



(12) 发明专利

(10) 授权公告号 CN 102522084 B

(45) 授权公告日 2013. 09. 18

(21) 申请号 201110436096. 2  
 (22) 申请日 2011. 12. 22  
 (73) 专利权人 广东威创视讯科技股份有限公司  
 地址 510663 广东省广州市广州高新技术产  
 业开发区彩频路 6 号  
 (72) 发明人 刘观发  
 (74) 专利代理机构 广州华进联合专利商标代理  
 有限公司 44224  
 代理人 王茹 曾旻辉

CN 1655235 A, 2005. 08. 17,  
 CN 1817025 A, 2006. 08. 09,  
 CN 1279462 A, 2001. 01. 10,  
 CN 1842787 A, 2006. 10. 04,  
 CN 1327573 A, 2001. 12. 19,  
 CN 100486284 C, 2009. 05. 06,  
 US 6339754 B1, 2002. 01. 15, 全文.  
 US 6385586 B1, 2002. 05. 07, 全文.

审查员 刘红梅

(51) Int. Cl.

G10L 15/26(2006. 01)

G10L 15/08(2006. 01)

(56) 对比文件

CN 1497932 A, 2004. 05. 19,  
 CN 101421728 A, 2009. 04. 29,

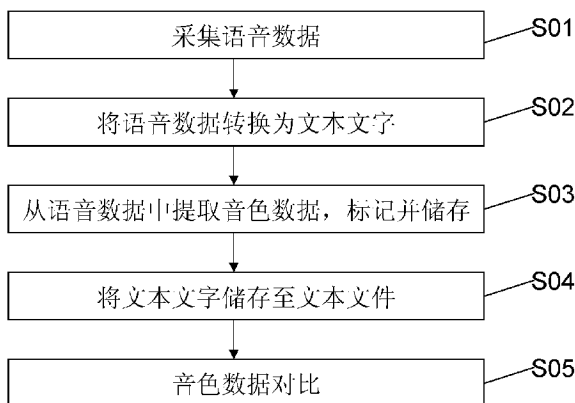
权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种将语音数据转换为文本文件的方法和系统

(57) 摘要

本发明涉及语音识别技术领域,特别涉及一种将语音数据转换为文本文件的方法和系统。所述方法包括:开始采集语音数据,将语音数据转换为文本文字;提取出第一份音色数据,标记所述音色数据并储存;将文本文字储存为对应所述已标记的音色数据的文本文件;从预设的信号采集时间间隔内采集的语音数据中提取出音色数据,并将所述音色数据与已标记的所有音色数据逐一进行对比;如果音色数据相同,则将转换成的文本文字储存至与所述已标记的音色数据对应的文本文件;如果音色数据不同,则标记所述音色数据并储存,将转换成的文本文字储存为对应所述音色数据的另一文本文件。本发明能区分不同发言人的发言内容,简化了录音后的人工整理工作,省时省力。



1. 一种将语音数据转换为文本文件的方法,其特征在于,包括如下步骤:

开始采集语音数据,将语音数据转换为文本文字;

从第一个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记所述音色数据并储存;

将第一个预设的信号采集时间间隔内采集的语音数据转换的文本文字储存为对应所述已标记的音色数据的文本文件;

从第二个所述信号采集时间间隔开始,从每一所述信号采集时间间隔内采集的语音数据中提取出音色数据,并将所述音色数据与已标记的所有音色数据逐一进行对比;

如果音色数据相同,则将所述语音数据转换成的文本文字存储至与所述已标记的音色数据对应的文本文件;如果音色数据不同,则标记所述音色数据并储存,将转换成的文本文字储存为对应所述音色数据的另一文本文件。

2. 根据权利要求1所述一种将语音数据转换为文本文件的方法,其特征在于,所述音色数据为声纹模型。

3. 根据权利要求1或2所述一种将语音数据转换为文本文件的方法,其特征在于,所述音色数据对比为声纹模型匹配。

4. 一种将语音数据转换为文本文件的系统,其特征在于,包括:

