



(12) 发明专利申请

(10) 申请公布号 CN 103095911 A

(43) 申请公布日 2013. 05. 08

(21) 申请号 201210549627. 3

(22) 申请日 2012. 12. 18

(71) 申请人 苏州思必驰信息科技有限公司

地址 215123 江苏省苏州市工业园区仁爱路
258 号独墅湖图书馆 C106

(72) 发明人 雷雄国 王艳龙 王欢良 俞凯
邹平

(51) Int. Cl.

H04M 1/725(2006. 01)

G10L 15/28(2013. 01)

G10L 15/20(2006. 01)

权利要求书2页 说明书4页 附图2页

(54) 发明名称

一种通过语音唤醒寻找手机的方法及系统

(57) 摘要

本发明公开了一种通过语音唤醒技术来寻找手机的方法及系统。所述系统应用于智能手机上,包括:一语音端点检测(VAD)模块,负责实时检测手机麦克风数据,检测是否有用户在说话及其说话的开始时间点;一语音唤醒模块,负责对语音端点检测模块检测到的语音进行实时解码,检测用户是否说了唤醒词;一自定义唤醒词模块,负责根据用户需求,自定义唤醒词并生成相应的资源。本发明通过智能语音唤醒技术检测到用户在寻找手机,并在检测到唤醒词后启动手机铃声和/或震动,从而能够方便、快捷地找到手机。本发明还提供了用户自定义唤醒词的功能,根据用户自身喜好定制个性化的唤醒词,让寻找手机更有乐趣。

1. 一种通过语音唤醒寻找手机的系统,其特征在于,包括:

语音唤醒模块,用于实时检测语音数据中的唤醒词并控制手机播放铃声和 / 或震动提示用户手机具体方位;

自定义唤醒词模块,用于输入唤醒词文本,并向云端自定义唤醒词模块发送请求,完成唤醒词资源包的下载。

云端自定义唤醒词模块,用于接收自定义唤醒词模块发送的请求并进行处理,提供唤醒词资源包的下载。

2. 如权利要求 1 所述的通过语音唤醒寻找手机的系统,其特征在于:

所述语音唤醒模块包括,

实时录音模块,用于调用手机 API 接口获取麦克风数据;

VAD 模块,用于检测从实时录音模块中获取的数据中是否存在语音信号并进行提取;

特征提取模块,用于将语音信号进行长时谱减分析和短时谱特征提取;

唤醒词检测模块,用于将特征提取模块提取得到的声学特征发送给解码器进行维特比解码,检测是否有唤醒词出现;

反馈控制模块,用于根据预先设定调用手机响应接口,控制铃声和 / 或手机震动。

3. 如权利要求 1 所述的通过语音唤醒寻找手机的系统,其特征在于:

所述自定义唤醒词模块,支持一个唤醒词和 / 或多个唤醒词。

4. 如权利要求 1 所述的通过语音唤醒寻找手机的系统,其特征在于:

所述云端自定义唤醒词模块包括,

唤醒词文本接收模块,用于接收自定义唤醒词模块发送的唤醒词文本请求;

语音库,用于存储常用音素和音元字节;

噪声库,用于存储各种实际环境下的噪声数据;

模型训练模块,用于采用基于统计的隐马尔科夫模型进行音素建模和 VAD 建模,采用上下文相关的建模方法对状态数进行聚类,得到上下文相关的三元音素模型及 VAD 模型;

模型裁剪模块,用于通过分析输入文本的上下文关系,将模型训练模块建立的音素模型进行裁剪;

解码网络扩展模块,用于采用基于加权有限状态转换器的方法,结合模型训练模块建立的音素模型,将唤醒词文本转换为语音识别解码网络;

资源包下载模块,用于提供唤醒词资源包的下载。

5. 如权利要求 4 所述的通过语音唤醒识别手机的系统,其特征在于:

所述解码网络扩展模块既可以部署在云端,也可以部署在本地。

6. 如权利要求 1-5 之一所述的通过语音唤醒寻找手机的系统,其特征在于:

通过远距离语音信号处理和远距离语音声学模型训练提高语音识别正确率,

其中,所述通过远距离语音信号处理包括:通过长时谱分析算法、谱减法去除混响信号带来的谱激变,然后,在提取出声学特征后,采用减均值、方差规整并进行自回归滑动平均模型算法去除由于环境噪声带来的谱激变;

