

藏语情感语音数据库构建

彭毛扎西^{1,2,3} 才智杰^{1,2,†} 才让卓玛⁴

1. 青海师范大学计算机学院, 西宁 810016; 2. 省部共建藏语智能信息处理及应用国家重点实验室, 西宁 810008;
3. 西宁大学计算机与信息科学学院, 西宁 810022; 4. 西南民族大学计算机科学与技术学院, 成都 610041;
† 通信作者, E-mail: czjqhsd@163.com

摘要 针对目前藏语语音情感类型划分不够细致, 语音情感分析数据库规模较小的问题, 在分析汉、英等语言语音情感类型划分方案及数据库基础上, 提出一种藏语情感语音数据库构建方案, 包括藏语语音情感分类、情感语音采集、情感语音标注以及有效性分析等。根据此方案, 建立面向藏语语音情感分析的情感类型集(TESSCS-9), 用录音法和剪辑法采集2786句藏语情感语音, 并对其进行标注, 利用改进的模糊综合评价法评估情感语音得到含2745句藏语情感语音数据库(TESDB-2745), 为藏语语音情感分析奠定了基础。

关键词 语音信号处理; 藏语; 情感语音; 数据库

Construction of Tibetan Emotional Speech Database

PENGMAO Zhaxi^{1,2,3}, CAI Zhijie^{1,2,†}, CAIRANG Zhuoma⁴

1. The College of Computer, Qinghai Normal University, Xining 810016; 2. The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Xining 810008; 3. School of Computer and Information Science, Xining University, Xining 810022; 4. School of Computer Science and Technology, Southwest Minzu University, Chengdu 610041; † Corresponding author, E-mail: czjqhsd@163.com

Abstract The classification of Tibetan speech emotion types is not detailed enough, and the database size of speech emotion analysis is also small. Based on the analysis of the classification of speech emotion types and database of Chinese, English and other languages, this paper designs a construction scheme of Tibetan emotional speech database, including Tibetan speech emotional classification, emotional speech collection, emotional speech tagging and effectiveness analysis, etc. According to this scheme, an emotion type set (TESSCS-9) for Tibetan speech emotion analysis is established. 2786 Tibetan emotional speech is collected by recording and editing methods, and annotated. The improved fuzzy comprehensive evaluation method is used to evaluate the emotional speech, so as to obtain a 2745 Tibetan emotional speech database (TESDB-2745), laying the foundation for Tibetan speech emotion analysis.

Key words speech signal processing; Tibetan; emotional speech; database

语音是人类最常用、最便捷的交流方式之一^[1], 利用语音进行自然、和谐的人机交互是语音信号处理领域的关键技术^[2]。语音中不仅包含语义信息, 而且包含丰富的情感信息。语音情感信息对人机交互起着重要的作用^[3], 因此语音信号处理既要挖掘语音中的语义信息, 还要分析语音中的情感信息。语音情感类型划分及数据库构建是语音情感分析的

基础。

汉、英等语言针对各自的语言及语音特点划分情感语音类型, 建立了情感语音数据库^[4], 在此基础上研究语音情感分析技术, 取得了较好的成果。藏语是中华文化的重要组成部分, 随着信息技术的发展, 藏文信息处理技术也取得长足的进步。藏语语音情感分析技术刚刚起步, 目前对藏语语音情感

类型的划分还不够细致^[5],用于藏语语音情感分析的数据库规模也比较小,制约了藏语语音情感分析技术的发展。

本文在分析汉、英等语言情感类型划分方案的基础上,采用离散情感类型划分方法,基于语音的音高和音长等韵律学特征,结合藏语的语音特点划分藏语语音情感类型,建立面向藏语语音情感分析的情感类型集(TESCS-9)。用录音法和剪辑法采集 2745 句藏语情感语音,并对其进行标注。运用改进的模糊综合评价法对建立的情感语音及标注进行有效性分析,构建藏语情感语音数据库 TESDB-2745,旨在为藏语语音情感分析提供数据资源。

1 研究现状

自 20 世纪 70 年代,学者们开始研究情感类型划分。情感类型划分是构建情感数据库的基础,根据情感表征方式的不同,可分为离散和维度两种划分方法^[6]。离散情感类型划分方法根据情感的面部表情、语言声调表情和身体姿态表情,将情感划分为以形容词标签形式描述情感状态的离散类型。维度情感类型划分方法根据情感激烈程度以及情感正负面程度等心理属性,以多维情感空间中点的形式描述情感状态的连续类型。维度情感划分方法需要将情感状态量化,在现有藏语语音处理技术条件下较难实现。离散情感划分方法简单直观,并且容易操作,是目前情感类型划分常采用的方法^[7]。

