**通用盲文翻译系统**

**软件设计说明书**

目录

[第一章 软件介绍与说明 3](#_Toc53997019)

[1.1软件简介 3](#_Toc53997020)

[1.2阅读对象 4](#_Toc53997021)

[1.3编写目的 4](#_Toc53997022)

[1.4 软件使用环境 4](#_Toc53997023)

[1.5 编程语言 4](#_Toc53997024)

[第二章 软件特点及设计 4](#_Toc53997025)

[2.1软件特点 4](#_Toc53997026)

[2.2国家通用盲文翻译系统的设计 4](#_Toc53997027)

[2.3软件的具体实现 5](#_Toc53997028)

[2.3.1软件GUI界面设计 5](#_Toc53997029)

[2.3.2盲文分词模块设计 7](#_Toc53997030)

[2.3.3汉字拼音化及修正模块设计 7](#_Toc53997031)

[2.3.4通用盲文声调省写模块设计 10](#_Toc53997032)

[2.3.5转换为盲文ASCII码模块设计 15](#_Toc53997033)

[2.3.6转换为盲文点序模块设计 16](#_Toc53997034)

[第三章 程序打包与使用 17](#_Toc53997035)

[第四章 总结 18](#_Toc53997036)

# 第一章 软件介绍与说明

为了使使用者对此软件有详尽的了解，加深使用者对于国家通用盲文翻译系统的理解，特编制了本使用说明书。通过阅读本说明书，使用者可以了解到该软件如何使用以及国家通用盲文翻译系统具体实现的流程是如何进行的。

此软件是针对盲文学习者以及开发使用盲文学习机的工作者而开发的。此软件包括：

* GUI操作界面设计模块：用来连接使用者与执行程序。
* 中文分词模块：用来对文本分词得到较符合国家通用盲文规则的文本，并输出。
* 汉拼转换模块：用来将汉字转换为拼音。
* 汉语拼音到盲文拼音的修正模块：将得到的汉字拼音按照盲文拼音修正，符合盲文规则。
* 基于通用盲文的声调省写模块：将盲文拼音的音调按照通用盲文规则处理，得到符合国家通用盲文规则的盲文拼音。
* 拼音转盲文ASCII码模块：将得到的符合国家通用盲文规则的盲文拼音以及其他字符转换为盲文ASCII码。
* 盲文ASCII码转为盲文点序模块：将盲文ASCII码转换为点序，输出。

本软件实现了在计算机上将中文文本翻译成符合国家通用盲文声调省写规则的盲文点序，输出分词结果、盲文ASCII码和盲文点序，得到的分词结果和点序结果可以作为嵌入式盲文学习机的输入内容，有助于盲文学习者和盲文学习机研发人员学习及科研。

本说明书可作为对本软件用户的参考资料。

## 1.1软件简介

随着社会的不断发展，对于公民文化水平的要求越来越高，同时对于视力障碍的残疾人士的文化程度也提出了新的要求，于是在各方努力研究下一款好学使用性高的国家通用盲文被设计了出来。同时随着信息技术的不断发展，已经有条件让盲文学习者告别厚重的盲文纸质书籍，也具备了通用盲文的推广学习的客观条件，所以迫切需要一款按照通用盲文规则的汉盲翻译系统来应用于盲文学习领域和嵌入式盲文学习机的开发中去。

本软件构建了一个符合国家通用盲文规则的汉盲翻译系统，较以往的翻译系统区别在于，更合理的分词、符合规则的声调省写、高效的翻译算法。设计的简便的GUI操作窗口，能让用户直接得到需要的文件。

本软件的主要功能是：对中文文本进行汉盲翻译得到的点序结果符合国家通用盲文。

## 1.2阅读对象

盲文学习人员、汉盲翻译程序开发人员、嵌入式盲文学习开发人员。

## 1.3编写目的

此软件运用Anaconda编译环境开发，采用python3语言，在windows系统上操作，用户只需运行软件，双击所要翻译的txt文件，即可得到符合国家通用盲文声调省写规则的盲文点序txt文件。

