软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2015,26(6):1409−1420 [doi: 10.13328/j.cnki.jos.004597]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

藏语语音合成单元选择*

才让卓玛[1,2], 李永明[1], 才智杰[2]

[1](陕西师范大学 计算机科学学院,陕西 西安 710062)

[2](青海师范大学 计算机学院,青海 西宁 810008)

通讯作者: 才让卓玛, E-mail: cr-zhuoma@163.com, http://www.snnu.edu.cn/

摘 要: 基于语料库的语音合成是国内外应用广泛的语音合成方法.在这种合成方法中,单元选择是语音合成的关键.通过分析藏语言文字的属性特征,设计了藏语语音合成系统模型,提出以构件、组合构件、字、词及句单元相融合的藏语语音合成方法,有效地保留了语音合成中大单元的完整性和小单元的灵活性与鲁棒性.同时,给出语音合成的单元选择策略与算法.实验数据表明:该策略与算法是有效和合理的,所选择的单元在封闭语料和开放语料上的覆盖率均达到预期目标.

关键词: 语音合成;单元选择;构件;组合构件;音素

中图法分类号: TP391

中文引用格式: 才让卓玛,李永明,才智杰.藏语语音合成单元选择.软件学报,2015,26(6):1409−1420. http://www.jos.org.cn/1000-9825/4597.htm

英文引用格式: Cairangzhuoma, Li YM, Cai ZJ. Unit selection in Tibetan speech synthesis. Ruan Jian Xue Bao/Journal of Software, 2015,26(6):1409−1420 (in Chinese). http://www.jos.org.cn/1000-9825/4597.htm

Unit Selection in Tibetan Speech Synthesis

Cairangzhuoma[1,2], LI Yong-Ming[1], CAI Zhi-Jie[2]

[1](College of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

[2](College of Computer Science, Qinghai Normal University, Xining 810008, China)

Abstract: Corpus-based speech synthesis is the most widely-used speech synthesis technology in home and abroad. In this type of synthesis method, unit selection is crucial to the speech synthesis. This paper designs a system model of Tibetan speech synthesis by analyzing the attributive characteristics of Tibetan text, and presents a mixed units mode with Tibetan components, combinational components, characters, words and sentences. The method effectively preserves the integrity of larger units and the flexibility and robustness of small units. At the same time, it provides the unit selection strategies and algorithms of Tibetan speech synthesis. Experimental data indicates that the strategies and algorithms are effective and the coverage of units reaches expected target in both the open corpus and closed corpus.

Key words: speech synthesis; unit selection; basic component; combinational component; phone

语音合成(speech synthesis)技术[1−4]旨在用计算机把文本信息转化为清晰、自然的语音,它涉及语音学、数字信号处理和计算机科学等学科,是信息处理领域的前沿技术,有着广泛的应用前景.在语音合成技术的发展过程中,早期的研究主要采用参数合成方法,如Holmes的并联共振峰合成器(1973)[5]和Klatt的串/并联共振峰合成器(1980)[6,7].后来,随着计算机技术的发展出现了波形拼接的合成方法,如基音同步叠加(PSOLA).

* 基金项目: 国家自然科学基金(61262051, 11271237, 61163018); 国家社会科学基金(13BYY141); 教育部"春晖计划"合作科研项目(Z2012093); "长江学者和创新团队发展计划"创新团队资助项目(IRT1068); 青海省科技厅应用基础研究计划基金(2011-Z-755, 2011-Z-754); 青海师范大学科研创新计划基金

收稿时间: 2013-11-28; 修改时间: 2014-01-14; 定稿时间: 2014-03-27

20 世纪 90 年代初,法语、德语、英语、日语等语种的语音合成系统已研制成功.从 20 世纪 80 年代起,汉语语音合成技术与国际研究同步发展[8],发表了高水平学术论文[9−18],研制出了一系列汉语语音合成系统,如联想佳音(1995)、清华大学 TH_SPEECH(1993)、中国科技大学 KDTALK(1995)、科大讯飞 Interphonic 5.0(2009)、北京宇音天下科技公司的嵌入式语音合成软件 emTTS(2012)、捷通华声 TTS(2013)等.这些系统能够将任意文本合成为连续自然的语音,具有多语种、多音色等特点,可应用于大小型声讯服务平台.尽管这些产品的语音输出质量和自然语音还有一定的差距,性能有待改进,但已创造出巨大的社会经济效益.

随着计算机运行速度的提高,内存容量的增大以及语料库建设技术的发展,基于语料库的语音合成(corpus-based speech synthesis)技术成为应用广泛的合成方法之一[19−30].该方法将所需语音单元预先录制建立语音库,通过拼接语音库中的语音单元合成自然语音,其语音合成的主要问题转化为对语音库中语音合成单元的获取、标注、索引和搜索.

藏语语音合成技术的研究近几年刚刚起步,在国家的大力支持下,藏语语料库规模日趋增大,分词、标注及字、词频统计技术也不断成熟[31−35],为基于语料库的藏语语音合成研究奠定了基础.文献[36,37]以藏语拉萨话和夏河话为研究对象,建立了藏语语音语料库,对藏语语音特征进行了分析;文献[38]设计了藏语语料标注规则,并用 Praat 对藏语语音语料进行了标注;文献[39,40]设计了藏语语音合成系统结构,并对文本分析模块进行了研究;文献[41]通过分析藏文传统拼读法与藏文音素拼读法的特征,提出了藏语音素合成音节的思想;文献[42]研究了分词、韵律标注规则和韵律标注等基于语料库的藏语语音合成技术;文献[43]在小规模藏语语料中采用 HMM 的汉语普通话语音合成框架.实现汉藏双语语音合成.

以上文献从不同角度研究了基于语料库的藏语语音合成技术,研究主要集中在分词、标注、文本分析和语音库建设等基础性工作,建立语音库时没有考虑语音单元选择问题.语音合成单元选择决定着语音库的大小,是建立结构合理、内容完整且规模适中语音库的基础,从而是基于语料库的语音合成的关键.

