



# (12)发明专利申请

(10)申请公布号 CN 111508523 A

(43)申请公布日 2020.08.07

(21)申请号 201910094375.1

(22)申请日 2019.01.30

(71)申请人 沪江教育科技(上海)股份有限公司

地址 201203 上海市浦东新区中国(上海)

自由贸易试验区博云路2号703室

(72)发明人 夏海荣 刘悦 张少飞 于佳玉

(74)专利代理机构 上海硕力知识产权代理事务

所(普通合伙) 31251

代理人 郭桂峰

(51)Int.Cl.

G10L 25/51(2013.01)

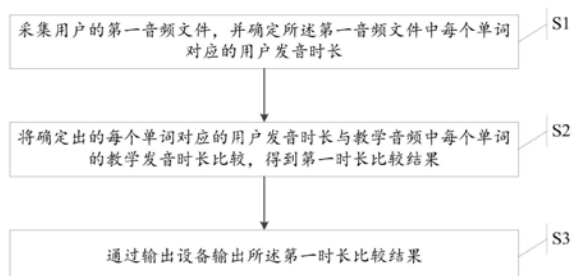
权利要求书2页 说明书7页 附图5页

## (54)发明名称

一种语音训练提示方法及系统

## (57)摘要

本发明公开了一种语音训练提示方法及系统,通过本发明所提供的方法,在用户进行朗读训练时,系统将用户的音频进行多维度分析,得到各个单词对应的时长、音量、基频等信息,然后与教学音频进行比较,并输出时长、音量、基频比较结果,从而用户可以通过图形化展示界面直接查看到差异,进而使得用户可以便捷快速的查看到调整位置以及调整方向,使用户的朗读训练更加高效。



1. 一种语音训练提示方法,其特征在于,所述方法包括:

采集用户的第一音频文件,并确定所述第一音频文件中每个单词对应的用户发音时长;

将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较,得到第一时长比较结果,其中,所述比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度;

通过输出设备输出所述第一时长比较结果。

2. 如权利要求1所述的方法,其特征在于,在通过输出设备输出所述第一时长比较结果之后,所述方法还包括:

采集用户基于所述第一时长比较结果的第二音频文件,并在所述音频文件中确定出每个单词对应的用户发音时长;

判定是否存在用户发音时长与教学发音时长的差值绝对值大于预设阈值的单词;

若存在,则输出第二时长比较结果,其中,所述第二时长比较结果包含每个单词的用户发音时长与教学发音时长的匹配度;

若不存在,则提示用户进入下一阶段训练。

3. 如权利要求2所述的方法,其特征在于,在提示用户进入下一阶段训练之后,所述方法还包括:

采集用户的第三音频文件,并确定所述第三音频文件中每个单词对应的用户发音音量;

将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较,得到第一音量比较结果,其中,所述第一音量比较结果包含了每个单词的用户发音音量与教学发音音量的匹配度;

通过输出设备输出所述第一音量比较结果。

4. 如权利要求3所述的方法,其特征在于,在通过输出设备输出所述第一音量比较结果之后,所述方法还包括:

判定是否存在用户发音音量与教学发音音量的差值绝对值大于预设阈值的单词;

若存在,则输出第二音量比较结果,其中,所述第二音量比较结果包含每个单词的用户发音音量与教学发音音量的匹配度;

若不存在,则提示用户进入下一阶段训练。

5. 如权利要求4所述的方法,其特征在于,在提示用户进入下一阶段训练之后,所述方法还包括:

采集用户的第四音频文件,并确定所述第四音频文件中每个单词对应的用户发音基频;

将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较,得到第一基频比较结果,其中,所述第一基频比较结果包含了每个单词的用户发音基频与教学发音基频的匹配度;

通过输出设备输出所述第一基频比较结果。

6. 如权利要求5所述的方法,其特征在于,在通过输出设备输出所述第一基频比较结果之后,所述方法还包括:

判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词；

若存在，则输出第二基频比较结果，其中，所述第二基频比较结果包含每个单词的用户发音基频与教学发音基频的匹配度；

若不存在，则提示用户训练完成。

7. 一种语音训练提示系统，其特征在于，所述系统包括：

采集模块，用于采集用户的第一音频文件，并确定所述第一音频文件中每个单词对应的用户发音时长；

处理模块，用于将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较，得到第一时长比较结果，其中，所述比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度；

输出模块，用于输出所述第一时长比较结果。

8. 如权利要求7所述的系统，其特征在于，所述采集模块，还用于采集用户的第三音频文件，并确定所述第三音频文件中每个单词对应的用户发音音量；

所述处理模块，还用于将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较，得到第一音量比较结果，其中，所述第一音量比较结果包含了每个单词的用户发音音量与教学发音音量的匹配度；

输出模块，还用于输出所述第一音量比较结果。

9. 如权利要求7所述的系统，其特征在于，所述采集模块，还用于采集用户的第四音频文件，并确定所述第四音频文件中每个单词对应的用户发音基频；

处理模块，还用于将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较，得到第一基频比较结果，其中，所述第一基频比较结果包含了每个单词的用户发音基频与教学发音基频的匹配度；