## 1.4 软件使用环境

本软件在使用时对硬件、软件和设备有如下要求：PC台式机或者笔记本电脑。

## 1.5 编程语言

本软件利用Anaconda编译环境开发。开发语言：python3。

# 第二章 软件特点及设计

## 2.1软件特点

本软件基于国家通用盲文方案设计，实现了对中文文本的汉盲翻译得到的结果符合国家通用盲文规则，有利于盲文信息化和通用盲文的推广。

## 2.2国家通用盲文翻译系统的设计

该软件的基本功能是对一个txt文本进行汉盲翻译，得到三份输出结果，包括通用盲文分词结果、通用盲文ASCII结果、通用盲文点序结果。该软件的构成设计如图2-1所示，包括交互界面设计以及执行程序设计，其中执行程序设计包括了盲文分词模块、汉字拼音化及修正模块、通用盲文声调省写模块、转换为盲文ASCII码模块、转换为盲文点序模块。盲文的阅读与中文文本不同，中文文本根据阅读者的经验判断进行语义切分，而盲文阅读者往往是视力不佳或者全盲难以切分中文词语甚至断句，所以盲文被设计为为需要分词，通过空格切分词语，便于阅读。盲文是用不同排列的点阵来代表拼音，再用拼音来拼出汉语，所以需要汉字转换为拼音模块，与此同时，盲文拼音与汉语拼音在某些个例上写法不同，所以需要对得到的汉语体系的拼音进行修正得到符合盲文体系的拼音。国内的盲文研究机构为了使盲文学习者减小学习压力、解决盲文书籍厚重难堪的问题，经过统计研究编写了一套盲文音调省写规则，于是本软件根据国家通用盲文的规则编写算法并执行，最终的盲文结果符合国家通用盲文规则。盲文是用64个盲文ASCII码对应带按需点序来表达所有的字符，所以需要将各种字符包括拼音、标点符号、数字和其他的特殊字符转换为盲文ASCII码，再转换为盲文点序，才算得到了符合国家通用盲文规则的盲文点序。



图2-1 翻译系统基本模块划分

## 2.3软件的具体实现

### 2.3.1软件GUI界面设计

翻译系统要给用户清晰的指引，所以需要设计一个人机交互界面，本系统用户界面的设计和开发基于EasyGUI依赖库，对应的人机交互框图如2-2所示。用户在运行程序后，会跳出选择窗口如图2-3所示，用户点击确认后系统对文本执行翻译任务，翻译成功弹出提示框如图2-4所示，在选择翻译的文件所在的目录下生成结果文件，如图2-5所示。



图2-2翻译系统人际交互框架

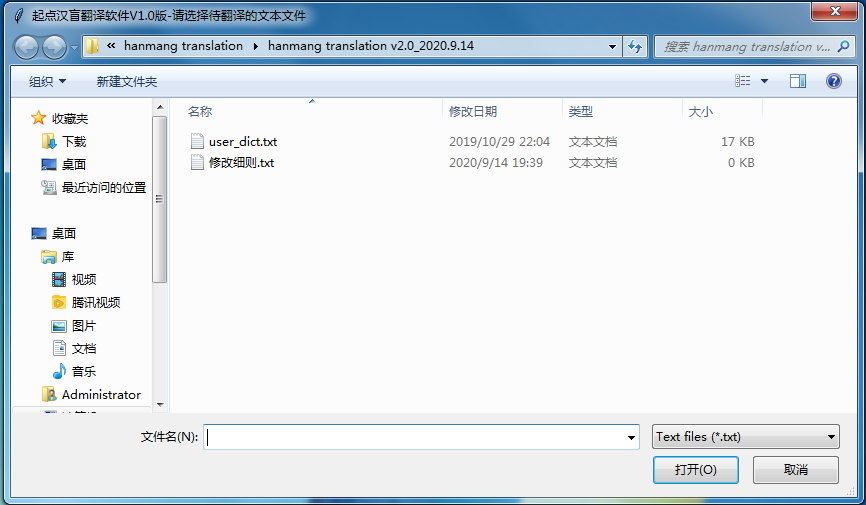


图2-3 选择翻译文件窗口



图2-4 翻译完成提示窗口

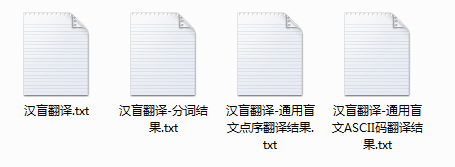


图2-5 翻译的目标文件与生成的结果文件

### 2.3.2盲文分词模块设计

该翻译系统选择了规则和统计相结合的jieba分词系统。Jieba系统是一个强大的中文分词系统，有全模式、精确模式和搜索引擎模式三种切分模式。精确模式是准确地将文字序列切分为接近语义的词语，该模式下的分词效果更加贴近语义本身，更适用于面向盲文的分词处理。如图2-6所示是盲文分词的处理算法整体流程图。