目前主要采用离散情感类型划分方法对英、汉语言的情感类型进行划分。文献[8-16]先后对英语的情感类型进行划分。我国很早以前就有关于情感类型划分的研究,《中庸》中将情感类型划分为喜、怒、哀和乐 4 种,《礼记》中将情感类型划分为喜、怒、哀、惧、爱、恶和欲 7 种。许小颖等^[17]将汉语的情感类型划分为基本情感、基于心理感受的情感和基于表现力的情感 3 个系统,其中基本情感类型划分为好(爱和敬)、恶、喜(乐)、怒、哀、惧和欲 7 种。赵力等^[18]在研究情感识别时,将汉语语音情感类型划分为欢快、愤怒、惊奇和悲伤 4 种。蒋丹宁等^[19]将汉语语音情感类型划分为中立、愤怒、恐惧、高兴、悲伤和惊奇 6 种。英、汉语音情感类型的分类见表 1。

在语音情感类型划分的基础上,学者们开始构建情感语音数据库。Ambrus^[20]构建情感语音数据库 Maribor,包含英语、斯洛文尼亚语、法语和西班牙语 4 种语言,情感分为厌恶、惊讶、高兴、恐惧、生气、悲伤和中性 7 种类型。McGilloway 等^[21]构建英语情感语音数据库 Belfast,包含生气、悲伤、高兴、恐惧和中性 5 种情感类型。Burkhardt 等^[22]构建德语情感语音数据库 EMODB,包含中性、生气、恐惧、高兴、悲伤和厌恶 6 种情感类型。韩文静等^[23]构建汉语情感语料库 CASIA,包含高兴、悲哀、生气、惊吓和中性等 5 种情感类型。除上述情感数据库外,还有日语、丹麦语、瑞

表 1 英汉情感类型分类
Table 1 Classification of emotion types in English and Chinese

文献	情感类型	类型数量
[8]	容忍、生气、期望、厌恶、愉悦、恐惧、悲伤、惊奇	8
[9]	害怕、愤怒、苦恼、欢乐、厌恶、惊奇、关爱、羞愧	8
[10]	快乐、悲哀、恐惧、愤怒、疼痛、厌恶、轻快、成功、失败、内疚、悔恨、骄傲、羞耻	13
[11]	愤怒、轻蔑、厌恶、悲伤、恐惧、内疚、兴趣、愉快、害羞、惊奇、痛苦	11
[12]	愤怒、厌恶、恐惧、高兴、悲伤、惊奇	6
[13]	愤怒、喜爱、高兴、害怕、悲伤、惊讶	6
[14]	愤怒、厌恶、害怕、高兴、悲哀	5
[15]	中性、愤怒、恐惧、高兴、悲伤	5
[16]	愤怒、恐惧、高兴、悲伤、厌恶、惊奇	6
《中庸》	喜、怒、哀、乐	4
《礼记》	喜、怒、哀、惧、爱、恶、欲	7
[17]	好(爱、敬)、恶、喜(乐)、怒、哀、惧、欲	7
[18]	愤怒、高兴、悲伤、惊奇	4
[19]	中性、愤怒、恐惧、高兴、悲伤、惊奇	6

典语、法语以及土耳其语情感语料库等^[24-26]。情感语音数据库的具体信息见表2。

目前,关于少数民族语音情感的类型划分处于探索阶段。塔什甫拉提·尼扎木丁等^[27]将维语的语音情感类型划分为高兴、生气、难过、平静、惊讶和害怕6种,建立相应的情感语音数据库,并基于此语料库进行语音情感识别。杜楠楠等^[28]将维吾尔语的语音情感类型划分为高兴、生气、悲伤和中性等5种类型,并研究情感语音韵律转换技术。单禹^[29]将蒙古语的语音情感类型划分为高兴、生气、悲伤、惊奇、恐惧、厌恶、无聊和中性8种,建立蒙古语情感语音库和蒙古语情感语音合成系统。在藏语研究方面,Guo等^[5]将藏语拉萨方言的情感类型划分为高兴、惊讶、悲伤和中性4种,并分析4种情感类型的音高、能量和时长等特征。这种划分方式比较常见,但情感覆盖不全面。次仁罗增^[30]将藏语文本情感类型划分为乐、好、怒、哀、惧、恶和惊7种,据此构建相应的藏文文本情感词典,并根据语音识别的文本结果分析藏语的语音情感类型,但未划分藏语语音情感类型,只能作

为藏语语音情感类型划分的参考。

2 藏语情感语音数据库的构建

2.1 构建方案

本文通过藏语语音情感分类、情感语音采集、情感语音标注以及有效性分析等过程,构建藏语情感语音数据库,其中情感语音采集使用录音法和剪辑法两种方式,数据库构建方案如图1所示。

