



(12) 发明专利申请

(10) 申请公布号 CN 101827256 A

(43) 申请公布日 2010.09.08

(21) 申请号 200910079537.0

(22) 申请日 2009.03.06

(71) 申请人 新奥特硅谷视频技术有限责任公司
地址 100080 北京市海淀区西草场一号硅谷
电脑城 7001

(72) 发明人 商文彬 羊美华 羊文广 黄琼

(74) 专利代理机构 北京天悦专利代理事务所
(普通合伙) 11311

代理人 田明 任晓航

(51) Int. Cl.

H04N 7/18(2006.01)

H04L 9/32(2006.01)

H04L 29/06(2006.01)

G10L 15/02(2006.01)

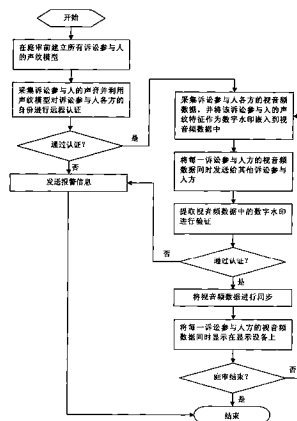
权利要求书 2 页 说明书 6 页 附图 2 页

(54) 发明名称

一种基于声纹认证和数字水印的远程数字化
庭审方法

(57) 摘要

本发明公开了一种基于声纹认证和数字水印
的远程数字化庭审方法,涉及远程身份认证、远程
视音频传输及同步等技术领域。传统庭审方式需
要诉讼参与人各方在法院提供的审判庭中进行,
诉讼成本高、效率低。本发明通过声纹识别对诉
讼参与人各方的身份进行远程认证;采集诉讼参
与人各方视音频数据,将该诉讼参与人声纹特征
作为数字水印嵌入到视音频数据中;将每一诉讼
参与人方的视音频数据同时发送给其他诉讼参与
人方,其他诉讼参与人方提取视音频数据中的数
字水印进行验证,验证通过后将视音频数据进行
同步,最后将每一诉讼参与人方的视音频数据同
时显示在显示设备上。本发明实现了远程的数字
化庭审,诉讼成本低,审理效率高,安全性好。



1. 一种基于声纹认证和数字水印的远程数字化庭审方法,包括以下步骤:

在庭审前建立所有诉讼参与人的声纹模型;

在庭审过程中,采集诉讼参与人的声音并利用声纹模型对诉讼参与人各方的身份进行远程认证,所述诉讼参与人各方处于不同的地点,通过网络相互连接;

采集诉讼参与人各方的视音频数据,并将该诉讼参与人的声纹特征作为数字水印嵌入到该诉讼参与人方的视音频数据中;

将每一诉讼参与人方的视音频数据同时发送给其他诉讼参与人方,其他诉讼参与人方提取视音频数据中的数字水印进行验证,验证通过后将视音频数据进行同步,最后将每一诉讼参与人方的视音频数据同时在显示设备上显示。

2. 如权利要求 1 所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于,所述建立诉讼参与人的声纹模型的方法为:首先通过样本声纹特征建立通用的背景声纹模型,然后提取诉讼参与人的声纹特征,根据该诉讼参与人的声纹特征修改背景声纹模型,将修改后的背景声纹模型作为该诉讼参与人的声纹模型。

3. 如权利要求 2 所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于,所述采集诉讼参与人的声音并利用声纹模型对诉讼参与人各方的身份进行远程认证的过程包括以下步骤:

采集诉讼参与人的语音,提取有效语音段;

从有效语音段中提取声纹特征;

将声纹特征与预先建立的该诉讼参与人的声纹模型进行匹配。

4. 如权利要求 3 所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于:每当诉讼参与人正式发言时,便通过声纹对其身份验证一次,如果未通过验证,则向所有诉讼参与人方发送报警信息。

5. 如权利要求 1 至 4 之一所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于,所述采集诉讼参与人各方的视音频数据的过程包括以下步骤:

对视频数据和音频数据分别进行采集;

将采集的视频数据和音频数据进行缓存;

