

ICS 35.040
L 71



中华人民共和国国家标准

GB/T 21023—2007

中文语音识别系统通用技术规范

General specification for Chinese speech recognition system

2007-06-29 发布

2007-11-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会

发布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 语音识别系统分类及表述	2
4.1 基于用途的分类	3
4.2 基于词汇量的分类	3
4.3 基于所能识别的说话人群的分类	3
4.4 基于系统工作模式的分类	3
4.5 语音识别系统的应用环境	4
4.6 语音识别系统使用背景环境	4
4.7 语音识别系统表述规范	4
5 语音识别系统性能指标	5
5.1 概述	5
5.2 识别性能分类指标	5
5.3 响应时间指标	6
5.4 系统分类指标要求	6
6 语音输入和识别输出	6
6.1 语音识别输入准则	6
6.2 语音识别输出准则	6
7 语音识别系统性能测试方法	6
7.1 概述	6
7.2 测试语料设计	6
7.3 测试语音录制	7
7.4 基于语音识别标准库的测试方法	7
7.5 基于现场口呼的测试方法	7
7.6 测试报告内容	7
8 语音识别标准库及规范	7
8.1 概述	7
8.2 语音识别标准库语音特性的描述	7
8.3 语音识别标准库的标注	8
8.4 语音识别标准库的来源	8
8.5 语音识别标准库的数据转换	8
附录 A (资料性附录) 听写系统有关符号读法	9
附录 B (资料性附录) 语音识别用无调汉语拼音表	11
参考文献	12

前 言

请注意本标准的某些内容有可能涉及专利。本标准的发布机构不承担识别这些专利的责任。

本标准的附录 A、附录 B 是资料性附录。

本标准由中华人民共和国信息产业部提出。

本标准由全国信息技术标准化技术委员会归口。

本标准由中国电子技术标准化研究所、中国科学院自动化研究所、清华大学、中国社会科学院语言研究所起草。

本标准主要起草人：徐波、孙甲松、李爱军、徐明星、黄泰翼、鲍怀翘、尹波、吴志刚。

中文语音识别系统通用技术规范

1 范围

本标准规定了中文语音识别系统的基本术语、分类、技术指标、测试方法、输入/输出以及中文语音识别标准库等。

本标准适用于各种计算机、网络和终端设备配置的中文语音识别系统。本标准的制定和实施主要用于指导中文语音识别系统的研制、应用和测试。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB 18030—2005 信息技术 中文编码字符集

