**语音调节报告**

---研究中心

# 1 项目目的

一般来说，语音合成（Text-to-speech）数据库制作需要耗费3~6个月的时间。期间，发音人的录音状态需要保持一致，即音色、能量、语速皆不能有大的差异。而这对于发音人来说，是相当困难的。

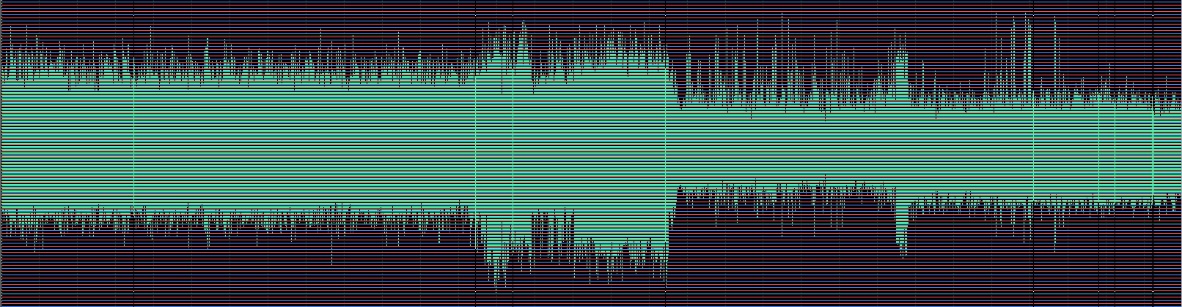
本算法，旨在解决TTS数据库能量不一致问题。首先分析标准语音，获知相关能量描述，根据能量调整准则，对所有的语音样本进行时域幅值调整，以最大限度地保证数据库整体能量的一致性。

通过反复沟通，语音合成部王红潞要求的调节后语音标准如下：

1. 在不发生截幅的条件下，调节后语音样本时域波形的宽度（也即信号时域中的最大值和最小值的距离）与语音模板库中语音文件的平均宽度近似相等；
2. 将样本文件调节后最大值和最小值不能都相等（合成后在波形极值处不能体现出明显的直线）。

# 2 录音结果分析

如前所述，发音人长期录音会得到音色、能量、语速不同的语音样本。这里将四种特殊的语音样本特性做出分析，如图1所示为录音人录的约4000条语音的时域波形。



D

C

B

A

图1 录音样本集合成波形

图1中给出四类语音表现形式，下面分析：

⑴ A类语音，能量集中，时域波形平滑，方差较小；

⑵ B类语音，能量不集中，整体偏高，方差较大；

⑶ C类语音，发生零漂，方差最大；

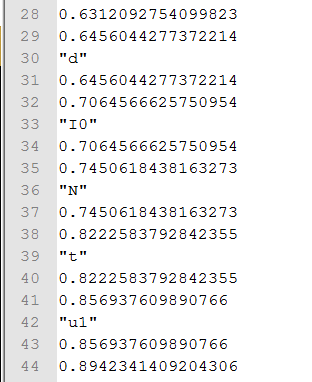
⑷ D类语音，能量不集中，整体偏低，个别采样点值较大，方差较大；

其中A、B、D类语音都未发生零漂现象。

# 3 算法思想

## 3.1 元音平均能量计算

每个语音都有其对应的切音文件interval，我们可以通过读取interval文件中每个元音的时间段信息，来找到语音中的采样点。



**图3-1 切音文件示例**

例如：“I0”对应时间段为0.7064~0.74506秒。

采样点个数

N个元音

**图3-2元音对应采样点**

如图3-2所示，语音中包含n个元音，我们记第i个元音中包含的采样点数为，第i个元音段中个采样点的值分别为 ，其中j=1~ ；那么我们可以得到此语音的平均能量:

## 3.2 幅度调整：

首先，我们选取20句左右的模板语音（能量分布比较均匀的标准语音），根据3.1节中平均能量计算方法，来计算所有模板句子的平均能量值M，此值作为衡量其他所有语音的标准能量值。

