

InterPhonic 6.5

语音合成系统

产品白皮书

安徽科大讯飞信息科技股份有限公司
ANHUI USTC iFLYTEK CO.,LTD.

目录

INTERPHONIC 6.5	1
1 公司简介	1
2 产品简介	3
2.1 系统概述	3
2.2 包含组件	4
2.3 支持音库	4
2.4 产品特性	5
2.4.1 系统尺寸:	5
2.4.2 硬件要求	5
2.4.3 技术指标	6
2.5 产品特点	6
2.6 产品应用	7
2.7 文档和相关资料	7
3 产品功能	8
3.1 功能概述	8
3.2 新增功能	9
3.3 功能特性	10
3.3.1 iFlyTTS接口	10
3.3.2 工具	10
3.3.3 定制资源包	11
3.3.4 CSSML文本标记语言	11
4 产品性能	12
4.1 性能的度量	12
4.2 影响性能的因素	12
4.3 测试环境	13
4.3.1 硬件标准测试环境	13
4.3.2 软件标准测试环境	14
4.3.3 基准环境	14
4.4 网络调用时的性能	14
5 开发接口简介	15
5.1 语音合成接口简介	15
5.1.1 iFLYTTS接口	15
5.1.2 QuickTTS接口	15
5.1.3 TTSCOM接口	16
5.1.4 SAPI5.0接口	16
5.2 资源管理接口简介	17
5.2.1 TTS-BGS接口	17
5.2.2 TTS-VPT接口	17

5.3 开发注意事项	18
6 技术支持	19

1 公司简介

安徽科大讯飞信息科技股份有限公司是一家专业从事智能语音及语言技术研究、软件及芯片产品开发、语音信息服务及电子政务系统集成的国家级骨干软件企业。是我国众多软件企业中为数极少掌握核心技术并拥有自主知识产权的企业之一，其智能语音核心技术代表了世界的最高水平。2008年，科大讯飞在深圳证券交易所挂牌上市，股票代码：002230。

科大讯飞成立于1999年，公司注册资本8036万元，主要股东包括：中国科学技术大学科技实业总公司、上海广信、联想投资、盈富泰克等。是我国产业化实体中，在语音技术领域基础研究时间最长、资产规模最大、历届评测成绩最好、专业人才最多及市场占有率最高的公司。

语音技术实现了人机语音交互，使人与机器之间沟通变得像人与人沟通一样简单。语音技术主要包括语音合成和语音识别两项关键技术。让机器说话，用的是语音合成技术；让机器听懂人说话，用的是语音识别技术。此外，语音技术还包括语音编码、音色转换、口语评测、语音消噪和增强等技术，有着广阔应用空间。

科大讯飞作为中国最大的智能语音技术提供商，在智能语音技术领域有着长期的研究积累，并在中文语音合成、语音识别、口语评测等多项技术上拥有国际领先的成果。科大讯飞是我国唯一以语音技术为产业化方向的“国家863计划成果产业化基地”、“国家规划布局内重点软件企业”、“国家火炬计划重点高新技术企业”、“国家高技术产业化示范工程”，并被信息产业部确定为中文语音交互技术标准工作组组长单位，牵头制定中文语音技术标准。2003年，科大讯飞获迄今中国语音产业唯一的“国家科技进步奖（二等）”，2005年获中国信息产业自主创新最高荣誉“信息产业重大技术发明奖”。2006年至2010年连续五届英文语音合成国际大赛（Blizzard Challenge）中蝉联大赛第一名。

基于拥有自主知识产权的世界领先智能语音技术，科大讯飞已推出从大型电信级应用到小型嵌入式应用，从电信、金融等行业到企业和家庭用户，从PC到手机到MP3/MP4/MP和玩具，能够满足不同应用环境的多种产品。科大讯飞占有中文语音技术市场60%以上市场份额，语音合成产品市场份额达到70%以上，在电信、金融、电力、社保等主流行业的份额更达80%以上，开发伙伴超过500家，以讯飞为核心的中文语音产业链已初具规模。

科大讯飞在语音技术核心研究和产业化方面的突出成绩引起了社会各界的广泛关注，吴邦国、贾庆林、李长春、习近平、贺国强等多位党和国家领导人都曾亲临科大讯飞视察，对科大讯飞做出的创新工作均给予充分肯定。国家各行业部委和权威机构也已经初步形成了科大讯飞“中文语音产业国家队”的基本共识。

