

InterSound™ 4.0 嵌入式语音合成系统

技术白皮书

安徽中科大讯飞信息科技有限公司

目录

1. 概述.....	2
2. 系统特性.....	3
2.1 高品质的语音合成效果.....	3
2.2 极低的资源需求.....	4
2.3 良好的系统移植性.....	4
2.4 灵活的系统架构.....	4
2.5 高度可裁减性.....	5
2.6 强大的定制能力.....	5
2.7 丰富的控制接口.....	6
2.8 可靠的开发支持.....	6
3. 技术特点.....	7
3.1 TTS合成效果指标.....	7
3.1.1 整体TTS特性.....	7
3.1.2 非受限文本合成引擎合成效果指标.....	8
3.2 文本标注.....	8
3.3 资源占用指标.....	8
3.3.1 InterSound™ 4.0 KM资源占用指标.....	8
3.3.2 InterSound™ 4.0 KB资源占用指标.....	9
4. 应用范围.....	10
5. 用户价值体现.....	11
6. 服务与支持.....	12

1. 概述

随着现代科学技术的快速发展，新的科技成果被源源不断的应用到电子产品之中，我们在日常生活中接触到的电子设备变得越来越智能化和小型化，如何让人们更方便的使用各种电子设备，与机器进行便捷的信息交流和准确的信息获取，使机器具备更加人性化的使用界面，成为电子产品发展的一个潮流。语音作为一种自然、灵活、便捷的交互方式，逐渐成为重要的人机交互手段。

如今，大容量、高自然度的语音合成技术已经比较成熟，广泛的应用于声讯服务、行业企业、计算机等领域，为人们的生活带来了便利。然而，语音合成技术在嵌入式环境下却依然没有得到广泛应用，约束其发展的主要原因有几个：

- 在嵌入式环境下，硬件和软件资源往往有很大的限制，传统的基于大语料库的语音合成技术需要非常高的运算资源和存储资源，无法满足嵌入式环境下对资源的苛刻需求。
- 一些简化过的轻量级语音合成算法虽然在资源占用上可以满足嵌入式环境的要求，但是由于合成算法过于简单，语音合成效果不尽人意，往往无法达到人们可以接受的程度，最终影响了使用效果。
- 嵌入式环境的硬件和软件平台千变万化，存在无数种硬件平台、操作系统和软件环境，传统的语音合成系统在适用性和移植性上也无法满足多样的嵌入式环境的需求。

为了解决语音合成技术在嵌入式环境下的应用问题，科大讯飞公司凭借着深厚的语音技术积累以及多年的嵌入式应用开发经验，针对上述关键点进行了长时间的研究和技术攻关，最终成功推出了真正适用于嵌入式应用的中文语音合成系统 InterSound™ 4.0，在系统资源占用、语音合成效果、系统移植性等多个关键点取得了突破性的进展，为语音合成技术在嵌入式环境下的应用推广迈出了重要的一步。

2. 系统特性

InterSound™ 4.0 具备多项独特的功能以及优异的性能，以下将进行详细的介绍：

2.1 新特性

InterSound™ 4.0 版本的新特性包括：

- 语种可扩充，包括多语混读模式，允许集成第三方引擎
- 受限文本范围可扩充
- 汉语拼音输入识别
- 支持流式文本输入
- 处理状态控制
- 处理进度通知
- 同步回显支持汉语拼音输出
- 允许关闭音频输出
- 支持 3 种发音风格
- 同时支持 4 个发音人
- 支持标点符号朗读
- 支持用户自定义汉语词库
- 支持用户自定义提示音
- 提示音支持汉字、字母、数字、符号混合的句子或短语，最大可跨 8 个 Token

2.2 高品质的语音合成效果

InterSound™ 4.0 合成系统具有高品质的语音合成效果，其原因来源于多个方面。

- 基于最新的嵌入式任意文本语音合成研究成果

InterSound™ 4.0 在嵌入式环境下的任意文本语音合成算法上取得了重大突破，新的语音合成算法在合成效果上已经接近传统大语料库语音合成算法的效果，合成效果与以往的嵌入式语音合成系统相比，更加自然和流畅。

- 集成了智能的文本分析算法

InterSound™ 4.0 的文本处理前端集成了多种智能的文本分析算法，将规则和统

计等处理方法进行了有机的结合，具备了较强的分词、多音字识别、姓氏识别、韵律环境预测能力，以及自动识别和处理日期时间、数值金额、电话号码等文本的能力，并能通过添加外部规则的方式，快速增加对新的文本格式的支持能力。