输出模块，还用于输出所述第一基频比较结果。

10. 如权利要求7所述的系统，其特征在于，所述处理模块，还用于判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词；

输出模块，还用于存在差异单词，则输出第二基频比较结果，其中，所述第二基频比较结果包含每个单词的用户发音基频与教学发音基频的匹配度；若不存在差异单词，则提示用户训练完成。

## 一种语音训练提示方法及系统

### 技术领域

[0001] 本申请涉及音频信息处理技术领域,尤其涉及一种语音训练提示方法及系统。

### 背景技术

[0002] 朗读是语言学习中的一个重要的学习方法:朗读可以提升学习者的发音的准确性和流利度,也能提高学习者对句子乃至篇章的理解能力,从而加强对重读、语调等韵律特征的正确使用。

[0003] 在朗读中,学习者可能会出现下面的错误或不当:单词的发音错误或不适当(包括元音、辅音、音节边界、重读、连读、省读等),词内及词间的不流利(包括不合适的时长和停顿),缺乏音高能力等韵律变化(重读的遗漏或者误用)、缺乏正确地进行语法语义相关的语调变化(例如句末的声调或者降调)、不能正确地理解语句并通过短语(Phrasing)来控制语音输出的节奏。

[0004] 目前,较为传统的方案通过如下两种方式来练习朗读:

[0005] 方式一:有声字典

[0006] 独立电子辞典设备、或者桌面软件、运行于移动设备中的软件(包括微信小程序、网页等)。在用户查询单词后,有声字典提供该词的传统释义,同时提供该词的可以播放的发音音频(真人录音或者计算机合成语言)。学习者通过播放该音频,了解该单词的发音,并可能进行口头模仿。有声字典也可能提供若干与单词相关的例句,这些例句也同样可能带有可以播放的音频。

[0007] 方式二:有声读物

[0008] 可以是独立发行的音频文件(mp3等)、图书的配套光盘、早期的录音磁带,也可以是某个内容平台上的节目形式:例如PodCast、喜马拉雅FM、微信公众号等等。学习者使用有声读物的方式通常为“听”。学习者当然也可以自行进行模仿。

[0009] 虽然,上述的方案可以引导用户进行朗读训练,但是方式一以及方式二不能对用户的朗读水平进行评价,学习者不能得到及时的反馈。

### 发明内容

[0010] 本发明提供了一种语音训练提示方法及系统,用以提升用户的朗读的学习效率。

[0011] 其具体的技术方案如下:

[0012] 一种语音训练提示方法,所述方法包括:

[0013] 采集用户的第一音频文件,并确定所述第一音频文件中每个单词对应的用户发音时长;

[0014] 将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较,得到第一时长比较结果,其中,所述比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度;

[0015] 通过输出设备输出所述第一时长比较结果。

- [0016] 可选的,在通过输出设备输出所述第一时长比较结果之后,所述方法还包括:
- [0017] 采集用户基于所述第一时长比较结果的第二音频文件,并在所述音频文件中确定出每个单词对应的用户发音时长;
- [0018] 判定是否存在用户发音时长与教学发音时长的差值绝对值大于预设阈值的单词;
- [0019] 若存在,则输出第二时长比较结果,其中,所述第二时长比较结果包含每个单词的用户发音时长与教学发音时长的匹配度;
- [0020] 若不存在,则提示用户进入下一阶段训练。
- [0021] 可选的,在提示用户进入下一阶段训练之后,所述方法还包括:
- [0022] 采集用户的第三音频文件,并确定所述第三音频文件中每个单词对应的用户发音音量;
- [0023] 将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较,得到第一音量比较结果,其中,所述第一音量比较结果包含了每个单词的用户发音音量与教学发音音量的匹配度;
- [0024] 通过输出设备输出所述第一音量比较结果。
- [0025] 可选的,在通过输出设备输出所述第一音量比较结果之后,所述方法还包括:
- [0026] 判定是否存在用户发音音量与教学发音音量的差值绝对值大于预设阈值的单词;
- [0027] 若存在,则输出第二音量比较结果,其中,所述第二音量比较结果包含每个单词的用户发音音量与教学发音音量的匹配度;
- [0028] 若不存在,则提示用户进入下一阶段训练。
- [0029] 可选的,在提示用户进入下一阶段训练之后,所述方法还包括:
- [0030] 采集用户的第四音频文件,并确定所述第四音频文件中每个单词对应的用户发音基频;
- [0031] 将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较,得到第一基频比较结果,其中,所述第一基频比较结果包含了每个单词的用户发音基频与教学发音基频的匹配度;
- [0032] 通过输出设备输出所述第一基频比较结果。
- [0033] 可选的,在通过输出设备输出所述第一基频比较结果之后,所述方法还包括:
- [0034] 判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词;
- [0035] 若存在,则输出第二基频比较结果,其中,所述第二基频比较结果包含每个单词的用户发音基频与教学发音基频的匹配度;
- [0036] 若不存在,则提示用户训练完成。
- [0037] 一种语音训练提示系统,所述系统包括:
- [0038] 采集模块,用于采集用户的第一音频文件,并确定所述第一音频文件中每个单词对应的用户发音时长;
- [0039] 处理模块,用于将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较,得到第一时长比较结果,其中,所述比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度;
- [0040] 输出模块,用于输出所述第一时长比较结果。
- [0041] 可选的,所述采集模块,还用于采集用户的第三音频文件,并确定所述第三音频文