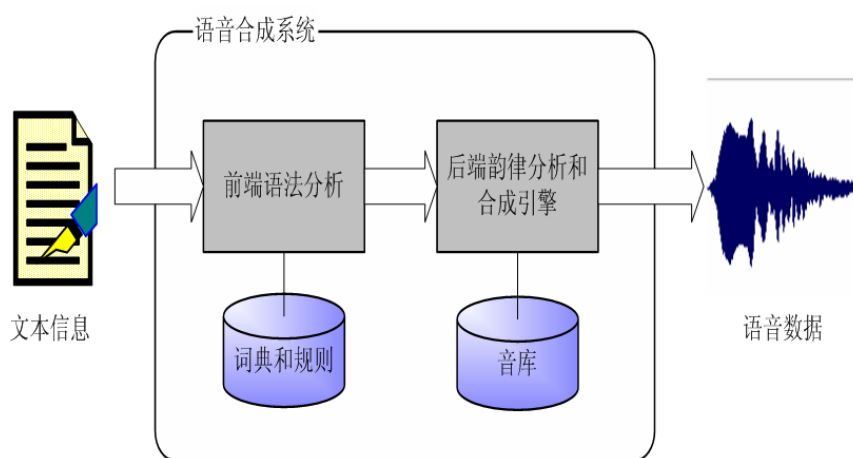
公司荣誉

- 国家863计划成果产业化基地
- 国家规划布局内重点软件企业
- 国家火炬计划重点高新技术企业
- 国家高技术产业化示范工程基地
- 国家级优秀软件企业
- 中国电子政务IT100强企业
- 中国中文语音交互技术标准工作组组长
- 国家博士后科研工作站

2 产品简介

2.1 系统概述

语音合成（Text-To-Speech），又称为文语转化，是将文本信息转化为语音数据的技术，涉及声学、语言学、数字信号处理、多媒体等多种前沿的高新科技，是中文信息处理领域的一项前沿技术。语音合成技术的原理示意如下：



图一：语音合成技术原理示意图

可以从图中看到，语音合成系统的合成引擎在完成文本到语音数据的转化过程中可以简单分解为两个步骤的处理。文本先经过前端的语法分析，通过科大讯飞公司精心制作的词典和规则的处理，得到格式规范，携带语法层次的信息，传送到后端。后端在前端分析的结果基础上，经过韵律方面的分析处理，得到语音的时长、音高等韵律信息，再根据这些信息在音库中挑选最合适的语音单元，语音单元再经过调整和拼接，就能得到最终的语音数据。

在整个转化处理的过程中牵涉到大量的中英文语法和韵律知识的运用，以及语法和语义分析的算法，最佳路径搜索，单元挑选和调整的算法，语音数据编码方面的知识。语音合成技术涵盖了语音文字的社会科学、数字信号处理等自然科学，是跨学科的高新技术。

InterPhonic语音合成系统是由安徽科大讯飞信息科技股份有限公司自主研发的中英文语音合成系统，主要面向行业级语音信息服务等应用，将文本信息动态实时地转换到流畅自然的语音数据。

InterPhonic

6.5语音合成系统以先进的大语料和TrainableTTS这两种语音合成技术为基础，能提供可比拟真人发音的高自然度、高流畅性、面向任意文本篇章的连续合成语音合成系统。InterPh

onic 6.5语音合成系统致力于建立和改善人—

机语音界面，为大容量语音服务提供高效稳定的语音合成功能，并提供从电信级、企业级到桌面级的全套应用解决方案，是新概念声讯服务、语音网站、多媒体办公教学的核心动力。

本文档针对希望了解InterPhonic

6.5语音合成系统的详细功能和性能特性的客户，详细阐述了InterPhonic产品的各项指标。通过本文，读者能够了解到世界顶尖级语音合成系统的整体功能，理解语音合成系统的概念和重要特性。

2.2 包含组件

InterPhonic

6.5语音合成系统是科大讯飞在长期收集客户需求和反馈的基础上，充分考虑集成开发和实际使用，面向大规模语音应用开发的产品。这套产品从功能上可以分成如下的组件：

- 语音合成引擎：提供语音合成功能的核心功能，包括了合成需要的各种资源文件、配置和应用程序如：资源管理中心和合成效果演示程序等；
- 开发接口：包括QuickTTS开发接口，TTS COM接口，SAPI接口、iFlyTTS等便于集成开发开发的接口封装，提供提示音和背景音资源管理接口及其开发文档和示例程序；
- 语音合成工具包：多种语音合成应用工具的集合。

2.3 支持音库

InterPhonic

6.5语音合成系统允许同时安装不同发音人、不同声音数据采样率的多种音库。系统当前支持的部分音库列表如下：

发音人	支持语种	支持采样率
小琪	中文及中英混读	8k/16k
小燕	中文及中英混读	8k/16k
小美	粤语及粤英混读	8k/16k
宇峰	中文、中英混读及英文	8k/16k
小倩	中文及中英混读	8k/16k
Henry	纯英文男声	8k/16k