- 对受限文本可定制分析算法和音库

InterSound™ 4.0 系统中引入了受限定制引擎的概念，可针对文本变化范围可事先预知的特定文本制定专门的文本分析策略和音库。定制处理算法与通用合成算法相比，合成效果更为准确和自然。

2.3 极低的资源需求

InterSound™ 4.0 在系统架构、代码实现、资源结构等方面进行了大量的优化工作。

- 整体系统架构的设计紧凑高效；
- 系统开发过程中，在合成算法、数据结构、代码编写等不同层面都进行了深入优化；
- 资源的结构设计也非常精巧，既降低了资源的尺寸，又提升了资源访问的效率。

最终的优化结果，使得 InterSound™ 4.0 系统在运算资源和存储资源的需求方面，都较传统的合成系统有了极大的降低。

2.4 良好的系统移植性

InterSound™ 4.0 系统对不同的硬件和软件环境有极强的适应能力和移植能力，系统对与具体环境相关的部分进行了抽象设计，如果要移植到不同的环境下，只要重新实现少量接口，即可快速实现不同环境下的版本。

- 目前 InterSound™ 4.0 系统在多种处理器芯片环境下，多种操作系统甚至无操作系统的软件环境下，都已经实现了移植版本。支持的操作系统包括：Win32，WinCE，HOPEN，Nucleus，Embedded Linux， μ COS，Symbian OS 等。支持的处理器芯片包括：X86 系列，ARM 系列，MIPS 系列，EPSON，TI DSP，AD DSP，SigmaTel 等 32 位和 16 位处理器。
- InterSound™ 4.0 可兼容不同的字节序环境，如 Little-Endian 和 Big-Endian 字节序。
- InterSound™ 4.0 的资源位置可以任意配置，既可配置到支持线性访问的 Nor Flash，P-ROM 等存储器中，也可配置到分块访问的 NandFlash 中，如果环境支持文件系统，还可以以文件的方式配置。这使得系统对不同嵌入式环境的支持能力大大增强。

2.5 灵活的系统架构

InterSound™ 4.0 具有非常灵活的系统架构，体现在：

- InterSound™ 4.0 采用了面向对象以及模块化设计的思想，具有非常灵活的系统架构。在统一的系统架构下，可挂接多种的语音合成处理引擎，各个引擎遵守标准的调用接口和数据结构定义，在共享了数据处理方法的同时，又使得各引擎可以任意搭配和混合使用，针对不同的应用领域，组合出不同的功能，满足不同场合下的要求。
- InterSound™ 4.0 可以集成第三方的 TTS 引擎在系统中，这给需要多语种的用户提供了统一的操作界面和整体处理策略，既方便用户使用，有助于充分保证混合语言合成时的效果。
- InterSound™ 4.0 的系统架构实现了代码和资源的分离，可更好的适应某些硬件应用场合，极少变动的系统代码可以事先 Mask 到 ROM 介质或者芯片内部，而可能会经常变动的库资源，可存储在可擦写介质中，让用户可以随时更新。
- InterSound™ 4.0 还允许用户在运行态可以动态插入自定义的数据处理模块，使得用户可以方便的增加新的数据处理算法，比如支持新的编解码算法，支持新的声音处理效果（如调节语速，调节音量，增加音效处理等），使得系统本身具备较强的可扩展性。

2.6 高度可裁减性

- 算法可裁减，有多种不同复杂度的语音合成算法可以选择，代码实现可繁可简，同时对动态 RAM 资源的需求也会有所变化。
- 功能可裁减，有多种语音合成引擎可以选择，对不需要的功能，可以不配置。
- 尺寸可裁减，支持多种语音合成算法和语音编解码算法，对不同的效果要求，资源的尺寸可大可小，具有很大的伸缩性。

2.7 强大的定制能力

InterSound™ 4.0 的定制能力体现在多个方面。

- InterSound™ 4.0 灵活的系统架构，使得功能的任意配置成为可能，用户可以决定是否采用或者去除某个处理算法或模块。此外，InterSound™ 4.0 的资源本身具有一定的自描述能力，用户可以通过控制资源的内容和大小来定制 InterSound™ 系统的功能。
- 用户自定义词典功能的引入，使得 InterSound™ 系统可以针对特定的文本环境定制专门的词典，在特定文本环境下，文本处理正确率大大改善。
- 受限定制引擎的引入，使得 InterSound™ 系统可以针对特定的文本范围录制专门的音库，在特定文本范围下的语音合成自然度方面表现更佳。
- 提示音功能可以根据不同的用户的需求，对一些固定的文本内容，由播音员事先录制成提示音，当合成系统遇到这些固定文本的时候，直接使用预先录制的自然语音进行回放，达到较好的语音效果。如果将预录音引擎配合受限定制引擎使用，可以在一些特定领域下接近真人播报的效果。

