a[m]:=ch1;
      end;
      break;
      end;
    end;
  end;
  until (trim(a[m])='ༀ') or (k=ds[mds]);
  zwzgs:=m;
fengci(a);
if k=ds[mds] then    readln(inputfile,ch);
    end;
  end;
end.
procedure fengci(const a1:sz);
begin
m:=0;
for i:=0 to zwzgs do
  begin
  if (copy(a1[i],length(a1[i])-1,2)='འི') then
    begin
    for j:=m to i do
        s:=s+a1[j]+'ཿ';
    s:=copy(s,1,length(s)-2);
            //writing case-auxiliary into file
        m:=i+1;
    cazao2(s);
    end
  else
    begin
        if seach1(a1[i])=1 then
          begin
          for j:=m to i do
              s:=s+a1[j];
          s:=copy(s,1,length(s)-length(a1[i]));
          seach2(s); // writing case-auxiliary
          m:=i+1;
          end
        else
          if a1[i]='ༀ' then
            begin
            for j:=m to i-1 do
                s:=s+a1[j];
            s:=copy(s,1,length(s)-2);
            seach2(s); //writing Punctuation
          end;
      end;
  end;
```

Step3: Identification of abbreviated word

From the above Tibetan text, the highest frequency of the abbreviated word is འི , so we show the identify algorithm of འི as follows.

Table 3 The special words in word segmentation table

No.	word	remarks	No.	word	remarks
1	གྲོང་ཁྱེར		31	ཚེ	
2	ག་གེ་མོའི	abbreviated	32	སྒྲོག	
3	སྒྲོག་འཕྲིན་ཅུའུ		33	རིན་གོང	
4	ཡིས	case-auxiliary	34	ཡིན	
5	ནེར		35	ལ	
6	སྟོང		36	སྒྲོག	
7	གཉིས		37	བཀོལ	
8	མོའི	abbreviated	38	ན	
9	སྐྲ		39	སྟོང་ལ	
10	ལྥ་པའི	abbreviated	40	ཅུ་ཚོད	
11	ཚེས		41	རེར	
12	གསུམ་བཅུ		42	སྐོར	

Table 3 (continued)

13	ཉིན	43	ཐིག	
14	ནས	44	དང	
15	བཟུང	45	གྲངས་ཆུང	
16	ཚེ	46	ཝ	
17	གསོང	47	དུག	
18	སྒྲོག	48	ཡིན	
19	སྐྱུད་པའི	abbreviated	49	ཚེ	
20	ཚོང་ལྥ	50	སྒྲོག	
21	སྐྱེ	51	རིན་གོང	
22	ཉིན	52	དི	
23	རེའི	abbreviated	53	གསོང	
24	ལྥ་རྡོའི	abbreviated	54	སྒྲོག	
25	དུས་ཚོད	55	རིན་གོང	
25	བརྒྱུད	56	ལས	
27	ནས	57	རྒྱ་ཚའི	abbreviated
28	ཕྱི་རྡོའི	abbreviated	58	ལྥ་བཅུ	
29	དུས་ཚོད	59	ཡི	case-auxiliary
30	བཅུ་བར	60	མཚོ	

```
Typedefine struct index1
{int length;// Describes the number of word
     int    head1;
     //describes the word's first address in the
corpus}ID1;
     Typedefine struct index2
     {int length;    //describes the number of words
     String [] WH;
             //describes the first character of words
     int head2; //describes the words' address which
             //first character is WH
}ID2;
     Typedefine struct data
     {sting[]    Word;    //stored words
     Char G1;//stored first-order parts of speech
     Char G2;// stored-second parts of speech
}DT;
```

Parts of index tables in structure of corpus are as follows (table 4 to table 6).

Finally, we identify all Abbreviated words and get segmentation word as follows.