图2-6 盲文分词的处理算法整体流程图

### 2.3.3汉字拼音化及修正模块设计

该翻译系统选择调用pypinyin拼音库获得汉语拼音，得到的汉字拼音字符串数据类型为声母+韵母+声调数字，其中声调数字5代表轻声，并且切分汉语拼音的声母、韵母、声调，将其单独存放。汉字的拼音体系和盲文的拼音体系存在一定的区别于联系，于是不仅要对获取的拼音进行切分还要按照盲文声母、韵母做一些面向盲文的拼音修正。如图2-7所示是汉字拼音化以及修正模块的设计总体流程图，包括汉拼转换、拼音字符串切分、拼音修正，最终得到符合盲文的拼音字符串。



图2-7 汉字拼音化以及修正的总体流程图

汉字拼音声母韵母切分算法流程如图 2-8所示。首先，需要判断返回的拼音字符串是否为空，若为空，需要在消息窗口提示，当前汉 字返回的拼音为空。若不为空，则需要从拼音字符串的第一位开始判断，当前字符是否存 在于韵母表中，若在韵母表中能够匹配到当前字符，则说明当前字符位置为声母韵母的切 分标志位，若不能在韵母表中成功匹配，则继续匹配拼音字符串中的下个字符，直到找到 拼音字符串中的韵母第一个字母为止。找到切分标志位后，首先需要判断标志位是否为 0， 标志位等于 0 说明当前汉字字符是单独韵母拼读的，不存在声母符号。若标志位不为 0， 则为正常多数汉字的情况，拼音由声母、韵母、声调组合而成。拼音字符串的第一位到标 浙江理工大学硕士学位论文 基于未登录词识别的汉盲翻译系统研究 25 志位为声母符号，标志位到字符串倒数第二位为韵母符号，最后数字位代表声调，将切分 得到的声母值、韵母值、声调值返回给调用该函数的位置。



图2-8 汉字拼音声母韵母切分算法流程图

在研究对比了汉语拼音与盲文拼音声母韵母体系的不同后，得出结论，需要对以下几种情况作出修正处理：

1. 汉字“他，她，它，的，么，嗯”等字符和汉字拼音为“yo”的特殊化处理。

2. 当汉字拼音声母为“y,w”时的特殊化处理。

3. 当经过声母“y,w”情形修正处理后的韵母为“iou,uen,uei”三者之一的特殊化处理。

4. 当汉字韵母的第一个字母是“u”，且声母是“j,q,x”情形下的特殊化处理。

5. 当汉字的声母为“z,c,s,zh,ch,sh,r”其中之一和韵母 i 组合时的特殊化处理。

在以上列出的几个区别的基础上，对各种情况的算法进行了设计，并整合按合理的顺序设计出最终的面向盲文的拼音修正算法，最终流程图如图2-9所示。



图2-9面向盲文的拼音修正整体流程图

### 2.3.4通用盲文声调省写模块设计

国家通用盲文的标调规则有如下规定：

基本规则中包含两个规则：1.1每个音节都标调；1.2变调音节标本调；

声调省写规则中包含八条规则：2.1声母为f的音节，省写阴平符号；2.2声母为p、m、t、n、h、q、ch、r、c的音节，省写阳平符号；音节tou2的声调符号不省写；2.3声母为b、d、l、g、k、j、x、zh、sh、z、s的音节，省写去声符号；音节le4、zi4的声调符号不省写；2.4韵母自成音节，省去去声符号，第2.5和2.6条规定除外；2.5音节yi1、er2、wo3、ye3、you3的声调符号省写；音节yi4、er4、wo4、ye4、you4的声调符号不省写；2.6音节o1、o2、o3、o4的声调符号省写；音节e1、e2、e3、e4的声调符号不省写；2.7声母自成音节后面连写韵母自成音节时，声母自成音节不适用省写规则2.2和2.3；2.8需要声调符号提供点位参照或者区分音义时，声调符号应不省写；