对于要调整的语音，我们首先计算它调整之前的平均能量值W：

然后我们可以得到它的调整因子：

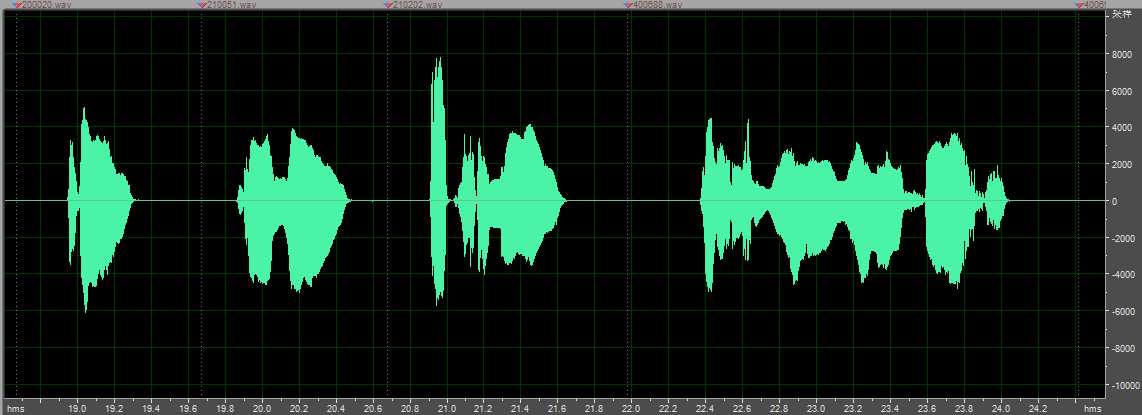
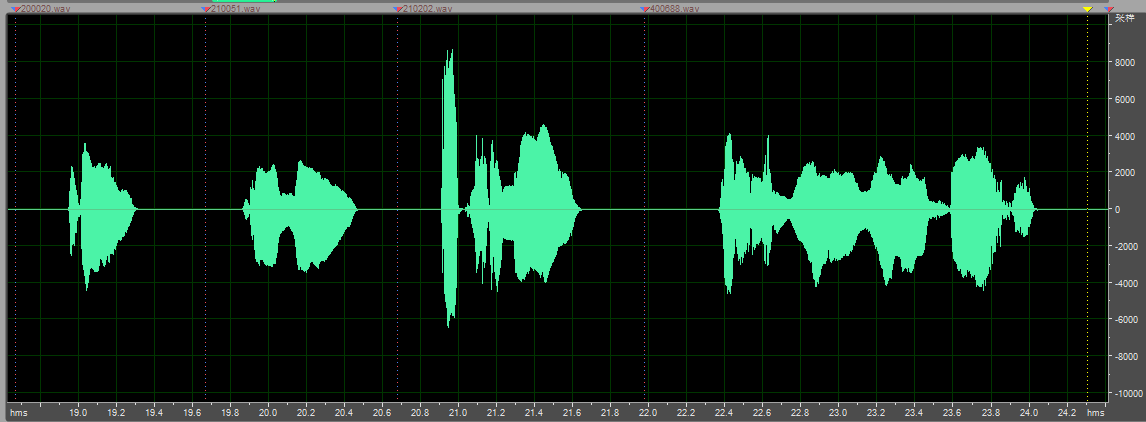
待调整语音的每个采样点值为，我们将这个语音的每个采样点调整为，这样调整之后，此语音的平均能量值为:

也就是说：调整之后的语音平均能量值与标准模板语音的平均能量值相等。

# 实验结果

我们算法能够保证调整后所有语音的元音平均能量值与标准模板句相等，也就是最终语音库所有语音的平均能量值都保持在同一个水平，语音听起来不会出现声音一会儿大、一会儿小的情况。

同时，我们考虑到有些语音调整之后可能出现部分截幅，所以算法设计时保证不出现截幅的情况下，使得能量值尽量接近。



# 5 代码说明

## 5.1 软件开发环境

开发环境：Windows平台下的VS2010，编程语言为C#

## 5.2 GUI与使用说明

开发的语音能量调节工具GUI图如下图所示：

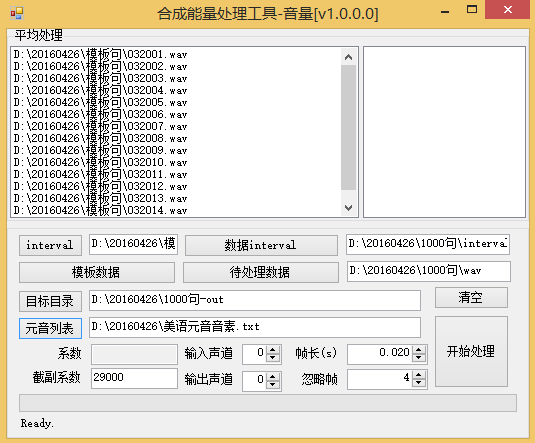


图5 语音能量调节工具界面图

GUI说明：

模板数据button --->选择模板库文件（wav）

Interval ---> 模板数据对应的interval文件所在的目录

待处理数据button --->选择被调节的所有语音文件所在的跟目录（里面的目录可以迭代）

数据interval ---> 带处理的数据对应的interval文件所在的根目录（里面的目录可以迭代）

目标目录button --->选择语音文件调节存放的目录（最终会在这个目录下生成与输入目录相同目录结构的子目录）

元音列表 ---> 选取元音列表集合

清空button --->清空选择的模板库文件和语音文件

开始处理button --->运行、开始处理

系数text --->显示该语音的调节系数

截幅系数text --->显示发生截幅后，处理得到的调节系数

输入输出声道button --->暂无需调节

帧长忽略帧button --->暂无需调节

Ready为显示处理的进度条

## 5.2 代码主要函数

C#程序编写的函数、无DLL，算法主要函数如下：

函数1： double amp\_temp = ExtractAmpFeats\_int(data, sampleRate, list\_item);

输入：data-----------------样本时域信号

SampleRate--------采样率

List\_item-------从interval文件中读取元音时间段

输出：样本句的平均能量值。

函数2：int overflowCnt = AdjustAmpX(ref data, sampleRate, targetAmp, list\_item);

输入：data-----------------样本时域信号（输出：调节好的样本时域信号）；

SampleRate--------采样率

targetAmp ------模板句的平均能量值