专题 多语种智能信息处理

藏语方言语音合成数据集

仁曾卓玛[1]，朱丽平[1,2*]

1. 中央民族大学信息工程学院，北京 100081

2. 国家语言资源监测与研究少数民族语言中心，北京 100081

摘要：本研究构建并公开了藏语卫藏、安多和康巴三大方言的语音合成数据集。本数据集来源于喜马拉雅 FM 听音软件，内容包含新闻、法律知识、生活常识、小故事等。数据集中的音频由专业的播音员录播而成，能够保证发音的准确性，通过用专业的音频软件切割播音内容，提供 10 多小时约 8.02 GB 的音频及对应的文本数据，共 4684 条句子。文本经过藏语专业人员审核，能够保证语法的正确性。从字丁、音标、语音现象的均衡覆盖率等多方面对数据集的质量评估表明，本数据集对各方言语言特征覆盖率高，且语音现象覆盖均衡。本数据集可以为研究藏语方言语音合成提供数据支撑，同时还可以作为研究藏语三大方言语音发音、停顿、韵律、节奏等语音特征的语料库。

关键词：语音合成；安多方言；卫藏方言；康巴方言；数据集

数据库（集）基本信息简介

数据库（集）名称	藏语方言语音合成数据集
数据作者	朱丽平，仁曾卓玛，加如，次仁罗布
数据通信作者	朱丽平（2007014@muc.edu.cn）
数据时间范围	2021年
地理区域	西藏拉萨地区、那曲、昌都，青海玉树，青海藏牧区，四川甘孜，阿坝藏区，甘肃甘南等地区
数据量	8.02 GB(压缩后4.3 GB)
数据格式	*.wav, *.txt
数据服务系统网址	http://www.doi.org/10.11922/sciencedb.j00001.00357
基金项目	国家社科基金项目（17BGL199）
数据库（集）组成	数据集被压缩为Tibetan Dialect Speech Synthesis Data Set.zip，其中包含3个大文件夹(ad、kb、wz)，大文件夹下各4个中文件夹(如ad下有ad1、ad2、ad3、ad4)，中文件夹包含多个小文件夹(如ad1下有ad1-1、ad1-2...ad1-10)，小文件夹中 *.wav 是语音数据，*.txt是文本数据。公共包含：（1）4684个*.wav藏语方言音频文件，总时长为10多小时，数据量为8.02 GB，其中，安多2.45 GB、康巴2.13 GB、卫藏3.34 GB；(2) 65个*.txt文本，数据量为1.631 MB，共4684条句子。

ISSN 2096-2223
CN 11-6035/N

文献 CSTR：
32001.14.11-6035.csd.2021.0104.zh
文献 DOI：
10.11922/11-6035.csd.2021.0104.zh
数据 DOI：
10.11922/sciencedb.j00001.00357
文献分类：信息科学

收稿日期：2021-12-31
开放同评：2022-02-08
录用日期：2022-05-17
发表日期：2022-06-28

* 论文通信作者
朱丽平：2007014@muc.edu.cn

引　言

随着科技快速发展，越来越多的智能产品进入人们的生活，在提供极大便利的同时，使人们的生活方式变得更加丰富[1]。语音合成，又称文语转换（Text to Speech，TTS）[2]技术，作为实现人机交互的重要方法，同样得到迅速发展，但目前在市面上却极少见到支持少数民族语言的语音合成技术产品。

藏族是人口较多的少数民族，拥有本民族的语言和文字。中国社会科学院民族语言调查组在对藏语言进行调查考察研究的基础上提出的卫藏方言、康区方言和安多方言的三分说[3]。目前大部分偏远藏区的基础教育薄弱，80%左右的藏族人只会听和说藏语而不识文字[4]。因此藏语音合成技术的发展对于藏族人民使用科技发展产物起到至关重要的作用。目前科大讯飞实现了藏语（拉萨话女声）语音合成，然而第六次人口普查数据[5]显示，约 700 万藏族人口中只有近 45%的人属于西藏户籍，其余藏族人生活中并不使用拉萨语，所以该产品在实际使用中远远不能满足现实需求。

