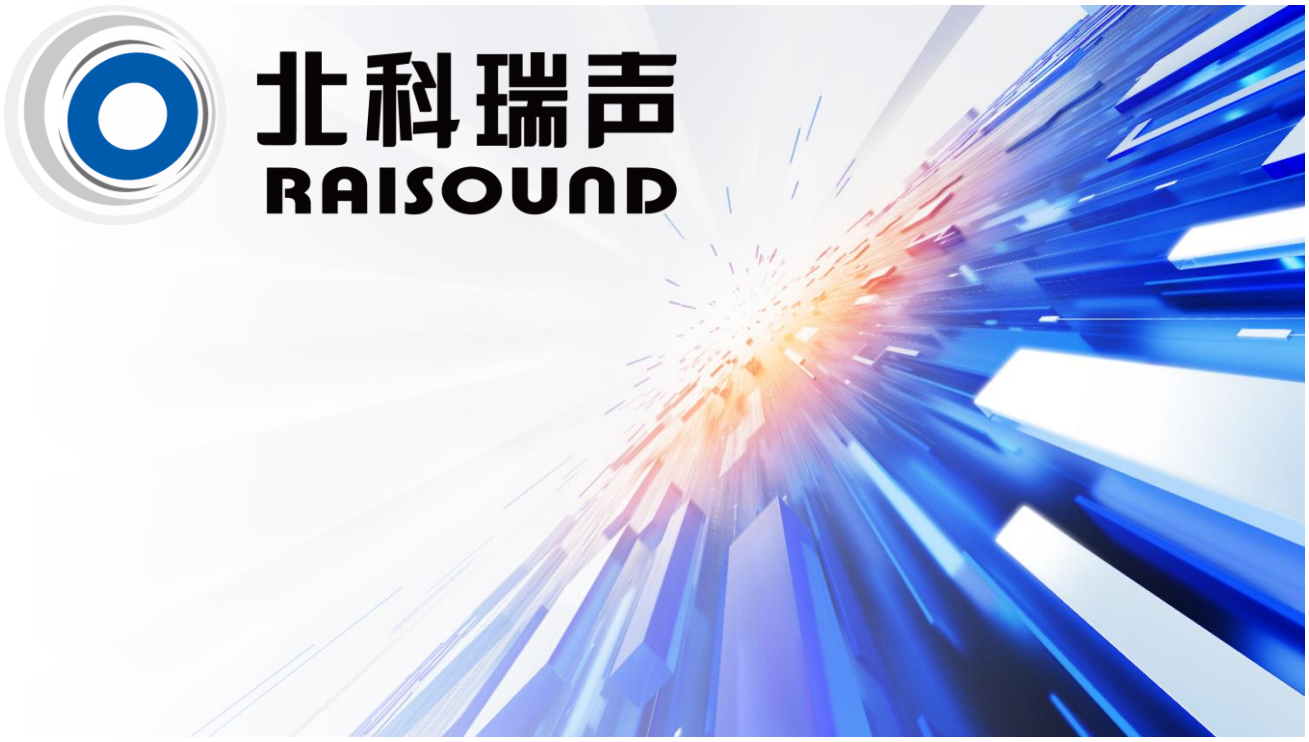

AI 语音云开放平台

产品规格书 V1.0

目 录

版权声明

免责声明

前言

1.1 文档目的

本文档作为北科瑞声 AI 语音云开放平台（以下简称“平台”）的产品规格说明书，详细阐述了平台的整体架构、核心功能、技术参数、部署方案、接入配置及安全要求等内容，为平台的开发、测试、部署及第三方接入提供标准化依据，确保各方对平台的功能、性能及使用规范达成一致认知。

1.2 文档范围

本文档覆盖平台的全部核心能力，包括平台概述、技术优势、系统架构、核心功能模块、技术参数、接入配置、部署方案、安全指标等内容，适用于平台开发人员、测试人员、第三方接入开发者及相关业务负责人。

1.3 适用对象

1. 平台开发团队：用于指导平台开发、迭代及优化工作；
2. 测试团队：用于制定测试用例、开展功能及性能测试；
3. 第三方接入开发者：用于了解平台能力、接入方式及开发规范；
4. 业务负责人：用于全面了解平台功能、性能及应用价值，支撑业务决策。