所述远距离语音声学模型训练包括:在训练数据中针对性的增加远距离录音数据,进行 HMM 状态数、音素模型聚类算法调整。

7. 如权利要求 1-5 之一所述的通过语音唤醒寻找手机的系统,其特征在于:

所述智能手机包括两种工作模式,模式一允许在任意时间检测到唤醒词即可命令反馈控制模块进行下一步动作,模式二要求在句首检测到唤醒词才可命令反馈控制模块进行下一步动作。

8. 一种通过语音唤醒寻找手机的方法,其特征在于,包括:

用户使用手机上的自定义唤醒词模块输入唤醒词文本,并向云端自定义唤醒词模块发送请求,云端自定义唤醒词模块对请求进行处理后提供唤醒词资源包的下载,所述自定义唤醒词模块下载唤醒词资源包;

手机上的语音唤醒模块实时检测语音数据并提取其中的唤醒词,控制手机播放铃声和/或震动提示用户手机具体方位。

9. 如权利要求 8 所述的通过语音唤醒寻找手机的系统,其特征在于:

所述语音唤醒模块实时检测语音数据并提取其中的唤醒词进一步包括,

实时录音模块调用手机 API 接口获取麦克风数据;

VAD 模块检测从实时录音模块中获取的数据中是否存在语音信号并进行提取;

特征提取模块对语音信号进行长时谱减分析和短时谱特征提取;

唤醒词检测模块将提取得到的信号声学特征发送给解码器进行维特比解码,检测是否有唤醒词出现;

如果有检测词出现,反馈控制模块根据预先设定调用手机响应接口,控制铃声和/或手机震动。

10. 如权利要求 8 所述的通过语音唤醒寻找手机的方法,其特征在于:

所述云端自定义唤醒词模块对请求进行处理后提供唤醒词资源包的下载进一步包括,唤醒词文本接收模块接收自定义唤醒词模块发送的唤醒词文本请求;

模型训练模块采用基于统计的隐马尔科夫模型音素建模和 VAD 建模,采用上下文相关的建模方法对状态数进行聚类,得到上下文相关的三元音素模型及 VAD 模型;

模型裁剪模块通过分析输入文本的上下文关系,将模型训练模块建立的音素模型进行裁剪;

解码网络扩展模块采用基于加权有限状态转换器的方法,结合模型训练模块建立的音素模型,将唤醒词文本转换为语音识别解码网络;

资源包下载模块提供唤醒词资源包的下载。

11. 如权利要求 8-10 之一所述的通过语音唤醒寻找手机的方法,其特征在于:

通过远距离语音信号处理和远距离语音声学模型训练提高语音识别正确率,

其中,所述通过远距离语音信号处理包括:通过长时谱分析算法、谱减法去除混响信号带来的谱激变,然后,在提取出声学特征后,采用减均值、方差规整并进行自回归滑动平均模型算法去除由于环境噪声带来的谱激变;

所述远距离语音声学模型训练包括:在训练数据中针对性的增加远距离录音数据,进行 HMM 状态数、音素模型聚类算法调整。

12. 如权利要求 8-10 之一所述的通过语音唤醒寻找手机的方法,其特征在于:

所述方法包括两种工作模式,模式一允许在任意时间检测到唤醒词即可命令反馈控制模块进行下一步动作,模式二要求在句首检测到唤醒词才可命令反馈控制模块进行下一步动作。

一种通过语音唤醒寻找手机的方法及系统

技术领域

[0001] 本发明涉及远距离语音识别领域,由其涉及一种语音唤醒识别手机的方法及系统。

背景技术

[0002] 在日常使用手机的过程中,经常会发生到处找手机找不到的情况。一般情况下,会通过另外一部电话拨打该手机的电话号码的方式来找手机。这种方式寻找手机需要满足一定的前提条件,存在一定的局限性。比如:没有第二部手机发起主动呼叫时,或者用户不记得自己的手机号的情况下,则无法通过上述方式找到手机。

[0003] 已公开的专利文献,如公开号为 CN102136855A 和 CN101132196A 的专利,都涉及到了采用近距离无线通信技术来寻找手机的方法。但这类方法需要额外增加一个与手机独立的硬件设备,而且需要在手机硬件内部增加相应的通讯硬件设备。这种体系结构有一定的局限性:一是必须在手机的硬件设计时考虑增加该功能,实现起来技术复杂、开发测试周期较长;二是增加了手机设计和生产的成本;三是额外的增加了第二个外部设备,用户需要随身携带,使用起来非常不方便。因此,很少在实际的手机中见到有基于这类专利的应用。