从近年发表的文献[6-8]中发现，当前西藏大学、青海师范大学、西北师范大学等高校师生在研究藏语语音合成问题时，研究者使用的数据集都是通过网上搜集文本数据，人工在设备上录音的方式完成的。这将会使研究者将大量时间和精力花在准备语料上，且由于条件有限，语料在质量上和数量上都有提升空间。文献[9]、[10]提供了藏语语音数据，但是数据量只有 26MB 和 666 段，且只有卫藏方言没有安多和康巴方言。

为提高语音合成数据集的创建效率，弥补藏语安多和康巴方言语音合成语料的不足，本研究从喜马拉雅 FM 听音软件里中国西藏网的“藏语播报”专辑，同时获取音频及对应文本创建藏语方言语音合成数据集，并从语言现象的覆盖率、三大方言的语音特征等方面对数据集的质量进行了分析评估。本数据集共 8.02 GB，其中安多方言 2.45 GB、康巴方言 2.13 GB、卫藏方言 3.34 GB，压缩后总数据集大小为 4.3 GB，内容包含新闻、故事、法律、生活常识等，为藏语语音合成研究和藏语三大方言语音学研究提供数据支撑。

1　数据采集和处理方法

1.1　语料选择的标准和依据

语音合成语料选择的基本标准是音频内容清晰、发音纯正、音素需覆盖均衡，音素、音律要准确，文本数据的字词句和语法要准确。根据这一标准，通过大量查阅音频资料并对比研究发现，喜马拉雅 FM 听音 APP 里的“藏语播报”专辑适合作为语音合成语料。该专辑内容多包含了近几年国内的重要新闻、有趣的小故事、普及法律知识和生活常识等，且每个音频都有相对应的文本内容。同一个文本语料对应有安多、卫藏和康巴三种方言音频，均由专业的播音员使用专业的设备在专用的录音棚内录制，并由专业人员剪辑，通过层层审核而成的。音频相较于普通录音内容，在读音、断句、语法等的准确性和音频的清晰度及完整性是毋庸置疑的。

1.2　数据采集方法

采集的数据包括藏语三大方言音频数据及其对应的文本数据，根据录音及播报时间将音频及其对应文本数据均分为四部分。

1.2.1 文本数据

本研究通过 Python3.8 网络爬虫技术从喜马拉雅 FM 听音 APP“藏语播报”专辑各音频播放界面爬取文本内容。每一篇文本保存三份，分别以安多、卫藏和康巴的拼音缩写和按音频播报顺序编号命名，例如，卫藏第三部分第一篇文本命名为“wz3-1.txt。由于第一部分前 32 个没有对应的康巴方言的音频，所以这部分文本只有两份，其余三个部分的文本均各有三份。

1.2.2 音频数据

直接从 APP 获取到的音频是 xm 格式的文件，该文件格式不属于音频文件，xm 到 wav 的格式转换方法繁杂，且转换后的音频会受损。经尝试发现该音频可以通过 Python 网络爬虫技术爬取，但本研究采用了更简单的获取方法，用安卓手机点击下载即可获得高质量的音频内容。每个音频根据所属方言和序号重命名，例如，卫藏第三部分第一个音频“wz3-1”。

1.3 数据处理方法

1.3.1 文本数据的处理

将每一篇完整的纯文本文字内容，根据完整的语义断句，划分多个具有实际意义的短句。每个语义完整的句子前添加标签，如，卫藏第三部分第一篇第一句“wz3-1-1:”。

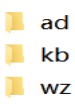
1.3.2 音频数据的处理

每一段待处理的音频为 10 分钟左右的完整录音，根据已切割好的文本短句对音频进行切割，得到各文本短句对应的音频语句。本研究主要使用了 Adobe Audition 软件和迅捷音频转换器完成音频切割。Adobe Audition 软件简称 Au，是由 Adobe 公司开发的一个专门的音频编辑和混合环境，能提供先进的音频剪辑、混合、控制和效果处理功能[11]；迅捷音频转换器是一款功能丰富的音频格式处理软件，支持数多种不同的音频格式，速度快、批量操作效率高，同时涵盖音频转换、音频剪切、音频合并、音频提取、音频录制、音频变速等多种功能。音频切割以保证每个独立句子的语义完整性为原则，切割后的音频长度为 0.1–20 秒，其中 0.3–5 秒之间的最多，音频格式为 wav。