件中每个单词对应的用户发音音量；

[0042] 所述处理模块,还用于将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较,得到第一音量比较结果,其中,所述第一音量比较结果包含了每个单词的用户发音音量与教学发音音量的匹配度；

[0043] 输出模块,还用于输出所述第一音量比较结果。

[0044] 可选的,所述采集模块,还用于采集用户的第四音频文件,并确定所述第四音频文件中每个单词对应的用户发音基频；

[0045] 处理模块,还用于将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较,得到第一基频比较结果,其中,所述第一基频比较结果包含了每个单词的用户发音基频与教学发音基频的匹配度；

[0046] 输出模块,还用于输出所述第一基频比较结果。

[0047] 可选的,所述处理模块,还用于判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词；

[0048] 输出模块,还用于存在差异单词,则输出第二基频比较结果,其中,所述第二基频比较结果包含每个单词的用户发音基频与教学发音基频的匹配度;若不存在差异单词,则提示用户训练完成。

[0049] 通过本发明实施例所提供的方法,在用户进行朗读训练时,系统将用户的音频进行多维度分析,得到各个单词对应的时长、音量、基频等信息,然后与教学音频进行比较,并输出时长、音量、基频比较结果,从而用户可以通过图形化展示界面直接查看到差异,进而使得用户可以便捷快速的查看到调整位置以及调整方向,使用户的朗读训练更加高效。

## 附图说明

[0050] 图1为本发明实施例中一种语音训练提示方法的流程图；

[0051] 图2为本发明实施例中基于第一音频文件对应时长分析结果示意图；

[0052] 图3为本发明实施例中第一音频文件对应的时长与教学发音时长之间的比较结果示意图；

[0053] 图4为本发明实施例中时长分析结果的图形化展示界面示意图；

[0054] 图5为本发明实施例中基于第三音频文件对应音量分析结果示意图；

[0055] 图6为本发明实施例中第三音频文件对应的音量与教学发音音量之间的比较结果示意图；

[0056] 图7为本发明实施例中音量分析结果的图形化展示界面示意图；

[0057] 图8为本发明实施例中第四音频文件对应的基频检测结果示意图；

[0058] 图9为本发明实施例中基频分析结果的图形化展示界面示意图；

[0059] 图10为本发明实施例中一种语音训练提示系统的结构示意图。

## 具体实施方式

[0060] 下面通过附图以及具体实施例对本发明技术方案做详细的说明,应当理解,本发明实施例以及实施例中的具体技术特征只是对本发明技术方案的说明,而不是限定,在不冲突的情况下,本发明实施例以及实施例中的具体技术特征可以相互组合。

[0061] 首先,对本发明实施例中所使用的名词进行解释:

[0062] 基频:语音在产生中气流冲击声带所产生的特定的频率;

[0063] 语调:指的是一个句子或者句子的片段的发音的音高轨迹的趋势。一般情况下,陈述句和特殊疑问句使用降调,而一般疑问句使用升调;

[0064] 时长:一个特定发音单元所需要的时间,通常用秒、毫秒来表示,在系统中,也用帧长(Frame length)来表示;

[0065] 如图1所示为本发明实施例中一种语音训练提示方法的流程图,该方法包括:

[0066] S1,采集用户的第一音频文件,并确定所述第一音频文件中每个单词对应的用户发音时长;

[0067] S2,将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较,得到第一时长比较结果;

[0068] 这里需要说明是,该比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度;

[0069] S3,通过输出设备输出所述第一时长比较结果。

[0070] 在语音训练时,首先采集用户的音频信息,从而得到第一音频文件,在得到第一音频文件之后,确定第一音频文件中每个单词对应的用户发音时长。

[0071] 如图2所示为本发明实施例中基于第一音频文件对应时长分析结果,在图2中,该第一音频文件对应有7个单词,每个单词有一个对应的时长。

[0072] 另外,在系统中还保存了教学发音时长,教学发音时长为基于教学音频得到的分析结果。系统在得到第一音频文件之后,将该第一音频文件中的发音时长与教学音频中的每个单词的教学发音时长比较如图3所示为第一音频文件对应的时长与教学发音时长之间的比较结果示意图。图3所示的示意图将通过输出设备进行显示,从而用户可以在显示界面中直接查看到发音时长与教学发音时长之间的差异,进而用户可以根据图形化展示界面进行发音时长的调整。

[0073] 在本发明实施例中,为了方便用户观察时长方面的差异,系统对单词的边界,在教学音频和用户音频中单词的左右边界,通过连线来加强时长的差异化。

[0074] 进一步,在本发明实施例中,系统给显著差异的单词进行特别标识处理,例如使用红色边框,或者是蓝色边框等。另外,在系统对停顿的区域进行特别标识处理,例如使用灰色方块等等。当然,此处只是一个示例,并不是限定。