Mary	纯英文女声	8k/16k
------	-------	--------

用户可以使用“资源管理中心”对当前安装的音库进行管理，进行设置缺省音库及注册/注销音库等操作。

用户可以采用下列方式来选择合成使用的音库：

1. 使用“InterPhonic资源管理中心”程序，在“音库管理”目录中，改变缺省的音库选项。

2. 使用iFly

TTS

SDK进行开发的过程中，使用相关接口函数的参数指定来进行设置。

3. 在合成演示程序主界面中直接进行音库选择。

2.4 产品特性

2.4.1 系统尺寸：

项目	安装包尺寸	运行库尺寸
运行库	349M	708M
安装包	367M	1G

2.4.2 硬件要求

服务器端	推荐配置： P4至强 3.0G双核双CPU； 2G内存； 140G SCSI硬盘； 100M Ethernet网卡； Windows2000/2003/XP/VISTA操作系统。	
客户端	最低配置	P4 2.4G CPU； 512M内存； 100M Ethernet 网卡； Windows2000/2003/XP/Vista操作系统。
	推荐配置	P4 3.0 CPU； 1G内存； 80G硬盘；

		100MEthernet网卡; Windows2003/XP/Vista操作系统。
--	--	--

2.4.3 技术指标

项 目		指 标
支持操作系统		Windows2000/2003/XP/Vista/2008
支持语种		汉语普通话、英语、粤语
支持文本控制标记		支持TCCM文本标记，CSSML
支持的输出语音数据格式	采样率	8K，16K
	声音数据格式	16bit/8bit PCM，aLaw/uLaw，ADPCM
支持速度调整		是，范围是：0.5~1.5
支持中英文混读		是
英文字符串处理		按照英文习惯发音
支持字符集		GB2312，GBK，BIG5，UNICODE，GB18030
识别汉字数量		支持GBK字库，共21003个简繁体汉字
系统词库词组数量		约300000条
支持用户定义词库		是
语音数据缓冲区设置		支持64K~2M Bytes
断句参数设置		支持回车符处理参数和最大断句长度参数
标点符号发音设置		支持标点符号发音方式的选择和具体发音的设置
数字串发音设置		支持数字串发音方式的选择
发音停顿风格设置		支持发音停顿风格的选择
节奏正确率		97%
句子可懂度		99%
自然度		4.5（5分制）

2.5 产品特点

语音合成产品对核心技术的依赖程度很高，为保证在核心技术上的长期领先优势，同时保证有效的将核心技术转化为功能、稳定性、易用性等各方面全面达到用户要求的高品

质产品，科大讯飞通过产学研相结合模式，保证研发水平和产品品质的同步提高。公司同国内语音研究领域具有领先水平的中科学声学所、中国科大和社科院语言所分别合作建立了联合实验室，将一些前瞻性的研究放在联合实验室完成，通过强强联合、优势互补，保证了核心技术的不断进步和领先优势。科大讯飞公司通过自主地研究和开发，在中文语音合成技术的多个领域都有独创性的技术成果。

- 独创的语料信息统计模型；
- 前后端一致性的语料库设计方法、和语料库的自动构建方法；
- 听感量化思想指导下，以变长韵律模板为基础的高精度韵律模型；
- 高鲁棒性的智能化文本分析处理技术；
- 基于听感损失最小的语料库裁减技术；
- 特定语种知识和系统建模方法分离的多语种语音合成系统框架；
- 面向特定领域应用的定制语音合成技术；
- Hmm-based波形拼接技术。

2.6 产品应用

语音合成技术是一种能够在任何时间、任何地点向任何人提供语音信息服务的高效便捷手段，非常符合信息时代海量数据、动态更新和个性化查询的需求。InterPhonic

6.5语音合成系统提供高效、灵活的服务，可以在多种领域内使用，如：

- ⇒ PC语音互动式娱乐和教学
- ⇒ 电信级、企业级呼叫中心平台
- ⇒ United Message Service (UMS)和Voice Portal 等新兴语音服务系统

InterPhonic的应用有利于改变原有的语音信息制作方式，即替代传统的简单拼接、人工录音或人工坐席，在规模语音应用系统中，多路并发的实时语音合成服务，不仅将显著提高信息更新的时效性，同时提供的语音信息内容与范围将得以大大拓展。InterPhonic语音合成系统高质量的语音效果、灵活的应用方式、经过实际应用验证的稳定性，已经使众多的客户成功地应用自助语音服务来取代传统的人工服务，在应用效果得以保障的前提下，提供了更高自动化程度，更低的成本、更高品质的服务。

2.7 文档和相关资料

语音合成系统的开发和使用可以参阅以下文档和资料：

- 随产品提供的《InterPhonic 6.5开发手册》；
- 在线帮助文档《InterPhonic 6.5帮助中心》，位于软件安装目录的Help目录下；

- 开发接口使用示例程序，位于安装目录的TTS SDK目录下；

3 产品功能

3.1 功能概述

InterPhonic语音合成系统是科大讯飞公司推出的新一代文语转化引擎，采用最先进的中文文本、韵律分析算法和大语料库的合成方法，合成语音已经接近真人的自然效果。InterPhonic语音合成系统具有的主要功能有：