2 数据样本描述

音频文件的采样率为 44.1kHz，声道为双声道立体声。分为 ad(安多)、kb(康巴)、wz(卫藏)三个音频文件夹，每个文件夹按音频录制时间分成 4 部分内容，ab 和 wz 各有 157 段音频、kb 有 125 段音频（缺第一部分的前 32 段音频），处理前的每段音频有 7–15 分钟。

图 1 为数据集结构及整体数据内容展示，切割后的音频长度为 0.1–20 秒，其中 0.3–5 秒之间的最多，音频格式为 wav，文本格式为 txt。

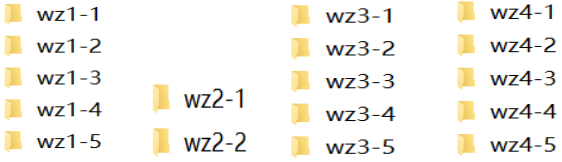


(a): 3 个大文件夹

(a): 3 large Refined-folders

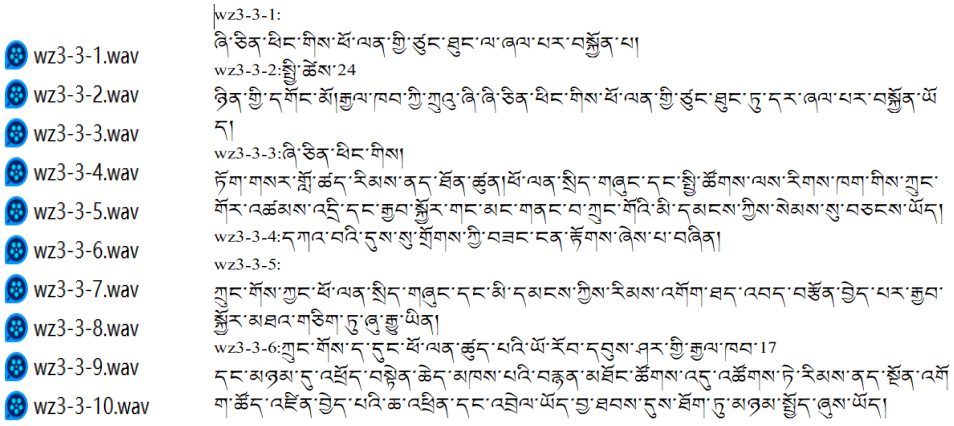
(b): 大文件夹包含的中文件夹

(b): Middle folders included in large folders



(c): 中文件夹包含多个小文件夹

(c): Multiple small folders included in middle-size folders



(d): 小文件夹里的音频及文本数据

(d): Audio and text data in small folders

图 1　数据集

Figure1　Dataset

3　数据质量控制和评估

（1）字丁和语音现象覆盖均衡性评估

为评估创建的语音合成数据集是否覆盖藏语基本音素组合，本研究根据藏文独特音素及构字法，通过统计数据集中藏文字丁的占比，分析各方言的语音现象的覆盖率；通过以基字为分类标准，统计同一个基字的字数来分析语音现象覆盖的均衡性。

郭须·扎巴军奶教授用 14 年时间研究发现藏文字丁共有 18531[12]，萨迦·索南孜摩编著的《正确读字法注疏》中表示藏文字丁共 18745 在平常能接触到的有 8000 多(包括牧民、农民、寺庙、学校等的领域性常用字)，去除特定领域的常用字，在日常生活和交流中常用的藏语字丁约 3000 左右。对语料进行机器统计并经过人工审核校对后发现，本数据集共有 137946 个字丁，其中卫藏 68796 个、