གྲོང་ཁྱེར་/nn ག་གེ་མོ་ ་ཨི་/tz འི་/gz སྒྲོག་འཕྲིན་ཅུའུ་/nn ཡིས་/gxཉེ་ར་/mj

སྟོང་/nn གཉིས་/mjལོ་ནའི་/tt/gz སྐྲ་/tt ལྥ་པ་/mxའི་/gz ཚེས་གསུམ་བཅུ་/mjཉིན་/tt

ནས་/cn བཟུང་/vt ཚེ་/vt གསོང་/vt སྒྲོག་/nn སྐྱུད་པ་/vtའི་/gz ཚོད་ལྥ་/nvཔེ་/cn།

ཉིན་རེའི་/tt/gz ལྥ་རྡོའི་/tt/gz དུས་ཚོད་/nnབརྒྱུད་/mjནས་/cn ཕྱི་རྡོའི་/gzདུས་

ཚོད་/nnབཅུ་བ་རེ་/gl/slའཕྲིག་/nnརིན་གོང་/nnཡིན་/upལ/glསྒྲོག་/nnབཀོལ/

ན་/glསྟོང་ལ་/qd རྒྱ་ཚོད་/ttརེ་ར་/rzར་/gl སྐོར་ཐིག་/nnདང་/cdགྲངས་ཆུང་/nnལྥ་/mj

དུག་/mjཡིན་/upཚེ་/vtསྒྲོག་/nnརིན་གོང་/nnའི་/yyགསོང་ར་སྒྲོག་/nnརིན་གོང་/nn

ལས་/glབརྒྱུ་ཚའི་/qlལྥ་བཅུ་/mjཡི་/gzམཚོ/ad།

The comparison between abbreviated word and its restoration as is shown in table 7.

Table 4 Index table 1

length(n)	Head1 f(n)=(n-1)*30+1
1	1
2	31
3	61
……	……
40	118

4.   The Design of Large-scale Speech Corpus

Speech corpus is used to store the sound file attributes and rhythm properties of the Tibetan letters, words and sentences, to extract the acoustic parameters of these sounds, and is integrated into a complete speech corpus. Tibetan characters' gender is divided into Male, female and neutral by the strength of sound, the amount of air, the muscle tightness of phonation organs and sound or not sound.

Considering the effect of the gender on speech, we design the structure of speech corpus as is shown in table 8.

4.1   Design of speech corpus

Make use of Corpus-based TTS method, we need pre-recorded commonly used words, phrases and sentences, and these words, phrases and sentences should cover all speeches in different contexts so as to es-

Table 5 Index table 2

length(n)	WH	Head2
1	ཀ	1
1	ཁ	2
1	ག	3
1	......	......
2	ཀ	450
2	ཁ	478
2	ག	529
2	......	......

Table 6 Index table 3

word	G1	G2
ཀ	n	n
ཁ	n	n
ག	h	h
......	...	...
ཀཁ	n	n
ཀཚ	n	n
......	...	...
ཁཁ	a	s
ཁཆ	n	n
......	...	...

Table 7 The table of abbreviated word and its restoration

No.	Abbreviated word	Restoration
2	གགེཚ	གགེཚ
8	ཚཚ	ཚཚ
10	ཁཚ	ཁཚ
19	ཁཚ	ཁཚ
23	རཚ	རཚ
24	ཚཚ	ཚཚ
28	ཁཚ	ཁཚ
57	ཁཚ	ཁཚ

Table 8 The structure of speech corpus

Index
Words, phrases, sentences
Sound File Name
The gender property of root letters
The gender property of former letters
The gender property of after letters
The gender property of backmost letters
The Position property of words and phrases

tablish a complete speech corpus. Therefore, how to determine the Tibetan words, phrases and sentences is one of the key technologies for the establishment of speech corpus.

According to reference[6] records : the statistics of the frequency and information entropy of Tibetan characters and syllables are based on the corpus of 20,000,000 syllables. The experiment shows : there are 5,334 standard Tibetan syllables, of which the syllables including one syllable, two syllables, three syllables and four syllables are 475, 3,061, 902 and 896 respectively. Therefore, to design speech corpus, we first have to determine the selection standard and the selection range of speech corpus through the Tibetan corpus. To do this we should use a minimum of corpus covering as much as possible the natural language phenomenon, thus to guarantee the fact that the speech corpus is not too large and guarantee the quality of speech.

Continuous sentences can fully reflect the rhythm structure of language, especially the phenomena of tone's information and the change of continuous speech.