1.4 术语与定义

1. AI 语音云开放平台：集成高精度语音识别、语音合成、声纹识别、离线意图命令词识别能力，提供统一移动端语音服务接口及多平台 SDK，支持离线使用及私有化部署的智能语音服务系统；

2. SDK：软件开发工具包，为第三方平台及用户提供快速接入和集成平台能力的工具集合；

3. API：应用程序接口，为第三方系统提供调用平台语音能力的标准化接口；

4. 实时语音识别：将实时采集的语音流即时转换为文本的技术；

5. 非实时语音识别（语音文件识别）：对已录制的音频文件进行语音转文本的技术；

6. 语音合成：将文本转换为自然、流畅语音的技术；

7. 声纹识别：通过分析语音中包含的个人身份信息，实现身份认证的技术；

8. 离线意图命令词识别：在无网络或弱网环境下，识别用户语音指令意图及特定命令词的技术；

9. MOS：平均意见得分，用于评价语音合成自然度的标准，得分越高，合成效果越自然；

10. 私有化部署：将平台部署于用户本地服务器，实现数据本地存储、本地管理的部署模式；

11. 端云融合：结合终端设备与云端服务的优势，实现离线场景与在线场景的无缝切换，确保服务连续性；

12. 微服务架构：将平台功能分解为离散的服务模块，实现模块解耦，便于扩展、维护及迭代；

13. 热词：特定行业、场景下的专业词汇、人名、地名等，通过自定义热词可提升语音识别准确率；

14. 强纠：对语音合成中特定文本的读音进行强制纠正的配置功能。

一、平台介绍

1.1 方案概述

北科瑞声 AI 语音云开放平台，基于 AI+智能语音能力赋能行业，通过搭载强大的声学模型、语言模型和专业识别引擎，为行业用户（含第三方平台及个人用户）提供智能语音交互的深度服务。平台集成了高精度语音识别、语音合成、声纹识别、离线意图命令词识别四大核心能力，采用国际领先的流式端到端语音语言一体化建模算法，支持多终端接入及多场景应用，而这套集成了四大核心能力的智能语音服务系统，可向第三方平台及用户提供统一的移动端语音服务能力接口。

为便于第三方平台及用户快速接入和集成，平台不仅通过多种形式的 API/SDK 提供服务，还配备多平台 SDK 开发套件，同时内含离线语音包，支持离线使用，可确保在无网络或弱网环境下亦能实现高效的人机交互，全面提升智能化作业体验和效率。平台可广泛应用于各类需要语音交互的场景，尤其适配对数据安全、国产化适配有较高要求的行业，通过私有化部署模式，适配国产化数据库、操作系统、芯片及中间件，保障数据安全与服务稳定，助力行业信息系统智能化升级，提升用户工作效率。

1.2 技术优势

北科瑞声深耕智能语音交互多年，在自然语音和语言领域拥有全链条核心技术，在智能语音私域大模型领域形成独有的技术优势，结合平台核心能力，具体优势如下：

1.2.1 私域语音大模型优势

支持内网私有化部署，智能语音平台可在封闭网络环境下，稳定提供语音识别、语音合成、声纹识别、离线意图命令词识别等多种智能语音功能；融合私域大语言模型 LLM，实现多种智能语音生成能力；非实时转写效率极高，1 小时音频可在 1 分钟内完成转写，中文普通话识别准确率达 98%及以上，满足行业高效作业需求。

1.2.2 热词快速自学习能力

支持快速热词添加，万级别热词可实现秒级快速自学习，无需复杂配置；针同行业的术语、人名、地名、机构名等，一次学习即可永久生效，有效提升特定场景下的语音识别准确率，适配行业个性化需求。

1.2.3 高效语音处理能力

具备高效的实时语音转写与非实时语音转写能力，支持多种音频格式，可实现转写内容的快速要点提取和总结；支持离线语音转写及离线意图命令词识别，无网络或弱网环境下仍能稳定工作；语音合成自然度高，中文合成效果 MOS 不低于 4.5 分，满足各类场景的语音输出需求。

1.2.4 千人千档个性化服务

每个用户可配置个性化云服务引擎，系统可自动学习重口音用户的口音和口语表达习惯，不断进行智能优化，有效提升语音识别准确率；同时支持个性化语音合成，可根据用户需求自定义音色、语速等参数，提升用户交互体验。

1.2.5 一站式解决方案

平台整合语音识别、语音合成、声纹识别、离线意图命令词识别等多项核心能力，开发者可同时获得所需的多项服务能力，一站式解决了需要到不同技术供应商获取服务的繁琐过程，让智能语音交互技术的集成更简单、更实用，降低开发成本与周期。

1.2.6 全方位信息安全保障

采用端云融合和私有云内网部署模式，数据传输全程加密，确保数据信息交互安全；具备完善的权限管理和控制机制，用户数据加密存储，实现数据可用不可见；建立可控信息销毁机制，全面保障用户隐私与数据安全，适配对数据安全有严格要求的行业。

1.2.7 多平台与国产化适配支持

支持主流操作系统接入，提供 WebAPI 覆盖 Android、iOS、Windows、Linux、鸿蒙等系统平台；同时支持麒麟、UOS 等主流国产自主操作系统，适配鲲鹏/飞腾等国产化芯片，可满足对国产化软硬件适配的要求；提供适用于 Android、iOS、鸿蒙平台的标准化 SDK 开发工具包，便于多终端快速接入。