[0075] 进一步,在通过输出设备输出第一时长比较结果之后,该方法还包括:采集用户基于第一时长比较结果的第二音频文件,并在音频文件中确定出每个单词对应的用户发音时长;判定是否存在用户发音时长与教学发音时长的差值绝对值大于预设阈值的单词;若存在,则输出第二时长比较结果;若不存在,则提示用户进入下一阶段训练。

[0076] 简单来讲,图3所示图形化展示界面中,用户发音的语速偏慢,在“dont”、“like”、“comedy”、“shows”都花费了更长的时间。同时,在“sorry”之后引入了偏长的停顿。

[0077] 在该比较结果的基础上,用户可以根据系统的提示,对发音进行调整,在调整之后,得到如图4所示图形化展示界面,在图4中用户发音时长与教学发音时长差异小于预设阈值,比如差异小于5%,则系统确定用户达标,从而可以进行下一阶段的训练,若是差异大于5%,则系统提示用户继续当前阶段的训练。

[0078] 进一步,在完成发音时长的训练之后,采集用户的第三音频文件,并确定第三音频文件中每个单词对应的用户发音音量;将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较,得到第一音量比较结果,通过输出设备输出第一音量比较结果。

[0079] 在音量训练时,首先采集用户的音频信息,从而得到第三音频文件,在得到第三音频文件之后,确定第三音频文件中每个单词对应的用户发音音量。

[0080] 如图5所示为本发明实施例中基于第三音频文件对应音量分析结果,在图5中,该第三音频文件对应应有7个单词,每个单词有一个对应的音量。

[0081] 另外,在系统中还保存了教学发音音量,教学发音音量为基于教学音频得到的分析结果。系统在得到第三音频文件之后,将该第三音频文件中的发音音量与教学音频中的每个单词的教学发音音量比较如图6所示为第三音频文件对应的音量与教学发音音量之间的比较结果示意图。图6所示的示意图将通过输出设备进行显示,从而用户可以在显示界面中直接查看到发音音量与教学发音音量之间的差异,进而用户可以根据图形化展示界面进行发音音量的调整。

[0082] 在本发明实施例中,为了方便用户观察音量方面的差异,系统对单词的边界,在教学音频和用户音频中单词的上下边界,通过连线来加强音量的差异化。

[0083] 进一步,在通过输出设备输出第一音量比较结果之后,该方法还包括:采集用户基于第一音量比较结果的第四音频文件,并在音频文件中确定出每个单词对应的用户发音音量;判定是否存在用户发音音量与教学发音音量的差值绝对值大于预设阈值的单词;若存在,则输出第二音量比较结果;若不存在,则提示用户进入下一阶段训练。

[0084] 简单来讲,图6所示图形化展示界面中,用户在“comedy”单词上发音偏高,在“shows”单词上发音偏低。

[0085] 在该比较结果的基础上,用户可以根据系统的提示,对音量发音错误的单词进行调整,在调整之后,得到如图7所示图形化展示界面,在图7中用户发音音量与教学发音音量的差异小于预设阈值,比如差异小于5%,则系统确定用户达标,从而可以进行下一阶段的训练,若是差异大于5%,则系统提示用户继续当前阶段的训练。

[0086] 进一步,在用户完成音量训练之后,方法还包括:采集用户的第四音频文件,并确定所述第四音频文件中每个单词对应的用户发音基频;将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较,得到第一基频比较结果;通过输出设备输出所述第一基频比较结果。判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词;若存在,则输出第二基频比较结果;若不存在,则提示用户训练完成。

[0087] 具体来讲,如图8所示为用户第四音频文件对应的基频检测结果,在图8中可以查看到用户在朗读时,每个单词对应的基频。将第四音频文件对应的基频与系统中教学音频对应的基频进行比较,比较得到的图形化展示界面如图9所示。

[0088] 在图9中用户可以在图形化展示界面中直接查看到发音基频与教学发音基频之间的差异,进而用户可以根据图形化展示界面进行发音基频的调整。

[0089] 进一步,在本发明实施例中通过如下方法来给出基频调整指示:

[0090] 1、获取单元的总长度L;

[0091] 2、初始化输出列表H;

[0092] 3、设置下标为 $i=1$ ;

[0093] 4、当 $i < 1$ :

[0094] 计算示范音频的当前单元的平均基频值与前一个基频值之间的差值 $\delta_{ti} = f_{ti} - f_{ti-1}$ ;

[0095] 计算示范音频的当前单元的平均基频值与前一个基频值之间的差值 $\delta_{si} = f_{si} - f_{si-1}$ ;

[0096] 如果差值 $\delta_{ti} = f_{ti} - f_{ti-1}$ 与 $\delta_{si} = f_{si} - f_{si-1}$ 的符号相反,也即 $\delta_{si} \times \delta_{ti} < 0$ :加入提示列表。

[0097] 更新 $i = i + 1$ 。

[0098] 4、输出提示列表;

[0099] 另外,在本发明实施例中,系统允许用户查看任何一个多音节词的详细信息,以及在示例和用户之间进行比较,详细信息包括:单词的音节化的描述,音节化描述包括构成词典发音的由音素构成的音节组合、词内重音;按照音节进行展示的基频曲线;每个音节的时长信息。