3 术语和定义

下列术语和定义适用于本标准。

3.1

语音识别 **speech recognition**

将人类的声音信号转化为文字或者指令的过程。

3.2

语音识别系统 **speech recognition system**

具有语音识别功能的开发工具、软件、装置或应用。

3.3

词汇表 **vocabulary list**

语音识别系统所包含和处理的所有词汇。

3.4

激活词汇 **active word**

语音识别系统在某个时间点能同时识别的所有词汇。

3.5

发音 **utterance**

用户输入的一个语音单位,可以是词、短语或句子。发音单元之间需有有意、明显停顿。

3.6

孤立字语音 **isolated word speech**

以字为发音单元,字与字之间需有有意、明显停顿的语音。

3.7

连接词语音 **connected word speech**

以词为发音单元,词与词之间需有有意、明显停顿的语音。

3.8

连续语音 **continuous speech**

词与词之间无需刻意停顿的自然发音语音。

注:包括朗读语音和自然语音。

3.9

汉语通用语 Mandarin and its dialect variation

普通话以及普通话的地方变体。

注：为中国人跨地区交流时最常见之普通话。

3.10

同构语音数据 isomorphic speech data

除说话人不同外,其他语音属性相同或基本相同的语音数据集合。

3.11

语音识别工作模式 working mode for speech recognition

语音识别系统对说话方式、内容以及句型等的限制。

3.12

标准配置 standard configuration

正常运行语音识别所需要的基本计算机配置。

注：如 CPU、内存、网络、麦克风、A/D 精度要求等,由生产厂家给出。

3.13

系统自适应 speaker adaption

语音识别系统通过用户发音完成对识别模型参数修改的自适应过程。

3.14

用户自适应 user adaption

用户熟悉应用系统、了解自动语音识别系统特性的学习过程。

3.15

删除错误 deletion error

用户所发语音在语音识别结果中没有出现的错误。

3.16

插入错误 insertion error

用户没有发的语音在识别结果中出现的错误。

3.17

替换错误 substitution error

用户所发语音被识别成其他语音的语音识别错误。

3.18

被测系统 system for testing

参加评估的语音识别系统。

3.19

测试系统 testing system

对被测系统进行评估的系统和体系。

3.20

副语言学 paralinguistic feature

言语中非音段的语音特征以及种种非语音信号。

注：这些行为被充分用来对总的交际效应作出贡献。

4 语音识别系统分类及表述

语音识别系统可以基于不同的用途、不同的词汇量、所能适应的不同人群、不同的工作模式、不同的应用平台进行分类。基本分类可以融合,例如某个系统可以是“基于 PC 平台的超大词汇量特定人群书

面语音识别系统”。

4.1 基于用途的分类

语音识别系统从用途上可以分为:文字输入类、声音检测识别类和对话类三类系统。

4.1.1 文字输入类

以文字录入为主要目的,要求把语音转化成文字的系统。

注:如听写机系统。

4.1.2 声音检测识别类

主要以检测连续或者孤立语流中的特定命令或者关键词为主要目的的系统。

注:如命令控制、声控玩具、关键词识别系统等。

4.1.3 口语对话类

主要采用语音方式进行信息获取为目的的系统。

注:如股票查询系统、电话订票系统、地理信息查询等。

4.2 基于词汇量的分类

语音识别系统的词汇量按照最大激活词汇量标定。语音识别系统从所能处理的词汇量来分类,可以分为小词汇量、中小词汇量、中词汇量、中大词汇量、大词汇量、超大词汇量六种。

4.2.1 小词汇量

系统所能识别的词汇量小于 127 的系统。

4.2.2 中小词汇量

系统所能识别的词汇量在 128~1023 之间的系统。

4.2.3 中词汇量

系统所能识别的词汇量介于 1024~4095 的系统。

4.2.4 中大词汇量

系统所能识别的词汇量介于 4096~16383 之间的系统。

4.2.5 大词汇量

系统所能识别的词汇量介于 16384~65535 的系统。

4.2.6 超大词汇量

系统所能识别的词汇量大于 65536 的系统。

4.3 基于所能识别的说话人群的分类

语音识别从所能识别的说话人群来区分,可以分为特定人识别、特定人群识别和非特定人识别。

4.3.1 特定人识别

需要经过专门学习(训练)并建立该发音人专用模型的系统。

4.3.2 特定人群识别

系统只能识别具有某种发音特性的人群的识别系统。

注:如标准普通话、成年人的标准普通话、儿童口音的标准普通话、上海口音的普通话等。

4.3.3 非特定人识别

面向一般用户通用语的识别系统,对说话人语音或口音没有专门限制,不需要使用者进行训练的系统。

4.4 基于系统工作模式的分类

语音识别系统从其工作模式来分,可以分为孤立字语音识别、连接词语音识别、有限书面语语音识别、有限口语语音识别、无限书面语语音识别、无限口语语音识别以及关键词检测七种。