本文借鉴国内外语音合成方法,通过分析藏语言文字的属性特征,设计了藏语语音合成系统模型,提出以构件、组合构件、字、词及句单元相融合的混合单元模式藏语语音合成方法,有效地保留了大单元的完整性和小单元的灵活性与鲁棒性,并给出单元选择策略和算法.在封闭语料和开放语料上对单元选择算法的有效性和合理性进行了测试,实验数据表明:该策略与算法有效、合理,组合构件、字、词及句单元对封闭语料和开放语料的覆盖率均达 80%以上,各类合成单元在开放语料上的覆盖率之和可达 100%.

1 藏语语音合成系统模型

藏语语音合成系统通过对藏语文本特征进行分析,利用计算机技术把藏语文本信息实时地转换为连续自然的藏语语音单元序列.因此,藏语文本特征分析和语音合成系统模型设计是藏语语音合成技术的基础.

1.1 藏语文本特征分析

藏语言文字是一种以辅音字母和元音字母为构件的拼音文字,以音节为单位,一般一个字为一个音节,各音节间用音节符“་”分隔.辅音可以单独构成音节(如ས),也可与其他辅音或元音合成音节(如ས་ས);元音不能单独构成音节,也不能与其他元音合成音节.每个音节至少包含一个构件(基字),最多不超过 7 个构件(前加字、上加字、基字、下加字、后加字、再后加字和元音).音节以基字(辅音字母)为中心,前加字、后加字和再后加字与基字横向拼写,上加字、下加字和元音与基字纵向拼写.组成音节的各个字母称作基本构件(basic component),简称构件;上加字、下加字和元音与基字纵向拼写而成的字母组合称作组合构件(combinational component),组合构件可进一步分解为构件.例如,藏文字བསྒྲས中前加字ད、后加字ག和再后加字ས等是构件,སྒྲ为组合构件;组合构件སྒྲི可进一步分解为上加字ས、基字ག、元音ྲ、下加字ྀ等构件,如图 1 所示.

综上所述,藏文字以音节为基本单位,每个音节由构件组成,藏文文法[44]指出:“字成词,词成句,句达意”.因此,藏文文本从大到小依次由句级、词级、字级、组合构件级和构件级这 5 个层次构成,藏语文本层次结构如图 2 所示.

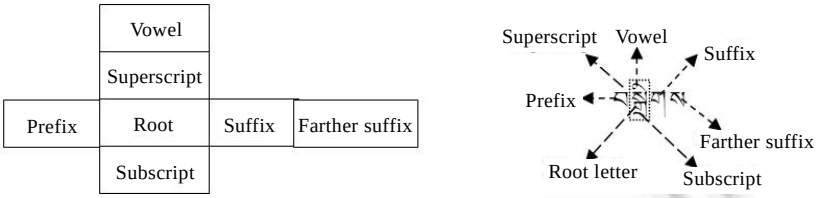


Fig.1 Structure of Tibetan characters and its example

图 1 藏语字结构及实例

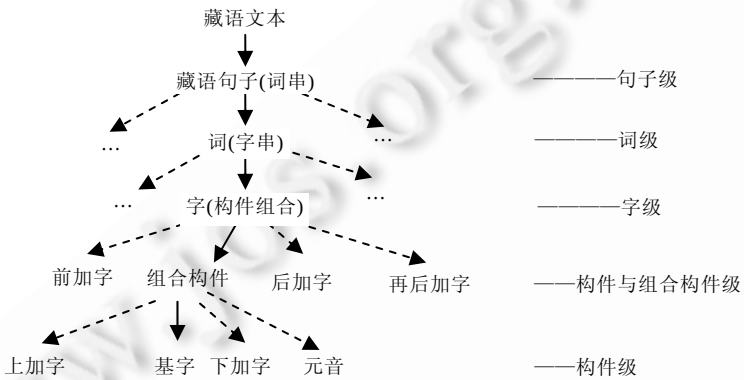


Fig.2 Structural levels of Tibetan text

图 2 藏语文本层次结构

1.2 藏语语音合成系统模型

藏语语音合成系统包括合成文本的切分标注、合成单元选择、根据各类规则库从语音库中搜索出最佳单元进行合成并输出.因此,将藏语语音合成系统分为自然语言处理和语音合成两个模块,其模型如图 3 所示.

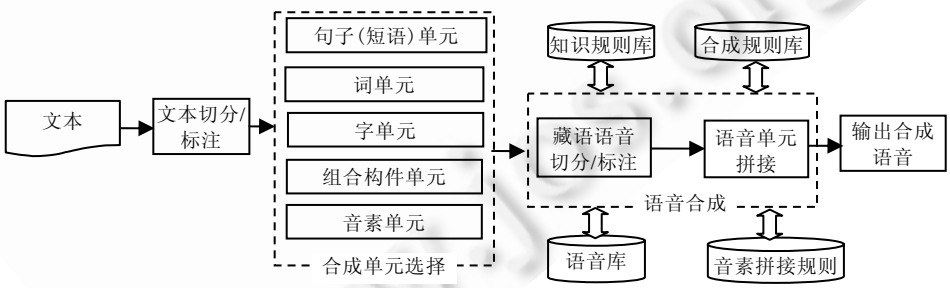


Fig.3 System model of Tibetan speech synthesis

图 3 藏语语音合成系统模型

自然语言处理模块主要对待合成文本进行切分与标注,提供与输入文本相对应的语音信息和韵律信息,建立文本候选单元库.由藏语文本层次特征可知:句语音可分解为词串语音,词语音可分解为字串语音,字语音(即音节)还可分解为各个构件的语音(本文称作音素).所以在藏语语音合成时,只要正确切分每一个构件的语音,总结出音素合成音节的规则,便可用它们合成所有的藏文字音节.因此,可以将构件的语音看作语音结构的最小单位(即音素),音节看作语音结构的基本单位.语料库中抽取构件、组合构件、字、词及句作为语音合成文本单元,建立多级语音合成单元库.语音合成模块利用自然语言处理模块提供的语言学参数,从语料库的候选单元中选择出最优语音单元序列将其拼接,并通过语音信号处理算法对波形进行调整,最终输出连续自然的语音.