1. 高质量语音——
将输入文本实时转换为流畅、清晰、自然和具有表现力的语音数据；
2. 多语种服务——
整合了多语种语音合成引擎，可提供中文、中英文混读、纯正英文、粤语、粤英文混读的语音合成服务；
3. 多音色服务——
提供丰富、风格多样化的音色选择，如浑厚淳正的男声，温柔甜美的女声，标准地道的英语男女声等等，所有音色库均继承科大讯飞语音合成技术一贯的优良品质。用户可根据不同应用业务需要，选择最适合应用场景的语音风格，并支持实时动态的音色切换；
4. 高精度文本分析技术——
保证了对文本中未登录词（如地名）、多音字、特殊符号（如标点、数字）、韵律短语等智能分析和处理；
5. 多字符集支持——支持输入GB2312、GBK、Big5、Unicode和UTF-8等多种字符集，普通文本和带有CSSML标注等多种格式的文本信息；
6. 多种数据输出格式——
支持输出多种采用率的线性Wav，A/U率Wav和Vox等格式的语音数据；
7. 提供预录音合成模板——
对合成文本中符合语音模板固定成分的文本使用发音人预录语音，非固定成分使用合成语音。这种方法有利于改进定制领域合成效果，简化定制流程，加快定制速度，同时也使得预录音的使用更自然、更灵活，满足更广泛的应用需求。
8. 灵活的接口——
提供了iFlyTTS接口、QuickTTS接口、COM接口、SAPI5.0接口、背景音管理接口和预录音管理接口，便于在多种环境下进行系统的集成；

9. 语音调整功能——

开发接口提供了音量、语速、音高（基频）等多种合成参数动态调整功能；

10. 配置和管理工具——

合成引擎提供了统一进行配置和管理的工具，完成了全局参数配置、用户词典、用户规则、定制资源包管理等功能；

11. 效果优化——

合成引擎提供了以定制资源包和CSSML为代表的多种针对实际应用环境进行合成效果优化的方法；

12. 一致的访问方式——

能以Client/Server方式访问远程的语音合成服务，并且提供与本地调用相同的开发接口，实现了完全透明的访问；

13. 背景音和预录音——

InterPhonic合成引擎提供了背景音和预录音的功能，大大提高了合成语音的自然度和表现力。

3.2 新增功能

InterPhonic

6.5语音合成系统是科大讯飞公司在业界享有盛誉的InterPhonic产品系列的最新产品，相比于之前的产品有多方面的改进和提升：

1. 合成效果提升：

1) 新的发音人及音色：InterPhonic

6.5语音合成系统中添加了在Blizzard国际英文大赛时采用的Henry音库。至此，InterPhonic

6.5语音合成系统可以提供小琪、燕平、宇峰、小美、谢婧、terry、王茹、Henry共8个不同发音人录制的音库，能够支持中文、中英混读、英文、粤英混读，每种音库分别提供8K和16K两种不同采样率的音库安装包。在合成过程中，如果安装了多个音库，用户可以根据需要实时切换并试听不同音色、不同采样率的合成效果。

2) 提升音库效果：采用超大规模语料库的方法，提升了通用语音库的合成效果；并针对一些音库做了优化，提供更自然流畅的合成语音；在教育领域的应用方面，我们处理了拼音、香港地区繁简字体的问题，使得InterPhonic

6.5在更广的领域里具有更深的应用价值。

3) 英文效果改进：InterPhonic

6.5语音合成系统提高了英文文本分析的准确率，使得英文合成的效果较之之前的产品更胜一筹。

4) 数字数值和姓名短语的改进：针对主流应用环境的普遍需求，科大讯飞广泛收集实际应用语料，并进行细致分析和专业的优化，InterPhonic

6.5语音合成系统在数字数值、姓名短语等最常见应用下效果提升显著，合成效果更加清晰准确、节奏感更强。

2. 语音合成模板

在InterPhonic

6.5语音合成系统中，增加了几百个常用的合成模板，并在“资源管理中心”增加了合成模板的管理功能和用户自己定制模板的功能，方便用户管理已有模板，并根据自己业务的需要决定还应增加哪些合成模板，以提高合成效果，降低模板制作的周期。

3. 多种合成技术

InterPhonic

6.5系统中整合了大语料库和可训练的参数合成等目前最先进的语音合成技术，并根据不同语种和不同风格发音人的特点综合利用这些合成技术的优点，使得InterPhonic

6.5系统在发音的自然度、流畅度、音库制作速度等方面都有了显著的提升。

4. 音频格式的转换

InterPhonic

6.5语音合成系统支持与当前发言人采样率和位数不一致的wav的外部调用，可以将与当前发言人采样率和位数不一致的wav以及与当前发言人格式不一致的alaw/mulaw转化成一致的格式。更方便用户体验灵活迅捷的语音合成功能。