当缓存的视频数据和音频数据达到预先设定的大小时,对视频数据和音频数据分别进行编码。

6. 如权利要求 5 所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于:将诉讼参与人的声纹特征嵌入到编码前的原始视频和音频码流中;或者在视频和音频数据编码过程中,将声纹特征作为数字水印嵌入到视频和音频数据中;或者将声纹特征作为数字水印嵌入到编码后的视频和音频码流中。

7. 如权利要求 1 至 4 之一所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于:所述方法还包括对解码后的视音频数据进行处理,所述对解码后的视音频数据进行处理包括增强视频清晰度、音频降噪和对视频中人物头像作马赛克处理。

8. 如权利要求 1 至 4 之一所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征在于:将显示设备分成多个显示区,每个显示区用于显示诉讼参与人一方的视音频。

9. 如权利要求 8 所述的一种基于声纹认证和数字水印的远程数字化庭审方法,其特征

在于：当显示设备的某一显示区无显示内容时，则与发言的诉讼参与人方对应的显示区被相应放大；当该诉讼参与人端发言结束后，自动缩小到原始尺寸。

一种基于声纹认证和数字水印的远程数字化庭审方法

技术领域

[0001] 本发明涉及远程身份认证、远程视音频传输及同步等技术领域，具体涉及一种基于声纹认证和数字水印的远程数字化庭审方法，适用于法院对案件进行庭审的过程中。

背景技术

[0002] 根据现行法律的规定，诉讼案件主要分类民事案件、行政案件和刑事案件。相应的，对案件的审理便分为民事案件的审理、行政案件的审理和刑事案件的审理。案件类型的不同以及案件具体情况的不同会使得庭审过程中的诉讼参与者有所不同。例如，民事案件诉讼参与者中必须有审判员、原告、被告，还可能有第三人、证人、鉴定人等；刑事案件的诉讼参与者还必须有公诉人。庭审程序可以大致概括如下几个阶段：对诉讼参与者身份进行确认阶段（该阶段在庭审前进行）；法庭调查阶段（该阶段内原、被告分别进行案件陈述并举证，对对方提供的证据进行质证，如果有证人，则还需要证人出庭作证）；法庭辩论阶段（由原、被告对争议焦点进行辩论）；最后陈述阶段（原、被告分别进行最后陈述）。

[0003] 随着科技的发展，各级法院逐渐开始试用科技手段支持庭审工作。例如，采用多媒体设备来辅助庭审、将庭审过程进行网络直播等。这种应用基本上处于设备的操作使用层次，没有从根本上改变传统的庭审方式，庭审的过程仍然需要诉讼参与者各方在法院提供的审判庭中进行。一方面，诉讼成本较高，因为法院需要提供审判庭，其他诉讼参与者需要交通费用等；另一方面，在某些情况下也会影响案件的审判质量，例如证人因为某种原因而不能出庭作证或者害怕打击报复而不愿出庭作证等。此外，由于法院审判庭资源的有限性、以及当事人因正当理由无法到庭等原因，造成审理期限的拖延，影响了案件的审判效率。

发明内容

[0004] 针对现有技术中存在的缺陷，本发明的目的是提供一种基于声纹认证和数字水印的、诉讼成本低、审判效率高、安全性好的远程数字化庭审方法。

[0005] 为了实现上述目的，本发明采用的技术方案如下：一种基于声纹认证和数字水印的远程数字化庭审方法，包括以下步骤：

[0006] 在庭审前建立所有诉讼参与人的声纹模型；

[0007] 在庭审过程中，采集诉讼参与人的声音并利用声纹模型对诉讼参与者各方的身份进行远程认证，所述诉讼参与者各方处于不同的地点，通过网络相互连接；

[0008] 采集诉讼参与者各方的视音频数据，并将该诉讼参与人的声纹特征作为数字水印嵌入到该诉讼参与者方的视音频数据中；

[0009] 将每一诉讼参与者方的视音频数据同时发送给其他诉讼参与者方，其他诉讼参与者方提取视音频数据中的数字水印进行验证，验证通过后将视音频数据进行同步，最后将每一诉讼参与者方的视音频数据同时在显示设备上显示。

[0010] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，所述建立诉讼

参与人的声纹模型的方法为：首先通过样本声纹特征建立通用的背景声纹模型，然后提取诉讼参与人的声纹特征，根据该诉讼参与人的声纹特征修改背景声纹模型，将修改后的背景声纹模型作为该诉讼参与人的声纹模型。