安多 41199 个、康巴有 27951 个，去重后卫藏有 2169 个、安多有 1864 个、康巴有 1743 个字丁。去重后的字丁数在三种方言常用字丁中的占比分别为卫藏 72%、安多 62%、康巴 58%，对日常交流使用的字丁覆盖率均超过 50%，但由于数据集规模不够大，并未覆盖所有字丁，后续将通过持续更新数据内容，进一步提高语音现象的覆盖率。

根据文献[13-14]藏语三大方言的不同发音特征进行总结得到藏语三大方言元音音标和辅音音标，分别如表 1 和表 2 所示。

表 1　藏语三大方言元音音标

Table1　Vowel symbols in three Tibetan dialects

元音	音标（卫藏）	音标（安多）	音标（康巴）
ཨ	a	a	a
ཨི	i	ɜ	i
ཨུ	u	ɜ	u
ཨེ	e	e	e
ཨོ	o	o	o/o:

表 2　藏语三大方言辅音音标

Table2　Consonant symbols in three Tibetan dialects

辅音	音标卫藏	音标安多	音标康巴	辅音	音标卫藏	音标安多	音标康巴
ཀ	ka	ka	ka	མ	ma	ma	ma
ཁ	kha	kha	kha	ཙ	tsa	tsa	tsa
ག	kha	ka	ka	ཚ	tsha	tsha	tsha
ང	ŋa	ŋa	ŋa	ཛ	tsha	dza	tsa
ཅ	tɕa	tɕa	tɕa	ཝ	wa	ʁa	wa
ཆ	tɕha	tɕha	tɕha	ཞ	ɕa	ɕa	xa
ཇ	tɕha	tɕa	tɕa	ཟ	sa	sa	sa
ཉ	noa	ɲa	ɲa	འ	ʔa	ʔa	ɦa
ཏ	ta	ta	ta	ཡ	ja	ja	ja
ཐ	tha	tha	tha	ར	ra	ra	ra
ད	tha	ta	ta	ལ	ɕa	la	la
ན	na	na	na	ཤ	ha	xha	xha
པ	pa	pa	pa	ས	ʔa	sa	sha
ཕ	pha	hpa/ha	pha	ཧ	ha	ha	ha
བ	pha	wa	pa	ཨ	ʔa	ʔa	ʔa

图 2 为藏文字的音节结构，这是一个完整的藏文字丁，包含构字规则的所有结构，其中除了元音以外的加字和基字都属于辅音字母，发音顺序为前加字、上加字、基字、下加字、元音、后加字、再后加字，最终发出的音为整个字丁的音。

以表 2 中的辅音字母为基字，与加字、元音字母组合，采用机器分类结合人工审核校对方法对藏语三大方言语音现象均衡性进行统计分析，基字及每个基字在卫藏、安多和康巴数据集中出现的次数进行统计分析，结果如表 3 所示，单位(次)。

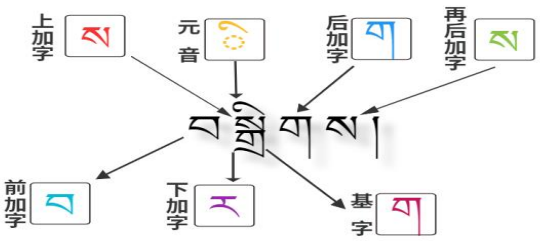


图 2　藏文字的音节结构

Figure2　Syllable structure of Tibetan characters

表 3　语音现象均衡性分析

Table 3　Equilibrium analysis of voice phenomena

序号	基字	卫藏	安多	康巴	序号	基字	卫藏	安多	康巴
1	ཀ	3983	2367	1593	16	ཨ	2905	1513	1112
2	ཁ	3012	1847	1221	17	ཪ	915	490	303
3	ག	7820	4769	3110	18	ཫ	2180	1462	881
4	ང	1204	682	474	19	ཬ	651	420	293
5	ཅ	1478	877	661	20	ཟ	84	52	35
6	ཆ	1693	1050	694	21	཭	2081	1123	916
7	ཇ	861	568	379	22	཮	841	465	306
8	ཉ	1172	736	474	23	཯	142	23	46
9	ཏ	2453	1444	1026	24	ཡ	2223	1301	918
10	ཐ	1930	1177	805	25	ར	2427	1530	954
11	ད	6205	3773	2586	26	ཱ	3105	1795	1240
12	ན	3173	1888	1285	27	ི	828	512	359
13	པ	5047	3025	2106	28	ཱི	2984	1834	1271
14	ཕ	1406	808	573	29	ུ	660	377	222
15	བ	5102	3160	2031	30	ཱུ	231	131	77