4.4.1 孤立字语音识别

语音识别系统能识别单个字,系统所需要识别的发音的每个音节前后都应有明显的发音停顿。

4.4.2 连接词语音识别

语音识别系统能识别单个词或短语发音,系统所需要识别的发音单位(词或短语)前后都应有明显

的发音停顿。

4.4.3 有限书面语语音识别

语音识别系统能识别在词汇、语法、语义以及语用等受到限制的书面语连续语音。

4.4.4 有限口语语音识别

语音识别系统能识别在词汇、语法、语义以及语用等受到限制的口语化语音。

4.4.5 无限书面语语音识别

语音识别系统能识别以朗读方式进行的没有特别语言限制的连续语音。

注：如听写机系统。

4.4.6 无限口语语音识别

语音识别系统能识别口语化的没有特别语言限制的连续语音。

4.4.7 关键词检测

语音识别系统能从口语连续语流中找到系统所定义的词汇或短语。

4.5 语音识别系统的应用环境

应用环境包括语音识别系统的具体平台以及所需要的标准配置。语音识别从应用平台来看,有 PC 平台、电话平台、嵌入式平台以及分布式平台等。

4.5.1 PC 平台

语音通过麦克风、PC 上的声卡等输入设备完成语音采样并在 PC 上完成识别的系统。

4.5.2 电话平台

语音通过电话通道、声卡和传输并在相应的服务端完成识别的系统。

4.5.3 嵌入式平台

识别系统是嵌入在某个信息终端产品上(如 HPC、PDA、手机和信息家电等),通过特定设备输入语音并完成识别的系统。

4.5.4 分布式平台

语音识别特征提取在前端嵌入式设备上完成,语音识别其他计算在后台系统和计算机上完成的系统。

4.6 语音识别系统使用背景环境

使用背景环境以信噪比来衡量,信噪比定义为:

$$SNR = 10\lg(P_s/P_n)$$

式中:

P_s ——信号的功率;

P_n ——噪声的功率。

4.7 语音识别系统表述规范

各类语音识别系统应按 4.1~4.6 所述的用途分类、词汇量规模、所适用的说话人群、工作模式、应用环境以及使用背景环境,在系统说明书以及提供测试时分别加以详细说明:

- a) 该系统的用途分类,不同类型的系统在性能和指标上有不同的分类要求。
- b) 应提供系统所能处理的词汇量和词汇清单;在词汇和词汇量不能精确给定的条件下(如有些系统完全由用户定义词汇),系统应按 4.2 给出系统所能处理的词汇量等级。
- c) 应按 4.3 给出其所能识别的说话人人群的具体限制及适用范围。对发音人普通话的标准程度的要求,可参照汉语普通话水平等级进行定义。
- d) 应按 4.4 给出系统的工作模式;对于有限语音识别,系统应给出有限语音的详细说明。
- e) 每种平台具有独特的语音识别器所要处理的语音特性:不同平台上麦克风与话者的距离限制,麦克风性能要求,支持的 A/D 转换精度和采样率等。语音识别系统应按照 4.5 所涉及的内容描述其具体的应用环境。

f) 应按 4.6 给出能正常工作的信噪比范围及在此范围内每隔 5dB 的性能变化曲线。

5 语音识别系统性能指标

5.1 概述

语音识别系统的性能评价按照系统的用途分类建立测试标准,多种用途的需求分别满足多种指标的要求。这些要求只同系统的用途有关,而同系统的其他限制(如词汇量、工作模式、应用条件等)无关,但系统给出的具体指标应明确在何种条件下成立。

5.2 识别性能分类指标

语音识别系统采用分类测试的原则,定义相应的分类指标。

5.2.1 文字输入类语音识别系统性能指标定义

该类应用用户主要关心最后输入的效率,因而应同时考虑系统插入、删除和替换等错误。

设正确文本字数为 N ,识别结果文本字数为 M ,按照识别结果文本与正确文本根据“最小代价匹配”原则运用动态规划算法,得到正确识别字数 M_c 、删除错误字数 D 、插入错误字数 I 和替换错误字数 S ,则有:

$$N = M_c + S + D;$$

$$M = M_c + S + I;$$

定义以下性能指标:

$$\text{字匹配率: } M_{cr} = (M_c / N) \times 100\%$$

$$\text{替代错误率: } S_{er} = (S / N) \times 100\%$$

$$\text{插入错误率: } I_{er} = (I / N) \times 100\%$$

$$\text{删除错误率: } D_{er} = (D / N) \times 100\%$$

$$\text{字错误率: } WER = ((S + I + D) / N) \times 100\%$$

$$\text{字准确率: } WCR = ((M_c - I) / N) \times 100\% = 100\% - WER$$

以 WER 或 WCR 度量系统性能指标。

5.2.2 声音检测识别类语音识别系统性能指标定义

该类型系统应兼顾动作的可靠性问题,指标定义如下:

假设关键词表的词汇量为 KW (个),检测语音长度为 HR (小时),出现关键词 N (次), C 为每小时每个关键词最大容忍的误报个数(一般取 10),系统报出关键词 M (个),其中,正确 FD (个),错误 FA (个), $FD + FA = M$ 。则:

$$\text{漏报个数: } FR = N - FD$$

$$\text{误报率: } F_{ar} = (FA / (KW \times HR \times C)) \times 100\%$$

$$\text{漏报率: } F_{rr} = (FR / N) \times 100\%$$

$$\text{检出率: } F_{dr} = (FD / N) \times 100\%$$

DET 曲线:以 F_{ar} 为横轴, F_{rr} 为纵轴画出的曲线。

ROC 曲线:以 F_{ar} 为横轴, F_{dr} 为纵轴画出的曲线。

等错率 EER :DET 曲线上 $F_{ar} = F_{rr}$ 时, F_{ar} 或 F_{rr} 的值

质量因数 FOM :ROC 曲线上 $F_{ar} = 0\%, 10\%, 20\%, \dots, 100\%$ 时, F_{dr} 的算术平均值

一般以 EER 或 FOM 值度量系统性能指标。

5.2.3 对话类语音识别系统性能指标定义

该类系统性能定义为:

N = 系统完成一个信息查询并获取到正确信息所需的平均语轮数;

对话效率 = $(1/N) \times 100\%$;

对话正确响应率 = 系统对信息咨询过程每次语音输入的平均正确响应率。

5.3 响应时间指标

系统的响应时间统一采用实时系数概念。在系统的标准配置条件下,假设发音从 T_s 开始,发音结束时间为 T_e ,识别结束时间为 T_r ,则实时系数定义为:

$$\text{实时系数} = (T_r - T_s) / (T_e - T_s)$$

5.4 系统分类指标要求

按照 4.7 表述要求给出语音识别系统的技术指标,具体测试条件选项如第 7 章所述。在 15 dB 使用背景环境下,不同用途的系统的技术指标应达到如下要求:

- 文字输入类系统:字准确率应大于或等于 85%。
- 声音检测识别类系统:等错误率小于或等于 30%。
- 口语对话类系统:对话效率大于或等于 40%;对话正确响应率应大于或等于 75%。

所有系统的实时系数应小于或等于 1.5。

6 语音输入和识别输出

6.1 语音识别输入准则

语音识别输入准则应为:

- 发音单元之间的间隔不超过 2 s;发音单元的持续时间不应小于 0.2 s;停顿时间超过 2 s,认为一次语音输入结束。
- 汉语语音识别系统应支持汉语通用语,从信息社会发展角度支持推广普通话。
- 语音识别系统对于大于 0.2 s 的语音在 SNR 大于 15 dB 的背景环境条件下能做出灵敏的反应。
- 对于文本中的非汉字字符包括数字、电话号码、标点符号,其可参照附录 A 所定义的方法朗读。

6.2 语音识别输出准则

语音识别输出准则应为:

- 文字类输出的语音识别结果为汉字内码、系统所定义的符号,其输出应满足 GB 18030—2005。
- 声音检测识别类语音识别输出结果可以是命令词、关键词或预定义代码(Pre-defined ID)。
- 对于声音检测识别类以及对话类识别系统,应给出系统对识别的可信度衡量,可信度范围为从 0~100 的整数。
- 声音检测识别类以及对话类系统输出应是或应能转换为语音合成系统可以识别的符号,以用于人机语音合成。