2.2 构建步骤

根据藏语情感语音数据库构建方案,本文构建面向语音情感分析的藏语情感语音数据库(Tibetan emotional speech database, TESDB-2745),数据库的具体信息见表3。TESDB-2745藏语情感语音数据库的构建过程分为4个步骤:1)藏语语音情感分类;2)情感语音采集;3)情感语音标注;4)情感语音的有效性分析。

2.2.1 藏语语音情感分类

藏语语音情感分类是藏语情感语音数据库构建的基础。在进行语音情感分类时,需要在保证各情感类别互斥的同时兼顾情感类别的覆盖率。英、汉

表2 情感语音数据库信息
Table 2 Emotional speech database information

数据库	语种	情感类型	规模
Maribor ^[20]	英语、斯洛文尼亚语、法语和西班牙语	厌恶、惊讶、高兴、恐惧、生气、悲伤、中性	每类情感约186句
Belfast ^[21]	英语	生气、悲伤、高兴、恐惧、中性	每类情感约320句
EMO-DB ^[22]	德语	中性、生气、恐惧、高兴、悲伤、厌恶	每类情感约100句
CASIA ^[23]	汉语	高兴、悲哀、生气、惊吓、中性	每类情感约100句

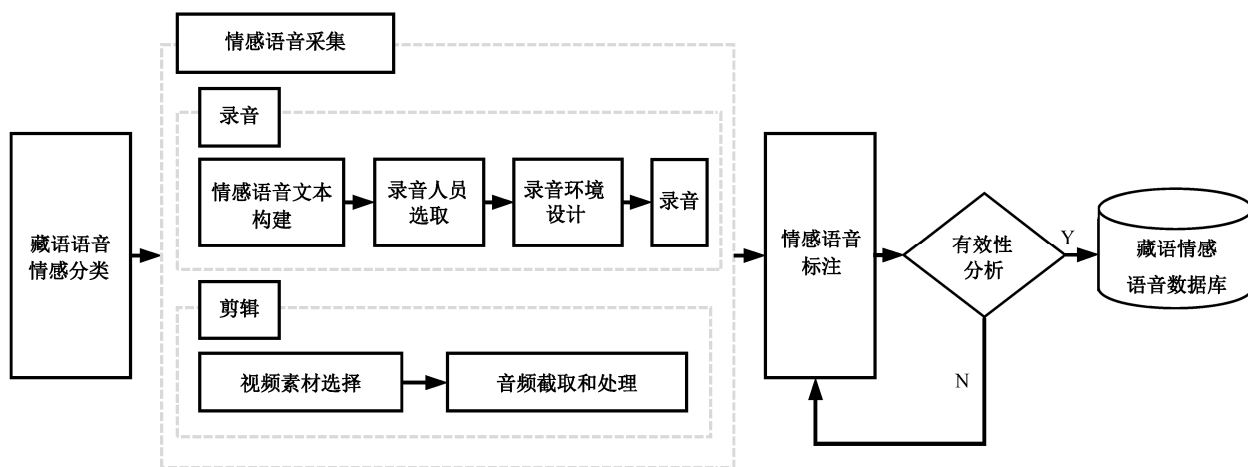


图1 藏语情感语音数据库构建方案
Fig. 1 Construction scheme of Tibetan emotional speech database

表 3 TESDB-2745 数据库信息
Table 3 TESDB-2745 database information

类别	各类型数量/句									合计/句
	高兴	愤怒	悲伤	恐惧	厌恶	惊讶	中性	夸张	焦急	
男声数量	258	349	185	260	169	177	134	96	59	
女声数量	141	153	289	86	88	92	88	50	71	
录制数量	99	99	93	100	101	97	100	100	100	2745
剪辑数量	300	403	381	246	156	172	122	46	30	
总数量	399	502	474	346	257	269	222	146	130	

等语言的语音情感分类基本上已经确定,而藏语语音情感具有较强的情境依赖性^[31],并与文化背景紧密联系^[32],因此藏语语音情感分类必须结合藏语语音的特点,不能完全套用其他语种的分类方式。本文在分析英、汉等语言语音情感分类的基础上,结合藏语语音自身的特点,对藏语语音情感进行分类。