表 3 中所有字丁被分为 30 类。以第 3、11、15、13 字母为基字的字数最多，而以第 20、23、30、29 字母为基字的字数较少，但这并不能说语音现象的均衡性不好，用藏文字的构字法分析后发现，第 3、11、15、13 等字数多的字母可以和上加字、下加字、前加字、后加字、后后加字组合，且基字与加字元音字组合就有更多的字丁，而 20、23、30 等字母很少有前加字、上加字、下加字的字丁，且日常生活中极少用。

综上所述，本数据集的语言特征覆盖率高，且语音现象覆盖均衡，在使用本数据进行藏语方言

语音合成研究时，可减少因训练数据缺失或稀疏引起模型泛化性能较差等问题。

（2）文本内容专业性评估

数据来源为喜马拉雅 FM 听音 APP“藏语播报”专辑内容，该内容转自中国西藏网，中国西藏网以中、英、藏、德、法 5 个文种向海内外网友介绍以西藏为主藏区信息的国家级重点新闻网站，是中国最大的涉藏专题综合性网站。所以本研究所用的文本在内容、结构及语法等方面均能够确保准确性和可靠性，且文本覆盖面广，涵盖日常生活、新闻、法律等各方面的内容。

（3）音频质量进行评估

本数据集的音频内容是由专业的播音员，在专业的设备上录制，并由专门的团队剪辑层层审核之后形成的，所以音频在发音、断句等的准确性以及语音的清晰度方面是毋庸置疑的。

（4）数据集创建人员评估

整个数据处理是由以藏语为母语的在校大学生和研究生完成，完成后的数据由相互审核完成初审核。初审核主要内容包括文本和音频校对、音频内容的完整性检查，修改有误标签、重新提取不完整的音频内容等，然后由专业人员做审核校对，最终得到可用于卫藏、安多和康巴方言合成语音的数据集。

4 数据价值

本数据集可用于卫藏、安多康巴藏语方言语音合成的训练集及测试集；同时可以根据其发音特点、停顿特点、韵律节奏特点[9]将其作为藏语三大方言语音学研究的语料库；结合机器翻译和人工审核方法，还可以将其拓展为藏语方言与其他语言之间的语音翻译数据集。

致 谢

感谢中国政法大学的戚肖克老师对数据集应用的建议，藏学研究院的贡保加同学文本语义断句方面的指导，感谢信息工程学院李宁同学在语音切割方面、郑怡扬同学在数据下载方面提供的指导。

本研究所使用的数据来源为，喜马拉雅 FM 听音软件里中国西藏网的“藏语播报”专辑，本数据只可用于科研相关内容，不可用于商业等其他交易内容。

数据作者分工职责

朱丽平（1970—），女，湖南省株洲市人，博士，教授，研究方向为语音翻译。主要承担工作：总体规划设计，数据集选择与采集指导，质量控制与协调管理。

仁曾卓玛（1995—），女，甘肃省甘南藏族自治州人，本科，中央民族大学硕士研究生，研究方向为语音处理。主要承担工作：卫藏第三四部分数据处理、安多第三四部分数据处理、康巴第三四部分数据处理。

加如（2000—），男，甘肃省甘南藏族自治州人，高中，中央民族大学本科生。主要承担工作：卫藏第一部分数据处理、安多第一部分数据处理。

次仁罗布（2000—），男，西藏自治区拉萨市人，高中，中央民族大学本科生。主要承担工作：卫藏第二部分数据处理、安多第二部分数据处理、康巴第二部分数据处理。

参考文献

[1] 张书祎. 试论人工智能对人类生活的实际影响[J]. 数码设计,2019(7): 114-115. [ZHANG S Y.On the actual influence of artificial intelligence on human life[J].Peak Data Science,2019(7):114-115.]