3.3 功能特性

InterPhonic6.5作为一套完整的语音合成系统，除了基本的语音合成功能外还包含很多功能特性，以下简要介绍该系统重要的功能特性：

3.3.1 iFlyTTS接口

科大讯飞语音合成引擎以Windows动态链接库的形式提供了“iFly TTS SDK开发接口”。这是所有合成引擎最底层的开发接口，提供最全面的功能。可以使用任何支持从动态链接库中得到和使用接口函数语言（如VC，C++Builder，Delphi等）进行合成引擎的集成开发。合成引擎提供了完善的接口说明文档和Sample程序，便于程序开发人员查阅和参考。

iFly TTS SDK开发接口有如下技术特点：

- 满足实时性要求较高的应用需求
 - 接口功能全面、性能较高，支持多路合成
 - 以动态链接库的形式提供（iFlyTTS.dll）
 - 本地调用和网络调用都封装为相同的接口标准，可以实现透明的访问
- 接口函数和功能简介请参阅在线帮助文档《InterPhonic 6.5帮助中心》：

3.3.2 工具

InterPhonic6.5提供了多种语音应用的工具，包括：

- 加密锁检测程序

加密锁使用的目的是防止非法拷贝和控制License的数量。使用加密锁检测程序检测的内容包括语音合成系统产品名称、版本、授权数量和授权期限。

- TTSDoctor

TTS

Doctor是科大讯飞为本公司的语音合成系统软件而配备的自动化技术支持工具，用户安装InterPhonic语音合成系统的同时，该工具也一并安装到用户机器上。

使用TTSDoctor可以实现的主要功能有：查看已安装的TTS产品信息；查看系统信心；辅助判断出错原因；获得技术支持。

用户使用TTS

Doctor收集本机信息，并以*.ttd文件形式保存，作为自己的维护参考；当用户在使用InterPhonic语音合成系统的过程中遇到问题时，可以把这个文件发送给我们的技术支持人员，以获得迅速有效的技术支持。

- CEditor

CSSML

Editor文本编辑器是科大讯飞公司为CSSML文本标记语言专门定制的可视化文本编辑工具，利用它可以方便地从其它格式文本（.txt或.xml）生成CSSML格式的文本(.cssml)，从而提升文本的语音合成效果。

通过使用CSSML

Editor文本编辑器，可以实现的目标包括：利用CSSML标记修正合成效果欠佳的文本；检查CSSML语法是否规范；转换文本格式。

- 资源管理中心

资源管理中心程序是InterPhonic

6.5语音合成系统的管理中枢，管理着所有对系统和语音合成有影响的参数和资源。资源管理中心中的设置会影响注册表和资源文件等全局资源，因此如果进行了修改，更新的内容

在合成系统下次启动之后才能生效。

3.3.3 定制资源包

在实际的语音应用中，合成文本有时属于某个特定领域（如银行领域），其内容和形式具有显著的特征。针对这种特性，科大讯飞公司提出了领域定制的概念，领域定制是指在特定领域应用时使用适合的定制资源包来提高合成语音效果。“定制资源包”是一组语音合成引擎在特定领域应用时使用的资源的集合，包含音库、词库、外部规则库、自定义符号库等等。使用定制资源包有助于提升特定应用领域的语音合成效果。

3.3.4 CSSML文本标记语言

SSML（Speech Synthesis Markup Language）是W3C国际化标准化组织制定的语音合成文本标记语言的标准,SSML语言使用XML的形式定义了描述合成文本处理方式的多种标记，大致可以分成三类：定义语法层次的标记（如段落、句子、词组等）、定义语法特征的标记（如文本读法等）和定义韵律特征的标记（如语速、音高、音量等）。但SSML主要针对英文的标记制定，并不完全满足中文语音合成的应用需要。

CSSML（Chinese Speech Synthesis Markup Language）是由科大讯飞提出并牵头制订的中文语音数据描述规范。该标准得到了国家863专家组、国家信标委和技术监督局的极大关注与支持，目前已经有国家信息产业部下文成立标准工作组，推进此项标准的制定。

CSSML基于VoiceXML，同时考虑到中文合成领域内的特性，通过对SSML规范中的标记进行必要扩展，使之更好的用于中文语音交互的使用，并且保持对SSML 1.0语言规范的完全兼容。目前CSSML可以对音标、轻重读、韵律层次、语气语调、背景音等多种特性进行灵活标注，并具有良好的扩展特性，有效提升了语音合成系统的实际应用能力。

通过CSSML对合成参数的设置，还可以对指定的文本设置特殊的合成参数（而不是像使用合成引擎提供的“资源管理工具”，一旦设置对所用合成文本都产生作用），从而对合成效果实现更灵活的控制。

InterPhonic

6.5系统能够自动识别处理规范格式的CSSML文本，正确支持其中的各项标记。用户可以在调用相关接口函数时指定输入的文本为CSSML格式。通过CSSML Editor对需要合成的文本进行可视化编辑，是一种行之有效的效果优化方案。

4 产品性能

4.1 性能的度量

语音合成系统的基本功能是将文本信息转化为语音数据，这一转化过程的快慢会影响到使用语音合成系统提供语音服务的应用程序的表现。如果转化速度过慢，就可能出现播放的语音不连续或者是等待时间过长等问题。因此，语音合成系统的性能特征直接关系到系统应用的效果。