1.2.8 高可用与可扩展能力

平台具备高可用设计，支持横向扩展，并可根据业务增长动态扩容；支持服务节点集群部署，关键组件无单点故障，确保服务连续性；除目前开放的核心语音能力外，可根据智能

语音技术发展及开发者需求，逐步开放更多拓展能力，打造全方位的智能语音交互开放平台。

1.3 建设原则

1.3.1 智能性

1. 运用智能语音技术，实现对多种场景的实时语音和历史语音大数据处理，采用端云融合架构，兼顾在线与离线场景的交互需求；

2. 平台整合语音前端处理、语音识别、声纹识别、语音合成、语意识别、离线意图命令词识别等多种能力，实现高效、智能的人机交互；

3. 基于 AI 算法持续优化，自动学习用户习惯、优化识别与合成效果，提升智能化水平。

1.3.2 安全性

1. 实现端+云私有化部署，支持部署于用户本地服务器，数据传输与存储全程加密，具备高安全性和隐秘性；

2. 实现国产信创平台全面支持，适配国产化数据库、国产化操作系统、国产化芯片、国产化中间件，满足行业安全合规要求；

3. 具备完善的权限管理、数据脱敏、日志审计及异常监测机制，防止数据泄露与滥用，保障用户隐私与系统安全。

1.3.3 可扩展性

1. 支持 API 接口文档对外开放，语音能力可快速被第三方平台及用户集成，降低接入门槛；

2. 采用微服务架构，与其他信息系统松耦合，架构灵活，能够实现和其他产品的无缝组合，适配不同行业的业务场景；

3. 提供符合行业标准的数据接口和开发工具，支持功能模块的横向扩展与纵向升级，可根据业务需求动态扩容；

4. 支持热词自定义、音色定制等个性化配置，适配不同行业、不同用户的个性化需求。

1.3.4 可靠性

1. 具备在规定条件和时间内完成用户所要求功能的能力，能长期稳定运行，满足 7×24 小时不间断服务需求；

2. 结构设计简洁合理，支持冗余备份，关键组件无单点故障，可靠性高；

3. 系统参数配置简单，调整频率低，自动化程度高，使用方便，操作简单，降低运维成本。

1.4 系统架构

北科瑞声 AI 语音云开放平台采用终端和云端融合的架构，整合语音前端处理、语音识别、语音合成、声纹识别、离线意图命令词识别、语义理解等多项功能，形成完整的语音交互能力，可实现多种接入方式、多种场景下的语音应用。平台整体采用微服务架构（Microservice Architecture），将功能分解到各个离散的服务中，实现解决方案的解耦，提升系统的可扩展性、可维护性及稳定性。

1.4.1 架构核心组成

1. 统一入口访问：通过 Nginx 实现负载均衡，反向代理到 Gateway，确保请求的均匀分发，提升系统并发处理能力；

2. 统一授权管理：Gateway 通过 Zookeeper 获取 Auth 服务，实现统一授权管理，确保只有授权用户及接口可访问平台资源，保障系统安全；

3. 统一服务注册中心：通过 Apache Zookeeper 实现微服务统一注册、发现及动态服务配置管理、配置分发，确保服务间的高效通信与协同；

4. 动态底层服务发现：底层服务统一接入连接池，实现动态服务健康检查，及时发现并剔除故障服务，保障服务连续性；

5. 数据缓存机制：MySQL 数据统一接入 Redis 缓存，通过延迟双删策略实现数据一致性，提升数据读取速度，减少数据库压力；

6. 核心能力模块：包含语音识别引擎、语音合成引擎、声纹识别引擎、离线意图命令词识别引擎、意图识别引擎等，提供平台核心语音服务能力；

7. 后台管理系统：负责平台的监控、配置、运维及日志管理，支撑平台的日常运行；

8. 接入层：提供 API 接口及多平台 SDK，支持第三方平台及用户快速接入，适配不同终端设备与操作系统。

1.4.2 架构优势

1. 解耦性强：微服务架构将各个功能模块独立拆分，模块间依赖度低，便于单独开发、测试、迭代及维护，降低系统复杂度；

2. 可扩展性好：支持服务节点的横向扩展，可根据业务增长动态增加服务实例，提升系统并发处理能力；

3. 可靠性高：关键组件无单点故障，支持冗余备份，动态健康检查机制可及时处理故障服务，确保系统 7×24 小时稳定运行；

4. 灵活性高：端云融合架构兼顾在线与离线场景，多平台接入支持适配不同终端设备，可快速响应不同行业、不同用户的需求；

5. 安全性高：统一授权管理、数据加密存储及传输，结合私有化部署模式，全方位保障数据安全与系统安全。

1.5 平台核心组成

1.5.1 语音识别引擎

自动语音识别技术（Automatic Speech Recognition，ASR）是一种将人的语音转换为文本的技术，是多学科交叉的领域，与声学、语音学、语言学、数字信号处理理论、信息论、计算机科学等众多学科紧密相连。其核心目标是让计算机能够“听写”出不同人所说出的连续语音，为后续的意图理解、指令执行提供基础。