发明内容

[0004] 本发明的目的在于提供一种通过语音唤醒技术实现的更高效自然、方便快捷的寻找手机的方法及系统。

[0005] 本发明提供一种通过语音唤醒技术寻找手机的方法,包括:

[0006] 建立一个覆盖全国各方言区口音的语音库和各种实际环境下的噪声数据库。采用中的语音库训练音素模型,并通过状态聚类方法得到上下文相关的三元音素模型;采用语音库及噪声数据库训练 VAD 模型。根据使用者提供的唤醒词文本,通过自适应方法从音素模型中生成定制音素模型。

[0007] 根据使用者提供的唤醒词文本,通过语音识别解码网络扩展方法,生成定制的唤醒词检测所需要的解码网络资源。根据使用者的实际需求,本发明通过在语音识别网络标识多个唤醒词对应文本的方法,以支持使用者定义多个唤醒词,这样使用者将自己常用且熟悉的词定义成唤醒词,通过说不同的唤醒词都可以寻找到手机,避免使用者忘记单个唤醒词带来的不便。

[0008] 采用 VAD 模型,对手机麦克风采集的语音逐帧计算语音和噪声的似然比,并根据似然比判断是否是语音,如果是静音或者环境噪声则舍弃,如果是语音则将语音数据进行实时检测,采用音素模型及解码网络资源进行实时解码,检测语音中是否出现唤醒词。

[0009] 检测出唤醒词后,调用智能手机的相应接口,让手机播放铃声和/或震动,以便使用者可以方便的知道手机所在的位置。当使用者找到手机后,手动停止播放铃声和/或震动。

[0010] 本发明提供两种唤醒模式,唤醒模式一允许使用者在任意时间说出唤醒词来寻找

手机,在该模式工作状态下,只要使用者说出唤醒词即可以实现手机唤醒;唤醒模式二要求唤醒词在句首才能够有效进行寻找手机,在该模式工作状态下,可以避免在随意聊天时无意中说到了唤醒词导致的误唤醒操作。使用者可以动态地设置和切换两种唤醒模式,十分方便。

[0011] 远距离唤醒是本发明的一个重要技术特征,和传统的语音处理技术相比,由于使用者说话时离手机设备的麦克风的距离一般在 0.2 米~10 米范围内,而传统语音处理技术,这个距离一般在 0.2 米以内,因此,在进行语音处理时,远距离语音中不仅受到周围环境噪声的影响,更重要的是语音信号的混响会导致语音唤醒的正确率大幅度下降。针对远距离语音信号的这一特点,本发明采用了针对性的算法研究,以大幅提升远距离情况下语音唤醒的成功率。具体算法主要包括远距离语音信号处理和远距离语音声学模型训练两部分,详细描述如下:

[0012] 远距离语音信号处理算法包括两部分:首先进行前端处理,传统语音信号处理中的采用的短时谱分析无法解决混响带来的问题,本算法通过长时谱分析算法、谱减法去除混响信号带来的谱激变;然后,在提取出声学特征后,采用减均值、方差规整并进行自回归滑动平均模型算法去除由于环境噪声带来的谱激变。

[0013] 远距离语音声学模型训练流程,首先在训练数据中针对性的增加远距离录音数据,使得训练出来的声学模型能够与实际使用环境相匹配。同时,针对远距离进行了 HMM 状态数、音素模型聚类算法调整,进一步提升远距离语音下的性能。

[0014] 本发明提供一种通过语音唤醒技术寻找手机的方法和系统,所述系统包括:

[0015] 语音唤醒模块,用于实时检测语音数据中的唤醒词并控制手机播放铃声和/或震动提示用户手机具体方位;

[0016] 自定义唤醒词模块,用于输入唤醒词文本,并向云端自定义唤醒词模块发送请求,完成唤醒词资源包的下载。

[0017] 云端自定义唤醒词模块,用于接收自定义唤醒词模块发送的请求并进行处理,提供唤醒词资源包的下载。

[0018] 本发明的优点:一是不需要增加额外的硬件,直接将系统安装到手机上便可以使用;二是使用者直接通过说话来寻找手机,提供了一种非常自然、快捷的寻找手机的方法;三是使用者可以自定义个性化的说法来寻找手机,让找手机的过程充满乐趣。