对于多路并发的语音合成系统，最直观的性能度量方法是测量每一路语音合成服务合成一定量的文本需要的时间，合成需要的时间越短，合成系统的性能越高。科大讯飞公司在长期与电话语音应用的集成商和运营商合作的过程中了解到，在语音业务中合成响应时间最好在3秒左右，最大不能超过6秒，否则用户会觉得等待时间过长，而播放合成语音一般以句为单位，文本长度在50汉字（100字节）左右。因此科大讯飞性能测试中使用“合成时间”来表示合成的快慢，定义为平均合成50个汉字（100字节）的时间值。根据电话语音应用的实际要求，对语音合成系统的性能能否达到“实时”进行了如下的定义：在多路并发合成不同文本的情况下，如果每一路满足最长合成时间小于6秒，平均合成时间约为3秒，就认为在该并发路数下，合成系统可以满足实时性要求。

为了准确的测量合成系统的性能指标，科大讯飞公司使用自行开发的测试工具MTTester进行不间断、高密度、多路并发的合成服务测试，在测试过程中获得的数据作为合成的性能数据。测试工具的每一路合成都使用不同的测试文本，在并发的情况下测量获得一个文本文件的所有语音数据需要的总时间，再平均到50个汉字需要的时间为“合成时间”。随着并发路数的增加，系统的CPU和内存的消耗越来越大，每一路的“合成时间”也会逐步增加。当合成时间满足实时性要求的临界条件时，可以得到在该测试环境下的“最大并发合成路数”。“最大并发合成路数”同样也表征了的合成系统的性能，值越大性能越高。

InterPhonic

6.5系统总体性能是按照吞吐量来计算的，吞吐量跟所具有的license个数密切相关，依上所述，一路平均合成效率为2秒50个汉字(即1秒25个汉字)，则合成系统的总体合成吞吐量为：1秒 25个汉字× license个数。

4.2 影响性能的因素

InterPhonic语音合成产品为了实现最好的合成效果，采用了复杂度非常高的算法，也使用了大量的资源文件（如音库、词库等）。因此合成系统的性能指标是受软件和硬件多

方面因素影响的。下面将对影响合成系统性能的因素进行一些分析和说明。

1. 语音合成系统本身的算法复杂度等因素：因为每个版本的语音合成系统的内部算法和资源使用都不尽相同，不同版本合成系统的性能表现是有差别的，有的系统之间的性能差别会非常大。

2. 测试服务器的CPU运算能力：因为合成系统内部进行了大量的逻辑和数学运算，运行语音合成的服务器的CPU运算能力（简单可以用CPU的主频来表示）和CPU的数量，都会极大的影响合成系统的性能表现。CPU主频越高，数量越多，合成系统的性能越高；

3. 内存：语音合成系统在服务的过程中，使用了大量的内存进行中间数据的存贮，并且内存中的数据进行了非常频繁的操作和计算，因此内存的可用大小和存取速度对合成系统的性能也有显著的影响，如果内存容量低于要求的大小，合成系统不能表现出最佳的性能；内存的存取速度越快，合成系统的性能也会越好。InterPhonic

6.5建议运行合成服务的计算机推荐配置2G的内存。

4. 操作系统：不同的操作系统对进程和线程的调度、内存管理、网络服务等方面的实现都存在差异。根据科大讯飞测试的结果，服务器版（如Advance Server3.0版）的产品，其性能表现和服务稳定性往往高于桌面级的产品（如Redhat9.0版）。因此科大讯飞公司建议在大规模的语音合成服务应用中使用服务器版。

5. 硬盘：因为合成系统使用的资源文件的尺寸都非常大，达到了Gbytes的量级，在合成过程中（特别是在开始提供合成服务的一段时间内），合成系统需要频繁的与硬盘交换数据，因此硬盘的随机访问速度也会对合成性能造成影响。建议客户在使用语音合成服务的计算机上配置转速高于7200转的IDE硬盘或者使用高速的SCSI硬盘。

6. 网络环境：在通过网络调用实现远程语音合成服务的应用中，网络环境的配置也会极大地影响合成系统的性能表现。在一台并发运行100路以上语音合成服务的计算机上，其网络交换的数据量会达到几十兆bps甚至更高。因此建议服务端和客户端都配置100Mbps以上的网卡，并且最好都在同一个局域网内。另外强烈建议使用交换速率和带宽更高的100Mbps网络交换机，而不是100Mbps的集线器。如果局域网带宽为10Mbps，语音合成网络服务的效率会极大地受到限制。