[0100] 综上所述,通过本发明实施例所提供的方法,在用户进行朗读训练时,系统将用户的音频进行多维度分析,得到各个单词对应的时长、音量、基频等信息,然后与教学音频进行比较,并输出时长、音量、基频比较结果,从而用户可以通过图形化展示界面直接查看到差异,进而使得用户可以便捷快速的查看到调整位置以及调整方向,使用户的朗读训练更加高效。

[0101] 对应本发明实施例所提供的本方法,本发明实施例中还提供了一种语音训练提示系统,如图10所示为本发明实施例中一种语音训练提示系统的结构示意图,该系统包括:

[0102] 采集模块101,用于采集用户的第一音频文件,并确定所述第一音频文件中每个单词对应的用户发音时长;

[0103] 处理模块102,用于将确定出的每个单词对应的用户发音时长与教学音频中每个单词的教学发音时长比较,得到第一时长比较结果,其中,所述比较结果包含了每个单词的用户发音时长与教学发音时长的匹配度;

[0104] 输出模块103,用于输出所述第一时长比较结果。

[0105] 进一步,在本发明实施例中,所述采集模块101,还用于采集用户的第三音频文件,并确定所述第三音频文件中每个单词对应的用户发音音量;

[0106] 所述处理模块102,还用于将确定出的每个单词对应的用户发音音量与教学音频中每个单词的教学发音音量比较,得到第一音量比较结果,其中,所述第一音量比较结果包含了每个单词的用户发音音量与教学发音音量的匹配度;

[0107] 输出模块103,还用于输出所述第一音量比较结果。

[0108] 进一步,在本发明实施例中,所述采集模块101,还用于采集用户的第四音频文件,并确定所述第四音频文件中每个单词对应的用户发音基频;

[0109] 处理模块102,还用于将确定出的每个单词对应的用户发音基频与教学音频中每个单词的教学发音基频比较,得到第一基频比较结果,其中,所述第一基频比较结果包含了每个单词的用户发音基频与教学发音基频的匹配度;

[0110] 输出模块103,还用于输出所述第一基频比较结果。

[0111] 进一步,在本发明实施例中,所述处理模块102,还用于判定是否存在用户发音基频与教学发音基频的差值绝对值大于预设阈值的单词;

[0112] 输出模块103,还用于存在差异单词,则输出第二基频比较结果,其中,所述第二基频比较结果包含每个单词的用户发音基频与教学发音基频的匹配度;若不存在差异单词,则提示用户训练完成。

[0113] 尽管已描述了本申请的优选实施例,但本领域内的普通技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本申请范围的所有变更和修改。