[0011] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，所述采集诉讼参与人的声音并利用声纹模型对诉讼参与人各方的身份进行远程认证的过程包括以下步骤：

[0012] 采集诉讼参与人的语音，提取有效语音段；

[0013] 从有效语音段中提取声纹特征；

[0014] 将声纹特征与预先建立的该诉讼参与人的声纹模型进行匹配。

[0015] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，每当诉讼参与人正式发言时，便通过声纹对其身份验证一次，如果未通过验证，则向所有诉讼参与人方发送报警信息。

[0016] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，所述采集诉讼参与人各方的视音频数据的过程包括以下步骤：

[0017] 对视频数据和音频数据分别进行采集；

[0018] 将采集的视频数据和音频数据进行缓存；

[0019] 当缓存的视频数据和音频数据达到预先设定的大小时，对视频数据和音频数据分别进行编码。

[0020] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，将诉讼参与人的声纹特征嵌入到编码前的原始视频和音频码流中；或者在视频和音频数据编码过程中，将声纹特征作为数字水印嵌入到视频和音频数据中；或者将声纹特征作为数字水印嵌入到编码后的视频和音频码流中。

[0021] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，该方法还包括对解码后的视音频数据进行处理步骤，所述对解码后的视音频数据进行处理包括增强视频清晰度、音频降噪和对视频中人物头像作马赛克处理。

[0022] 如上所述的一种基于声纹认证和数字水印的远程数字化庭审方法，将显示设备分成多个显示区，每个显示区用于显示诉讼参与人一方的视音频。当某一显示区无显示内容时，则与发言的诉讼参与人方对应的显示区被相应放大；当该诉讼参与人端发言结束后，自动缩小到原始尺寸。

[0023] 采用本发明所述方法，庭审过程中诉讼参与人各方可以处于不同的地理位置，无需统一到法院提供的审判法庭进行诉讼，从而节省了各方的诉讼成本。由于庭审过程不受地域的限制，避免了因交通等问题造成的诉讼延误现象，缩短了法院审理案件的期限，提高了审判效率。此外，通过声纹认证确认诉讼参与人的身份，并将其作为数字水印嵌入到视、音频数据中，不仅能够保证诉讼参与人身份的真实性，无需增加额外的认证设备，而且能够确保视音频数据在网络传输时的安全性。

附图说明

[0024] 图 1 是本发明所述方法流程图；

[0025] 图 2 是声纹认证流程图；

[0026] 图 3 是显示设备分区示意图。

具体实施方式

[0027] 下面结合具体实施方式和附图对本发明进行详细描述。

[0028] 本发明的核心思想是：在法院庭审过程中，诉讼参与人各方可以处在不同的地点，通过网络相互连接，利用视音频采集设备采集各方的视音频，将各方的视音频同时显示在各方的显示设备上。为了确保诉讼参与人身份的真实性和视音频数据在网络传输过程中的安全性，采用声纹认证的方式并将声纹特征作为水印嵌入到视音频数据中，防止视音频数据在网络传输过程中被篡改。

[0029] 声纹是用电声学仪器显示的携带语言信息的声波频谱。人类语言的产生是人体语言中枢与发音器官之间一个复杂的生理物理过程，人在讲话时使用的发生器官（舌、牙齿、喉头、肺、鼻腔）在尺寸和形态方面每个人的差异很大，所以任何两个人的声纹图谱都有差异。但每个人的语音声学特征既有相对稳定性，又有变异性，不是绝对的、一成不变的。尽管如此，由于每个人的发音器官都不尽相同，因此在一般情况下，仍然能够区分不同的人的声音或判断是否是同一人的声音。

[0030] 在法院庭审过程中，采用语音识别诉讼参加人身份无需增加额外的硬件设备，而且能够在诉讼参与人发言时随时收集语音进行验证，无需诉讼参与人的配合，因此本发明采用声纹认证方式来保证诉讼参与人身份的真实性。

[0031] 图 1 出示了本发明所述方法流程，主要包括以下步骤。

[0032] (1) 在庭审前建立所有诉讼参与人的声纹模型。

[0033] 在庭审前，首先需要确认诉讼参与人各方的身份。诉讼参与人根据具体的案件类型和案件特点的不同而不同，包括审判人员、原告、被告、公诉人、第三人、证人等。庭审时诉讼参与人各方可以处于不同的地点，通过网络进行远程身份认证。

