**TTSmate V1 系统需求说明书**

**1. 引言**

**1.1 编写目的**

本文档旨在详细说明 TTSmate V1 智能语音合成客户端的系统需求，为开发测试和最终用户提供一份清晰准确的功能和性能参照。

**1.2 背景**

**1.2.1软件名称**

TTSmate V1 - AI 驱动的文本转语音工具

**1.2.2项目背景**

在和朋友进行实时聊天的时候，因为室友睡觉而无法说话，只能使用文字和他们交流。然后就想到如果可以把文字转换成语音并输入到实时聊天频道中，那样就可以进行比较实时的聊天（当无法说话时）。这个软件也可以用作线上会议当中使用，可以用于分析，总结并输出咨询给参会成员，可以起到一个开头的作用，也作为线上桌游例如狼人杀、剧本杀等等游戏的旁白。

**1.2.3开发者**

张宏毅（项目经理、软件开发、需求分析、功能构思），蔡耀明(需求分析、文档编写)

**1.2.4用户**

需要快速将文本思路转换为语音的个人用户，包括但不限于线上社交爱好者、线上会议参与者、桌游玩家等。

**1.3 定义**

TTS：Text - to - Speech，即文本转语音技术，可将文字信息转化为声音信号。

DeepSeek API：一种具备强大文本生成能力的人工智能服务接口，能够根据输入的主题生成相关文本内容。

Baidu TTS API：百度提供的文本转语音云服务，可将输入的文本合成为高质量的语音。

GUI：图形用户界面，为用户提供直观的操作界面。

**2. 任务概述**

**2.1 目标**

TTSmate V1 旨在打造一个轻量级响应迅速的桌面应用程序，其核心流程是：

1.用户输入一个简短的主题或提示词

2.应用程序通过调用 DeepSeek API 围绕该主题智能生成一段文本（限制在 100 字以内）

3.应用程序将生成的文本发送到 Baidu TTS API 将其合成为高质量的 MP3 音频

4.应用程序将接收到的音频数据直接在用户的默认音频设备上播放

该工具解决了从创意到语音输出的整个流程，为用户提供了一个无缝的 "创意 - 文本 - 语音" 转换体验，有效解决了用户在特定场景下的交流和信息输出需求。

**2.2 用户特点**

目标用户：熟悉基本电脑操作，对文本转语音有实际需求的个人用户，涵盖线上社交、会议、游戏等多个领域。

技术水平：初级到中级，能够按照详细说明完成 config.toml 文件中 API 密钥的配置。

使用场景：

线上社交场景：在不方便语音交流时，快速将文字消息转换为语音发送给聊天对象。

线上会议场景：辅助会议组织者进行开场发言，或对会议要点进行语音总结和分享。

线上桌游场景：作为狼人杀、剧本杀等游戏的专业旁白，增强游戏的沉浸感。

内容创作场景：快速验证文本的语音效果，为视频、音频等媒体内容生成旁白草稿。

**3. 功能与性能需求**

**3.1 核心功能**

**3.1.1 灵活的文本源**

提供一个复选框，用户可通过勾选与否来决定是否启用 “使用 DeepSeek 生成文案” 功能。

当启用该功能时，应用程序会将用户在输入框内输入的内容视作 “主题”，随后调用 DeepSeek API 围绕此主题生成用于语音合成的文本。

若禁用该功能，应用程序会把用户输入的内容直接认定为 “最终文本”，并将其用于后续的语音合成流程。

**3.1.2可配置的 AI 行为**

提供一个下拉菜单，菜单选项来自 config.toml 文件中预先定义的多个提示词模板，用户可从中选择一个合适的模板，以指导 DeepSeek API 生成符合特定需求的文本。