北科瑞声的语音识别引擎基于目前最先进的端到端深度神经网络框架，通过海量的语音和文本数据训练获得高性能的语音识别模型，支持实时语音识别和语音文件识别，具备中英文混合识别、智能标点与数字规整、热词优化、强纠等能力，识别准确率高、响应速度快，同时支持离线识别，适配无网络或弱网场景。

1.5.2 语音合成引擎

语音合成技术（Text-to-Speech，TTS）是将文本转换为自然、流畅语音的技术，核心是通过声学模型与语言模型的协同，还原人类语音的韵律、语气及发音特点。平台语音合成引擎采用平均意见得分（MOS）标准，中文合成效果 MOS 不低于 4.5 分，合成语音自然、清晰，可满足各类场景的语音输出需求。

引擎支持多种音色（包含男声、女声），并支持中文和英文的合成，可根据用户需求动态调整音量、语速等参数，支持对特定文本读音进行强制纠正的配置功能，同时支持流式/非流式语音合成，适配实时交互与非实时播报等不同场景，支持离线合成，确保无网络环境下的正常使用。

1.5.3 声纹识别能力引擎

声纹识别技术（Voiceprint Recognition）是通过对声音中包含的个人身份相关信息进行分析，实现身份认证的技术，其核心是提取语音中的声纹特征，与预设的声纹库进行比对，完成身份确认。平台声纹识别引擎可通过采集端硬件设备进行高清拾音和音频信息传输，实现远程身份确认；同时可与语音识别技术结合，提升声纹认证的准确性和语音识别的性能。

引擎支持说话人注册、说话人确认、说话人识别等核心功能，可实现多人语音场景下的说话人分离，从多人混合的语音信号中，分离出不同说话人的语音成分，支持有监督和无监督说话人分离，适配会议、访谈等多场景的身份识别需求。

1.5.4 意图识别引擎

意图识别引擎是融合语音识别与自然语言理解技术的核心能力，通过语音识别将用户语音指令转换为文本，再依托自然语言处理技术解析用户意图，实现信息系统对用户需求的精准理解，从而高效回应用户意图与指令。该引擎具备场景意图识别与分类、领域分类、实体识别、关键词聚类提取、多轮对话管理等能力，可准确识别用户意图，显著提升语音交互的智能化水平。

同时引擎具备离线意图命令词识别能力，支持在无网络环境下，本地识别用户语音指令意图及特定命令词，无需依赖云端即可完成终端本地处理，具备响应快、准确率高的特点，还可按行业需求自定义命令词适配不同场景。其与在线意图识别引擎协同工作，既能保障无网络环境下语音交互服务正常运行，又能整体提升平台智能化作业的体验与效率。

1.5.6 后台管理系统

后台管理系统是平台运维、监控及配置的核心载体，负责平台的日常运行管理，主要功能包括：

1. 首页监控：实时显示服务使用情况（服务并发量、服务调用次数、用户登录次数）、硬件使用情况（CPU 使用率、内存使用率、磁盘使用率）以及系统日志（接口调用日志系、统操作日志），便于运维人员实时掌握平台运行状态；

2. 实时语音转写运行监控与管理：实时统计语音转写服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息，提供行业模型、语种、关键词、热词、敏感词的配置功能，可根据需求优化转写效果；

3. 非实时语音转写运行监控与管理：实时统计非实时语音转写服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息，提供行业模型、语种、关键词、热词、敏感词的配置功能，可根据需求优化转写效果；

4. 流式/离线语音合成运行监控与管理：实时统计流式/离线语音合成服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息；

5. 声纹识别运行监控与管理：实时统计声纹识别服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息；

二、产品详细介绍

2.1 平台核心功能介绍

2.1.1 语音识别功能

平台语音识别功能基于高性能语音识别引擎，支持实时语音识别、非实时语音识别（语音文件识别）、离线语音识别三种模式，覆盖多种场景需求，具体功能如下：

1. 实时语音识别：支持语音实时流的即时转写，首字响应时间快，可用于实时交互、会议字幕、实时指令识别等场景；支持带口音普通话、粤语等方言识别，适配不同用户的发音习惯；

2. 非实时语音识别：支持对已录制的音频文件进行转写，单个音频文件时长可超过 10 个小时，转写效率极高，1 小时音频可在 1 分钟内完成转写，输出规范文稿；支持 mp3、mp4、wav、amr、m4a、pcm、ogg、opus 等多种音频格式；支持通过 URL 上传音频文件，支持异步回调和轮询两种方式获取转写结果；

3. 离线语音识别：支持在无网络或弱网环境下，对本地音频进行识别，无需依赖云端服务，响应速度快，识别准确率与在线模式保持一致；支持离线意图命令词识别，可自定义命令词，适配离线场景下的指令控制；