[0034] 确定了诉讼参与人后，需要在庭审前建立所有诉讼参与人的声纹模型，存入数据库中，作为庭审时验证诉讼参与人身份的基准模型。

[0035] 声纹识别可以分为文本相关的和文本无关的两种。文本相关的声纹识别需要被识别人按照规定的内容发音，精确地建立声纹模型，而识别时也必须按照规定的内容发音，因此可以达到较好的识别效果，但需要被识别人的配合，如果被识别人的发音与规定的内容不符合，则无法正确识别。而与文本无关的识别则不规定说话人的发音内容，该识别方式建立模型较为困难，但无需被识别人的配合。

[0036] 由于本实施方式，当诉讼参与人正式发言时，便收集该诉讼参与人的语音进行验证，这样能够保证整个庭审过程中诉讼参与人身份的真实性，因此采用与文本无关的识别方式。

[0037] 建立模型时，首先搜集特定信道下足够多的人的样本语音，例如，搜集男女各 100 人、每人不少于 1 分钟的样本语音，用于近似描述所有人的平均声纹特征分布。然后将这些语音提取声纹特征参数后，放在一起进行迭代聚类，得到若干个高斯混合，作为通用的声纹背景模型 (UBM)。

[0038] 提取语音特征参数之前需要对语音进行预处理，因为原始语音中除了含有说话人自身发出的有意义的内容之外，还有停顿静音或背景噪声干扰等无意义部分，所以应先判

断各个语音段的有效性,去除掉无意义的部分,只从有效语音段中提取声纹特征。提取声纹特征可采用如下所述方法:将有效语音采样点分为一组具有一定长度的小窗(即分帧),相邻窗间有局部重叠。各帧内部首先施加海明(Hamming)窗函数,并进行高频提升,然后做快速傅立叶变换(FFT),求出频谱。再对频谱施加按Me1刻度规律排列的、含有若干三角形滤波器的滤波器组,将得到的各个滤波器的输出值进行离散余弦变换(DCT),得到一组系数,为Me1频率倒谱参数(MFCC)矢量,然后把相邻帧的MFCC矢量逐维求差,并将得到的差矢量拼接到原MFCC矢量上,得到的参数作为声纹特征序列。

[0039] 建立了声纹背景模型后,采集诉讼参加人的足够长的有效语音,例如采集1分钟的有效语音,利用上述方法提取出声纹特征,再根据最大后验概率(MAP)的自适应算法,用提取出的声纹特征修正UBM,从而得到一个新的高斯混合模型(每个人的声纹特征可以看作是在高维特征空间的一种概率分布,可以用多个高斯分布来拟合,这类集合称作高斯混合模型),作为该诉讼参与人的声纹模型,存入数据库中,作为庭审时通过语音识别该诉讼参加人的匹配基准。

[0040] 在庭审前,通过上述方法建立所有诉讼参与人的声纹模型。

[0041] 图2出示了声纹认证的方法流程,包括以下步骤:首先采集诉讼参与人的语音,提取有效语音段;然后从有效语音段中提取声纹特征;将提取的声纹特征通过网络发送给存储声纹模型的服务器,最后在服务器上将声纹特征与预先建立的该诉讼参与人的声纹模型进行匹配。

[0042] 本实施方式在庭审过程中,每当诉讼参与人正式发言时,便通过声纹对其身份验证一次。验证的过程在后台进行,如果验证失败,则向诉讼参与人各方发送报警信息。

[0043] (2) 采集诉讼参与人各方的视音频数据,并将该诉讼参与人的声纹特征作为数字水印嵌入到该诉讼参与人所在方的视音频数据中。

[0044] 诉讼参与人各方均设有视频采集设备和音频采集设备,视频和音频采用单独的通道进行采集。首先对采集到的视频数据和音频数据进行缓存,当缓存的数据的大小达到预先设定值时,对视频数据和音频数据进行编码,然后将编码后的视频数据和音频数据分发给诉讼参与人各方。在视频数据和音频数据中均包含时间点信息,作为视音频同步时的参考。