简写规则中包含两个规则：3.1“的、么、你、他、她、它”简写；3.2“的、么、你、他、它”后面连写韵母自成音节时，不能简写。

根据上述的各调规则进行统一盲文声调省写算法的设计，一一展开分析其中基本规则在获取拼音的时候已经解决，简写规则中的两个小规则也在拼音修正模块中已经解决，这小节主要对声调省写规则中的八条规则进行逐一分析以及实现的算法介绍。

声调省写规则2.1实现，对于规则2.1只需要判断经过修正后得到的盲文拼音声母是否是f以及声调是否是阴平声调即声调是一声。如图2-10是声调省写规则2.1实现流程图,判断当前字符的声母是否为”f”且声调是”1”,如果满足条件,将声调置空,否则声调不做变化执行结束。例如“分开”，翻译得到盲文ASCII码为“f0k[a”，盲文点序为“110100001011101000010101100000”，其中“分”字的音节符合上述规则，声调省写。



图2-10 音调省写规则2.1实现流程图

声调省写规则2.2的实现，该条规则不仅要判断符合的声母和音调，还要提出一个特殊的个例。如图2-11流程图,判断当前字符的声母是否为p、m、t、n、h、q、ch、r、c之一且声调字符串为2,如果不满足条件,则声调不做变化执行结束;否则继续判断,判断判断声母字符串为“t“，韵母字符串为”ou“,如果不满足则将声调置空,否则声调不做变化执行结束。例如“和平”，翻译得到的盲文ASCII码结果为“h5p\*”，盲文点序为“110010010001111100100001”，其中“和”、“平”两个字都符合上述规则，声调省写。例如“田头”，翻译得到盲文ASCII码的结果为“t%t(1”，盲文点序为” 011110100101011110111011010000”，其中“田”满足上述规则，声调省写，“头”满足上述规则声调不省写。



图2-11 音调省写规则2.2实现流程图

声调省写规则2.3的实现，该条规则不仅要判断腐恶规则的声母和音调，还要提出两个特例。如图2-12是声调省写规则2.3实现流程图，开始，判断当前字符声母字符串为b、d、l、ɡ、k、j、x、zh、sh、z、s之一且声调为4,如果条件不满足, 则声调不做变化执行结束,否则进入下一个判断。判断声母字符串为”l”且韵母字符串为”e”,如果条件满足,则声调不做变化执行结束,否则进入下一个判断。 判断声母字符串为”z“且韵母字符串为“i”,如果条件满足,则声调不做变化执行结束,否则将声调置空。最后执行结束。例如”干劲”，翻译得到的盲文ASCII码为“gvag\* “，盲文点序为” 110110111001100000110110100001 “，”干“和”劲“均满足上述规则，音调省写。例如“快乐”，翻译得到的盲文ASCII码为“kyl52”，盲文点序为“101000101111111000010001011000”，“快”字满足上述规则，声调省写，“乐”字满足上述规则，声调不省写。



图2-12 音调省写规则2.3实现流程图

声调省写规则2.4的实现，如图2-13是声调省写规则2.4的实现流程图,开始,为pinyin字符串赋值pinyin=韵母字符串+声调字符串,然后进入一个判断,判断当前字符的声母字符串为空且声调为4声,如果条件不满足则声调不做变化执行结束,反之如果条件满足进入下一个判断。判断韵母字符串为'i','er','uo','ie','iu'之一或pinyin字符串为'e1','e2','e3','e4'之一,如果条件满足则声调不做变化执行结束,反之条件满足,将声调置空,执行结束。例如“爱国”，翻译得到的盲文ASCII码为“[go1”，盲文点序为“010101110110101010010000”，“爱”字满足上述规则，声调省写；例如“物联网”，翻译得到盲文ASCII码为“ul%17'”，盲文点序为“101001000000111000100101010000011011001000”，其中“物“字符合上述规则声调省写。