附图说明

[0019] 图 1 是本发明实施例寻找手机的系统结构图

[0020] 图 2 是本发明实施例寻找手机的云端自定义唤醒词的系統结构图

[0021] 图 3 是本发明实施例寻找手机的方法流程图

[0022] 图 4 是本发明实施例寻找手机的自定义唤醒词的方法流程图

具体实施方式

[0023] 下面结合图例,给出通过语音唤醒寻找手机的方法及其系统更详细的技术特征以及一些典型的实施案例。

[0024] 一种通过语音唤醒寻找手机的方法和系统。所述系统由一语音唤醒模块、自定义

唤醒词模块和云端自定义唤醒词系统组成。

[0025] 如图 1 所示,所述系统包括语音唤醒模块 11、自定义唤醒词模块 12、唤醒词资源包 13。在寻找手机时,使用者与手机的距离相对于正常使用语音识别系统而言比较远的,一般情况下在 0.2 米到 10 米的范围内。在远距离范围内,使用者只需要喊出唤醒词,系统检测到语音并分析出语音中包含唤醒词后,即可启动手机铃声和 / 或震动,从而迅速地找到手机。实际系统存在两种唤醒模式:模式一只要使用者说出唤醒词即可以实现手机唤醒;模式二要求唤醒词在句首才能够有效进行寻找手机,这主要是考虑避免在随意聊天时无意中说到了唤醒词导致的误唤醒操作,使用者可以动态地设置和切换两种唤醒模式,十分方便。

[0026] 本实施例所述的语音唤醒模块 11,包括实时录音模块 111、VAD 模块 112、特征提取模块 113、唤醒词检测模块 114 和反馈控制模块 115。其中所述实时录音模块 111 通过调用手机通用 API 接口获取麦克风数据;VAD 模块 112 采用基于能量和模型的方法检测从实时录音模块 111 中获取的数据中是否存在语音信号,并从数据中将语音信号提取出来;特征提取模块 113 负责将语音信号进行长时谱减分析和短时谱特征提取;唤醒词检测模块 114 通过将语音的声学特征送入解码器进行维特比解码,检测是否包含有唤醒词出现;反馈控制模块 115 负责检测到关键词后控制手机向用户进行反馈,即播放铃声和 / 或使手机震动等。

[0027] 本实施例的特征提取模块 113 中,用于训练音素单元 HMM 模型的声学特征逐帧提取,首先,采用长时谱减法去除远距离混响带来的频谱激变影响,其次,每 25ms 数据提取出一帧的预感知线性预测 (PLP, Perceptual Linear Prediction) 特征,帧移为 10ms。并采用减均值、方差规整和自回归滑动平均模型去除环境噪声影响。在本实施例建立噪声数据库,噪声数据库要求覆盖手机实际使用过程中各类实际噪声环境。录音设备覆盖各类常见的智能手机麦克风。

[0028] 在本实施例所述的自定义唤醒词模块 12,用于输入唤醒词文本数据,并向云端自定义唤醒词模块的 HTTP 服务 21 发送处理请求,在云端自定义唤醒词模块完成处理后,进行资源包 13 的下载及存储。本模块支持多个唤醒词文本输入。

[0029] 本实施例所述的唤醒词资源包 13,包含声学模型及解码网络等资源。

[0030] 如图 2 所示,所述云端自定义唤醒词系统包括 HTTP 服务 21、后台服务 22。当用户需要设置个性化找手机的唤醒词时,用户可以在手机上输入唤醒词内容文本,并提交到云端自定义唤醒词系统,即可方便地下载个性化唤醒资源包,同时,该模块支持多个唤醒词的自定义资源生成。

[0031] 本实施例所述的 Http 服务 21,包括用于接收自定义唤醒词模块 12 发送请求的唤醒词文本输入 211 和资源包下载 212。

[0032] 在本实施例所述的后台服务 22,包括语音库 221、模型训练 222、模型裁减 223 和解码网络扩展 224。

[0033] 在本实施例建立语音库 221 中,语音库 221 的录音文本要求覆盖中英文所有的音素和音节单元,常用音节的分布相对均衡。录音人要求覆盖全国各大言区,录音人性别均衡,年龄呈高斯分布。

[0034] 在本实施例的模型训练 222 中,包括音素建模和 VAD 建模,采用了基于统计的隐马尔科夫模型 (HMM, Hidden Markov Model) 进行建模。同时,在音素模型中,进一步采用上下