4. 中英文混合识别：支持中文夹杂英文、数字的混合识别，可准确识别文本中的中英文词汇及数字，适配多语言交互场景；

5. 智能标点与数字规整：支持智能标点预测，自动为转写文本添加标点符号；支持数字规整，可将语音中的数字转换为规范格式（如电话号码、车牌号、日期、金额等），提升文本可读性；

6. 热词与强纠：支持自定义热词功能，可添加行业术语、人名、地名、机构名等热词，提升特定词汇识别准确率；支持热词快速自学习，万级别热词秒级生效；支持对特定文本读音进行强制纠正，确保识别与转写的准确性；

7. 说话人分离：支持多人语音场景下的说话人分离，可通过对说话人语音特征提取和比对，区分不同说话人，实现不同说话人文本的分离，支持 2-4 人说话角色分离和标记，适配会议、访谈等场景；

8. 优化功能：支持语气词、停顿词、重复词过滤，智能分段，提高转写文本质量；支持识别空语音，避免无效转写；支持可定制格式的转写内容导出，满足不同用户的文档需求；

9. 行业模型适配：根据用户使用场景，提供多种行业语言模型灵活切换，具备通用行业语言模型，可适配不同行业的个性化需求。

2.1.2 语音合成功能

平台语音合成功能基于高性能语音合成引擎，支持流式/非流式语音合成，合成语音自然、清晰，适配多种场景的语音输出需求，具体功能如下：

1. 实时语音合成：支持流式语音合成，首帧响应时间快，可用于实时交互、语音播报等场景；支持非流式语音合成，适用于长文本播报场景；

2. 音色与语种支持：支持多种音色（包含男声、女声），接口入参可指定音色，满足不同场景的语音输出需求；支持中文和英文的合成，可实现中英文混合合成；

3. 个性化定制：支持个性化定制语音库，可基于大规模录音语料库进行声学模型训练、建模、统计，打造专属音色；支持合成语音的音量、语速等多种合成参数动态调整，适配不同用户的听觉需求；

4. 特殊文本处理：支持针对不同数字类型（如金额、日期、电话号码等）的语音合成发音方式自定义，确保合成语音的准确性；支持 UTF8 合成文本的编码，兼容多种文本格式；

5. 输出格式支持：支持输出多种采样率的语音，包含 8000Hz、16000Hz 等，适配不同终端设备的播放需求；

6. 离线合成支持：支持离线语音合成，内置离线语音包，无网络或弱网环境下仍能正常合成语音，确保服务连续性。

2.1.3 声纹识别功能

平台声纹识别功能基于高精度声纹识别引擎，可实现身份认证、说话人分离等功能，适配多种场景的身份确认需求，具体功能如下：

1. 说话人注册：支持创建、修改、更新、删除每个用户的声纹信息，管理员可管理所有用户的声纹信息，建立完善的声纹库；

2. 说话人确认：用于确认说话人是否为声明身份，可结合随机产生的字符串的语音识别，提高身份确认的准确性，适用于身份验证、权限管控等场景；

3. 说话人识别：用于对於待确认的语音身份信息的搜索和确认，可从声纹库中匹配对应的说话人，实现快速身份识别；

4. 说话人分离：与语音识别结合，从多人混合的语音信号中，对每个说话人的语音进行声纹特征提取，分离出不同说话人的语音成分，支持有监督和无监督说话人分离，适配会议、访谈等多场景；

5. 远程身份确认：可通过采集端硬件设备进行高清拾音和音频信息传输，实现远程身份确认，适用于远程办公、远程授权等场景。

2.1.5 意图识别功能

平台意图识别功能基于意图识别引擎，结合语音识别技术，实现对用户语音指令的精准解析，具体功能如下：

1. 场景意图识别与分类：识别输入文本中的意图，并将其归类到相应的意图类别中，如指令控制、业务咨询、信息查询等；

2. 领域分类：支持识别输入文本其所属领域，例如医疗、政务、闲聊等，适配不同行业的业务场景；

3. 实体识别：支持识别和提取文本中的实体，例如人名、地名、日期、金额、机构名等，为意图解析提供支撑；

4. 关键词聚类提取：可提取文本中的核心关键词，实现对用户需求的快速把握；

5. 多轮对话管理：支持当无法通过一次识别判断输入文本的意图时，通过多轮问答逐步定位意图，提升交互的准确性和智能化水平。

6. 离线意图解析：在无网络环境下，可快速解析用户语音指令的核心意图，执行相应的操作，无需依赖云端服务，响应速度快；

73. 离线语音包支持：内置离线语音包，可在终端设备本地完成识别处理，确保无网络环境下的正常使用；支持离线语音包的更新与升级，持续优化识别效果；

2.1.6 后台管理功能

后台管理系统为平台的运维、监控及配置提供全方位支持，具体功能如下：

1. 首页监控：实时显示服务使用情况（服务并发量、服务调用次数、用户登录次数）、硬件使用情况（CPU 使用率、内存使用率、磁盘使用率）以及系统日志（接口调用日志系、统操作日志），便于运维人员实时掌握平台运行状态；