4.3 测试环境

4.3.1 硬件标准测试环境

服务端：Intel Xeon 2.4G ×2 CPU 1G RAM 140G SCSI HD 100M Ethernet网卡

客户端：Pentium IV 2.4、512M DRAM、80G HD

4.3.2 软件标准测试环境

服务端：redhat advance server3.0

InterPhonic 6.5语音合成系统（企业版，安装8K音库）和iFly TTS SDK标准接口

客户端：Microsoft Windows XP操作系统

4.3.3 基准环境

“硬件标准测试环境”，本地测试使用服务端的硬件环境；

“软件标准测试环境”，本地测试使用服务端的软件环境；

合成参数全部取缺省值；

合成语料是包含数字、英文和中文的5K左右大小的GB2312混合文本；

4.4 网络调用时的性能

下表所列数据为InterPhonic

6.5在网络环境下不同并发线程时的合成函数响应时间和吞吐量等测试参数。经测试，各音库的性能表现基本相似，此处所列数据均为小燕音库的测试数据。

测试方法：在服务中将SES启用，客户端用isp客户端，端口指定为13606。

授权数量	最长合成时间(s)	最短合成时间(s)	平均合成时间(s)	客户端吞吐量(byte)	系统占用的内存峰值 (mb)		CPU占用时间%	
					服务器端	客户端	服务器端	客户端
1	1.714	1.653	1.683	59	213	2	1.1	0.1
4	1.720	1.658	1.683	238	221	5	2.8	0.4
10	1.735	1.514	1.656	604	282	6	9.2	0.7
30	1.841	1.539	1.693	1772	385	9	31.6	2.9
60	2.656	1.624	1.874	3202	528	13	70.3	7.3
90	3.992	1.792	2.391	3764	678	18	86.8	10.3

线程数量与授权路数一致。

5 开发接口简介

科大讯飞语音合成开发包提供了多个能够使用讯飞TTS系统进行资源管理和语音应用的开发接口，通过这些接口用户可以开发出具有TTS功能的各种企业级、桌面级应用系统。这些接口都需要语音合成引擎的支持，并且各个接口的功能和面向的开发者不同，适合多种开发环境，开发者可以根据具体的应用场合进行选择。

5.1 语音合成接口简介

5.1.1 iFlyTTS接口

iFlyTTS接口是语音合成引擎提供的基础引擎支持功能的开发接口，提供功能全面、性能出色的语音合成开发功能。iFlyTTS接口能够满足实时性要求较高的应用需求，接口功能全面、性能较高，支持多路合成、网络合成和各种参数设置。

它以动态链接库iFlyTTS.dll的形式提供，是最常用的TTS系统的开发接口，这个接口对开发人员的技术开发能力有较高的要求。这个接口支持C/C++、Delphi、C++ Builder等语言的调用，因为使用了大量的结构和指针，不支持VB、C#、Java等高级语言的调用。

目前iFlyTTS接口还增加了多引擎管理接口和语音资源管理接口两套新的应用接口。多引擎管理接口可以用来管理、查询位于本地或网络中的所有语音引擎及每个引擎拥有的Voice信息，控制语音服务引擎、查询引擎的静态属性及动态统计信息，以实现多合成引擎的语音服务需要。语音资源管理接口是讯飞语音应用平台客户端提供的应用程序接口，用来管理语音应用平台上的所有资源，包含控制语音服务、获取网络中语音服务节点的信息，监控节点状态，并实现动态负载均衡的功能。

接口组件：

开发组件	组件组成	说明
头文件	iFly_TTS.h、TTSErrcode.h	接口函数的声明
动态引入库	iFlyTTS.lib	包含接口函数的引入
运行时刻库	iFlyTTS.dll	相关动态链接库

接口位置：

类型	存放位置
接口位置	安装目录的.\ttssdk\bin目录下
接口示例	安装目录的.\ttssdk\samples\iflytts目录下

5.1.2 QuickTTS接口

QuickTTS接口是在iFlyTTS.dll接口的基础上，给用户提供的进行TTS系统快速开发的接口。QuickTTS接口提供的函数使用一些基本的数据类型，对外提供更加简单易用的函数接口形式，用户只需调用一个函数就可以实现文本到语音文件的转换，适合VB、C#等语言的调用，这个接口还支持Java接口，可以实现Java环境下调用TTS系统进行合成。接口也支持多路合成、网络合成和各种参数设置。。

QuickTTS接口以QuickTTS.dll的形式提供，调用时需要QuickTTS.dll和iFlyTTS两个dll库的支持，用户还可以在接口的配置文件QuickTTS.cfg中进行的接口的日志配置。

接口组件:

开发组件	列表	说明
头文件	QuickTTS.h	QuickTTS接口的声明
动态引入库	QuickTTS.lib	包含接口函数的引入
运行时刻库	iFlyTTS.dll、QuickTTS.dll	相关动态链接库

接口位置:

类型	存放位置
接口位置	安装目录的.\ttssdk\bin目录下
接口示例	安装目录的.\ttssdk\samples\QuickTTS目录下

5.1.3 TTSCOM接口

TTSCOM接口是在iFlyTTS.dll接口的基础上, 使用ATL开发的一套基于COM技术的语音合成插件, 以供二次开发商使用, 特别是提供给使用VB和ASP等开发人员的一套现代化的基于组件的开发工具。

TTSCOM接口以iFlyTTSCtrl.ocx的形式提供, 首次使用前及更改iFlyTTSCtrl.ocx的路径时需要对这个接口进行注册, 用户可以使用讯飞提供的TTSCOM“配置工具”(CtrlConfigure.exe)或者直接用系统的Regsvr32命令进行注册。调用时需要指定iFlyTTS.dll接口的路径, 这个可以在讯飞的“配置工具”中指定, 或者直接在程序代码中指定。