文相关的建模方法,对状态数进行聚类。

[0035] 在本实施例的模型裁减 223 中,通过分析唤醒词文本输入 211 的上下文关系,将模型训练 222 中建立的通用音素模型进行裁减。

[0036] 在本实施例的解码网络扩展 224 中,自定义唤醒词资源模块采用了基于加权有限状态转换器 (WFST, Weighted Finite State Transducer) 的方法,结合模型训练 222 中建立的音素模型,将用户提供的唤醒词文本转化为语音识别解码网络,该转换功能由部署在云端系统提供,也可以集成在本地系统中实现。

[0037] 如图 3 所示,使用者在寻找手机时,在距离手机 10 米以内的范围内,说出唤醒词,系统经过 VAD 检测出有语音数据后,立即进行实时的唤醒词检测,一旦检测到用户说了唤醒词,系统自动开启手机铃声和 / 或振动,方便使用者确定手机的具体方位。

[0038] 所述云端自定义唤醒词模块对请求进行处理后提供资源包下载的过程如图 4 所示:

[0039] 首先,建立语音库和噪声数据库,提取声学特征,训练音素模型并得到上下文相关的三元音素模型,同时训练 VAD 模型;然后,根据自定义唤醒词模块 12 发送的自定义唤醒词文本,提取出唤醒词对应的发音序列,构造自定义的音素模型、识别网络和发音词典,生成自定义唤醒词资源包供自定义唤醒词模块 12 下载。

[0040] 以上所述,仅为本发明的优选实施例,并不用以限制本发明,凡依本发明权利要求及说明书内容所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