[2] TAYLOR P. Text-to-Speech Synthesis[M]. Cambridge:Cambridge University Press, 2009. DOI:10.1017/cbo9780511816338.

[3] 拉龙东智. 藏语语音识别技术研究[D]. 西藏: 西藏大学, 2015. [LALONG D Z.Research on Tibetan Speech Recognition Technology [D]. Tibet: Tibet University, 2015.]

[4] 南措吉, 才让卓玛, 都格草. 基于 BLSTM 和 CTC 的藏语语音识别[J]. 青海师范大学学报(自然科学版), 2019, 35(4): 26-33. DOI:10.16229/j.cnki.issn1001-7542.2019.04.005. [NAN C J,CAI R,DU GC.Tibetan speech recognition based on BLSTM-CTC[J].Journal of Qinghai Normal University (Natural Science Edition),2019,35(4):26-33.DOI:10.16229/j.cnki.issn 1001-7542.2019.04.005.]

[5] 国务院第六次全国人口普查办公室，中国 2010 年第六次全国人口普查主要数据[M]. 北京: 中国统计出版社,2011. [The Sixth National Bureau of public security, the main data of the Sixth National Bureau of public security in 2010 [M] Beijing: China Statistics Press, 2011.]

[6] 谢香腾. 藏语卫藏方言的语音合成技术研究[D].西藏大学,2021.DOI:10.27735/d.cnki.gxzdx.2021.000236. [XIE X T.Research on speech synthesis technology of Tibetan Satellite Tibetan dialect [D]. Tibet University, 2021. DOI:10.27735/d.cnki. Gxzdx. 2021.000236.]

[7] 王智浩.嵌入式藏语语音合成系统的研究[D].西北师范大学,2021.DOI:10.27410/d.cnki.gxbfu.2021.000504. [WANG Z H.Study on embedded Tibetan speech synthesis system [D]. Northwest Normal University, 2021. DOI:10.27410/d. cnki. Gxbfu. 2021.000504.]

[8] 都格草. 基于神经网络的藏语语音合成技术研究[D]. 青海师范大学,2019.DOI:10.27778/d.cnki.gqhzy.2019.000190. [DOU G C.Research on Tibetan Speech Synthesis Based on Neural Network [D]. Qinghai Normal University, 2019. DOI:10.27778/d.cnki. Gqhzy. 2019.000190.]

[9] 韦向峰, 袁毅, 张全, 吐尔逊·卡得. 基于端点检测的蒙藏维语音片段数据集[J/OL]. 中国科学数据, 2019, 4(4). (2019-07-22). DOI: 10.11922/csdata.2019.0024.zh. [WEI X F,YUAN Y,ZHANG Q,et al.A data set of Mongolian,Tibetan and Uyghur speech fragments based on voice activity detection[J].China Scientific Data,2019,4(4).DOI: 10.11922/csdata.2019.0024.zh.]

[10] 韦向峰, 袁毅, 张全, 池哲洁. 2015 年中国少数民族地区蒙藏维言语录音数据集[J/OL]. 中国科学数据, 2016, 1(2). DOI: 10.11922/csdata.120.2015.0024. [WEI X F, YUAN Y, ZHANG Q, et al. Mongolian, Tibetan, and Uyghur speech data from Chinese minority regions in 2015[J]. China Scientific Data, 2016, 1(2). DOI:10.11922/csdata. 120.2015.0024.]

[11] 黄雯静, 程敏熙. 用 Adobe Audition 软件探究影响倒水时声音特性的因素[J]. 物理通报, 2021(12): 123-127,131.DOI:10.3969/j.issn.0509-4038.2021.12.030. [HUANG W J, CHENG M X. Using Adobe Audition software to explore the factors affecting the sound characteristics when pouring water[J]. Physics Bulletin, 2021(12): 123 – 127, 131. DOI:10.3969/j.issn.0509-4038.2021.12.030.]

[12] 郭须·扎巴军奶编著.藏文文字研究[M].北京：中国藏学出版社,2015. [GX·ZHANBA JN.Tibetan Character Study [M]. Beijing: China Tibetan Press, 2015.]