在分析英、汉等语言的情感分类时,对表 1 中表示相同类型而采用不同术语表示的类型名称进行统一,并统计不同类型出现的次数。将表 1 中“高兴”“欢乐”“愉快”“快乐”“愉悦”“喜”和“乐”统一为“高兴”,将“愤怒”“生气”和“怒”统一为“愤怒”,将“悲哀”和“悲伤”统一为“悲伤”,将“害怕”“恐惧”和“惧”统一为“恐惧”,将“厌恶”“厌倦”和“恶”统一为“厌恶”,将“惊讶”和“惊奇”统一为“惊讶”,将“关爱”“喜爱”和“爱”统一为“喜爱”,将“羞愧”“羞耻”和“害羞”统一为“害羞”,将“疼痛”和“痛苦”统一为“痛苦”,将“骄傲”和“轻蔑”统一为“骄傲”,共得到 21 种不同的情感类型。术语统一后的英、汉语音情感语音类型及出现次数如表 4 所示。

由表 4 可见,在英、汉等语言情感分类中“高兴”“愤怒”“悲伤”“恐惧”“厌恶”“惊讶”和“中性”7 种类型出现次数最多(占总出现次数的 80.73%),是目前大部分学者较为认可的使用最广泛的情感类型。

在分析英、汉语音情感分类的基础上,本文选

择一部由藏族专业演员演绎,情感丰富饱满,且内容贴近真实生活的藏语本土影视剧《莫登帕果》(མོ་དོན་པ་ཀོ་མཁོ་),对其中所有语音(982 句)进行人工情感划分。经统计,影视剧中共出现 9 种情感类型,其中 7 种与英、汉语音情感类型中分布最广的前 7 种一致,另外在情感语音中还出现“夸张”和“焦急”两种情感类型,并且出现频率较高。用文本表示情感时,“夸张”和“焦急”两种情感类型不是很明显,但用语音表达时,这两种情感类型非常突出,而且不能归属于其他类型。例如,“ངས་ལག་པ་གཅིག་གིས་འཛོང་འཛོན་ལྷན་ལྷན་” (我用一只手臂抓住野牦牛)通过语音语调表达“我”力大无穷的强烈“夸张”情感,不属于其他 7 种情感类型;“སྐར་མ་ལྔ་ཡིས་རྒྱུག་རེའི་སྐོར་རྒྱུག་རྒྱ་རེད། ཁོ་ད་དུང་མ་ཐོན་ཐལ།” (仅剩五分钟就要关闭考场了,但他还没到)通过语音语调及语速表达“他”迟到的“焦急”情感,也不属于其他 7 种情感类型。影视剧《莫登帕果》中的语音情感类型及分布如图 2 所示。

本文对所有采集的情感语音从音高和音长两方面进行韵律特征分析,结果如表 5 所示。可以看出,9 种类型的韵律特征区分度明显。

我们通过分析英、汉等语言的情感分类以及藏语影视剧中语音的情感类型,构建面向藏语语音情感分析的藏语情感语音类型集(Tibetan Emotional Speech Classification Set, TESCS-9), TESCS-9 = {“高兴”“愤怒”“悲伤”“恐惧”“厌恶”“惊讶”“中性”“夸张”

表 4 英汉情感语音类型及出现次数分布
Table 4 Emotional speech type and frequency distribution in English and Chinese

类型	高兴	愤怒	悲伤	恐惧	厌恶	惊讶	中性	喜爱	害羞	痛苦	内疚
次数	17	17	16	12	10	10	6	3	3	2	2
频次	15.60	15.60	14.68	11.01	9.17	9.17	5.50	2.75	2.75	1.83	1.83
类型	骄傲	苦恼	轻快	成功	失败	悔恨	兴趣	容忍	期望	欲	
次数	2	1	1	1	1	1	1	1	1	1	
频次	1.83	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	

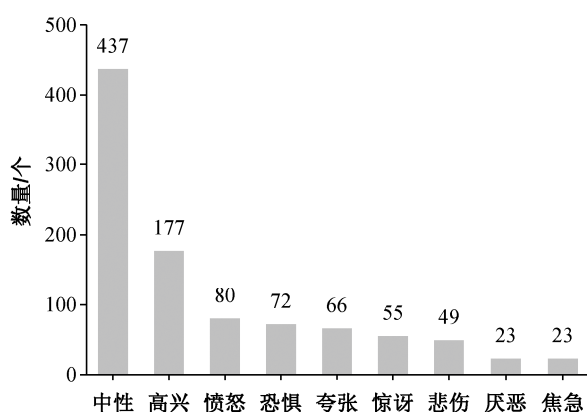


图2 藏语影视剧《莫登帕果》情感类型统计

Fig. 2 Statistical on the emotional types of Tibetan film “མོ་དུན་པ་ཀུ་མེ་ལོ་མེ་ལོ་”