[0114] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

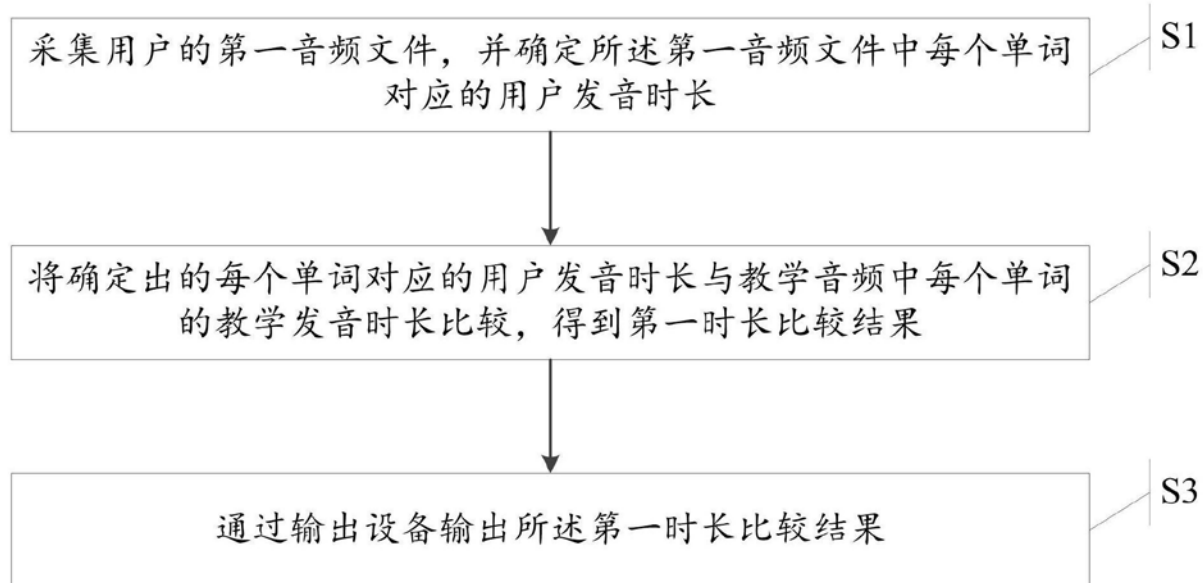


图1

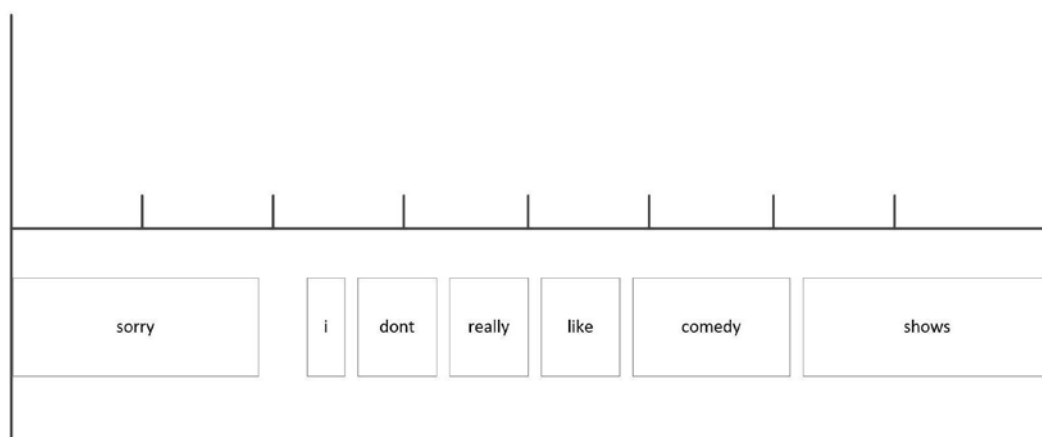


图2

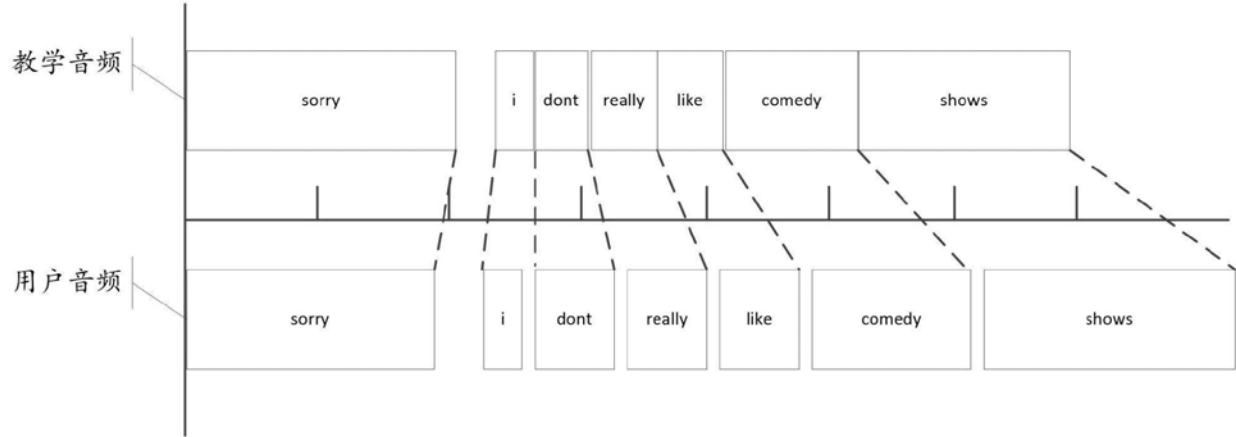