2 藏语语音合成单元选择

2.1 合成单元选择策略

语音库是语音合成的重要数据资源,一般情况下,选用的合成单元越大,合成的语音效果越好.极端的情况是预先录制所有句子语音,合成时直接选用原始的语音,然而这种作法对任何语言来说都不现实.由藏文音节与文本特征可知:藏文字的音节与各个构件的语音关系紧密,切分出每一个构件的语音便可用构件的语音去合成所有音节.然而,仅以音素作为藏语语音合成单元而不用字、词、句等大单元,那么大量的音素拼接会产生过多的语音拼接点,从而导致合成语音在某些拼接点处失真.因此,合成流畅、自然的藏语语音需要选择灵活多样的语音单元.为确保大单元语音的自然度和小单元语音的灵活性与鲁棒性,藏语语音合成采用以构件单元(basic component unit)、组合构件单元(combinational component unit)、字单元(character unit)、词单元(word unit)和句单元(sentence unit)相融合的混合单元模式,混合单元模式既能减少合成拼接点,又能减小语音库规模.合成时采用大单元优先原则,在大单元不存在时选择较小单元或音素单元.

例如,合成文本བོད་བདེ་མོང་ནི་མཚོ་སྔོན་ནས་ཡང་བའི་སློབ་ཁས་མ་སྐྱལ་ཐེལཔིན།(你好,我是来自青海的学生卓太)时,先对文本进行切分得到བོད་བདེ་མོ/se ང /ba.ནི/co.མཚོ/ch.སྔོན/j.ནས/ch.ཡང་བ/wo འི/co.སློབ་མ/wo.སྐྱལ/j.ཐེལ/j.པིན/ch| (ba 表示构件单元,co 表示组合构件单元,ch 表示字单元,wo 表示词单元,se 表示句单元,j 表示该文本的语音不在库中),从而用句单元བོད་བདེ་མོ,词单元ཡང་བ、སློབ་མ,字单元ནས、པིན,组合构件ནི、འི,构件ང 进行合成.合成单元库中不存在的字སྐྱལ和ཐེལ需要进行分解,根据分解时取大优先原则,将སྐྱལ由组合构件སྐྱ与后加字ལ合成;组合构件ཐེ在库中不存在,将ཐེལ用基字ཐ、元音ེ和后加字ལ进行合成.

语音合成单元选择的优劣,直接影响合成语音的自然度和流畅性.混合模式单元库的建立,既能减少拼接点使合成语音平滑,又能保证语音合成的自由度.因此在混合模式下,各层次合成单元的选择尤为重要.

构件是组成藏文字的最小单元,因而是最小的合成单元,包括 30 个辅音字母、4 个元音字母、5 个前加字、10 个后加字、2 个再后加字、3 个上加字和 4 个下加字,共计 58 个构件,建立构件单元库.为了便于语音合成,将基字和下加字构成的不需拼接语音的 36 个组合构件(ཀྱ་ཀྲ་ཁྱ་ཁྲ་གྱ་གྲ་གླ་གྭ་ཅྱ་ཆྱ་ཆྲ་ཇྱ་ཉྱ་ཏྲ་ཐྲ་དྲ་དྭ་ནྱ་པྱ་པྲ་ཕྱ་ཕྲ་བྱ་བྲ་མྱ་མྲ་ཙྭ་ཚྭ་ཞྭ་ཟྭ་འྱ་འྲ་ཡྲ་ལྱ་ཤྱ་ཤྲ་ཧྲ་ཨྱ)看成整体也纳入构件单元库.组合构件由基字与上加字或下加字或再下加字或元音上下叠加构成,因此,从语料中统计出构字能力较强的组合构件(即,语料中出现频次较高的组合构件)建立组合构件单元库.藏文字由构件或组合构件组成,是构成藏文文本的基本单位.因此,选择藏文字单元需从语料中统计出现代规范藏文字并建立频度统计表,剔出已选入构件单元库与组合构件单元库的单元(作基字的辅音字母与高频组合构件,如ས、ྷ),然后选取频次较高的藏文字作为字单元库.藏语中词的个数比较多,为了从海量的词库中筛选出合适的词单元,从统计出的词库中先剔出已选入字单元库、组合构件单元库和构件单元库的词(如ཀྱང་、སྤུ་、ཐྱོས་等),然后将词库中的高频词选入词单元库.句是最大的合成单元,为了找出常用且具有代表性的句单元,先建立较大规模的常用句库并进行切分,以句子所含词的个数作为调整参数建立备选句库,然后对备选句库中的句子以所含常用词条数作为调整参数,筛选出常用句建立句单元库.合成单元选择过程如图 4 所示.

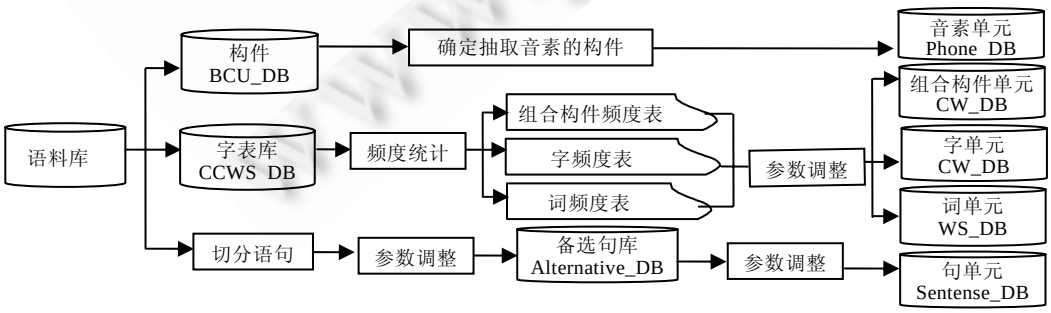


Fig.4 Process of unit selection

图 4 合成单元选择过程

2.2 语音合成单元选择算法

藏文是以辅音字母和元音字母为构件的拼音文字,构件是构成音节的最小单位.为了正确切分音节并保存各构件的属性特征建立构件单元库 *BCU_DB*(basic component unit DB),该库用来存放构件、构件类型及频次,其结构定义为:

Typedef *Struct BCU_DB*
{*string*[·] *BC*; //存放构件单元
Int TYPE; //描述构件位置特征}

其中:*BC* 中存放构件及非藏文字符;*TYPE* 描述构件的类型,取值范围为 1~96,其中,1~30 表示 *BC* 构件为 30 个辅音字母,31~34 表示 BC 构件为 4 个元音字母,35~37 表示 *BC* 构件为 3 个上加字,38~41 表示 *BC* 构件为 4 个下加字,42~46 表示 BC 构件为 5 个前加字,47~56 表示 *BC* 构件为 10 个后加字,57~58 表示 *BC* 构件为 2 个再后加字,59~94 表示由基字和下加字构成的 36 个组合构件,95~n(n≥95 可根据非现代藏文字个数取不同值,本系统中,n=96)表示 *BC* 中存放的非现代藏文字符.构件单元库只含 96 个合成单元,不需筛选按上述策略所叙方法直接人工建立.为提高系统性能,该库存放数据时应局部有序,即,*TYPE* 的取值在某一段内时按 *BC* 字段有序.