7 语音识别系统性能测试方法

7.1 概述

为保证语音识别系统测试的再现性,测试应尽量采用基于语音识别标准库的测试方法。无法采用基于语音识别标准库测试的,可采用基于现场口呼的测试方法。测试语料的设计与语音库的录制应保证与实际使用场景的一致性,测试的结果以满足规范的测试报告形式给出。

7.2 测试语料设计

测试语料应从词汇量覆盖、测试点覆盖以及领域覆盖等加以设计。测试集文本上分成若干组,每组可以由若干人发音组成。设计要求如下:

- 对于小词汇量每组测试集应包含所有词汇。
- 对于中小词汇量每组测试集的合集应覆盖系统的所有词汇量。
- 对于中大词汇量以上的测试集,每组测试集词汇的合集应考虑尽可能多地覆盖系统的词汇量。
- 无限连续语音识别应从不同领域、不同时期的语料中随机选择,同时考虑发音现象的覆盖性。

- e) 有限语言的语音识别应充分考虑句型、词汇、语义等的覆盖性。

7.3 测试语音录制

测试语音录制要求如下：

- a) 说话人的选择应在符合系统对说话人限制的条件下,尽可能选择具有代表性和统计分布规律的发音人,特别是考虑不同口音、不同年龄、不同语速、不同教育背景、不同说话韵律等因素。
- b) 无论是特定人还是非特定人的语音识别系统,其测试的发音人至少为 30 个人以上;每人发音测试语料中的一组或多组语料;不同发音人尽量采用不同语料组。
- c) 对于非特定人系统,特别应强调对含有一定口音的汉语语音适应能力和汉语声调问题。
- d) 测试语音的录制应与系统说明中的平台、采样率、输入通道等保持相对一致或接近;录音过程至少包括录音、标注和确认三个步骤,保证测试数据库的正确性。

7.4 基于语音识别标准库的测试方法

基于语音识别标准库测试指采用录制的语音数据对被测系统进行直接或间接的测试。方法说明如下：

- a) 直接测试指利用被测系统带有的应用程序输入/输出接口,直接利用语音识别标准库中的语音文件进行测试。
- b) 间接测试指测试系统利用高保真回放设备把语音识别标准库中的语音通过双方认可的方式输出到被测系统中。

被测系统应至少提供两种测试方法中的一个接口。

7.5 基于现场口呼的测试方法

现场口呼测试基本情况同测试语音的录制。

- a) 需要有两个以上识别结果记录者,记录被测系统对当前发音的输出结果;记录表应包括发音人、记录人、操作人、监督人、发音内容、语音识别结果等内容。
- b) 全部发音者测试结束后,统一按照性能标准进行指标评估。评估至少有两个人以上参与。
- c) 对于识别结果能以文件形式给出的,被测系统按照发音人还应给出文件形式的输出结果,以便自动评测。

7.6 测试报告内容

测试后应提交标准测试报告。报告应由以下几部分构成：

- a) 按照第 4 章规定的所有内容对被测系统的完整描述。
- b) 按照第 8 章描述的测试数据的语音属性、测试词汇以及测试说话人的选择及确定情况。
- c) 按照第 5 章定义的每个人识别各项指标以及平均识别指标。
- d) 测试过程的情况记录,采用的测试方法及运行过程的流畅性。
- e) 被测系统的配置情况。

8 语音识别标准库及规范

8.1 概述

语音识别标准库分为训练库、评估库以及测试库三种。训练库主要提供给语音识别研究和开发者训练相关声学模型;评估库用于提供给研究和开发者进行系统的自我验收和评估;测试库用于测试中心评估有关系统。训练库和评估库由评测机构定期向社会公布,并提供相应的存储介质;测试库则由评估机构管理。评估库与测试库应是同构语音数据。