接口组件:

开发组件	组件组成	说明
头文件	iFlyTTSCom.h	接口函数的声明
动态引入库	无	无
运行时刻库	iFlyTTSCtrl.ocx、iFlyTTS.dll	相关动态链接库

接口位置:

类型	存放位置
接口位置	安装目录的.\ttssdk\bin\ttscom目录下
接口示例	安装目录的.\ttssdk\samples\ttscom目录下

5.1.4 SAPI5.0接口

SAPI5.0接口是在iFlyTTS.dll接口的基础上, 按照微软Speech API 5.0接口规范开发并且完全符合其应用方法的用户开发接口, 该接口提供了对COM、Active X的全面支持, 还支持多线程和XML标记。

SAPI5.0接口以iFlySAPI.dll的形式提供, 使用前需要保证操作系统中已经安装了Microsoft SAPI

5.0及以上组件。用户首次使用这个接口时需要对接口进行注册, 可以使用讯飞提供的SAP

15.0 “注册工具”（RegSAPI11.exe）进行注册，注册时需要选择iFlyTTS.dll接口和注册文件RegisterSAPI.txt所在的位置，单击“注册”即可。如果用户不再使用这个接口，可以使用“注册工具”卸载SAPI5.0接口。

接口组件：

开发组件	组件组成	说明
头文件	sapi.h sapiiddk.h spcollec.h Spddkhlh.h spdebug.h sperror.h speventq.h sphelper.h spuihelp.h	接口函数的声明
动态引入库	sapi.lib	包含接口函数的引入
运行时刻库	iFlySAPI.dll、iFlyTTS.dll	相关动态链接库

接口位置：

类型	存放位置
接口位置	默认为C:\Program Files\Common Files\SAPI1.1目录下
接口示例	安装目录的.\tssdk\samples\Sapi5.0目录下

5.2 资源管理接口简介

5.2.1 TTS-BGS接口

管理TTS系统的背景音资源的接口（TTS-BGS SDK），用户可以根据这个接口编写管理背景音资源的应用程序。

这个接口只能在安装TTS系统的机器上使用，既只支持本地调用或者共享文件式的调用，不支持网络调用。并且接口针对于具有一些Win32编程经验的C/C++程序员。

接口组件：

开发组件	组件组成	说明
头文件	TTS_BGS.h iFly_TTS.h TTSErrcode.h	接口函数的声明
动态引入库	bgsmsgmt.lib	包含接口函数的引入
运行时刻库	bgsmsgmt.dll、aucodec.dll	相关动态链接库

接口位置：

类型	存放位置
接口位置	安装目录的.\ressdk\bin目录下
接口示例	安装目录的.\ressdk\samples\bgstest目录下

5.2.2 TTS-VPT接口

管理TTS系统的预录音资源的接口（TTS-VPT SDK），用户可以根据这个接口编写管理预录音资源的应用程序。

这个接口只能在安装TTS系统的机器上使用，既只支持本地调用或者共享文件式的调用，不支持网络调用。并且接口针对于具有一些Win32编程经验的C/C++程序员。

接口组件：

开发组件	组件组成	说明
头文件	TTS_VPT.h iFly_TTS.h TTSErrcode.h	接口函数的声明
动态引入库	vptmgmt.lib	包含接口函数的引入
运行时刻库	vptmgmt.dll、aucodec.dll	相关动态链接库

接口位置：

类型	存放位置
接口位置	安装目录的.\ressdk\bin目录下
接口示例	安装目录的.\ressdk\samples\vpttest目录下

5.3 开发注意事项

用户在用上述接口进行开发时，可以先参考一下系统的示例程序，把需要的.dll动态库文件和相关的.h头文件及.lib文件拷贝到程序目录下，或者按照说明注册相关的接口，然后运行一下示例程序，以了解接口的调用过程及结果，再进行开发。

关于开发接口的详细介绍和开发说明请参见《InterPhonic 6.5帮助中心》，如果用户在安装使用或者开发中遇到任何问题，请及时电话或者邮件联系我们。

6 技术支持

如果您在安装或使用中有任何问题或者建议，请与我们联系！！

联系时的对问题的描述应尽量包含以下内容：

- * 系统配置（包括CPU、内存、硬盘、操作系统及产品版本等信息）
- * 问题细节（包括问题的重现过程及合成的文本内容等）
- * 问题重现（包括详细的操作过程）

• 电话支持

请于周一～周五，北京时间9：00～17：00间，拨打电话0551—5331813获得技术支持信息。

• 电子邮件支持

请将问题的详细描述发至：tts_support@iflytek.com

• 信件支持

请将问题详细描述发至：安徽省合肥市望江西路666号高新区信息产业基地讯飞大厦

邮政编码：230088

或传真至：0551—5331801 / 5331802