为了便于描述组合构件、字及词单元选择算法,本文将组合构件、字与词的字表库定义为相同结构,统一的字表库 *CCWS_DB*(combination component and words DB)用于存放组合构件、字、词及出现的频次,并给出组合构件、字、词单元选择算法.字表库 *CCWS_DB* 其结构定义如下:

Typedef *Struct CCWS_DB*
{*String*[] *Str*; //存放组合构件、字及词
Int N; //描述频次
}

构件、组合构件、字、词单元选择算法

组合构件单元选择:CCU_ Algorithm {
While not *eof*(*File*){
Text=*Read*(*File*); //读入文本
S=*WordDive*(*Text*); //识字
TightenWordDecompose(*S*,*CCWS_DB*); //将 *S* 中的紧缩字分解后保存到字表库 *CCWS_DB* 中
If (*CCWS_DB.Str*=(‘·’ or ‘·’)) //紧缩字频次统计
Count(*CCWS_DB.Str*,*CCU_DB*);
Decompose(*S*,*CCWS_DB*); //*S* 中的藏文字构件分解后将组合构件保存到 *CCWS_DB* 中
Count(*CCWS_DB.Str*,*CCWS_DB*);} //组合构件频次统计
While not *eof*(*CCWS_DB*)
If (*CCWS_DB.N*>α) //α表示频次参数
Output(*CCWS_DB.Str*,*CCU_DB*);}

字单元选择:WU_Algorithm {
While not *eof*(*File*){
Text=*Read*(*File*); //读入文本
S_1=*WordDive*(*Text*); //识字
S_2=*Correcting*(S_1); //正字
Normalize(S_2,*Word*,*TightenWord*); //从 S_2 中识别紧缩字后将字和紧缩字分别存入 *Word* 和 *TightenWord*
Count(*Word*,*CCWS_DB*); //将 *Word* 中的字频统计后放入到库 *CCWS_DB* 中
Decompose(*TightenWord*,*CCWS_DB*); //将 *Tighten Word* 中的紧缩字分解后保存到 *CCWS_DB* 中
If (*TightenWord* =(‘·’ or ‘·’)) //成字性紧缩字频次统计

```
Count(TightenWord,CCWS_DB); }
While not eof(CCWS_DB)
If (CCWS_DB.N>β)                        //β表示选择参数,由频次确定
Output(CCWU_DB.Str,WU_DB);              //将符合选择参数的字单元保存到字单元库 WU_DB 中
}
词单元选择算法:WSU_Algorithm {
While not eof(TagFile){
Sentence=Read(TagFile);                 //从标注文本中读句子
Words=WordsDive(Sentence);              //识别词
Count(Words,CCWS_DB);                   //词频统计
}
While not eof(CCWS_DB)
If (CCWS_DB.N>γ)                        //γ表示高频词选择参数
Output(CCWS_DB.Str,WSU_DB);             //将 CCWS_DB 中符合选择参数的词保存到词单元库 WSU_DB 中
}
```

组合构件由基字与上加字或下加字或再下加字构成,选择组合构件单元时,从藏文文本 *File* 中通过识别紧缩字 *TightenWord* 与组合构件,并将它们存入字表库 *CCWS_DB* 中,统计出各组合构件与紧缩字的频次,将字表库 *CCWS_DB* 中频次大于参数α的组合构件作为满足条件的单元输出到组合构件单元库 *CCU_DB* 中;藏文字可以由构件或组合构件组成,单元选择时需要对读入的藏文文本 *File* 进行正字和规范化,将规范的藏文字和紧缩字分别存入变量 *Word* 和 *TightenWord* 中,然后统计每个藏文字在文本中出现的频次,并将频次大于参数β的藏文字作为满足条件的单元输出到字单元库 *WU_DB* 中;词单元的选择需要较多的文本信息,如词性标注、词的频次等,单元选择时,先对文本进行分词标注,生成标注文本,对标注文本中的词进行频次统计后,将频次大于参数γ的词输出到词单元库 *WSU_DB*.算法中,*File* 表示未加工藏语文本,*TagFile* 表示已标注藏语文本.

- 句单元选择算法

通过对《安多藏语会话读本》、《藏族谚语精选》及文字杂志与新闻报刊等中的简单句统计分析得出:简单句平均包含 6 个词,最少的 1 个词,如ཆོག(可以),སློབ་སྦྱོང(学习);最多的 15 个词,如:

སློབ་གྲོགས་/བཀྲ་ཤིས་/དང་/ངེད་/གཉིས་/མཉམ་དུ་/རྩལ་སྦྱོང་/ཐང་/གི་/ཁར་/ཀར་/དཔེ་ཆ་/བལྟ་/ར་/འགྲོ།

(我和扎西同学一起去操场边上看书).

藏语常用句一般由简单句构成,且句中所含词均为常用词.因此,可以用所含词的个数和常用词个数来刻画常用句.其过程为:首先,从标注文本 *TagFile* 中根据句长参数δ预选出备选句输出到备选句库 *Alternative_DB* 中,然后对 *Alternative_DB* 中句子统计常用词个数,如果考查的句子中所含常用词数达到指标参数θ,则将该句输出到句单元库 *SU_DB* 中.备选句库 *Alternative_DB* 结构定义及句单元选择算法如下:

```
Typedef Struct Alternative_DB
{String[ ] Sentense;
Int N;}
SU_Algorithm(TagFile) {
j=1;
While not eof(TagFile){
 Sentence=Read(TagFile);                //读句
 N=1;                                   //N 表示句子 Sentence 中所含词数
 For (i=1; i<=length(Sentence); i++)
    {ch=copy(Sentence,i,1);
```

```
        If ch〈〉'/' then W[N]=W[N]+ch;          //将句子 Sentence 中的词依次放入到数组 W 中
        Else N++;}
        C[j]=N;                                  //数组 C 中存放句子中所含词数
        j++;
  If (N<=δ)                                      //选择参数δ由句长确定
        Output(Sentence,Alternative_DB);         //筛选词数不超过δ的句放入到句库 Alternative_DB 中
 }
  While not eof(Alternative_DB){
     Sentence=Read(Alternative_DB.Sentence);     //读备选句
     j=1;
  For (i=1; i<=C[j]; i++)
   If (W[i]∈WSU_DB)                              //查看常用词在句中出现的个数
        k++;
  If (k>θ)                                       //参数θ由句中所含常用词个数确定
        Output(Sentence, SU_DB);                 //抽取次数大于参数θ的句子
  j++;}
  }
```

