



# 中华人民共和国国家标准

GB/T 36464.1—2020

---

## 信息技术 智能语音交互系统 第 1 部分：通用规范

Information technology—Intelligent speech interaction system—  
Part 1: General specifications

2020-04-28 发布

2020-11-01 实施

国家市场监督管理总局 发布  
国家标准化管理委员会

## 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 系统通用功能框架 .....	4
5 语音交互界面要求 .....	5
5.1 语音采集 .....	5
5.2 语音播报 .....	5
5.3 输入输出 .....	5
5.4 环境噪声适应能力 .....	5
6 数据资源要求 .....	5
6.1 音频数据 .....	5
6.2 文本数据 .....	5
7 前端处理要求 .....	6
7.1 语音唤醒 .....	6
7.2 声源定位 .....	6
7.3 声纹识别 .....	6
7.4 语音增强 .....	7
7.5 格式转换 .....	7
7.6 重采样 .....	7
8 语音处理要求 .....	7
8.1 语音识别 .....	7
8.2 语义理解 .....	8
8.3 语音合成 .....	8
8.4 端点检测 .....	8
8.5 语音编解码 .....	9
8.6 全双工交互 .....	9
8.7 情感计算 .....	9
9 服务接口要求 .....	9
10 应用业务处理要求 .....	9
附录 A(资料性附录) 部分参数及其计算方法 .....	10
A.1 概述 .....	10
A.2 拾音距离 .....	10
A.3 语音交互成功率 .....	10
A.4 语音唤醒 .....	10

A.5	语音识别 .....	11
A.6	语义理解 .....	11
A.7	语音合成 .....	12
A.8	语音质量 .....	12
A.9	声纹识别率 .....	13
A.10	语音编解压缩率 .....	13
A.11	语音增强 .....	13
A.12	声源定位 .....	13
A.13	语音打断成功率 .....	13
参考文献 .....		15



## 前 言

GB/T 36464《信息技术 智能语音交互系统》拟分为以下几个部分：

- 第1部分：通用规范；
- 第2部分：智能家居；
- 第3部分：智能客服；
- 第4部分：移动终端；
- 第5部分：车载终端；

.....

本部分为 GB/T 36464 的第1部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本部分由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本部分起草单位：中国电子技术标准化研究院、科大讯飞股份有限公司、沈阳新松机器人自动化股份有限公司、深圳市优必选科技股份有限公司、中国科学院自动化研究所、上海智臻智能网络科技股份有限公司、中国盲人协会、中国电信集团有限公司、百度在线网络技术(北京)有限公司、中科讯飞互联(北京)信息科技有限公司、上海计算机软件技术开发中心、北京小米移动软件有限公司、西宁市大数据服务管理局、中汽研(天津)汽车工程研究院有限公司、西宁大数据有限公司、极限元(杭州)智能科技有限公司、中国移动通信有限公司研究院、厦门盈趣科技股份有限公司、深圳市腾讯计算机系统有限公司、重庆中科云从科技有限公司、中国医学科学院生物医学工程研究所、广东省标准化研究院、苏州思必驰信息科技有限公司、出门问问信息科技有限公司、哈尔滨亿时代数码科技开发有限公司、四川赛闻检测股份有限公司。

本部分主要起草人：董建、马万钟、徐洋、张锋、袁杰、陶建华、王海坤、叶雷鸣、李洪亮、李庆忠、杨震、吕彦锋、宋文林、穆亚敏、潘榕、蔡立志、朱亚军、周伟、孟宪明、孙巍、温正棋、鲍薇、杨本植、陈建成、戴小兰、翟尤、李军、蒲江波、王佳胜、龙梦竹、陈威钢、赵婕、冯丽。