图3

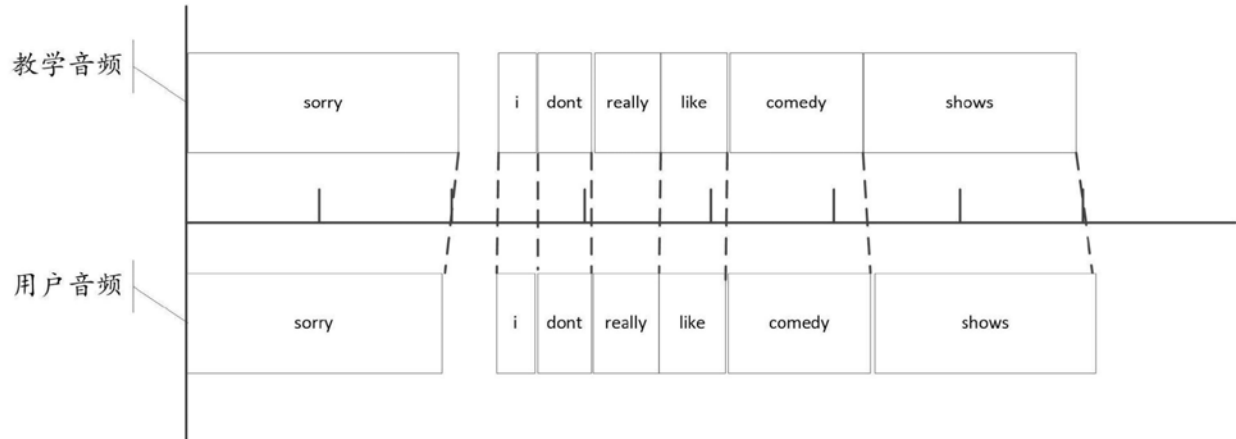


图4

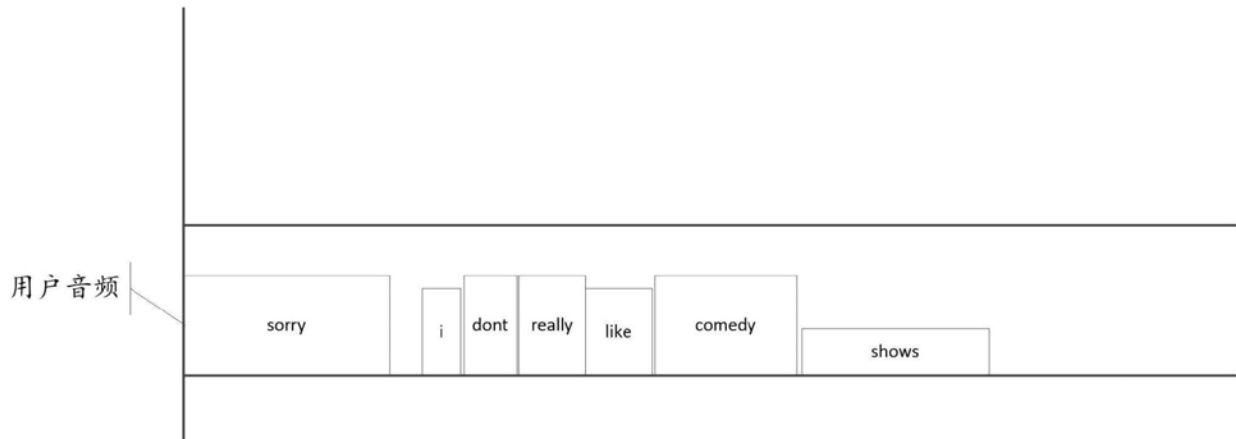


图5

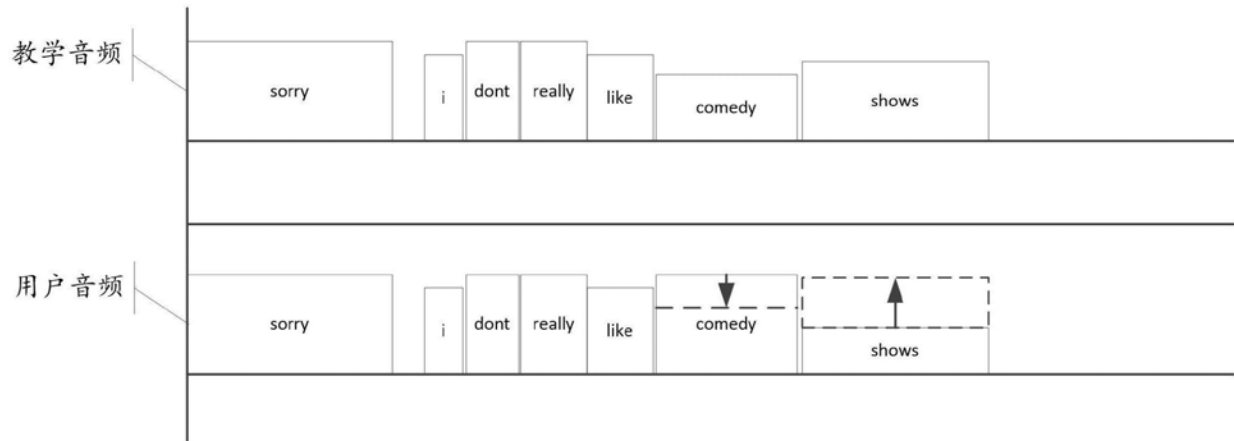


图6