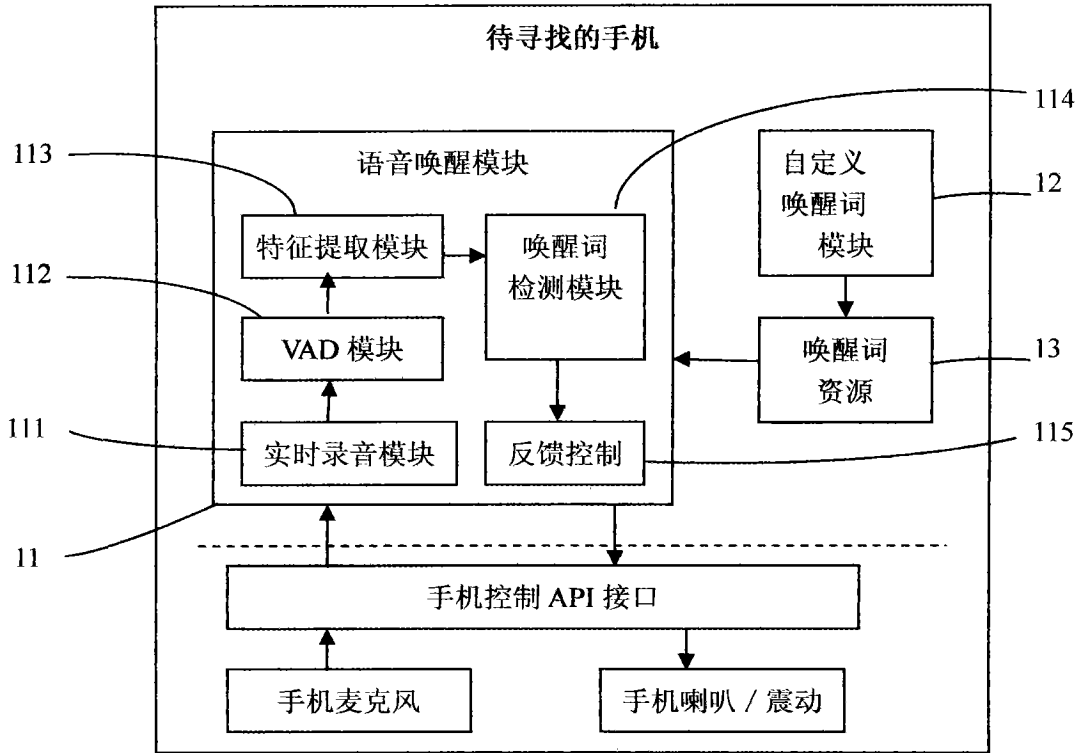


图 1

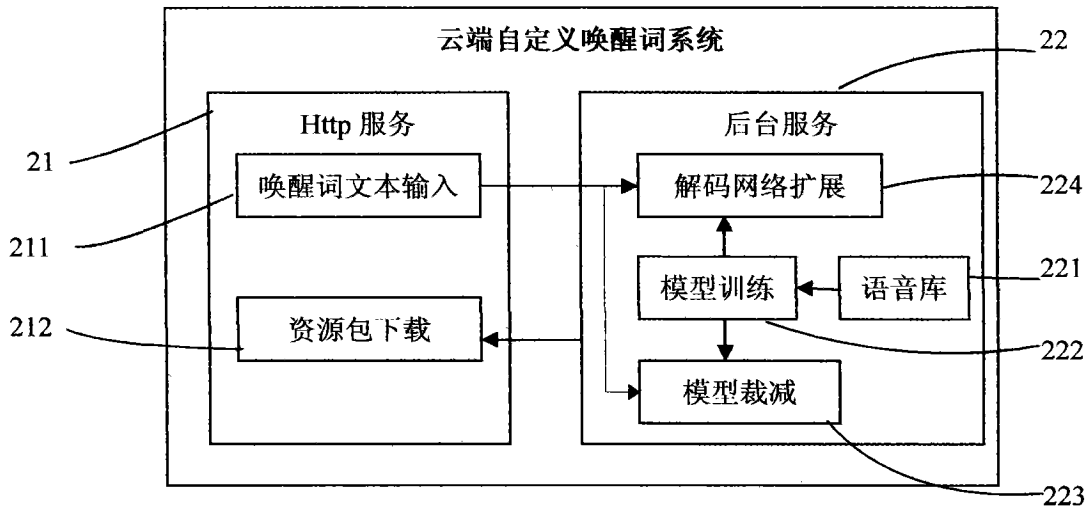


图 2

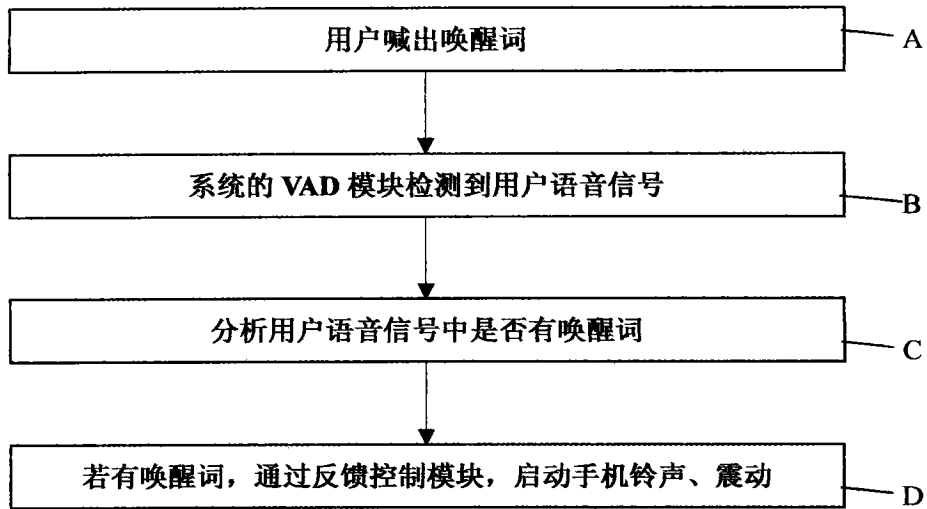


图 3

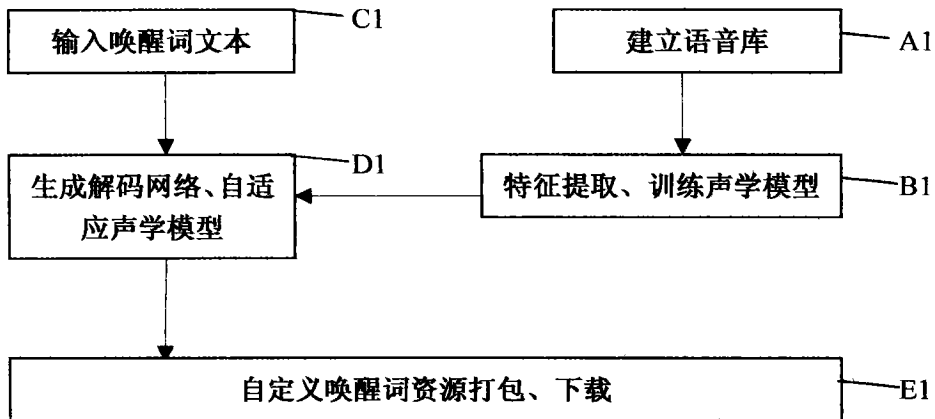


图 4