[13] 羊忠旦增. 藏语三大方言比较研究[D].中央民族大学,2013. [YANGZHONG DZ.A comparative study of the three Tibetan dialects [D].Minzu University of China, 2013.]

[14] 才让卓玛,李永明,才智杰.基于 Mealy 机的藏文字构件分解[J].电子学报,2015,43(05):935- -939. [CAIRANG Z M,LI YM,CAI ZJ.Tibetan Component Decomposition Based on Mealy Machine [J]. Journal of Electronics, 2015,43(05): 935 - 939.]

论文引用格式

仁曾卓玛, 朱丽平. 藏语方言语音合成数据集[J/OL]. 中国科学数据, 2022, 7(2). (2022-06-24). DOI: 10.11922/11-6035.csd.2021.0104.zh.

数据引用格式

朱丽平, 仁曾卓玛, 加如, 等. 藏语方言语音合成数据集[DS/OL]. Science Data Bank, 2022. (2022-06-20). DOI: 10.11922/sciencedb.j00001.00357.

A dataset of Tibetan dialect speech synthesis

ZHUOMA Renzeng[1], ZHU Liping[1,2*]

1. College of Information Engineering, Minzu University of China, Beijing, P. R. China

2. National Language Resource Monitoring & Research Center of Minority Languages, Minzu University of China, 100081, P. R. China

*Email: 2007014@muc.edu.cngan

Abstract: In this study, we compiled a Tibetan speech synthesis dataset of for three Tibetan dialects, Namely, Weizang, Amdo and Kangba. The dataset is derived from the audio in Himalayan FM listening software, which contains news, legal knowledge, common sense of life, stories, etc. and has a high coverage of dialectal phonetic phenomena. The audio in the dataset is recorded and broadcast by professional broadcasters to ensure the accuracy of pronunciation. By cutting the broadcast content with professional audio software, we compiled about 8GB of audio and corresponding text data in about 10 hours, totaling 4,684 sentences. The text is checked by Tibetan professionals to ensure the correctness of grammar. The quality assessment of the dataset in terms of balanced coverage of characters, phonetic symbols and the phonetic phenomena shows that the dataset has a high coverage of linguistic features of all dialects, with a balanced coverage of phonetic phenomena. This dataset is expected to provide data support for the study on Tibetan dialect phonetic synthesis, and can also be used as a corpus for the study on pronunciation, pause, rhythm and other characteristics of the three Tibetan dialects.

Keywords: speech synthesis; Amdo dialect; Weizang dialect; Kangba dialect; dataset

Dataset Profile

Title	A dataset of Tibetan dialect speech synthesis
Data corresponding author	Liping Zhu (2007014@muc.edu.cn)
Data authors	ZHU Liping, ZHUOMA Renzeng, JIA Ru, CIREN Luobu
Time range	2021
Geographical scope	Lhasa, Nagqu and Qamdo,Tibet,Yushu and Tibetan pastoral areas, Ganzi and Aba Tibetan areas, Sichuan,Gannan,Gansu,etc.
Data volume	8.02 GB (4.3 GB after compression)
Data format	*.wav, *.txt
Data service system	<http://www.doi.org/10.11922/sciencedb.j00001.00357>
Source of funding	National Social Science Foundation of China (17BGL199)
Dataset composition	The dataset is compressed into “Tibetan Dialect Speech Synthesis Data Set.zip”. It comprises 3 large-size folders (ad, kb and wz); each large folder contains 4 middle-size folders (e.g. ad1, ad2, ad3, and ad4 under ad), and each middle-size folder contains multiple small folders (e.g. ad1-1, ad1-2... ad1-10 under ad1). Among them, “*.wav” is voice data, and “*.txt” is text data. In total, the dataset includes: (1) 4,684 “*.wav “audio files, with a total length of more than 10 hours and a data volume of 8.01GB, including 2.45GB for Amdo, 2.13GB for Kangba and 3.34GB for Weizang; (2) 65 “*.txt” text files, with a total of 4684 sentences and a data volume of 1.631MB.