# 信息技术 智能语音交互系统

## 第1部分：通用规范

### 1 范围

GB/T 36464 的本部分给出了智能语音交互系统通用功能框架,规定了语音交互界面、数据资源、前端处理、语音处理、服务接口、应用业务处理等功能单元要求。

本部分适用于智能语音交互系统的通用设计、开发、应用和维护。

### 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 11460 信息技术 汉字字型要求和检测方法

GB 18030 信息技术 中文编码字符集

GB/T 21024—2007 中文语音合成系统通用技术规范

GB/T 34083—2017 中文语音识别互联网服务接口规范

GB/T 34145—2017 中文语音合成互联网服务接口规范

SJ/T 11380—2008 自动声纹识别(说话人识别)技术规范

### 3 术语和定义

下列术语和定义适用于本文件。

#### 3.1

**语音交互 speech interaction**

人类和功能单元之间通过语音进行的信息传递和交流活动。

[GB/T 36464.2—2018,定义 3.1]

#### 3.2

**语音交互系统 speech interaction system**

由功能单元(或其组合)、数据资源等组成的能够实现与人类之间进行语音交互的系统。

[GB/T 36464.2—2018,定义 3.2]

#### 3.3

**智能语音交互系统 smart speech interaction system**

以语音识别、语义理解、语音合成等全部或部分人工智能技术为基础,由智能软硬件组成,具备智能人机交互能力的语音交互系统。

#### 3.4

**人机交互 human machine interaction**

人类和功能单元之间,为完成确定任务,以一定的交互方式进行的信息传递和交流活动。

3.5

**功能单元 functional unit**

能够完成特定任务的硬件实体,或软件实体,或硬件实体和软件实体。

[GB/T 5271.1—2000,定义 01.01.40]

3.6

**语音合成 speech synthesis**

通过机械的、电子的方法合成人类语言的过程。

[GB/T 21024—2007,定义 3.1]

3.7

**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。

[GB/T 21023—2007,定义 3.1]

3.8

**关键字识别 keyword spotting**

针对连续语音流中的特定关键字进行识别和检出的过程。

注:关键字识别不需要识别全部文字,只需要识别和检测出关注的关键字及其出现位置。

3.9

**命令字识别 command word recognition**

一种基于语音识别语法的语音识别方式,是在语音识别语法规则限定的范围内,对于给定的语音输入,语音识别引擎给出语音识别语法覆盖范围内的文本或拒识作为识别结果。

[GB/T 34083—2017,定义 3.3]

3.10

**连续语音识别 continuous speech recognition**

识别任意的连续语音,并给出相对应的文本。

注:连续语音识别不限制用户说话的词汇、内容和方式,用户可以以任意说的形式输入语音。

[GB/T 34083—2017,定义 3.4]

3.11

**语义理解 semantic understanding**

使功能单元理解人说话的意图。

3.12

**语义词典 semantic dictionary**

用于语义理解的词汇工具。

3.13

**语义库 semantic library**

以语义网络表示的结构化知识的集合。

注:语义库包含具有语义信息的字、词、句等说法。

3.14

**声纹 voiceprint**

对语音中所蕴含的、能表征和标识说话人的语音特征,以及基于这些特征(参数)建立的数学模型的总称。

[SJ/T 11380—2008,定义 3.1.1]

3.15

**声纹识别 voiceprint recognition; VPR**

说话人识别 speaker recognition

根据待识别语音的声纹特征识别该段语音所对应的说话人的过程。

[SJ/T 11380—2008, 定义 3.1.6]

### 3.16

**声纹模型 voiceprint model**

对声纹特征进行描述的数学模型。

[SJ/T 11380—2008, 定义 3.1.3]

### 3.17

**语音唤醒 speech wakeup; voice trigger**

处于音频流监听状态的语音交互系统,在检测到特定的特征或事件出现后,切换到命令字识别、连续语音识别等其他处理状态的过程。

### 3.18

**唤醒命令字 wakeup command word**

用于唤醒处于关键字识别状态的语音交互系统所用的结构化关键字。

### 3.19

**误唤醒 false wakeup**

语音唤醒过程中出现的,无音频流或音频流中没有出现唤醒所需的特征或事件时,语音唤醒系统被唤醒的现象。

### 3.20

**语音压缩 speech compression**

用于在较窄的频带上传输或存储语音信号,或者采用比正常情况下所必需的比特量少的一种语音处理技术。

[GB/T 5271.29—2006, 定义 29.01.27]