图2-13 音调省写规则2.4实现流程图

声调省写规则2.5的实现，如图2-14是声调省写规则2.5的实现流程图,开始,为pinyin字符串赋值pinyin=韵母字符串+声调字符串,然后进入一个判断,判断当前字符的声母字符串为空且pinyin是'i1','er2','uo3','iu3','ie3之一,如果条件满足,则将声调置空然后执行结束,反之条件不满足直接执行结束。例如，“我又有”，翻译得到吗，盲文ASCII码为“o \2 \”，盲文点序为“101010110011011000110011”，其中“我”和“有”字符合上述规则，声调省写。“又”符合上述规则声调不省写。



图2-14 音调省写规则2.5实现流程图

声调省写规则2.6的实现，如图2-15是声调省写规则2.6的实现流程图。开始,为pinyin字符串赋值pinyin=韵母字符串+声调字符串,然后进入一个判断,判断声母字符串为空且pinyin字符串为'o1','o2','o3','o4'之一,如果条件满足,则将声调置空然后执行结束,反之条件不满足直接执行结束。例如“哦”，翻译得到盲文ASCII码“5”，盲文点序为“010001”，“哦”字符合上述规则，声调省写。例如“恶”，翻译得到盲文ASCII码为“52”，盲文点序为“0010001011000”，符合上述规则音调不省写。



图2-15 音调省写规则2.6实现流程图

在对上述几条规则实现的算法基础上，对音调省写规则算法的执行与否设计算法执行判断。根据规则2.7和规则本身来看，声调省写函数的调用不需要所有的都执行一遍，如图2-16是对字符是汉字字符获取拼音切分和修正后，执行哪几条音调省写规则的判断流程，也是声调省写规则2.7的具体实现流程图。开始，获取当前汉字字符修正后的拼音，判断1当前汉字声母是否为空且后一个汉字是汉字字符如果条件不满足则执行所有的声调省写判断，满足则入下一个判断。判断2当前汉字字符的声母是否是声母自成音节，如果条件不满足则执行声调省写规则123（根据判断1可以确定非韵母自成音节所以不需要执行声调省写规则456的判断），满足则进入下一个判断。判断3获取并判断后一个汉字字符是否是韵母自成音节，即修正后的拼音声母为空，如果不满足则执行声调省写规则23（根据判断12可以确定为声母自成音节所以不需要执行声调省写规则1），满足则对声调不处理，即满足声调省写规则7的条件不执行声调省写规则23。例如，“事业”，翻译得到盲文ASCII码为“:2e2”，盲文点序为“100011011000100010011000”，其中，“事”为声母自成音节，“业”为韵母自成音节，符合上述规则，“事”不适用规则2.3，声调不省写，。例如“驰骋”，翻译得到盲文ASCII码为“qq#'”，盲文点序为“111110111110001111001000”，其中“驰”为声母自成音节，但是“骋”非韵母自成音节，所以“驰“不符合上述规则，而符合规则2.2调省写。



图2-16 声调省写模块断流执行流程图

### 2.3.5转换为盲文ASCII码模块设计

在上述几个模块的基础上，转换为盲文ASCII码模块是相当于经过判断算法设计，程序执行完分词模块后的文本字符串，通过逻辑判断，实现对不同字符的操作判断调用不同的函数执行。如图2-17所示。因为声调省写规则2.7的需要，不仅要获取当前字符的拼音还要对当前字符往后一个字符的拼音进行获取用于判断当前汉字字符的声调是否需要省写，所以为了防止当遍历到最后一个文本字符时再获取往后一个字符而导致的列表溢出错误，在字符串的末端添加一个空格。然后开始进入字符串的遍历，取遍历到倒数第二个元素，也就是末端空格前一个元素。程序中获取当前元素后把它转换为字符形式。开始判断当前字符是否在特殊字符字典中，如果满足，匹配对应的盲文ASCII码添加到输出列表中；如果不满足，则进入下一个判断。判断当前字符是否符合汉字Unicode编码，如果满足，获取这个字符的拼音并切分声母韵母声调，各自赋值，等待后续的调用，然后根据盲文拼音对汉语拼音进行修正，接着调用声调省写规则的实现程序执行声调省写判断处理，最后匹配处理完的当前字符拼音ASCII码添加到输出列表中；如果不满足，进入下一个判断。判断当前字符是否是空格字符，如果是，将空方的盲文ASCII码添加到输出列表中；如果不满足就将该字符判断为非汉语体系字符，打印“当前字符非汉语体系字符＋‘该字符’”再添加一个空方的ASCII码添加到输出列表中。最后遍历结束，程序返回完整的ASCII码输出列表。上述得到的ASCII码最终用于翻译得到盲文点序，输出。