2. 实时语音转写运行监控与管理：实时统计语音转写服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息，提供行业模型、语种、关键词、热词、敏感词的配置功能，可根据需求优化转写效果；

3. 非实时语音转写运行监控与管理：实时统计非实时语音转写服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息，提供行业模型、语种、关键词、热词、敏感词的配置功能，可根据需求优化转写效果；

4. 流式/离线语音合成运行监控与管理：实时统计流式/离线语音合成服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息；

5. 声纹识别运行监控与管理：实时统计声纹识别服务的数据（昨日、今日以及累计系统使用量），显示服务接口信息；

2.2 接口对接

平台提供标准化的 API 接口及多平台 SDK 开发套件，便于第三方平台及用户快速接入和集成，实现语音识别、语音合成、声纹识别、离线意图命令词识别等核心能力的应用，具体对接方式如下：

2.2.1 接口类型

1. 语音识别接口：包括实时语音识别接口、非实时语音识别接口（语音文件识别接口）、离线语音识别接口，支持多种请求参数配置，返回规范的转写结果；

2. 语音合成接口：包括流式语音合成接口、非流式语音合成接口、离线语音合成接口，支持音色、语速、音量等参数配置，返回合成后的语音数据；

3. 声纹识别接口：包括说话人注册接口、说话人确认接口、说话人识别接口、说话人分离接口，支持声纹信息的管理与识别；

4. 离线意图命令词识别接口：支持离线命令词配置接口、离线意图识别接口，实现离线场景下的指令识别与解析；

2.2.2 SDK 支持

平台提供适用于 Android、iOS、鸿蒙平台的标准化 SDK 开发工具包，包含语音识别、语音合成、声纹识别及离线意图命令词识别等核心功能，SDK 具备以下特点：

1. 轻量化设计：SDK 体积小，占用终端设备资源少，不影响终端设备的正常运行；

2. 集成便捷：提供详细的开发文档、示例代码及调试工具，开发者可快速完成集成，降低开发成本与周期；

3. 兼容性强：适配不同版本的操作系统，支持 Android 6.0 及以上、iOS 9.0 及以上版本以及鸿蒙系统 2.0 及以上的移动设备接入并稳定运行；

4. 功能完整：SDK 包含平台全部核心能力，支持在线与离线模式切换，可满足不同场景的应用需求；

5. 稳定性高：经过严格的测试验证，确保在不同终端设备上稳定运行，减少崩溃与异常情况。

2.2.3 接口调用规范

1. 接口协议：采用 HTTP/HTTPS 协议，确保接口调用的安全性与稳定性；

2. 数据格式：请求与响应数据采用 JSON 格式，便于解析与处理；

3. 授权认证：所有接口调用均需进行授权认证，通过 API 密钥实现身份验证，确保接口安全；

4. 调用频率：支持高并发调用，平台提供 100 路并发访问能力，可根据业务需求动态扩容；

5. 错误处理：接口返回详细的错误码及错误信息，便于开发者排查问题，快速解决接口调用异常。

2.3 技术参数

平台技术参数涵盖系统集成、语音识别、语音合成、声纹识别、离线意图命令词识别、接入配置、高可用等多个维度，具体参数如下：

2.3.1 系统集成参数

1. 集成方式：提供 SDK 或 API 两种集成方式，可实现语音识别、语音合成、声纹识别及离线命令词识别能力的快速应用；

2. 兼容性：移动端 SDK 或 API 接口兼容各类移动终端，可满足用户移动终端使用需求，适配多平台、多版本操作系统。

2.3.2 语音识别参数

1. 识别准确率：支持语音实时识别和语音文件识别，在语音清晰、非特定人场景下，中文普通话识别准确率大于 98%，英文识别准确率大于 90%；带口音普通话识别准确率不低于 97%；