### 3.21

**压缩等级 compression level**

对语音压缩算法输出的压缩语音码流率的档次划分。

注:不同的档次划分决定了编解码算法的复杂程度,决定了解码后的语音质量。

### 3.22

**端点检测 voice activity detection**

一种用于分析、判断连续音频流中有效语音起始点和结束点的语音处理技术。

### 3.23

**麦克风阵列 microphone array**

由具有确定空间拓扑结构的多个麦克风组成的,对信号的空间特性进行采样并滤波的系统。

### 3.24

**语音增强 speech enhancement**

当有效语音信号被各种噪声干扰,甚至淹没后,从含有噪声背景的声音信号中提取有效语音信号,抑制、降低噪声干扰的技术。

### 3.25

**混响 reverberation**

语音信号在室内传播时,由于反射、衍射等原因,使得语音信号到达语音采集设备的传播路径除了直达路径外还存在着其他路径,从而产生接收信号幅度衰减、音质变差等降低语音信号质量的现象。

### 3.26

**有效语音信号 valid speech signal**

有效语音 valid speech

语音采集过程中所关注、期望的语音信号。

3.27

**噪声 noise**

语音采集过程中,采集到的由非有效语音信号源发出的,能干扰、影响对有效语音信号的理解或处理的声音信号。

3.28

**声源定位 acoustic source localization**

对发声物体位置的判断过程。

3.29

**语音打断 speech interruption**

语音交互系统在播放声音的过程中,当语音采集设备检测到有效语音输入时,中断播放声音,转到语音识别等其他处理过程。

[GB/T 36464.2—2018,定义 3.18]

3.30

**情感计算 affective computing**

在人机交互过程中对特定情感的收集、识别、决策和表示。

## 4 系统通用功能框架

智能语音交互系统(以下简称系统)包括语音交互界面、前端处理、语音处理、服务接口、应用业务处理和数据资源等功能单元,其中:

- a) 语音交互界面提供系统与人直接进行语音交互的人机界面,包括语音信号输入、输出以及由前端处理、语音处理支撑的语音能力;
- b) 数据资源包括系统处理的音频数据和文本数据;
- c) 前端处理提供语音唤醒、声源定位、声纹识别、语音增强、格式转换、重采样等功能;
- d) 语音处理提供语音识别、语义理解、语音合成、端点检测、语音编解码、全双工交互、情感计算等功能;
- e) 服务接口提供外部设备/设施调用系统语音服务的接口;
- f) 应用业务处理对语音处理的结果转换为对应的应用指令并反馈业务响应结果。