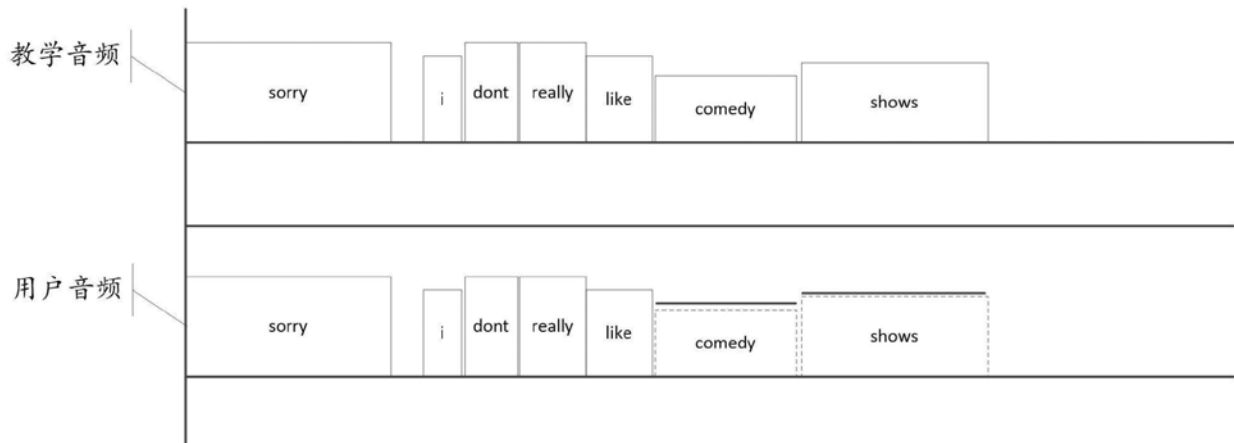


图7

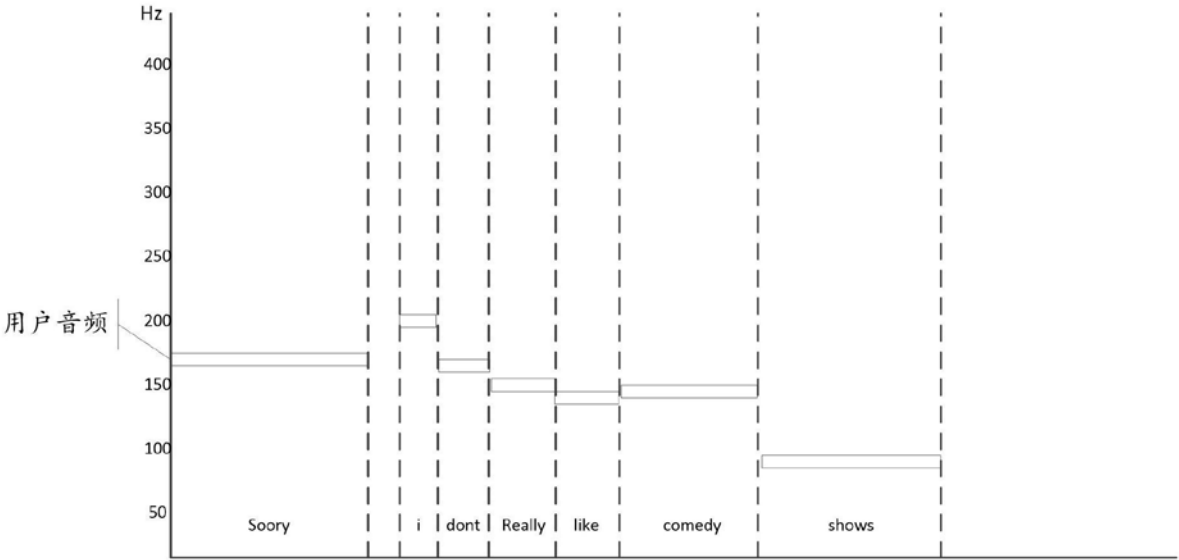


图8

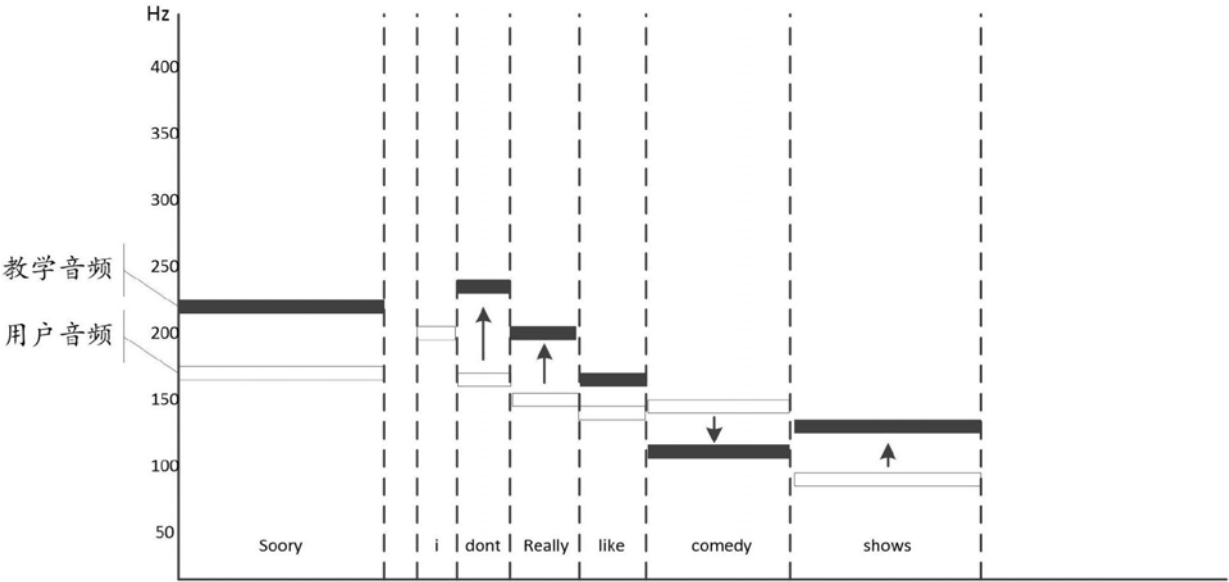


图9

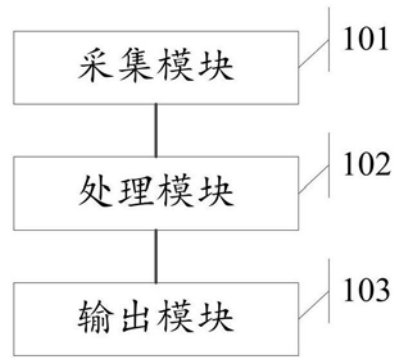


图10