2.3 实验数据分析

为了确保候选单元的适用性及合理性,实验语料内容涵盖社会科学、自然科学、工程科学等领域,主要来源于报刊杂志、教材与网络资源.其中,1.5G 语料作为选择构件、组合构件、字及词的实验语料,120KB(共 2 206 条句子)藏语简单句作为选择句单元的实验语料.从青海藏语广播电视网络上下载的藏文语料作为各类合成单元进行综合测试的开放语料.对于单元选择的结果,主要从两个方面进行考量:(1) 单元选择的效率;(2) 单元选择的合理性.单元选择的目标是:句单元在开放语上的覆盖率达 10%左右,词单元在开放语上的覆盖率达 20%左右,字单元在开放语上的覆盖率达 35%左右,组合构件单元在开放语上的覆盖率达 15%左右,组合构件、字、词及句等单元的综合覆盖率在开放语料上达 80%左右,剩余 20%左右的语料由构件单元合成,从而使所有单元总数在开放语料中的覆盖率达 100%.

2.3.1 单元选择算法在封闭语料中的测试

(1) 句单元选择算法测试

从 2 206 条藏语简单句的实验语料中抽取 1 000 个常用词建立了常用词库;根据句中所含词的个数参数δ,从 2 206 个简单句中抽取句子建立备选句库 *Alternative_DB*;将句中所含常用词的个数参数θ所限定的简单句存储到句单元库 *SU_DB* 中.参数δ,θ取不同值时,句单元抽取实验数据见表 1.

Table 1 Data of sentence unit selection

表 1 句单元选择数据表

参数θ / 单元数及比例 (%) / 参数δ	1		2		3		4		5		6	
1	2	0.09	0	0	0	0	0	0	0	0	0	0
2	51	2.31	4	0.18	0	0	0	0	0	0	0	0
3	171	7.75	21	0.95	51	2.31	0	0	0	0	0	0
4	358	16.23	86	3.90	21	0.95	1	0.05	0	0	0	0
5	540	24.48	161	7.30	86	3.90	3	0.14	0	0	0	0
6	755	34.22	278	12.60	161	7.30	6	0.27	0	0	0	0
7	951	43.11	392	17.77	278	12.60	15	0.68	4	0.18	1	0.05

由表 1 中的数据可见:随着句中词的个数参数δ值的增大,备选句库 *Alternative_DB* 中的句子个数增多;随着

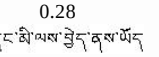
句中常用词个数参数θ值的增大,句库 *SU_DB* 中的句子个数减少.当参数δ=5,θ=1 时,抽取了 540 个句单元;当参数δ=5,θ=2 时,抽取了 161 个句单元;当参数δ=5,θ=3 时,抽取了 86 个句单元;当参数δ=5,θ=4 时,抽取了 3 个句单元;当参数δ=5,θ=5 时,没能抽取到句单元.当参数δ=6,θ=2 时,抽取了 278 个句单元,约占被抽取句子数的 12.60%,与我们预期的目标接近,且这 278 个句子内容也基本满足要求.

(2) 组合构件、字与词单元选择算法测试

在 1.5G 语料中,组合构件共出现 398 个,总频次为 815 528,其中,最高频次为 23 741,最低频次为 1,平均频率为 0.05%;字共出现 3 444 个,总频次为 1 244 416,其中,最高频次为 63 323,最低频次为 1,平均频率为 0.28%;词共出现 15 981 个,总频次为 962 195,其中,最高频次为 60 198,最低为 1 次,平均频率为 1.66%.为了不失一般性,参考组合构件、字及词的平均频率将各算法中的频度参数取α=0.05,β=0.28,γ=1.66,单元选择相关数据见表 2.

Table 2 Dada of combination component, character and words

表 2 组合构件、字与词单元选择数据表

参数	组合构件参数α	字参数β	词参数γ
参数值	0.05	0.28	1.66
示例			
单元数	208	69	571
单元在库中的比例(%)	52.26	2.00	3.57

算法抽出的 208 个组合构件单元、69 个字单元、571 个词单元与人工抽取的常用语组合构件、字、词基本吻合.但是,实验数据显示出两个问题:

- 第一,由于传统语法对字、词定义的模糊性,语料中存在构件、组合构件、字和词多兼的语法结构,如:  既是组合构件,同时也字、词;  既是字也是词,从而导致各库中的单元存在大量的交叉重复现象.经统计,既是高频构件又是高频组合构件的单元在组合构件库(398 个)中有 36 个,约占组合构件库的 9.05%;既是高频组合构件又是高频字的单元在字库(3 444 个)中有 281 个,约占字库的 8.16%;既是高频字又是高频词的单元在词库(15 982 个)中有 1 885 个,约占词库的 11.61%,交叉重复单元在库中共占到 28.81%;
- 第二,由于语料中字、词的频次分布不均匀,频率达到平均频率的字与词仅占各库的 2.00%和 3.57%.所以,尽管算法能够比较有效地抽取单元,但由于上述原因,抽出的单元难以达到实际需要.因此,我们对语料库进行了如下调整:
  1) 从组合构件库中剔出已选入构件库的由基字与下加字构成的 36 个组合构件;
  2) 从字库中剔出已选入构件库的 29 个基字、252 个成字的高频组合构件;
  3) 从词中剔出 10 个基字、177 个成字高频组合构件和 1 668 个高频字,从而消除了各库数据交叉重复的现象.