“焦急”}。

2.2.2 情感语音的采集

采集情感语音的方法有录音法和剪辑法。录音法是目前普遍采用的方法,其优点在于可以在短时间内获取所需的情感语音,并且可以控制录音的情感类型、文本内容和录音人员的性别等,但真实度和自然度不高。剪辑法是从各种媒体信息中截取包含情感的语音片段,优点是情感真实度高,有上下文内容关联信息,但耗时费力,并且会受到背景噪声和混响的影响。本文采集情感语音时,使用录音法和剪辑法相结合的方式。

1) 用录音法采集藏语情感语音,包括文本的构建、录音人员的选取和录音环境的设计等环节。进行文本构建时,对每种情感类型,选择特定情感较

明显、情感容易判断、内容贴近日常生活且情感丰富的10个句子,句子长度为15个音节左右。这样既容易表演,又能保证语音数据中情感的真实性。录音人员选择10名在校大学生(5名男性,5名女性),年龄在20~30岁之间,母语均为藏语,发音清晰,且录音当天没有患感冒等影响发音的疾病。为了提高录音人员的情感表达能力,我们进行有关情感语音类型及录音人员演技的培训。考虑到音频信号的真实应用场景和获取的便捷性,未严格要求统一的录音软件、硬件及录音环境。音频的采样位数、采样频率和声道数分别为16位、41 kHz和单声道,音频文件的存储格式规定为WAV格式,共采集900句藏语情感语音。

2) 用剪辑法采集藏语情感语音,包括视频素材的选择、音频的截取和处理等环节。选择视频素材时要考虑语料的规模,由于目前藏语本土影视剧数量较少,因此我们选择汉语电视连续剧的译制片和藏语本土影视剧作为剪辑素材。译制片的作品一般由专业配音演员配音,情感演绎较恰当。藏语本土影视剧一般由藏族专业演员演绎,情感丰富饱满,内容贴近真实生活。为了降低噪声对音频质量的影响,对影视剧音频进行降噪处理。通过辨听,选择情感特征明显的语句进行截取并保存,音频的采样位数、采样频率和声道数分别为16位、41 kHz和单声道,共采集1886句藏语情感语音。

2.2.3 情感语音的标注

对采集的每一条情感语音,从情感类型、采集方式、发音人性别和情感判断难易程度等方面进行

表5 藏语情感语音类型及韵律特征
Table 5 Tibetan emotional speech type and prosodic features

序号	类型	标签	韵律学特征		说明
			音高(强度)	音长(速度)	
1	高兴	Happy	高	略快	表示某种追求并如愿以偿时的愉悦情感
2	愤怒	Angry	高	快或慢	表示为实现愿望的行动受挫时的气愤情感
3	悲伤	Sad	低	略慢	表示分离、丧失和失败时的伤情感
4	恐惧	Fear	正常	很快	表示面临某种危险情境时的恐惧情感
5	厌恶	Disgust	低	非常快	表示遇到反感的事物时的讨厌情感
6	惊讶	Surprised	较高	快	表示发生意想不到的或不寻常的事情时的惊讶情感
7	中性	Neutral	正常	正常	表示不带任何情感
8	夸张	Exaggerate	高	较快	表示用夸张词语或动作表达内心的强烈情感
9	焦急	Anxiety	较高	很快	表示在应对某事件过程中与自己认知冲突时的焦虑情感

标注, 标注信息如表 6 所示。

2.2.4 情感语音的有效性分析

改进的模糊综合评价法(fuzzy comprehension evaluation method)是模糊评价法(MOS)与熵权法(the entropy weight method, EWM)相结合的, 将定性分析转化为定量分析的评价方法, 能对呈现模糊性的系统做出科学、合理和贴近实际的量化评价, 广泛应用于各个领域。任鹏辉等^[33]等建立一种面向语音情感计算的数据库, 并采用改进的模糊综合评价方法对其进行评价。宋静等^[34]等构建一个语音情感数据库 TYUT2.0, 并通过模糊综合评价与层次分析法、熵权法相结合进行评价。为了保证所采集情感语音的可靠性, 本文运用改进的模糊综合评价法对建立的情感语音及标注进行有效性分析, 步骤如下。

1) 建立综合因素评价集 $U=(u_1, u_2, u_3, u_4, u_5)$, $u_1 \sim u_5$ 分别代表准确度、自然度、清晰度、噪声影响度和类型标注准确度这 5 个影响情感语音的因素, 评价标准见表 7。

2) 建立评语集 $V=(v_1, v_2, v_3, v_4, v_5)$, V 是评价者对评价对象可能给出的各种评价组成的集合, $v_1 \sim v_5$ 分别表示好、较好、中、较差和差 5 个评语等级, 并约定 $v_1 \sim v_5$ 的值 $V_weight=[100, 85, 65, 35, 0]$ 。