8.2 语音识别标准库语音特性的描述

语音识别标准库除语音数据的文件载体外,还应提供关于语音库本身与语音数据的完整描述。需要描述的特性至少包括：

- a) 语音库所用语料的语体:对话语篇语音库、独白语篇语音库等。

- b) 语音库话语的自然程度:
 - 1) 朗读语音语料库:指有计划的、按照文字朗读的语音库,内容可以是:音节、词和短语、句子、对话语篇或独白语篇;
 - 2) 自然口语语音语料库:随意即兴式的自然口语音库,内容可以是对话语篇或独白语篇。
- c) 信号的频宽特性:如 0~4 k 的电话语音库等。
- d) 录音特性:录音时间、录音地点、录音通道、噪声环境等;其中:
 - 1) 录音时间用“××××年××月××日”表示;
 - 2) 录音通道包括 PC、电话、嵌入等;
 - 3) 麦克风指标包括类型、名称、灵敏度以及阻抗、低频、高频截止频率。
- e) 说话人特性:姓名、性别、籍贯、成长地、文化程度、年龄、口音的方言区。
- f) 数据特性:采样频率、采样精度、通道数目、通道噪声等。

8.3 语音识别标准库的标注

语音识别标准库的标注包括基本标注和扩展标注。所有语音识别标准库都应包含基本标注。基本标注包括:

- a) 汉字与音节标注:即语音信息的汉字表述和相应的音节表述。这些信息既可提供给识别系统使用,也能为语言学研究提供素材。
- b) 口语到文字转写工作的要点是口音和副语言学现象等的准确转写。文字标注应标明基本文字信息以及副语言学现象,基本标注中的副语言学现象可用通用副语言学符号表示。
- c) 音节标注采用标准普通话音节标注(参见附录 B),音节标注应为有调标注。声调标注中 0 表示轻声,1 表示阴平,2 表示阳平,3 表示上声,4 表示去声。

8.4 语音识别标准库的来源

语音识别标准库实行开放式管理,各种数据可以通用或专用协议方式加入标准库。

- a) 本标准实施后由国家投资的新的语音库,应符合本标准的各项要求,经评测机构验收合格后成为语音识别标准库。
- b) 已有语音库和非国家投资的语音库,经数据自我检验,保证信息完整后可提出申请;由评测机构对数据进行认定和验收后成为语音识别标准库。

标准语音库应定期通过网络等手段予以公布和发布。

8.5 语音识别标准库的数据转换

为提高语音识别标准库的利用效率,可依据有关国家标准和算法进行不同采样率、不同声学通道、不同麦克风以及不同采样精度数据库格式的转换。这些算法涉及滤波算法、通道模拟、采样率转换以及其他模拟场景的仿真。仿真目标环境(如电话通道等)应符合国家有关的通道标准。转换应提供有关技术资料,转化后的语音库由专家委员会组织鉴定后成为语音识别标准库。详细技术资料以及算法应同转化后的数据库一起公布。

附 录 A
(资料性附录)
听写系统有关符号读法

A.1 数字的读法和输出转换

A.1.1 数字序列的读法和输出转换

数字序列在标准书写格式上为数目不限的单个数字(0、1、2、3、4、5、6、7、8、9;零、一、二、三、四、五、六、七、八、九)排列构成的数。其标准读法为由左至右依次读出各个数字。例如:123456的读法为:一二三四五六。对于语音识别系统而言,遵循数字序列标准读法的语音输入,产生数字序列标准书写格式的识别输出。识别结果中单个数字采用何种形式(小写数字、大写数字),由具体应用规定。

A.1.2 整数的读法和输出转换

数学上,整数有其标准的书写格式和读法。例如:整数1024(大写:一千零二十四),其标准读法为:一千零二十四。对于语音识别系统而言,遵循整数标准读法的语音输入,产生整数标准书写格式的识别输出。识别结果以何种形式(小写、大写)表示,由具体应用规定。

A.1.3 小数的读法和输出转换

数学上,小数有其标准的书写格式和读法。例如:小数12.38(大写:十二点三八),其标准读法为:十二点三八。对于语音识别系统而言,遵循小数标准读法的语音输入,产生小数标准书写格式的识别输出。识别结果以何种形式(小写、大写)表示,由具体应用规定。