(3) 构件单元选择说明

在语音合成时,当合成对象在句、词、字单元库中不存在时,需要用最小的合成单元(构件单元)进行合成,而 30 个辅音字母、4 个元音、5 个前加字、10 个后加字、2 个再后加字、3 个上加字、4 个下加字(共计 58 个)和基字与下加字组成的不需合成语音的 36 个组合构件在音素合成音节的过程中都有可能出现,因此将这 94 个构件直接选择为构件单元.

(4) 各类合成单元综合抽取测试

从组合构件、字与词单元选择算法测试中的解决方案可知:构件单元库对组合构件单元、字单元、词单元的选取都有影响,组合构件单元库对字单元、词单元的选取有影响,字单元库对词单元的选取有影响.因此,各类单元依次抽取并进行综合测试较为合适.通过调整参数进行测试,最终将组合构件单元选择参数α取 0.001,字单元选择参数β取 0.001,词单元选择参数γ取 0.016,从而得到 336 个组合构件单元、1 732 个字单元、950 个词单元,各类单元在各库内所占比例分别达 84.42%,50.29%和 6.37%;句单元选择参数δ取 6,θ取 2 时,抽取了 278 个句

子,库内所占比例达 12.60%;组合构件、字、词及句单元在 24KB 测试语料上的覆盖率分别为 13.91%,39.90%,22.70%与 5.78%,共计 82.29%.由此可见,组合构件、字、词及句单元选择算法可行有效.各类单元选择相关数据见表 3.

Table 3 Test data of unit selection

表 3 各类单元选择测试数据表

	组合构件α	字参数β	词参数γ	句参数	
				δ	ε
参数值	0.001	0.001	0.016	6	2
单元数	336	1732	950	278	
示例	བྱ་གྱི་ག་དུ་མི་གྱེ	དང་ལས་ཐུབ་ནས་ཡོད	ཉ་རུཊ་དོན་ཚན་གཞི་ནས་ཐུག	ང་ཚོར་འཆར་གཞི་ཉང་ཡོ་འདུག	
库内比例 (%)	84.42	50.29	6.37	12.60	
单元覆盖率 (%)	13.91	39.90	22.70	5.78	

2.3.2 合成单元在开放语料中的综合测试

从网络上选取了 3 段不同风格与内容的开放语料(青海藏语广播电视网的新闻联播、人生感言及人物传记,语料大小分别为 15KB,22KB 和 31KB,共计 68KB),对抽取的句、词、字及组合构件单元在这 3 个开放语料上的覆盖率分别进行测试,各类单元在不同语料中的综合测试数据见表 4.

Table 4 Test data of unit selection in the opened corpus

表 4 各类单元在开放语料中的测试数据

语料	覆盖率				
	句子(%)	词(%)	字(%)	组合构件(%)	合计(%)
目标值	10.00	20.00	35.00	15.00	80.00
测试语料 1	16.40	14.20	39.70	13.14	83.44
测试语料 2	10.03	14.47	40.12	20.34	84.96
测试语料3	9.98	18.19	35.36	17.25	80.78

由表 4 的数据可知:句、词、字及组合构件单元在测试语 1(新闻联播)的覆盖率分别为 16.40%,14.20%,39.70%和 13.14%,综合覆盖率为 83.44%;在测试语 2(人生感言)的覆盖率分别为 10.03%,14.47%,40.12%和 20.34%,综合覆盖率为 84.96%;在测试语 3(人物传记)的覆盖率分别为 9.98%,18.19%,35.36%和 17.25%,综合覆盖率为 80.78%.各类单元在开放语料中的覆盖率情况如图 5 与图 6 所示.

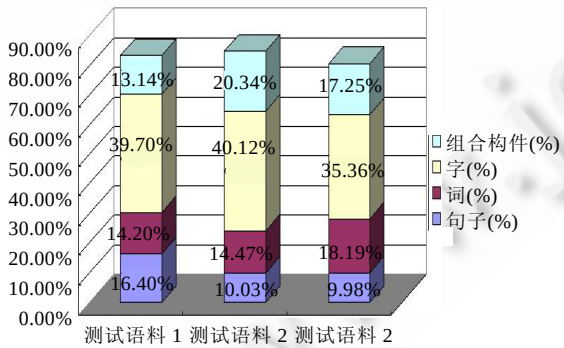


Fig.5 Coverage of units in opened corpus

图 5 合成单元在开放语料中的覆盖率

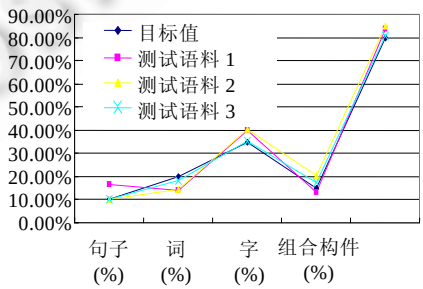


Fig.6 Distribution of units in different corpus

图 6 合成单元在不同语料的分布情况

为了评估系统最终合成的语音效果,以选择的单元建立语音库,开发了基于句、词、字、组合构件及构件相融合的藏语语音合成测试系统,在封闭语料和开放语料上对合成语音的自然度和可懂度进行测试.封闭语料中合成语音基本自然流畅,能够满足主观要求.开放语料中随机抽取了 500 句,经测试,其中 413 句的合成语音自

然流畅,87 句的合成语音存在不平滑现象.经分析,出现不平滑现象的主要原因有两点:其一,句单元选择的常用句语料规模小,覆盖领域不够全面,有些常用句未能抽取到句库中,致使合成时以小单元合成常用句,没能发挥大单元的优越性;其二,音素合成音节及平滑处理技术不够成熟,使得合成语音不够流畅.今后将通过收集不同领域、不同层面的文本以扩大语料,从中抽取常用句及高频短语优化句库;提高音素合成音节及语音的平滑处理技术,提高系统的整体性能.

3 结束语

本文通过分析藏语言文字的属性特征,设计了藏语语音合成系统模型,提出以构件、组合构件、字、词及句单元相融合的混合模式藏语语音合成方法,有效地保留了大单元的完整性和小单元的灵活性与鲁棒性,并给出选择单元的策略和算法.针对不同的语料,对句单元选择算法和组合构件、字及词单元选择算法分别进行了测试,并对测试过程中发现的测试语料交叉重复的现象进行了调整,从而避免了所选单元的交叉和重复.实验数据表明:该策略与算法是有效和合理的,组合构件、字、词及句单元对封闭语料与开放语料的覆盖率均达 80%以上,各类合成单元在开放语料上的覆盖率之和达 100%.下一步的工作是优化句库,提高音素合成音节及平滑处理技术,进一步增强合成单元的灵活性与鲁棒性.