系统的通用功能框架见图 1;部分参数定义及其计算方法参见附录 A。



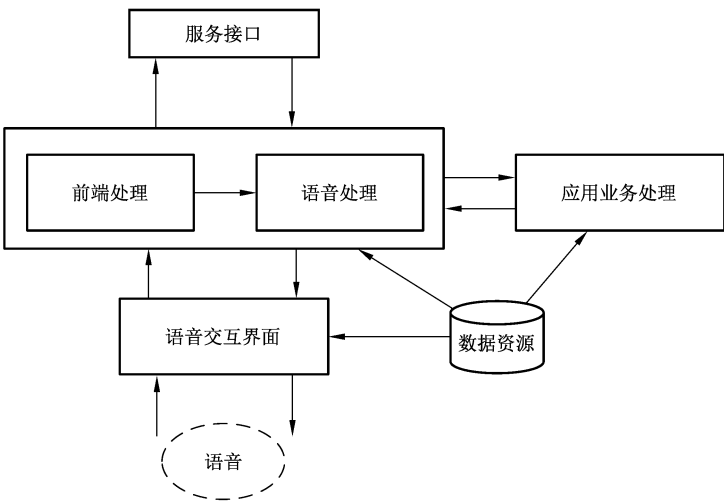


图 1 智能语音交互系统的通用功能框架

5 语音交互界面要求

5.1 语音采集

系统应通过传声器或麦克风阵列等具备语音采集能力的拾音设备对语音进行采集；应根据不同的拾音距离，提供近场拾音和/或远场拾音中的一种。

5.2 语音播报

系统应通过音频播放设备对语音进行播报；应支持播放语音所需的性能要求，例如音频格式、带宽等。

5.3 输入输出

系统应支持中文普通话输入输出；除中文普通话之外，对有方言和其他语种输入输出需求的，系统应支持地方方言、民族语言以及其他语种输入输出。

5.4 环境噪声适应能力

系统应在不同场景的典型环境噪声下能成功进行语音交互，确保人机交互可用。

注：不同场景的典型环境噪声参见 GB/T 36464 的第 2~5 部分。

6 数据资源要求

6.1 音频数据

系统的音频数据主要用于语音识别和语音合成，要求如下：

- a) 语音识别的输入音频数据格式应符合 GB/T 34083—2017 中表 1 的要求；
- b) 语音合成的输出音频数据格式应符合 GB/T 34145—2017 中表 2 的要求。

6.2 文本数据

系统的文本数据主要用于各功能单元之间的信息传输、交换和共享，要求如下：

- a) 应与具体操作系统和平台无关,并且可扩展;
- b) 宜是结构化数据,便于系统处理;
- c) 中文编码字符应符合 GB 18030 的规定,并依据 GB/T 11460 进行检测;
- d) 中文语音合成数据交换格式应符合 GB/T 21024—2007 中第 5 章的要求。

## 7 前端处理要求

### 7.1 语音唤醒

#### 7.1.1 命令字唤醒

系统应支持使用预先定义的命令字来对系统进行语音唤醒。

#### 7.1.2 命令字声纹唤醒

系统在语音唤醒过程中,宜支持使用文本相关声纹识别和命令字匹配,且在声纹确认成功后能唤醒系统。

#### 7.1.3 自定义唤醒命令字

系统应支持对用于语音唤醒的命令字进行自定义。



#### 7.1.4 多唤醒命令字

系统应支持使用不同的命令字进行语音唤醒;可根据指定的唤醒命令字进入对应的状态或模式。

#### 7.1.5 多音频流监听

系统在进行语音唤醒时,应支持同时监听多个音频流。

### 7.2 声源定位

系统应支持通过对声源的平面角、方位角和声源距离的计算,来对发声源进行定位。

### 7.3 声纹识别

#### 7.3.1 一般要求

系统应支持下列声纹识别功能:

- a) 文本相关的声纹识别;
- b) 文本无关的声纹识别;
- c) 指定文本的声纹识别;
- d) 声纹模型训练;
- e) 声纹模型自适应;
- f) 声纹确认;
- g) 声纹辨认;
- h) 声纹检出;
- i) 声纹追踪;
- j) 语言相关的声纹识别;

k) 语言无关的声纹识别。

上述功能描述和要求应符合 SJ/T 11380—2008 第 3 章的要求。

### 7.3.2 声纹文本获取

系统应支持获取指定文本或自定义文本,用于声纹模型训练、声纹模型自适应、声纹确认和声纹辨认。

## 7.4 语音增强

### 7.4.1 噪声抑制

系统应支持对输入语音中的背景噪声进行抑制,提高语音的信噪比。

### 7.4.2 混响消除

系统应支持对输入语音中的晚期混响进行抑制,提升语音信号的清晰度和可懂度。

## 7.5 格式转换

系统应支持将音频的格式转换到另外一种格式,以满足语音处理的要求。

## 7.6 重采样

系统应支持改变数字语音信号的采样率,以满足语音处理的要求。

# 8 语音处理要求

## 8.1 语音识别

### 8.1.1 一般要求

系统应支持下列全部或大部分语音识别功能:

- a) 中文语音识别服务;
- b) 多语种识别;
- c) 多方言识别;
- d) 多语种混读识别;
- e) 自定义语法;
- f) 个性化识别;
- g) 识别结果多候选;
- h) 自定义热词;
- i) 识别结果进阶;
- j) 语言信息识别;
- k) 说话者信息识别。

上述功能描述和要求应符合 GB/T 34083—2017 中 4.2 和 4.3 的要求。

### 8.1.2 语音识别方式

系统应支持近场音频处理和/或远场音频处理中的一种;应至少支持关键字识别、命令字识别、连续

语音识别中的一种。

## 8.2 语义理解

### 8.2.1 自定义语义词典

系统能支持应用程序自定义语义词典和用户自定义语义词典。

### 8.2.2 自定义语义库

系统能支持应用程序自定义语义库和用户自定义语义库。

### 8.2.3 模糊识别

系统应正确处理错别字、同义词、多字少字问题。

### 8.2.4 语义抽取

系统在交互过程中,应抽取语义要素和用户的关键意图。

### 8.2.5 语义排序

系统能在语义理解结果中给出多个排序后的理解结果,供用户进行选择或二次确认。

## 8.3 语音合成

系统应支持下列全部或大部分语音合成功能:

- a) 中文语音合成;
- b) 流式语音合成;
- c) 多种合成文本编码;
- d) 个性化合成;
- e) 多语种合成;
- f) 多方言合成;
- g) 多语种混读合成;
- h) 合成音频多音色;
- i) 用户自定义分词;
- j) 用户自定义读音;
- k) 合成文本位置信息;
- l) 文本分词和拼音信息;
- m) 音频时间信息。

上述功能描述和要求应符合 GB/T 34145—2017 中 4.2 和 4.3 的要求。

## 8.4 端点检测

### 8.4.1 单端点检测

系统应支持从一段连续的音频流中检测出第一个语音段的起始点和结束点。

### 8.4.2 多端点检测

系统应支持从一段连续的音频流中检测出多个语音段的起始点和结束点。

### 8.4.3 端点检测灵敏度设置

系统应支持设置语音等待超时时长和尾部静音长度,调整语音端点检测的灵敏度。

## 8.5 语音编解码

### 8.5.1 变码率编码

系统应支持通过设置编码等级或通过其他方式改变语音编码算法输出的编码语音的码流率。

### 8.5.2 压缩等级设置



系统应支持根据当前网络状况、系统性能等要求来设置语音编码算法的压缩等级。

## 8.6 全双工交互

系统宜支持全双工语音交互;在此状态下,应支持一次语音唤醒和随时语音打断,能进行上下文语境和开放式对话管理,能控制对话节奏和预测用户意图等。

## 8.7 情感计算

系统宜支持以语音信号为载体的情感计算。

## 9 服务接口要求

系统应具备可供外部调用的服务接口。其中,中文语音识别的互联网接口应符合 GB/T 34083 的规定,中文语音合成的互联网接口应符合 GB/T 34145 的规定。

## 10 应用业务处理要求

系统应支持将用户意图转换成应用和业务的控制命令或系统指令,实现应用和业务的响应。

## 附 录 A

### (资料性附录)

#### 部分参数及其计算方法

#### A.1 概述

本附录给出了用于描述智能语音交互系统的部分参数定义及其计算方法。

#### A.2 拾音距离

当声源与拾音设备之间的距离 $\leq 1$  m时,为近场;当声源与拾音设备之间的距离 $> 1$  m时,为远场。

#### A.3 语音交互成功率

在一定的时间段内,成功的语音交互会话总数占有效的语音交互会话总数的百分比。“成功的语音交互会话”指获取到完整的语音服务结果,期间未产生差错的语音交互会话;“有效的语音交互会话”指全部的语音交互会话去除由于用户终端故障或用户行为、参数错误导致的失败会话。