3) 建立评分表 $D=(d_{ij})_{5 \times 5}$, d_{ij} 表示 u_i 因素评价

表 6 藏语情感语音标注信息

Table 6 Tibetan emotional speech annotation information

标注信息	类别	标记符
情感类型	高兴	Happy
	愤怒	Angry
	悲伤	Sad
	恐惧	Fear
	厌恶	Disgust
	惊讶	Surprised
	中性	Neutral
	夸张	Exaggerate
	焦急	Anxiety
	采集方式	录音法
剪辑法		J
发音人性别	男	M
	女	W
情感判断难易程度	难	D
	中	M
	易	E

表 7 情感语音评价标准

Table 7 Emotional speech evaluation criterion

因素评价项	标准说明
准确度 (u_1)	情感语音是否能表达情感类型
自然度 (u_2)	语音的情感表现是否自然
清晰度 (u_3)	语音咬字是否清晰
噪声影响度 (u_4)	背景音、噪声等是否影响
类型标注准确度 (u_5)	情感类型标注是否正确

指标在 v_j 评语上的人数。

4) 计算评价矩阵 $R=(r_{ij})_{5 \times 5}$, $r_{ij}=d_{ij}/d$, d 为打分总人数。

5) 用熵权法确定客观权重 $W=(w_1, w_2, w_3, w_4, w_5)$, W 计算方法如下。

① 构建 m 个语音在因素 U 评价指标上的评价等级矩阵 $C=(c_{ij})_{m \times 5}$, c_{ij} 表示 n 个人对第 i 条语音第 j 个因素评价指标的百分制平均得分。

② 构建矩阵 C 的标准化数据矩阵 $Y=(y_{ij})_{m \times 5}$, $y_{ij} = \frac{c_{ij} - \min(c_i)}{\max(c_i) - \min(c_i)}$ 。

③ 构建因素评价指标比重矩阵 $P=(p_{ij})_{m \times 5}$, $p_{ij} = \frac{y_{ij}}{\sum_{i=1}^m y_{ij}}$ 。

④ 构建因素评价指标的信息熵 $E=(E_j)$, 其中 $E_j = -\ln(m)^{-1} \sum_{i=1}^m p_{ij} \ln p_{ij}$ ($j=1, 2, \dots, 5$)。

⑤ 计算差异性系数 $G=(g_j)$, $g_j=1-E_j$ ($j=1, 2, \dots, 5$)。

⑥ 确定客观权重 $W=(w_j)$, $w_j = \frac{1-E_j}{n - \sum E_j}$ ($j=1, 2, \dots, 5$)。

6) 计算模糊综合评价矩阵 $B1=W \times R$ 。

7) 计算模糊综合评价矩阵 $B1$ 的归一化矩阵 B 。

8) 计算该语音的总得分 $S=B \times V_weight$ 。

9) 设定阈值分数 P , 如果 $S \geq P$, 则保留此条语音; 否则不符合要求, 删除该语音。

建立评分表时, 组织 10 名人员对采集的 2786 条情感语音从好、较好、中、较差和差 5 个评语等级进行投票, 投票分布如表 8 和图 3 所示。可以看到, 评语等级投票“好”的人数约占 50%, 评语等级投票“较好”的人数约占 30%, 评语等级投票“中”的人数约占 10%, 这三项的投票累积人数占比达 90% 以上, 表明从评语等级投票结果来看, 采集的情感

表 8 评语等级投票统计
Table 8 Evaluation grades vote statistics

因素评价指标	票数				
	好	较好	中	较差	差
准确度	13628	9388	3057	1071	716
自然度	13413	9919	3350	989	189
清晰度	15730	8215	2858	896	161
噪声影响度	14408	8973	3219	1035	225
类型标注准确度	13556	8727	3529	1358	690