采集单元,用于采集语音数据;

转换单元,用于将所述采集单元采集的语音数据转换为文本文字;

提取单元,用于从所述采集单元在各个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记在第一个信号采集时间间隔内采集的语音数据中提取出的音色数据并储存;

存储单元,用于将所述转换单元转换的文本文字储存为对应所述提取单元中已标记的音色数据的文本文件;

对比单元,用于从第二个所述信号采集时间间隔开始,将所述提取单元提取的音色数据与所述提取单元中已标记的所有音色数据逐一进行对比;;

所述提取单元在上述对比的结果为音色数据不同时,标记对应的所述音色数据并储存;

所述存储单元在音色数据相同时,将所述转换单元转换的文本文字存储至与所述提取单元中已标记的音色数据对应的文本文件;在音色数据不同时,将转换成的文本文字储存为对应所述音色数据的另一文本文件。

5. 根据权利要求4所述的一种将语音数据转换为文本文件的系统,其特征在于,所述音色数据为声纹模型。

6. 根据权利要求4或5所述的一种将语音数据转换为文本文件的系统,其特征在于,所述对比单元中的音色数据对比为声纹模型匹配。

## 一种将语音数据转换为文本文件的方法和系统

### 技术领域

[0001] 本发明涉及语音识别技术领域,特别涉及一种将语音数据转换为文本文件的方法和系统。

### 背景技术

[0002] 在演讲、直播、新闻采访或者会议现场,都需要专职人员录制发言,再将其发言内容整理成文字。目前语音识别技术在当今科技中比较成熟,目前已有多种录音记录方法和装置,能在录音现场对语音数据实时记录并且智能转换为文本文件。

[0003] 然而在会议、直播等场合一般会有多位发言人,现有技术能将录制的语音数据统一转换为文本文件,但无法提供发言人 / 发言内容的精确对应关系,这样仍需耗费人力物力对发言内容进行整理,使得验证和校对文字记录非常困难。

### 发明内容

[0004] 本发明在于克服现有技术的缺陷和不足,提供一种将语音数据转换为文本文件的方法,用于解决现有技术中发言人 / 发言内容对应关系不精确的问题。

[0005] 本发明的另一目的还在于提供实现上述方法的系统。

[0006] 一种将语音数据转换为文本文件的方法,包括如下步骤:

[0007] 开始采集语音数据,将语音数据转换为文本文字;

[0008] 从第一个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记所述音色数据并储存;

[0009] 将第一个预设的信号采集时间间隔内采集的语音数据转换的文本文字储存为对应所述已标记的音色数据的文本文件;

[0010] 从第二个所述信号采集时间间隔开始,从每一所述信号采集时间间隔内采集的语音数据中提取出音色数据,并将所述音色数据与已标记的所有音色数据逐一进行对比;

[0011] 如果音色数据相同,则将所述语音数据转换成的文本文字存储至与所述已标记的音色数据对应的文本文件;如果音色数据不同,则标记所述音色数据并储存,将转换成的文本文字储存为对应所述音色数据的另一文本文件。

[0012] 优选地,所述音色数据为声纹模型。

[0013] 优选地,所述音色数据对比为声纹模型匹配。

[0014] 一种将语音数据转换为文本文件的系统,包括:

[0015] 采集单元,用于采集语音数据;

[0016] 转换单元,用于将所述采集单元采集的语音数据转换为文本文字;

[0017] 提取单元,用于从所述采集单元在各个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记在第一个信号采集时间间隔内采集的语音数据中提取出的音色数据并储存;

[0018] 存储单元,用于将所述转换单元转换的文本文字储存为对应所述提取单元中已标

记的音色数据的文本文件；

[0019] 对比单元,用于从第二个所述信号采集时间间隔开始,将所述提取单元提取的音色数据与所述提取单元中已标记的所有音色数据逐一进行对比；

[0020] 所述提取单元在上述对比的结果为音色数据不同时,标记对应的所述音色数据并储存；

[0021] 所述存储单元在音色数据相同时,将所述转换单元转换的文本文字存储至与所述提取单元中已标记的音色数据对应的文本文件；在音色数据不同时,将转换成的文本文字储存为对应所述音色数据的另一文本文件。