Therefore, we select sentences through many ways, such as newspapers, textbooks, science books, literature, Buddhism and historical resources. We also try to make all sentences cover the basic syntax structure of Tibetan , which has an important factor on the rhythm.

The sentences are designed into two sections: simple sentences and compound sentences.

The simple sentences include four kinds of sentences: (1)declarative sentences; (2)interrogative sentences; (3)imperative sentences; (4)exclamatory sentences. Here are the examples.

(1) Declarative sentences
ང་ཚོའི་སྟེངས་འདིའི་འཕྲུལ་མི་ཚོགས་ཆེན་ལ་གྲས་རིམ་གཙོ་བོ་གསུམ་ཡོད།

(2) Interrogative sentences
སྦོབ་ཁང་ནང་དུ་སྦོབ་མ་དུ་ཡོད།

(3) Imperative sentences
སྦོལ་དུ་གར་ལགས། གཤས་གསར་པ་དེ་གཅིག་གཏོང་གནང་དང་།

(4) Exclamatory sentences
བུ་ཆུང་ཨ་དུན་པའི་སྐྱུ་འདི། དེ་ཆེན་ཨ་མས་གསན་ནེ་འདུ་དགའ་བ་ལ།

The compound sentences also include two kinds of sentences（with conjunctions or without conjunctions), here are the examples.

(1) The compound sentences with conjunction
ང་ཚོའི་སྦོབ་གྲར་སོག་རིགས་དང་། བོད་རིགས། ཧྥྱིའི་རིགས། ཡུ་གུར་ རིགས། ཁྲི་ནུན་རིགས། གྲཱང་རིགས། མིའོ་རིགས། དྲྱིས་རིགས་སོགས་ མི་རིགས་ལྡུ་བརྒྱ་ལྷག་ཚམ་གྱི་སྦོབ་གྲ་བ་ཡོད་པ་རེད།

(2) The compound sentences without conjunction
བགྲ་ཕིས་བདེ་ལེགས་སྦུན་སྦམ་ཚོགས། གཏུན་དུ་བདེ་བ་འཐོབ་པར་ཤོག

4.2 Speech tagging

The determination of rhythm information is related to linguistics, phonetics, signal processing, psychology, acoustics and other areas of knowledge. This process requires the computer to understand natural language correctly, and identifies speech units corresponding to each fundamental frequency, duration, pause, stress and other prosodic information. The various types of rhythmic elements are organized according to certain rules, thus to generate a sense of priorities and self-assured in listening and playing in the emotions function.

Typically, because of the same character or words in different location in the sentence, the pronunciation of the severity, length of pause are different too. etc: དགའ in དགའ་བ and in ཤེས་རྒྱལ་དགའ has different pronunciation of the severity. རྒྱ in རྒྱ་ཆེན and in མཚོང་རྒྱ has difference in both pronunciation of the severity and length of pause. Therefore, synthetic, natural, and clear voice must have a good rhythm rules. In order to extract more precise quantitative description of the rules of rhythm, data mining techniques is used in a semi-automatic way to find specific patterns, association rules, change rules, exception information structure, and a statistically significant event. All Tibetan characters and common words can mined from corpus. Firstly, we classify Tibetan phoneme, syllable, character, word and sentence, and establish the voice corpus of character, word, phrase and special sentence respectively; secondly, we analyze the Tibetan sentence structure, the pause and accentuate in sentences. Finally, we process the voice symbols of Tibetan, English and number, and we shouldn't tag the non-pronunciation characters, for example, space and newline, etc. Preprocessing must be run before voice tagging, so as to filter the nun-pronunciation characters and extract the useful characters.

5. Conclusion

In this paper, we represent the essence of TTS and compare the development of TTS at home and abroad. According to Tibetan written language, Tibetan pronunciation features and the mechanism of Tibetan, we research the text normalization, segmentation and voice tagging, etc，which are the key techniques of Tibetan TTS by using text analysis and processing.

References

[1] Zhao Li, *Voice Signal Processing*, M. Machinery industry press, 2008.

[2] Cai Lianhong，Huang Dezhi, and Cai Rui, *The Basic and Applied of Modern Speech Technology*, Tsinghua university press, 2003.11.