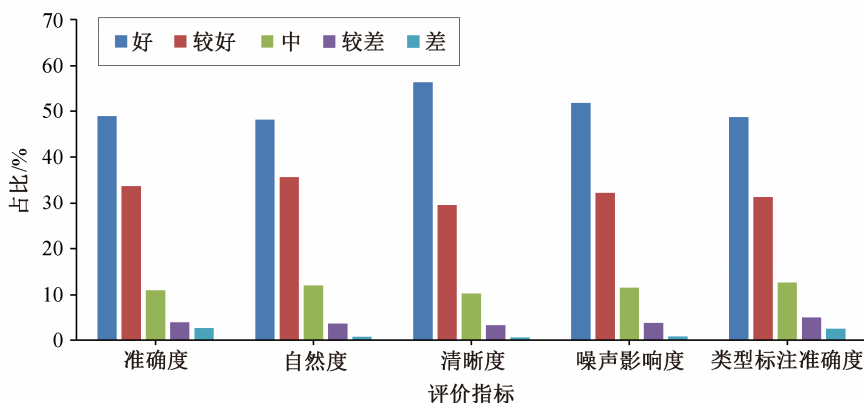


图 3 评语等级投票分布
Fig. 3 Evaluation grades voting distribution

语音质量比较好。

在计算客观权重时,组织 10 名人员对随机抽取的 20 条情感语音按准确度、自然度、清晰度、噪声影响度和类型标注准确度 5 项因素评价指标用百分制打分,评分分布如图 4 和表 9 所示。可以看到,因素评价指标评分在 [0, 60) 的人数约占 2%, 在 [60, 100] 的人数约占 98%, 大多数的评分在 [90, 100] 之

间,占比在 50% 以上,表明从因素评价指标得分来看,采集的情感语音质量也是比较好的。

在评语等级投票和因素评价指标评分的基础上,得到客观权重,进而计算每条语音的总得分 S 。当阈值 P 取 60(及格)时,2714 句情感语音 $S > P$,情感语音合格;72 句情感语音的 $S < P$,情感语音不合格。不合格的语音主要有两种类型:一种是情感类型标

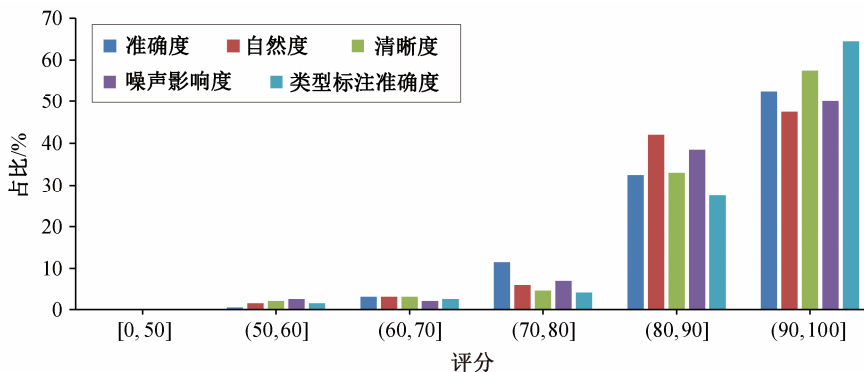


图 4 因素评价指标评分分布
Fig. 4 Factor evaluation index grade distribution

表 9 因素评价指标评分统计
Table 9 Factor evaluation index grade statistics

得分区间	打分人数				
	准确度	自然度	清晰度	噪声影响度	类型标注准确度
[0, 50)	0	0	0	0	0
[50, 60)	1	3	4	5	3
[60, 70)	6	6	6	4	5
[70, 80)	23	12	9	14	8
[80, 90)	65	84	66	77	55
[90, 100]	105	95	115	100	129

注不准确(31 句); 另一种是情感类型无法确定(41 句)。对情感类型标注不准确的语音重新标注, 删除无法确定情感类型的语音, 最终从 2786 句情感语音中获取 2745 句来构建 TESDB-2745 数据库。

3 结束语

本文在分析汉、英等语言语音情感类型划分方案及数据库的基础上, 设计了藏语情感语音数据库构建方案, 依照此方案构建藏语情感语音数据库(TESDB-2745)。

首先采用离散情感类型划分方法, 结合藏语语音的特点划分藏语语音情感类型, 建立面向藏语语音情感分析的情感类型集(TESCS-9); 然后采用录音法和剪辑法采集 2786 句藏语情感语音, 并对其进行情感类型标注; 最后运用改进的模糊综合评价法, 对情感语音及标注进行有效性分析, 从而建立藏语情感语音数据库 TESDB-2745。

TESDB-2745 基本上符合研究要求, 可用于藏语语音情感特征分析和语音情感识别等相关研究, 为藏语语音情感分析奠定了基础。

未来的工作中, 我们将不断扩充数据库规模, 并进行藏语情感语音声学特征分析及情感识别研究。

参考文献

[1] 罗德虎, 冉启武, 杨超, 等. 语音情感识别研究综述. 计算机工程与应用, 2022, 58(21): 40–52
 [2] 刘振焘, 徐建平, 吴敏, 等. 语音情感特征提取及其降维方法综述. 计算机学报, 2018, 41(12): 2833–2851
 [3] Khalil R A, Jones E, Babar M I, et al. Speech emotion recognition using deep learning techniques: a review.