2. 音频格式兼容性：支持多种音频格式，至少包括 MP3、WAV，同时兼容 mp4、amr、m4a、pcm、Ogg、opus 等格式；

3. 中英文混合识别：支持中文夹杂英文、数字的混合识别，识别准确率不低于 97%；

4. 智能标点与规整：支持智能标点预测与数字规整，数字规整支持电话号码、车牌号、日期、金额等多种格式；

5. 热词支持：支持自定义热词功能，万级别热词可实现秒级快速自学习；

6. 转写效率：非实时转写 1 小时音频 1 分钟转写完成，单个音频文件时长可超过 10 个小时；

7. 响应时间：实时语音转写首字响应时间小于等于 500ms，响应时间不超过 1 秒；

8. 说话人分离：支持 2-4 人说话角色分离和标记，分离准确率不低于 95%；

9. 离线识别：支持离线语音识别，离线识别准确率与在线模式保持一致，响应时间不超过 1 秒。

2.3.3 语音合成参数

1. 合成自然度：采用平均意见得分（MOS）标准，中文合成效果 MOS 不低于 4.5 分；

2. 音色与语种：支持多种音色（包含男声、女声），支持中文和英文的合成，可实现中英文混合合成；

3. 强纠功能：支持对特定文本读音进行强制纠正的配置功能；

4. 合成参数：支持合成语音的音量、语速等参数动态调整，语速可调范围为 50-200 词/分钟；

5. 输出格式：支持输出 8000Hz、16000Hz 等多种采样率的语音，支持 PCM、WAV 等输出格式；

6. 响应时间：实时语音合成首帧响应时间小于等于 200ms，整体响应时间不超过 1 秒；

7. 离线合成：支持离线语音合成，离线合成效果与在线模式保持一致。

2.3.4 声纹识别参数

1. 识别准确率：在语音清晰场景下，声纹识别准确率不低于 95%；
2. 说话人分离：支持有监督和无监督说话人分离，分离准确率不低于 95%；
3. 响应时间：声纹识别响应时间不超过 500ms；
4. 声纹库支持：支持大规模声纹库管理，可存储万级以上用户的声纹信息；
5. 拾音要求：支持高清拾音，适配普通麦克风、专业拾音设备等多种采集端硬件。

2.3.5 意图识别参数

1. 识别准确率：在语音清晰场景下，离线命令词识别准确率不低于 98%；
2. 响应时间：离线命令词识别响应时间不超过 300ms；
3. 命令词支持：支持自定义命令词，最多可支持千级以上命令词配置；
4. 离线语音包：内置离线语音包，体积小，支持更新与升级；
5. 兼容性：支持在无网络或弱网环境下稳定运行，适配多种移动终端。

2.3.6 接入配置参数

1. 移动设备系统版本兼容性：支持 Android 6.0 及以上、iOS 9.0 及以上版本以及鸿蒙系统 2.0 及以上的移动设备接入并稳定运行；

2. 多平台 SDK 支持：提供适用于 Android、iOS、鸿蒙平台的标准化 SDK 开发工具包，包含全部核心功能；

3. 并发能力：提供 100 路并发访问能力，支持横向扩展，可根据业务增长动态扩容至万路以上；

4. 国产化适配：支持服务私有化部署于本地服务器，能够适配信创要求的国产化数据库、国产化操作系统、国产化芯片、国产化中间件；支持麒麟、UOS 等主流国产自主操作系统，适配鲲鹏/飞腾等国产化芯片。

2.3.7 高可用参数

1. 部署模式：支持服务节点集群部署，关键组件无单点故障；
2. 扩展性：支持横向扩展，可根据业务增长动态扩容；
3. 稳定性：系统可用性不低于 99.9%，支持 7×24 小时不间断服务；