交互成功率的计算方法参见公式(A.1):

$$P_s = \frac{S}{S + F} \times 100\% \quad \dots\dots\dots (A.1)$$

式中:

$P_s$  ——交互成功率, %;

$S$  ——交互成功的次数;

$F$  ——交互失败的次数。

#### A.4 语音唤醒



##### A.4.1 唤醒率

某段时间内的语音唤醒操作中,成功唤醒的次数占语音唤醒总次数的比率。用于描述语音唤醒操作的正确响应情况,其计算方法参见公式(A.2):

$$\rho_{sw} = \frac{N_{sw}}{N_w} \times 100\% \quad \dots\dots\dots (A.2)$$

式中:

$\rho_{sw}$  ——唤醒率, %;

$N_{sw}$  ——成功唤醒次数;

$N_w$  ——语音唤醒操作次数。

##### A.4.2 误唤醒频度

误唤醒频度描述误唤醒操作在单位时间内出现的频度,其计算方法参见公式(A.3):

$$f_{FW} = \frac{N_{FW}}{T} \dots\dots\dots (A.3)$$

式中:

$f_{FW}$  ——误唤醒频度,单位为次每小时(次/h);

$N_{FW}$  ——在考察的时长内出现误唤醒的次数;

$T$  ——评价时长,单位为小时(h)。

## A.5 语音识别

### A.5.1 字准确率

字准确率的计算方法参见 GB/T 21023—2007 的 5.2.1 的规定。

### A.5.2 句识别率

句识别率的计算方法参见公式(A.4):

$$P_{sr} = \frac{N_{sr}}{N_{si}} \times 100\% \dots\dots\dots (A.4)$$

式中:

$P_{sr}$  ——句识别率,%;

$N_{sr}$  ——智能语音交互系统正确识别的句数;

$N_{si}$  ——标注总句数。

### A.5.3 起始响应时间

从检测到用户有效语音输入时起至得到第一部分识别结果时止中间经过的时间,单位为毫秒(ms),用于描述语音识别响应的实时性。

### A.5.4 结束响应时间

从检测到用户有效语音输入结束时起至得到最后一部分识别结果时止中间经过的时间,单位为毫秒(ms),用于描述语音识别响应的实时性。

## A.6 语义理解

### A.6.1 语义理解正确率

语义理解正确率计算方法参见公式(A.5):

$$R_{ss} = \frac{N_{ss}}{N} \times 100\% \dots\dots\dots (A.5)$$

式中:

$R_{ss}$  ——语义理解正确率,%;

$N_{ss}$  ——操作意图及语义要素均被正确判断的次数;

$N$  ——用户输入被正确识别出文本信息的总次数。

### A.6.2 响应正确率

语义理解响应正确率计算方法参见公式(A.6):

$$R_{SR} = \frac{N_{SS} + N_{SR}}{N} \times 100\% \quad \dots\dots\dots (A.6)$$

式中：

$R_{SR}$  ——语义理解响应正确率，%；

$N_{SS}$  ——操作意图及语义要素均被正确判断的次数；

$N_{SR}$  ——被正确拒识的次数；

$N$  ——用户输入被正确识别出文本信息的总次数。

## A.7 语音合成

### A.7.1 合成响应时间

合成响应时间描述了语音合成系统响应的及时性，指从用户写入合成文本时起至拿到第一块合成音频时止中间经过的时间，单位为毫秒(ms)。

### A.7.2 平均码流率

一次语音合成会话输出的合成音频数据总量除以语音合成会话的持续时间，单位为千比特每秒(kbit/s)。合成音频数据总量以解码后的音频数据量计算，语音合成会话的持续时间指从用户写入合成文本时起至拿到最后一块合成音频时止中间经过的时间。