[3] http://www.iflytek.com/Html/gyxf/hxjs/hc/

[4] Wang Yong-sheng and Li Mei, "English accent assignment based on morphological rules and machine learning," *Computer Applications*, Vol. 28, No.1, pp.88-91, 2008.

[5] Zhang Sen, Liu Lei, and Diao Luhong, "Problems on Large -Scale Speech Corpus and the Applications in TTS," *Chinese Journal of Computer*, Vol.33, No.4, pp.689-695, 2010.

[6] Weilan Wang and Wanjun Chen, "The frequency and information entropy of Tibetan character and syllable," *Terminology Standardization & Information Technology*, No.2, pp. 27-31, 2004.

[7] Weilan Wang, "The frequency - rank of language unit in modem Tibetan," *Science Technology and Engineering*, Vol.4, pp.413-417, 2004.

[8] Cai Rang Zhuo Ma and Cai Zhijie, "A decomposition algorithm for words components in the Tibetan word frequency statistics system," *Computer Engineering & Science*, Vol.33, No.3, pp.159-162, 2011.

[9] Cai Rang Zhuo Ma and Cai Zhijie, "The method of component decomposition for modern Tibetan characters," *Journal of Qinghai University*, Vol. 28, No.4, pp.83-86, 2010.

[10] Cai Rang Zhuo Ma and Cai Zhijie, "Study of corpus-based Tibetan TTS technology," *Journal of Qinghai Normal University*, Vol. 26.2, pp.66-69, 2010.

[11] Cai Zhijie, "The research and Development of Tibetan Text-to-Speech," *Symposium on Chinese Minority Language Information Processing and Language Resource Construction*, pp.343-346, 2005.

[12] Cai Zhijie, "Identification of abbreviated word in Tibetan word segmentation," *Journal of Chinese Information Processing*, Vol.23, No.1, pp.353-37, 2009.

[13] Wu Zhiyong, Cao Guangqi, Meng M. Helen, and Cai Lianhong, "A unified framework for multilingual text-to-speech synthesis with SSML specification as interface," *Tsinghua Science and Technology*, Vol.14, No.5, pp.623-630, 2009.

[14] Yi Lifu, Li Jian, Hao Jie, and Xiong Ziyu, "Improved grapheme-to-phoneme conversion for mandarin TTS," *Tsinghua Science and Technology*, Vol.14, No.5, pp.606-611.

[15] Zeng Lingping and Chai Peiqi, "English TTS system using variable-length concatenating units," *Tsinghua Science and Technology*, Vol.31, No.3, pp.180-182.

[16] Hemin and Gao Gu, "Computer engineering to speech system," *A cta Scientiarum Naturalium Universitatis NeiMongol*, Vol.35, No.1, pp.100-104, 2004.

[17] Li Yonghong, Kong Jiangping, and Yu Hongzhi, "Rules for the auto-transformation of Tibetan text to IPA," *Tsinghua Univ (Sci & Tech)*, Vol.48, No.S1, pp.621-626, 2008.

[18] Niu Fei and Dexi Jiacuo, "Introduction of Tibetan's Latin transliteration," *Journal of Qinghai Normal University*, No.3, pp.42-45, 2006.

[19] Shuang Zhiwei, Zhang Shile, and Qing Yong, "Warped source spectrum for voice conversion and similarity," *Tsinghua Univ (Sci & Tech)*, Vol.49, No.S1, pp.1408-1412.

[20] Cai Rang Zhuo Ma and Cai Zhijie, "The feature analysis of phonemic pronunciation in the Tibetan text to speech," *Computer Science and Service System*, IEEE Catalog number:CFP128M-PRT,6(2012),4011-4013.

[21] Li Youghong, Gu Lu, and Fang Huaping, "Design and labeling for Tibetan speech synthesis corpus based on HTS," *International Conference on Computer,Electrical, and Systems Sciences and Engineering(2011)*, pp.416-419.

[22] Gao Guo-hong, Li Xue-yong, and Lv Jin-na, "Design of the speech synthesis from text to speech based on the TTS technique," *Second ETP/IITA World Congress in Applied Computing, Computer Science,and Computer Engineering(2010)*, pp.172-174.