4. 备份机制：支持数据冗余备份，确保数据不丢失。

2.4 安全指标

平台高度重视数据安全与系统安全，严格遵循相关安全规范，具备完善的安全保障机制，具体安全指标如下：

2.4.1 访问控制

1. 实施严格的访问控制策略，确保只有授权用户可以访问系统，并根据用户角色限制其操作权限；

2. 支持强密码策略，要求密码复杂度符合行业标准，同时支持多因素身份验证，提升访问控制的安全性；

3. 支持会话管理，自动超时退出，防止未授权访问。

2.4.2 数据安全

1. 数据传输加密：采用 HTTPS、AES 等加密算法，对数据传输过程进行全程加密，防止数据被窃取、篡改；

2. 数据存储加密：对系统中的敏感数据（如用户信息、声纹信息、转写文本等）进行加密存储，确保数据安全；

3. 数据脱敏：对敏感数据进行脱敏处理，隐藏关键信息，防止信息泄露；

4. 数据备份与恢复：建立完善的数据备份机制，定期进行数据备份，支持数据快速恢复，确保数据不丢失；

5. 可控信息销毁：建立可控的信息销毁机制，对过期数据、无用数据进行安全销毁，保障用户隐私。

2.4.3 安全审计与监测

1. 实时监测系统活动，及时发现异常行为和潜在的安全风险，发出预警提示；

2. 定期进行安全审计，对系统日志、接口调用日志、操作日志等进行分析，发现安全问题并及时解决；

3. 支持安全事件追溯，可通过日志查询定位安全事件的原因及相关责任人。

2.4.4 网络与系统安全

1. 网络架构安全：优化网络拓扑结构，部署防火墙、入侵检测系统等安全设备，有效防范网络攻击和威胁；

2. 系统可维护性：及时安装系统补丁，支持 OTA 更新，更新安全软件，修复安全漏洞；

3. 日志管理：具备完善的系统日志和日志分析工具，记录系统的运行状态、错误信息和警告信息，帮助管理员及时发现问题并进行处理；

4. 病毒防护：支持部署病毒防护软件，定期进行病毒扫描，防止病毒感染。

2.4.5 国产化安全适配

1. 支持国产化操作系统、芯片、数据库、中间件，确保系统在国产化环境下的安全稳定运行；

2. 遵循国家相关安全标准，满足相关行业的安全合规要求。

2.5 平台部署方案

北科瑞声 AI 语音云开放平台支持私有云部署、云端部署两种模式，其中私有云部署可满足对数据安全有严格要求的行业需求，部署灵活、性能稳定，具体部署方案如下：

2.5.1 部署模式

1. 私有云部署：将平台部署于用户本地服务器，数据本地存储、本地管理，有效保障数据安全；支持多个服务器节点集群部署，关键组件无单点故障，确保服务连续性；单系统支持超千路并发，可扩展到万路以上，性能稳定，部署快捷；

2. 云端部署：部署于公有云服务器，支持多区域部署，可根据用户地理位置选择就近部署，提升响应速度；支持弹性扩容，根据业务需求动态调整服务器资源，降低运维成本。

2.5.2 部署架构

1. 整体服务采用 Docker 容器化部署，实现服务的隔离与快速部署，降低环境依赖；

2. 多服务隔离，使用 Zookeeper 实现服务注册发现，确保服务间的高效通信与协同；

3. 使用独立代理链接池服务注册内部底层服务，实现语音云对底层服务的注册发现与管理；

4. 底层数据存储：使用 MySQL 做静态数据存储（如用户信息、配置信息等），Redis 做动态数据缓存（如热词、会话数据等），ELK 实现日志存储与分析，Prometheus 做数据监控存储；

5. 网络安全部署：使用代理服务和公网防火墙将内网服务代理到公网（云端部署模式），私有云部署模式下可关闭公网访问，仅支持内网访问，提升安全性；

6. DevOps 部署：使用 Gitlab+GitlabRunner 实现自动发布构建、版本发布，提升部署效率，减少人工操作失误。

2.5.3 部署环境要求

2.5.3.1 硬件环境要求

1. 服务器配置（单节点）：

- CPU：Intel Xeon E5-2690 v4 及以上，或鲲鹏 920 及以上；
- 内存：32GB 及以上；
- 磁盘：1TB SSD 及以上；
- 网络：千兆网卡，支持内网高速通信；

2. 集群部署：建议至少 2 个服务器节点，实现冗余备份，确保服务稳定性；

3. 采集端设备：支持普通麦克风、专业拾音设备等，确保音频采集质量。

2.5.3.2 软件环境要求

1. 操作系统：支持主流国产化操作系统及通用操作系统，具体包括麒麟 V10 及以上、UOS V20 及以上等国产自主操作系统，以及 CentOS 7.6 及以上、Windows Server 2019 及以上通用操作系统，适配鲲鹏/飞腾等国产化芯片及 Intel/AMD 通用芯片。

2. 数据库：支持国产化数据库及通用数据库，适配达梦 8.0 及以上等国产数据库，以及 MySQL 8.0 及以上通用数据库，确保数据存储的稳定性与兼容性，满足国产化适配要求。

3. 中间件：支持国产化中间件及通用中间件，保障服务间的高效通信与协同运行。

4. 容器环境：支持 Docker 20.10 及以上版本，Docker Compose 2.0 及以上版本，用于实现服务的容器化部署、隔离与快速迭代，降低环境依赖，提升部署效率。

5. 依赖组件：需安装 JDK 1.8 及以上版本（推荐 OpenJDK 1.8）、Python 3.7 及以上版本，以及 Zookeeper 3.6 及以上版本、Redis 6.0 及以上版本，确保平台核心服务正常运行；E

LK（Elasticsearch 7.10 及以上、Logstash 7.10 及以上、Kibana 7.10 及以上）用于日志存储与分析，Prometheus 2.20 及以上版本用于数据监控存储。

6. 网络环境：私有云部署模式下，需确保内网环境稳定，支持服务器节点间高速通信，带宽不低于 1000Mbps；云端部署模式下，需确保公网网络通畅，延迟不高于 50ms，保障接口调用的响应速度与稳定性。

三、附则

3.1 文档修订

本文档将根据平台功能迭代、技术升级及业务需求变化，定期进行修订更新，修订后将标注修订版本、修订日期及修订内容，确保文档内容与平台实际情况保持一致。

3.2 责任说明

1. 本文档所阐述的平台功能、技术参数、部署方案等内容，均基于北科瑞声 AI 语音云开放平台 V1.0 版本制定，若平台版本升级，相关内容将相应调整。

2. 平台的部署、第三方接入，需严格遵循本文档规定的标准与规范，若因未按规范操作导致平台故障或功能异常，北科瑞声不承担相关责任。

3. 本文档的最终解释权归北科瑞声所有。

3.3 联系方式

若在平台使用、部署、接入过程中遇到问题，可通过以下方式联系技术支持：

技术支持电话：0755-86329312、13066937587