[0022] 优选地,所述音色数据为声纹模型。

[0023] 优选地,所述对比单元中的音色数据对比为声纹模型匹配。

[0024] 与现有技术相比,本发明的技术结合了语音转换技术和说话人识别技术,从而实现了发言人/发言内容的精确对应关系,本发明能有效区分每一位发言人的发言内容,简化了录音后的人工整理工作,方便快捷、省时省力。

#### 附图说明

[0025] 图 1 是本发明将语音数据转换为文本文件的方法在一实施例中的流程图。

[0026] 图 2 是本发明将语音数据转换为文本文件的系统在一实施例中的结构框图。

#### 具体实施方式

[0027] 下面结合实施例对本发明作进一步的说明,但并不作为对本发明限制的依据。

[0028] 如图 1 所示,是本发明实施例将语音数据转换为文本文件方法的流程图,包括如下步骤：

[0029] S01、采集语音数据；

[0030] S02、将语音数据转换为文本文字；

[0031] S03、从第一个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记所述音色数据并储存；

[0032] 该信号采集时间间隔可以根据需要来设置,在一较佳实施例中,该信号采集时间间隔可设置为 60s,则采集到的语音数据时长为 60s,从该语音数据中提取出音色数据；将该音色数据标记为音色数据 0001 后储存；

[0033] 具体地,所述音色数据为声纹模型；所述声纹模型包含一组声纹特征,该声纹特征主要有：谱包络参数语音特征,基音轮廓、共振峰频率带宽特征,线性预测系统,倒谱系数等；

[0034] S04、将第一个预设的信号采集时间间隔内采集的语音数据转换的文本文字储存为对应所述已标记的音色数据的文本文件；

[0035] 在一较佳实施例中,可将转换成的文本文字储存为文本文件 0001.txt,与音色数据 0001 相对应；

[0036] S05、从第二个所述信号采集时间间隔开始,从每一所述信号采集时间间隔内采集的语音数据中提取出音色数据,并将所述音色数据与已标记的所有音色数据逐一进行对比；

[0037] 如果音色数据相同,则将所述语音数据转换成的文本文字储存至与所述已标记的音色数据对应的文本文件;如果音色数据不同,则标记所述音色数据并储存,将转换成的文本文字储存为对应所述音色数据的另一文本文件;

[0038] 在一较佳实施例中,经过下一个信号采集时间间隔 60s,采集到另一份时长 60s 的语音数据,从该语音数据中提取出音色数据,将此音色数据与已标记的音色数据 0001 进行对比;如果音色数据对比相同,将语音数据转换为文本文字储存到文本文件 0001.txt 中;如果音色数据对比不同,将该音色数据标记为音色数据 0002,将该语音数据转换成的文本文字储存到另一份文本文件 0002.txt 中;

[0039] 具体地,所述音色数据对比为声纹模型匹配。这里可采用基于矢量量化聚类方法对声纹模型进行识别。

[0040] 相应地,本发明实施例还提供一种将语音数据转换为文本文件的系统,如图 2 所示,是该系统的一种结构示意图。

[0041] 在该实施例中,所述系统包括:

[0042] 采集单元,用于采集语音数据;

[0043] 转换单元,用于将所述采集单元采集的语音数据转换为文本文字;

[0044] 提取单元,用于从所述采集单元在各个预设的信号采集时间间隔内采集的语音数据中提取出音色数据,标记在第一个信号采集时间间隔内采集的语音数据中提取出的音色数据并储存;该信号采集时间间隔可以根据需要来设置,在一较佳实施例中,该信号采集时间间隔可设置为 60s,则所述采集单元采集到的语音数据时长为 60s,从该语音数据中提取出音色数据;把在第一个信号采集时间间隔内采集的音色数据标记为音色数据 0001 后储存;