下拉菜单中还包含 “自定义模板” 选项，用户若有特殊需求，可在图形用户界面（GUI）中临时编写自己的系统提示词，并将其应用于此次的文本生成过程。

**3.1.3云端语音合成**

集成百度 TTS 云服务 API，保证语音合成的高质量和稳定性。

支持对生成的文本进行语音合成，输出标准的 MP3 格式音频文件。

提供丰富的 UI 控件，允许用户灵活调整语音合成的参数，具体如下：

发音人 (Person)：提供多种预设的男女声以及精品声线供用户选择，满足不同场景和个人喜好的需求。

语速 (Speed)：用户可根据实际需求自由调节语音播放的速度。

音调 (Pitch)：支持用户调整语音的音调，使语音更加生动自然。

音量 (Volume)：用户可以方便地调节语音的音量大小。

**3.1.4音效板 (Soundboard)**

根据 config.toml 文件中的配置信息，在用户界面（UI）上动态生成一组按钮。

每个按钮对应一个预设的音效文件（支持 MP3、WAV 等常见格式），用户点击按钮即可立即播放相应的音效。

**3.1.5双通道音频播放**

内置基于 rodio 库开发的音频播放器。

该播放器使用两个独立的播放通道（Sink）：一个通道专门用于播放通过百度 TTS 合成的语音；另一个通道专门用于播放音效板中的预设音效。

通过这种设计，音效能够在 TTS 语音播放的同时进行播放，实现简单的音频叠加效果，为用户带来更丰富的音频体验。

**3.1.6状态显示**

在应用程序的界面上设置醒目的状态显示区域，为用户提供清晰、实时的状态反馈，例如：“就绪”“正在生成文本...”“正在合成语音...”“正在播放...”“错误：...” 等。

**3.1.7配置管理**

通过本地的 config.toml 文件集中管理应用程序的所有可配置项，具体包括：

1.访问 DeepSeek 和百度服务所需的 API 密钥。

2.应用程序的默认参数，如语音合成的默认语速、音量、音调等。

3.AI 提示词模板列表，为用户提供多样化的文本生成引导。

4.音效板中每个按钮对应的音效名称和文件路径，方便用户对音效进行管理和配置。

**3.2 用户界面 (UI)**

1.使用 egui 框架构建用户界面，确保应用程序在不同操作系统上具有良好的兼容性和轻量化的特点。

2.界面布局遵循简洁易用的设计原则，包含文本输入区、参数滑块 / 下拉菜单、执行按钮和状态显示区，方便用户进行操作和查看系统状态。

3.为了确保中文显示的美观和清晰，程序会自动尝试加载 Windows 系统自带的中文字体（如微软雅黑、等线等），并在其他操作系统上提供备选字体方案。

**3.3 性能要求**

响应时间：从用户点击生成按钮到系统开始播放语音，整个过程应在数秒内完成，具体时间会受到网络延迟的影响。系统应尽量优化算法和网络请求，减少不必要的等待时间。

资源占用：作为一款桌面工具，应用程序应严格控制内存和 CPU 使用率，确保在运行过程中不会对其他程序的正常运行造成明显影响。

**4. 运行环境与依赖**

**4.1 硬件与系统**

操作系统：已在 Windows 系统上进行了全面测试，由于采用了 Rust 和 eframe 技术，理论上可扩展至 macOS 和 Linux 系统，但需要进一步进行兼容性测试。

硬件要求：基础的 PC 硬件配置即可满足系统运行需求，但需要确保系统具备可用的音频输出设备。

网络：必须连接到互联网，以便应用程序能够正常访问 DeepSeek 和 Baidu 的云服务。

**4.2 外部依赖**

DeepSeek API Key：用户必须自行申请并在 config.toml 文件中正确配置该密钥，以确保能够正常使用文本生成功能。

Baidu API Key & Secret Key：用户需要自行申请百度的 API 密钥和 Secret Key，并在配置文件中进行准确配置，以实现语音合成功能。

**4.3 接口**

外部接口：

https://api.deepseek.com/chat/completions：用于调用 DeepSeek API 进行文本生成。

https://aip.baidubce.com/oauth/2.0/token：用于获取百度 API 的访问令牌。

https://tsn.baidu.com/text2audio：用于调用百度 TTS API 进行语音合成。

内部接口：

使用 cpal 库与操作系统的音频后端进行交互，确保音频数据能够正常输出到音频设备。

使用 rodio 库对音频数据进行解码处理，保证音频播放的流畅性。

**5. 尚需解决的问题**

跨平台兼容性：虽然理论上支持 macOS 和 Linux 系统，但尚未进行完整的测试，特别是在字体加载和音频设备的兼容性方面需要进一步优化。

错误处理：目前的错误处理机制主要是在状态栏显示简单的文本信息，后续需要进一步完善，提供更详细的错误日志和用户指引，方便用户快速定位和解决问题。

音频设备选择：当前系统固定使用系统默认的音频输出设备，未来计划增加让用户手动选择音频输出设备的功能，以满足不同用户的个性化需求。