IEEE Access, 2019, 7: 117327–117345
 [4] 韩文静, 李海峰. 情感语音数据库综述. 智能计算机与应用, 2013, 3(1): 5–7
 [5] Guo Dandan, Yu Hongzhi, Hu A, et al. Statistical analysis of acoustic characteristics of Tibetan Lhasa dialect speech emotion // SHS Web of Conferences. EDP Sciences. Seoul, 2016, 25: 01017
 [6] 高庆吉, 赵志华, 徐达, 等. 语音情感识别研究综述. 智能系统学报, 2020, 15(1): 1–13
 [7] 孙晓虎, 李洪均. 语音情感识别综述. 计算机工程与应用, 2020, 56(11): 1–9
 [8] Plutchik R. A general psychoevolutionary theory of emotion. New York: Academic Press, 1980: 3–33
 [9] Tomkins S. Affect imagery consciousness: Volume II: the negative affects. Berlin: Springer Publishing Company, 1963
 [10] Krech D, Crutchfield R S, Livson N. Elements of psychology. 2nd ed. New York: Knopf, 1974
 [11] Izard C E. Human emotions. New York: Plenum Press, 1977
 [12] Ekman P. Facial expression and emotion. American Psychologist, 1993, 48(4): 384–392
 [13] Shaver P, Schwartz J, Kirson D, et al. Emotion knowledge: further exploration of a prototype approach. Journal of Personality and Social Psychology, 1987, 52(6): 1061–1086
 [14] Johnson-Laird P N, Oatley K. The language of emotions: an analysis of a semantic field. Cognition and Emotion, 1989, 3(2): 81–123
 [15] Petrushin V A. Emotion recognition in speech signal: Experimental study, development, and application // Proc Sixth International Conference on Spoken Language Processing. Beijing, 2000: 222–225
 [16] Nwe T L, Foo S W, De Silva L C. Speech emotion reco-

- gnition using hidden Markov models. *Speech Communication*, 2003, 41(4): 603–623
- [17] 许小颖, 陶建华. 汉语情感系统中情感划分的研究 // 第一届中国情感计算及智能交互学术会议论文集. 北京, 2003: 215–221
- [18] 赵力, 钱向民, 邹采荣, 等. 语音信号中的情感识别研究. *软件学报*, 2001, 12(7): 1050–1055
- [19] 蒋丹宁, 蔡莲红. 基于语音声学特征的情感信息识别. *清华大学学报(自然科学版)*, 2006, 46(1): 86–89
- [20] Ambrus D C. Collecting and recording of an emotional speech database [D]. Maribor: University of Maribor, 2000
- [21] McGilloway S, Cowie R, Douglas-Cowie E, et al. Approaching automatic recognition of emotion from voice: a rough benchmark // *Proc the 2000 ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*. Belfast, 2000: 207–212
- [22] Burkhardt F, Paeschke A, Rolfes M, et al. A database of German emotional speech // *Interspeech 2005*. Lisbon, 2005: 1517–1520
- [23] 韩文静, 李海峰, 阮华斌, 等. 语音情感识别研究进展综述. *软件学报*, 2014, 25(1): 37–50
- [24] Ververidis D, Kotropoulos C. A state of the art review on emotional speech databases // *Proc the 2003 Richmedia Conference*. Lausanne, 2003: 109–119
- [25] Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods. *Speech Communication*, 2006, 48(9): 1162–1181
- [26] El Ayadi M, Kamel M S, Karray F. Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recognition*, 2011, 44(3): 572–587
- [27] 塔什甫拉提·尼扎木丁, 梁瑞宇, 谢跃, 等. 采用原子表示模型的维吾尔语语音情感识别. *信号处理*, 2020, 36(1): 9–17
- [28] 杜楠楠, 赵晖. 维吾尔语情感语音韵律转换研究. *计算机工程与应用*, 2016, 52(19): 154–160
- [29] 单禹. 蒙古语情感语音合成系统研究与实现[D]. 呼和浩特: 内蒙古大学计算机学院, 2021: 15–16
- [30] 次仁罗增. 藏语语音情感识别方法研究[D]. 拉萨: 西藏大学信息科学技术学院, 2019: 28–30
- [31] Kanwal S, Asghar S, Hussain A, et al. Identifying the evidence of speech emotional dialects using artificial intelligence: a cross-cultural study. *PLOS ONE*, 2022, 17(3): e0265199
- [32] Laukka P, Elflein H A. Cross-cultural emotion recognition and in-group advantage in vocal expression: a meta-analysis. *Emotion Review*, 2021, 13(1): 3–11
- [33] 任鹏辉, 张雪英, 孙颖. 面向语音情感计算的数据库的构建与应用研究. *电视技术*, 2012, 36(21): 89–92
- [34] 宋静, 张雪英, 孙颖, 等. 基于模糊综合评价法的情感语音数据库的建立. *现代电子技术*, 2016, 39(13): 51–54