References:

[1] Rouse M. Speech synthesis definition. http://whatis.techtarget.com/definition/speech-synthesis

[2] Parlikar A, Black AW. Data-Driven phrasing for speech synthesis in low-resource languages. In: Proc. of the ICASSP 2012. 2012. 3013−4016. [doi: 10.1109/ICASSP.2012.6288798]

[3] Chen LZ, Gales MJF, Braunschweiler N, Akamine M, Knill K. Integrated automatic expression prediction and speech synthesis from text. In: Proc. of the ICASSP 2013. 2013. 7977−7981. [doi: 10.1109/ICASSP.2013.6639218]

[4] Takamichi S, Toda T, Shiga Y, Sakti S, Neubig G, Nakamura S. Improvements to HMM-based speech synthesis based on parameter generation with rich context models. In: Proc. of the Interspeech 2013. 2013. 362−368.

[5] Holmes JN. The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer. IEEE Trans. on the Audio and Electroacoustics, 1973,21:298−305. [doi: 10.1109/TAU.1973.1162466]

[6] Klatt DH. Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America, 1980,67:971−995. [doi: 10.1121/1.383940]

[7] Lin CY, Jang JSR. A two-phase pitch marking method for TD-PSOLA synthesis. In: Proc. of the Interspeech 2004, Vol.1. 2004. 211−212.

[8] Feng Z, Sun JG, Zhang CS, Wang Y. Reseach advance of Chinese speech synthesis. Journal of Jilin University, 2007,25(2): 198−201 (in Chinese with English abstract).

[9] Tao JH, Zhao S, Cai LH. Study of Chinese speech synthesis system based on statistic prosody model. Journal of Chinese Information Processing, 2002,16(1):1−6 (in Chinese with English abstract).

[10] Dong MH, Lua KT. Using prosody database in Chinese speech synthesis. In: Proc. of the Interspeech 2000. 2000. 243−246.

[11] Zhang DJ, Chen ZX, Huang HY. Design and implementation of mapping address algorithm in Chinese text-to-speech system. Ruan Jian Xue Bao/Journal of Software, 2002,13(1):105−110 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/105.htm

[12] Chou FC, Tseng CY, Lee LS. A set of corpus-based text-to-speech synthesis technologies for mandarin Chinese. In: Proc. of the TASLP 2002, Vol.10. 2002. 481−494.

[13] Zhang W, Wu XR, Zhao ZW, Wang RH. Virtual non-uniform synthesis instances pruning approach for corpus-based speech synthesis system. Ruan Jian Xue Bao/Journal of Software, 2006,17(5):983−990 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/983.htm

[14] Zhang W, Wu XR, Liu J, Wang RH. A non-uniform clustering synthesis instances pruning approach for corpus-based TTS. Chinese Journal of Computers, 2007,30(11):2017−2024 (in Chinese with English abstract).

[15] Zhang S, Liu L, Diao LH. Problems on large-scale speech corpus and the applications in TTS. Chinese Journal of Computers, 2010, 33(4):687−696 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2010.00687]

[16] Li Y, Tao JH, Zhang M, Pan SF, Xu XY. Text-Based unstressed syllable prediction in mandarin. In: Proc. of the Interspeech 2010. 2010. 1752−1755.

[17] Li Y, Tao JH, Xu XY. Hierarchical stress modeling in mandarin text-to-speech. In: Proc. of the Interspeech 2011. 2011. 2013−2016.

[18] Yang CY, Ling ZH, Dai LR. Unsupervised prosodic phrase bouddary labeling of mandarin speech synthesis database using context-dependent HMM. In: Proc. of the ICASSP 2013. 2013. 6875−6879. [doi: 10.1109/ICASSP.2013.6638994]

[19] Hunt A, Black A. Unit selection in a concatenative speech synthesis system using a large speech database. In: Proc. of the ICASSP'96. 1996. 373−376. [doi: 10.1109/ICASSP.1996.541110]

[20] Black AW, Taylor PA. Automatically clustering similar units for units selection in speech synthesis. In: Proc. of the Eurospeech'97. 1997. 601−604.

[21] Charpentier FJ, Stella MG. Diphone synthesis using an overlap-add technique for speech vaveforms concatenation. In: Proc. of the ICASSP'86. 1986. 2015−2018.

[22] Hon H, Acero A, Huang X, Liu J, Plumpe M. Automatic generation of synthesis units for trainable text-to-speech systems. In: Proc.of the ICASSP 1998, Vol.1.1998. 293−296. [doi: 10.1109/ICASSP.1998.674425]

[23] Chu M, Peng H, Yang H, Chang E. Selection non-uniform units from a very large corpus for concatenative speech synthesizer. In: Proc. of the ICASSP 2001. 2001. 785−788. [doi: 10.1109/ICASSP.2001.941032]

[24] Kim SH, Lee YL, Hirose K. Unit generation based on phrase break strength and pruning for corpus-based text-to-speech. ETRI Journal, 2001,23(4):168−176. [doi: 10.4218/etrij.01.0101.0403]

[25] Kishore SP, Black AW. Unit size in unit selection speech synthesis. In: Proc. of the Eurospeech 2003. 2003. 1317−1320.

[26] Ling ZH, Hu Y, Shuang ZW, Wang RH. Compression of speech database by feature separation and pattern clustering using STRAIGHT. In: Proc. of the ICSLP 2004. 2004. 766−769.

[27] Bennett CL. Large scale evaluation of corpus-based synthesizers: Results and lessons from the blizzard challenge 2005. In: Proc. of the Interspeech 2005. 2005.

[28] Nakano Y, Tachibana M, Yamagishi J, Kobayashi T. Constrained structural maximum a posteriori linear regression for average-voice-based speech synthesis. In: Proc. of the Interspeech 2006. 2006. 2286−2289.