语音合成系统的平均码流率宜大于或等于合成音频格式的标称码流率。

## A.8 语音质量

### A.8.1 声强级

声强级用于度量声音的大小，其计算方法参见 GB/T 16404—1996 中 3.5 的规定。

### A.8.2 信噪比

信噪比用来描述智能语音交互系统使用的背景环境，其计算方法参见 GB/T 21023—2007 中 4.6 的规定。

### A.8.3 平均意见得分

平均意见得分(MOS)是语音质量的一种主观度量。MOS 量化分值参见表 A.1。

表 A.1 主观 MOS 量化分值

评分	主观测听效果
5	优秀,察觉不到任何不自然,音色接近播音员
4	较好,仅能察觉若干不自然
3	尚可,能察觉到不自然但可以接受,评分分界线
2	较差,明显察觉不自然并不愿意接受
1	极差,无法接受



## A.9 声纹识别率

声纹识别率包括错误拒绝率、错误接受率、漏识率、误警率等参数,其计算方法参见 SJ/T 11380—2008 中 3.3.2 的规定。

## A.10 语音编解压缩率

语音编解压缩率为语音压缩算法输出的压缩音频与输入的待压缩音频的码流率之比。

## A.11 语音增强



### A.11.1 信噪比改善

信噪比改善为语音增强功能单元输出语音信噪比与输入语音信噪比的比值。

### A.11.2 噪声抑制量

噪声抑制量的计算方法参见公式(A.7):

$$D_{NR} = 10 \log \frac{\sum_{n=0}^{N-1} |\nu_{in}(n)|^2}{\sum_{n=0}^{N-1} |\nu_{out}(n)|^2} \dots\dots\dots (A.7)$$

式中:

$D_{NR}$  ——噪声抑制量,单位为分贝(dB);

$\nu_{in}(n)$  ——输入信号中第  $n$  个噪声信号的振幅;

$\nu_{out}(n)$  ——输出信号中第  $n$  个噪声信号的振幅;

$N$  ——输入信号频谱频率分量的总数量。

## A.12 声源定位

### A.12.1 平面角定位误差

平面角定位误差为声源定位功能单元计算得到的声源位置平面角与真实值的差值。

### A.12.2 俯仰角定位误差

俯仰角定位误差为声源定位功能单元计算得到的声源位置俯仰角与真实值的差值。

### A.12.3 距离定位误差

距离定位误差为声源定位功能单元计算得到的声源位置距离与真实值的差值。

## A.13 语音打断成功率

对话管理中,语音打断成功率指某段时间内,语音打断操作被正确响应的次数占总次数的比率。其

计算方法参见公式(A.8)：

$$P_i = \frac{N_i}{N} \times 100\% \quad \dots\dots\dots (A.8)$$

式中：

$P_i$  ——语音打断成功率，%；

$N_i$  ——被语音交互系统正确响应的次数；

$N$  ——交互内容中需要执行打断操作的次数。



## 参 考 文 献

- [1] GB/T 5271.1—2000 信息技术 词汇 第1部分:基本术语
  - [2] GB/T 5271.29—2006 信息技术 词汇 第29部分:人工智能 语音识别与合成
  - [3] GB/T 16404—1996 声学 声强法测定噪声源的声功率级 第1部分:离散点上的测量
  - [4] GB/T 21023—2007 中文语音识别系统通用技术规范
  - [5] GB/T 36464.2—2018 信息技术 智能语音交互系统 第2部分:智能家居
  - [6] GB/T 36464.3—2018 信息技术 智能语音交互系统 第3部分:智能客服
  - [7] GB/T 36464.4—2018 信息技术 智能语音交互系统 第4部分:移动终端
  - [8] GB/T 36464.5—2018 信息技术 智能语音交互系统 第5部分:车载终端
  - [9] ISO/IEC DIS 30150-1 Information technology—Affective computing user interface(AUI)—Part 1:Model
-