图2-17任意字符转换为盲文ASCII码的算法流程图

### 2.3.6转换为盲文点序模块设计

在得到盲文ASCII码后，需要将盲文ASCII码转换为点序，将中文转换为盲文阅读器设备可有效读取的数据，是汉盲翻译最终关键环节。该软件中最后阶段将盲文ASCII码转换为盲文点序，本质是为了满足触觉点字呈现装置的编码通讯需求，同时也更加直观地反映了盲文符号的实际情况。盲文点序中，触觉凸点设计为“1”，触觉凹点设计为“0”。按照盲文单方内的“123456”的点位顺序，满方盲文符号可以表示为“111111”，空方表示为“000000”。将最终的点序连接起来，得到最终的翻译结果。如图2-18所示，获取盲文ASCII码后，通建立的盲文ASCII码和盲文点序的对应字典，采用键值对的形式，直接匹配到对应的键值，如果为空格，则添加空方点序“000000”，最终按照顺序输出为连续的盲文点序结果。



图2-18盲文ASCII码转点序算法流程图

# 第三章 程序打包与使用

为了使程序具有较好的可执行性，调用python的pyinstaller第三方库，通过对源文件的打包，Python可以在没有安装Python的环境中运行。生成后的程序在生成的dist文件夹中（如图3-1），其中的文件后缀为.exe的为可执行文件（如图3-2），其他文件是该文件的动态链接库。在没有安装Python的电脑上要运行国家通用盲文翻译系统只需复制TongYongBraille2.0文件夹到电脑上，双击TongYongBraille2.0.exe按提示操作即可。

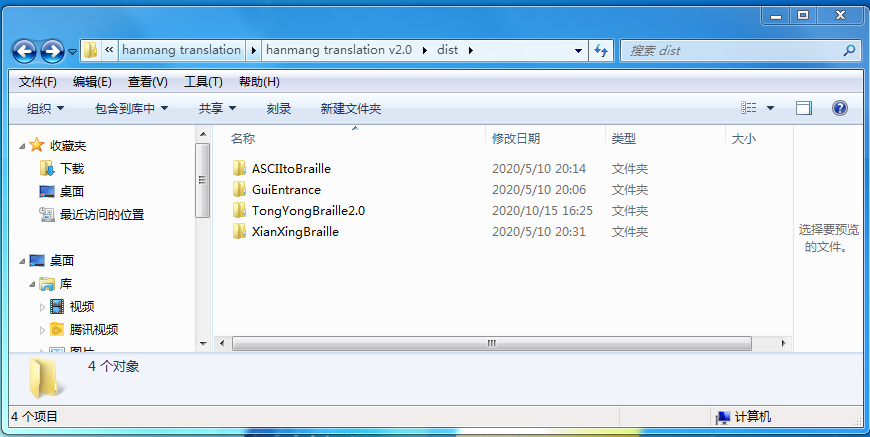


图3-1打包生成的dist文件夹

C:\Users\Administrator\Desktop\软著相关\exe.PNG

图3-2 通用盲文翻译系统可执行文件

# 第四章 总结

本翻译系统，通过对国家通用盲文规则的分析，以及盲文与汉语之间的区别，设计程序算法，运行后能得到基本符合盲文规则的盲文分词结果、国家通用盲文ASCII码结果和国家通用盲文点序结果，符合国家通用盲文的推广以及盲文信息化的大进程。

但是目前的翻译系统，在分词模块还未完全贴切人工的盲文分词，还可以进一步改进。同时该翻译系统对于“国家通用盲文音调省写规则2.8需要声调符号提供点位参照或者区分音义时，声调符号应不省写。”还未实现，可以继续改进算法，进一步完善。是结果完全符合国家通用盲文方案。