[29] Qian Y, Wu ZZ, Gao BY, Soong FK. Improved prosody generation by maximizing joint probability of state and longer units. In: Proc. of the TASLP 2011, Vol.19. 2011. 1702−1710. [doi: 10.1109/TASL.2010.2097248]

[30] Bhakat RK, Narendra NP, Rao KS. Corpus based emotional speech synthesis in Hindi. Pattern Recognition and Machine Intelligence, 2013,8251:390−395. [doi: 10.1007/978-3-642-45062-4_53]

[31] Stan A, Watts O, Mamiya Y, Giurgiu M, Clark RAJ, Yamagishi J, King S. TUNDRA: A multilingual corpus of found data for TTS research created with light supervision. In: Proc. of the Interspeech 2013. 2013. 2331−2335.

[32] Cai RJ. Tibetan corpus processing method. Computer Engineering and Applications, 2011,47(6):138-139 (in Chinese with English abstract).

[33] Cai ZJ, Cai RZM. Design of Tibetan character property analysis system based on Corpora. Computer Engineering, 2011,37(22): 269−272 (in Chinese with English abstract).

[34] Cai ZJ. Identification of abbreviated word in Tibetan word segmentation. Journal of Chinese Information Processing, 2009,23(1): 35−37 (in Chinese with English abstract).

[35] Cai RZM, Cai ZJ. A decomposition algorithm for words components in the Tibetan word frequency statistics system. Computer Engineering & Science, 2011,33(3):159−162 (in Chinese with English abstract).

[36] Cai ZJ, Cai RZM. Design of Tibetan part of speech tagging dictionary. Journal of Chinese Information Processing, 2010,24(5): 46−48 (in Chinese with English abstract).

[37] Kong JP. Perceptual phonetics of Tibetan. National Languages, 1995,(3):56−64 (in Chinese).

[38] Li YH, Yu HZ, Kong JP. Design and implementation of Tibetan continuous speech corpus. Computer Engineering and Application, 2010,46(13):233−235 (in Chinese with English abstract).

[39] Qu Z, Chun Y. Designing the rule of annotation of corpus data in synthesis of Tibetan speech. Journal of Tibet University, 2012, 27(1):62−66 (in Chinese with English abstract).

[40] Cai RZM, Cai ZJ. Study of corpus-based Tibetan TTS technology. Journal of Qinghai Normal University, 2010,26(2):67−69 (in Chinese with English abstract).

[41] Cai RZM, Cai ZJ. Study on the text analysis technology of Tibetan text to speech. Int'l Journal of Intelligent Engineering & Systems, 2012,5(4):10−17.

[42] Cai RZM, Cai ZJ. The feature analysis of phonemic pronunciation in the Tibetan text to speech. In: Proc. of the CSSS 2012. 2012. 4011−4013.

[43] Suo NZX. Study on the key technology of Tibetan speech synthesis. Lasa: Tibet University, 2011 (in Chinese).

[44] Wang HY, Yang HW, Gan ZY, Pei D. Realizing mandarin-Tibetan bilingual speech synthesis by speake adaptive training. Journal of Tsinghua University (Sci. &Tech.), 2013,53(6):776−780 (in Chinese with English abstract).

[45] LUSANG CCJC. Tibetan Grammar. Beijing: Ethnic Publishing House, 2007 (in Chinese).

附中文参考文献:

[8] 冯哲,孙吉贵,张长胜,王岩.汉语语音合成的研究进展.吉林大学学报,2007,25(2):198−201.

[9] 陶建华,赵晟,蔡莲红.基于统计韵律模型的汉语语音合成系统的研究.中文信息学报,2002,16(1):1−6.

[11] 张大军,陈肇雄,黄河燕.汉语文语转换系统地址映射算法的设计与实现.软件学报,2002,13(1):105−110. http://www.jos.org.cn/1000-9825/13/105.htm

[13] 张巍,吴晓如,赵志伟,王仁华.基于虚拟不定长的语音库裁剪方法.软件学报,2006,17(5):983−990. http://www.jos.org.cn/1000-9825/17/983.htm

[14] 张巍,吴晓如,刘江,王仁华.语音库裁剪的一种不定长递阶聚类方法.计算机学报,2007,30(11):2017−2024.

[15] 章森,刘磊,刁麓弘.大规模语音语料库及其在 TTS 中应用的几个问题.计算机学报,2010,33(4):687−696. [doi: 10.3724/SP.J.1016.2010.00687]

[32] 才让加.藏语语料库加工方法研究.计算机工程与应用,2011,47(6):138−139.

[33] 才智杰,才让卓玛.基于语料库的藏文字属性分析系统设计.计算机工程,2011,37(22):269−272.

[34] 才智杰.藏文自动分词系统中紧缩词的识别.中文信息学报,2009,23(1):35−37.

[35] 才让卓玛,才智杰.藏文字频统计系统中字构件分解算法.计算机工程与科学,2011,33(3):159−162.

[36] 才智杰,才让卓玛.班智达藏文标注词典设计.中文信息学报,2010,24(5):46−48.

[37] 孔江平.藏语(拉萨话)声调感知研究.民族语文,1995,(3):56−64.

[38] 李永宏,于洪志,孔江平.藏语连续语音语料库设计与实现.计算机工程与应用,2010,46(13):233−235.

[39] 曲珍,春燕.藏语语音合成中语料数据标注规则的设计.西藏大学学报(自然科学版),2012,27(1):62−66.

[40] 才让卓玛,才智杰.基于语料库的藏语 TTS 技术研究.青海师范大学学报(自然科学版),2010,26(2):67−69.

[43] 索南扎西.藏语语音合成关键技术研究.拉萨:西藏大学,2011.

[44] 王海燕,杨鸿武,甘振业,裴东.基于 HMM 的普通话语音合成框架来实现汉藏双语语音合成的方法.清华大学学报,2013,53(6):776−780.

[45] 色多五世罗桑崔臣嘉措.藏文文法根本颂色多氏大疏.北京:民族出版社,2007.

才让卓玛(1970－),女,青海都兰人,博士生,教授,主要研究领域为自然语言处理,藏文信息处理.

才智杰(1970－),男,教授,主要研究领域为藏文信息处理,藏语自然语言处理.

李永明(1966－),男,博士,教授,博士生导师,主要研究领域为拓扑学,智能系统分析,可计算与复杂性理论.