A.1.4 分数的读法和输出转换

数学上,分数有其标准的书写格式和读法。例如:分数 $\frac{2}{5}$ (大写:五分之二),其标准读法为:五分之二。对于语音识别系统而言,遵照分数标准读法的语音输入,产生分数标准书写格式的识别输出。识别结果以何种形式(小写、大写)表示,由具体应用规定。

A.1.5 货币的读法和输出转换

货币的标准书写格式分为大小写两种形式。小写货币的书写格式为整数或小数,如150元,47.28元。大写货币的书写格式为整数的大写形式(对于其小写书写格式为整数的情况)或带货币单位(元、角、分)的通俗写法。如:一百五十元,四十七元两角八分。货币的标准读法有两种:一是整数/小数的标准读法+“元”,如:四十七点二八元;二是带货币单位(元、角、分)通俗写法的直接读法,如:四十七元两角八分。对于语音识别系统而言,遵循货币标准读法的语音输入,产生货币标准书写格式的识别输出。识别结果以何种形式(小写、大写)表示,由具体应用规定。

A.2 电话号码的读法和输出转换

普通固定电话号码的标准书写形式为:[国际代号][区号]市话号码[分机号]([]中的内容表示非必选),其中各组成部分均为小写格式数字序列。在各组成部分代码之间还可以有选择的添加一个“-”号,以示区别,如0086-0551-3601363-815。普通寻呼号码与普通固定电话号码完全相同,只是将[分机号]换成[寻呼号]。移动电话号码则直接以若干位小写数字表示(小写格式数字序列)。

普通固定电话号码与普通寻呼号码的读法为:[国际代号]+停顿+[区号]+停顿+市话号码+停顿+[分机号/寻呼号],其中国际代号、区号、市话号码、分机号/寻呼号的读法均为数字序列读法,参见A.1.1中的数字序列的读法。

移动电话号码则直接按数字序列读法进行。

各种电话号码的识别输出结果均为其标准书写形式,可以根据具体应用的要求决定是否采用“-”连接各个部分。缺省情况下采用“-”连接各个部分。

A.3 标点符号的读法和输出转换

汉语中常用的标点符号大约有 20 多种,每种标点符号均有其标准的书写格式和读法,如表 A.1 所示。

表 A.1 标点符号读音表

符号	读法	符号	读法	符号	读法	符号	读法
。	句号	…	省略号	~	波浪号	[左方括号
,	逗号	—	破折号	(左[圆]括号]	右方括号
、	顿号	·	点号)	右[圆]括号	{	左大(花)括号
;	分号	‘	左单引号	<	左尖括号	}	右大(花)括号
:	冒号	’	右单引号	>	右尖括号		
?	问号	“	左双引号	《	左书名号		
!	感叹号	”	右双引号	》	右书名号		

注:对语音识别系统而言,遵循标点符号标准读法的语音输入,产生标点符号标准书写格式的识别输出。

A.4 单个英文字母、首字母缩写词的读法及其输出转换

26 个英文字母(大写:A~Z,小写 a~z)的读法遵循其国际音标发音。对语音识别系统而言,输出结果为该英文字母本身。输出结果采用大小写形式,取决于具体应用的要求。

首字母缩写词的标准书写格式有三种:

- a) 各个字母依次排列,字母之间无分隔符号,如:USA;
- b) 各个字母依次排列,字母之间以“.”分隔,整个词以“.”结束,如:U. S. A. ;
- c) 各个字母依次排列,字母之间无分隔符号,整个词以“.”结束,如:USA. 。

首字母缩写词的标准读法为:从左到右依次读出各个英文字母。对语音识别系统而言,遵循首字母缩写词标准读法的语音输入,产生首字母缩写词标准书写格式的识别输出。识别结果采用三种标准格式中的哪一种,取决于具体应用的要求。

A.5 单个英文单词的读法及其输出转换

语音识别系统在定义了一个英文单词的有限集合后,可以对该集合中单个英文单词进行识别。英文单词的读法遵循其国际音标发音。对语音识别系统而言,遵循英文单词国际音标发音的语音输入,输出的识别结果为该英文单词集合中识别到的英文单词本身。