- 讯飞公司具有多年积累下来的丰富的语音数据制作经验。针对 InterSound™ 4.0 的系统定制,有专门的数据定制小组负责受限引擎的制作以及预录音定制服务。可以按照用户的需求,定制专用的受限合成引擎和定制音库,录制和制作预录音音库。讯飞公司拥有自动高效的定制工具和丰富的定制经验,可以在极短的时间内满足用户的定制需求。

2.8 丰富的控制接口

讯飞公司具有多年的嵌入式应用开发经验,因此在进行系统设计的时候,充分考虑到了用户对语音合成技术的需求,在调用接口上有很多体贴的设计。

- 用户可以通过简短的文本标记,控制合成系统的最终表现,例如,通过“数字读法”的控制标记,用户可以任意制定遇到数的时候,是按数值方式还是号码方式来念,当然,也可以选择让系统自动判断。目前 InterSound™ 4.0 的基本系统支持“数的读法”,“一的读法”,“静音产生”,“多文本内码切换”,“定制引擎标记”等标记。增强型的 InterSound™ 系统,可以支持调节语速,调节音量等增强型的文本标记,甚至扩展出更多的标记。
- 处理进度通知功能。InterSound™ 4.0 系统首次实现了在合成文本的同时通知应用当前合成的文本位置的功能,这使得实现希望在合成语音的同时显示播放位置的应用提供了可能。
- 处理状态控制功能,允许用户随时暂停或停止合成,这为用户提供了更大的控制能力,配合处理进度通知功能,使用户可以实现前所未有的应用模式,例如中断当前合成插播内容,并在插播结束后从中断处恢复。
- 同步回显功能。InterSound™ 4.0 系统可以在播放声音的同时,同步显示该声音对应的文本,使得应用可以实现边听边看,先看后听等功能,对屏幕显示空间有限的嵌入式电子设备和一些教育软件应用有非常高的实用意义。
- InterSound™ 4.0 提供了对多种汉字编码方式的支持,包括 GB2312/GBK, Big5, Unicode (含 UTF16LE、UTF16BE、UTF8 等),充分考虑了不同地区用户的需求。

2.9 可靠的开发支持

- InterSound™ 4.0 的 API 开发接口非常简洁,易用,且功能强大。Standard API 和 Simple API 两层开发接口分别针对不同层次的用户,方便用户选用。
- 在 InterSound™ 4.0 的 SDK 开发包中,还提供了详细的开发文档以及演示程序。
- 除此以外,讯飞公司还专门配置了经验丰富的技术支持人员来支持 InterSound™ 4.0 的应用移植和开发,解决用户开发过程中遇到的困难,帮助用户设计语音应用方案。

如上特性,保证了 InterSound™ 4.0 系统比较容易移植并集成到各种应用环境下,为用户的应用程序带来了强大的语音界面,并给最终客户以全新的语音效果体验。

3. 技术特点

3.1 TTS 合成效果指标

3.1.1 整体 TTS 特性

项目	指标
文本类别	普通书面文本
识别汉语拼音串	支持
定制文本范围	支持汉语号码、数值、日期、时间，还可根据用户需要定制
提示音	支持
字符集	GB2312 / GBK / BIG5 / Unicode (UTF16LE / UTF16BE / UTF8)
语种	汉语普通话，可集成第三方其他语言引擎
中英文混读	是，英文按照字母或单词发音，也可集成第三方英文引擎
特殊符号识别	支持
系统词库词组数量	KB 引擎约 55000 条，KM 引擎约 20000 条
用户自定义词典	支持
多音字识别	支持普通多音字和姓氏多音字
发音人	天畅 / 文静 / 燕平 / 宇峰
发音风格	天畅仅支持一字一顿；文静、燕平、宇峰支持一字一顿和平铺直叙，其中燕平还支持有声有色
语音数据输出格式	PCM 系列或其它用户自定义的语音编码格式
速度调整	支持
音高调整	支持
音量调整	支持