[0045] 具体地,所述音色数据为声纹模型;所述声纹模型包含一组声纹特征,该声纹特征主要有:谱包络参数语音特征,基音轮廓、共振峰频率带宽特征,线性预测系统,倒谱系数等;

[0046] 存储单元,用于将所述转换单元转换的文本文字储存为对应所述提取单元中已标记的音色数据的文本文件;在一较佳实施例中,可将所述转换单元转换成的文本文字储存为与提取单元中的音色数据 0001 相对应的文本文件 0001.txt;

[0047] 对比单元,用于从第二个所述信号采集时间间隔开始,将所述提取单元提取的音色数据与所述提取单元中已标记的所有音色数据逐一进行对比;

[0048] 所述提取单元在上述对比的结果为音色数据不同时,标记对应的所述音色数据并储存;

[0049] 所述存储单元在音色数据相同时,将所述转换单元转换的文本文字存储至与所述提取单元中已标记的音色数据对应的文本文件;在音色数据不同时,将转换成的文本文字储存为对应所述音色数据的另一文本文件。

[0050] 在一较佳实施例中,经过下一个信号采集时间间隔 60s,所述采集单元采集到另一份时长 60s 的语音数据,所述提取单元从该语音数据中提取出音色数据;将此音色数据与已标记的音色数据 0001 进行对比;

[0051] 如果音色数据对比相同,所述存储单元将所述转换单元转换的文本文字储存到文本文件 0001.txt 中;

[0052] 如果音色数据对比不同,所述提取单元将该音色数据标记为音色数据 0002 并储存,所述存储单元将转换成的文本文字储存到对应音色数据 0002 的文本文件 0002.txt 中。

[0053] 以上所述的本发明实施方式,并不构成对本发明保护范围的限定。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明的权利要求保护范围之内。

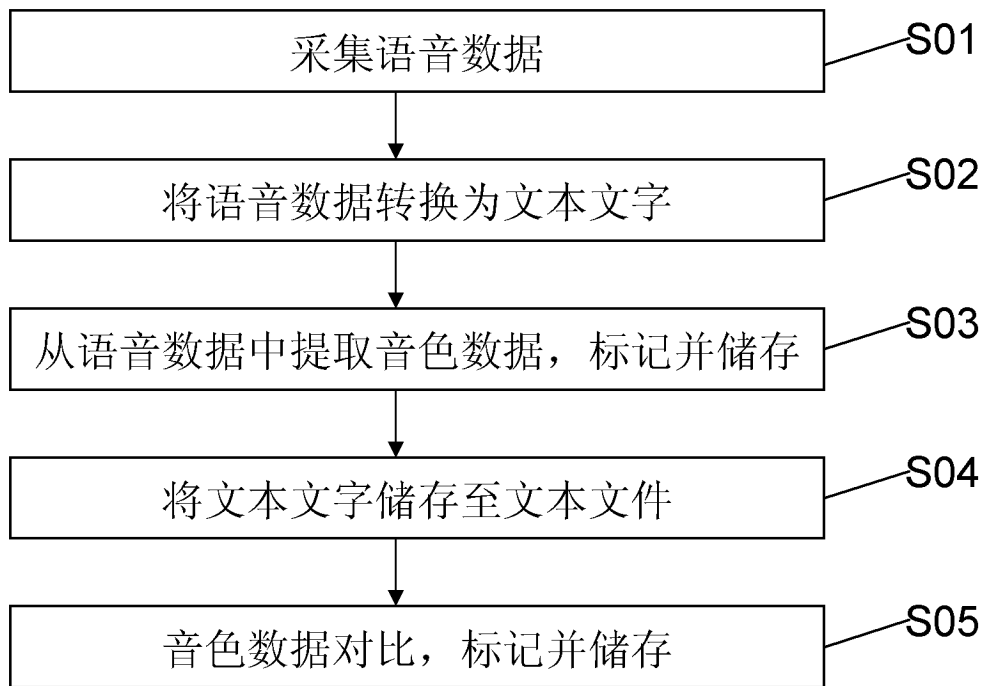


图 1



图 2