A.6 中英文混合短语或句子的读法及其输出转换

中英文混合短语或句子,其读法由左至右读出其中的每一个中文、英文单词或短语,英文单词的读法参照第 A.5 章中的描述。语音识别系统输出的识别结果为该中英文混合短语或句子。

附录 B

(资料性附录)

语音识别用无调汉语拼音表

表 B.1 语音识别用无调汉语拼音表

a	wo	du	pie	kou	nen	lo	jue	zhuang	chai	sao	fei
ai	ba	duan	pin	kong	nei	lou	jun	zhui	chan	se	feng
an	bai	dui	ping	ku	neng	long	za	zhun	chang	sen	fou
ang	ban	dun	pou	kua	ni	lu	zai	zhuo	chao	seng	fo
ao	bang	duo	po	kuan	nian	luan	zan	ca	che	si	fu
e	bao	ga	pu	kuai	niang	lun	zang	cai	chen	sou	ha
en	ben	gai	ta	kuang	niao	luo	zao	can	cheng	song	hai
ei	bei	gan	tai	kui	nie	lv	ze	cang	chi	su	han
er	beng	gang	tan	kun	nin	lue	zen	cao	chou	suan	hang
eng	bi	gao	tang	kuo	ning	ran	zei	ce	chong	sui	hao
yi	bian	ge	tao	ma	niu	rang	zeng	cen	chu	sun	he
ya	biao	gen	te	mai	nong	rao	zi	ceng	chuan	suo	hen
yan	bie	gei	tei	man	nou	re	zou	ci	chuai	sha	hei
yang	bin	geng	teng	mang	nu	ren	zong	cou	chuang	shai	heng
yao	bing	gou	ti	mao	nuan	reng	zu	cong	chui	shan	hou
ye	bo	gong	tian	me	nuo	ri	zuan	cu	chun	shang	hong
yin	bu	gu	tiao	men	nv	rou	zui	cuan	chuo	shao	hu
ying	da	gua	tie	mei	nue	rong	zun	cui	xi	she	hua
yong	dai	guan	ting	meng	la	ru	zuo	cun	xia	shei	huan
yo	dan	guai	tou	mi	lan	ruan	zha	cuo	xian	shen	huai
you	dang	guang	tong	mian	lang	ruì	zhai	qi	xiang	sheng	huang
yu	dao	gui	tu	miao	lai	run	zhan	qia	xiao	shi	hui
yuan	de	gun	tuan	mie	lao	ruo	zhang	qian	xie	shou	hun
yue	dei	guo	tui	min	le	ji	zhao	qiang	xin	shu	huo
yun	den	pa	tun	ming	lei	jia	zhe	qiao	xing	shua	
ou	deng	pai	tuo	miu	leng	jian	zhei	qie	xiong	shuan	
weng	di	pan	ka	mou	li	jiang	zhen	qin	xiu	shuai	
o	dia	pang	kai	mo	lia	jiao	zheng	qing	xu	shuang	
wu	dian	pao	kan	mu	lian	jie	zhi	qiong	xuan	shui	
wa	diao	pen	kang	na	liang	jin	zhou	qiu	xue	shun	
wan	die	pei	kao	nai	liao	jing	zhong	qu	xun	shuo	
wai	ding	peng	ke	nan	lie	jiong	zhu	quan	sa	fa	
wang	diu	pi	kei	nang	lin	jiu	zhua	que	sai	fan	
wei	dou	pian	ken	nao	ling	ju	zhuan	qun	san	fang	
wen	dong	piao	keng	ne	liu	juan	zhuai	cha	sang	fen	

参 考 文 献

- [1] 汉语拼音方案. 1958年2月11日第一届全国人民代表大会第五次会议批准
 - [2] 普通话异读词审音表. 国家语言文字工作委员会, 国家教育委员会, 广播电视部. 1985年12月联合修订
-