输入模式	支持完整缓冲区输入和流式回调输入
合成参数变更通知	支持
处理进度通知	支持
同步回显	支持
回显汉语拼音	支持
关闭输出	支持

3.1.2 非受限文本合成引擎合成效果指标

在 InterSound™ 4.0 系统中存在 KB 和 KM 两种非受限文本合成引擎，它们可以合成任意的中文文本。KB 引擎是指资源占用较大，合成效果较为自然流畅的合成引擎。KM 引擎是指资源占用较低，合成效果能满足基本需求的合成引擎。下表中的清晰度、自然度的评分按照 5 分制给出，5 分为最高分，0 分为最低分。

项目	平铺直叙风格(KB)	一字一顿风格(KM)
自然度	4.2	3.3
节奏正确率	95%	72%
句子可懂度	98%	96%

3.2 文本标注

InterSound™ 4.0 支持简单文本标注。文本标注在文本中用于设置处理策略，或执行一些特殊的功能，如静音一段时间等。

详细的文本标注列表参见《InterSound™ 4.0 简单文本标注用户手册》。

3.3 资源占用指标

3.3.1 InterSound™ 4.0 KM 资源占用指标

- RAM 及代码尺寸占用情况（单位为字节）

代码空间	RAM	
	堆空间	栈空间
20K	4K	< 1K

- 资源大小（单位为字节）

码流率	大小	运算量
256Kbps	15M	< 1MIPS
32Kbps	2.5M	< 1.5MIPS
5.3Kbps	800K	3.5MIPS

3.3.2 InterSound™ 4.0 KB 资源占用指标

- RAM 及代码尺寸占用情况（单位为字节）

代码空间	RAM	
	堆空间	栈空间
80K	9K	< 1K

- 资源大小（单位为字节）

码流率	大小	运算量
256Kbps	65M	< 2MIPS
32Kbps	8M	< 2.5MIPS
5.3Kbps	2.5M	5MIPS
谱参数合成	1M	12MIPS

注意：以上各项技术指标数据是 InterSound™ 4.0 系统的通用指标，反映了 InterSound™ 系统的总体指标。在不同的具体软硬件环境配置环境下，实际的技术指标会有所差别。其中 RAM 空间均不包括播音缓冲区。

4. 应用范围

- **汽车电子**

GPS，车内诊断系统，车载电话等

- **手机通讯设备**

Caller ID，语音电话本，短信播报，语音导航等

- **电子图书**

使得 eBook 产品具备语音能力，表现力更加丰富，具有更好的教育和学习意义。

- **行业应用**

资费查询，机场车站广播等

- **智能仪器**

给电子仪表加上自然流畅的语音报读功能，增强仪器的智能化，在一些特殊的较危险应用场合，比如电力检测等，可以解放使用者的双眼，增加安全系数。

- **PC 桌面软件**

可集成到一些常用的桌面软件中，如文本处理、文档校对，信息阅读、电子游戏等软件中。增加软件的实用性和娱乐性。

其他的应用范围还有智能家电、语音玩具、电子导游机，自动售货机等

5. 用户价值体现

- 增强用户满意度

InterSound™ 4.0 提供了当前最好的嵌入式语音合成技术，自然流畅的语音效果，将您的应用程序带来更为友好和人性化的使用界面，增加用户的满意度。

- 增强用户体验

InterSound™ 4.0 可集成到多种嵌入式设备中，让最终客户可以随时随地都享受到高品质的语音合成服务。

- 新的商业应用点

InterSound™ 4.0 将您的产品增加独特的亮点，可增强产品竞争力，对扩宽产品市场范围有极大的帮助。

6. 服务与支持

关于本开发包更加详细的信息请访问<http://www.iflytek.com/>，或者通过本公司在您所在地的代理商获得最新的资讯。

若您还有任何疑问，请与我们联系：

- **联系地址**

安徽省合肥市国家级高新技术产业开发区信息产业基地讯飞语音大厦

安徽中科大讯飞信息科技有限公司

嵌入式产品部

邮编：230088

- **联系电话**

0086-551-5331816

- **传真**

0086-551-5331801

0086-551-5331802

- **电子邮件**

embedded_support@iflytek.com

讯飞语音 沟通无限

科大讯飞公司凭借深厚的语音合成技术积累以及雄厚的研发能力，将持续不断的研究嵌入式语音合成技术，提供最新最好的代表用户需求的语音合成系统，不断追求为用户创造价值的目标，推动语音信息化时代的发展。