[0045] 需要说明的是:在视频和音频采集时,只有当诉讼参与人一方正式发言时,才需要同时采集发言一方的视频和音频,其他方则只采集视频信息即可,而无需采集音频信息。

[0046] 在上述过程中,将身份验证时提取的声纹特征作为水印,分别嵌入到视音频数据和音频数据中。数字水印可以在原始的视频和音频数据码流中嵌入,也可以在编码阶段嵌入,还可以在编码后嵌入。将水印直接嵌入在原始视频和音频流中不仅会增加视频和音频流的数据比特率,而且编码后水印容易出现丢失现象,也会降低视频和音频的质量。在编码过程中嵌入水印,一般不会增加视频和音频流的数据比特率,但会降低视频和音频的质量。在编码后嵌入水印,一方面计算复杂度低,不会造成视频和音频质量的下降;另一方面由于压缩比特率的限制而限制了嵌入水印的数据量的大小。综合以上原因并结合庭审的特点,本实施方式采用在编码过程中嵌入水印,因为庭审过程中对视频和音频的质量要求不高,但对视频和音频数据的网络传递的效率要求较高。

[0047] (3) 将每一诉讼参与人方的视音频数据同时发送给其他诉讼参与人方,其他诉讼

参与人方提取视音频数据中的数字水印进行验证,验证通过后将视音频数据进行同步,最后将每一诉讼参与人方的视音频数据同时在显示设备上显示。

[0048] 当诉讼参与人各方收到其他方的视音频时,在解码的过程中同时提取数字水印,然后与作为数字水印的原始声纹特征进行匹配。如果匹配吻合,则证明该段视音频在网络传输过程中没有被篡改。

[0049] 视音频数据解码后,可以对视音频数据再进行处理,例如增强视频清晰度、音频降噪和对视频中人物头像作马赛克处理等。在庭审过程中,如果证人不希望公开自己的相貌特征,则可以将其面部进行马赛克处理,这样能够起到对证人的保护作用。

[0050] 接下来对视频数据和音频数据进行同步处理,此处只需对发言一方的视频和音频进行同步,其他方仅采集的是视频数据,因此无需同步的操作。当诉讼参与人方(除本方外,即视频和音频所在方)对同一时间段的视、音频均同步完成后,统一向所有诉讼参与人方发送命令,在所有诉讼参与人方的显示设备上同时播放该段视音频。这样,能够保证所有诉讼参与人方的显示设备上显示的内容一致。

[0051] 在诉讼参与人各方可以设置多台显示设备,用于分别显示诉讼参与人各方的视音频;也可以将一台显示设备进行分区,每区显示诉讼参与人一方的视音频。如图3所示,显示设备分为审判人员显示区、原告显示区、被告显示区、证人显示区、证据显示区、报警信息显示区。当某一区无显示内容时,如证人显示区、报警信息显示区,其他显示区的大小被相应放大。本实施方式中,当某一诉讼参与人发言,并且某一显示区无显示内容时,则与诉讼参与人对应的显示区被相应的放大;当该诉讼参与人发言结束后,则自动缩小到原始尺寸。

[0052] 此外,在庭审的法庭调查阶段,需要原、被告双方向法庭出示证据,如果有证人,还需要证人提供证言。在有证人参加的庭审中,还需要设置证人一方,当证人作证完毕后,证人方便与其他诉讼参与人方切断。当原告或被告出示证据并对该证据所要证明的实事进行陈述时,在各方的显示设备的证据显示区内显示与原告或被告正在陈述相对应的数字化证据。此处所述的数字化证据是根据举证期限内原、被告提交的证据在庭审前制作的。例如,物证可以拍摄成图片,书证、鉴定结论、勘验笔录等可以扫描成电子文件。在证据展示时,可以控制诉讼参与人各方的显示设备同时显示。

[0053] 在证人提供证言后,或者庭审结束后,需要对书记员记录的庭审笔录进行确认。此时,需要签名或盖章的诉讼参与人方应设置数字签名设备或电子公章设备。法院方将庭审过程中记录的电子文件通过发送给需要签名或盖章的诉讼参与人方,然后该诉讼参与人方利用数字签名设备或电子公章设备在电子文件上进行签名或盖章后回传给法院方。为了保证电子文件在网络传输过程中被篡改,同样可以将诉讼参与人的声纹特征作为数字水印嵌入到电子文件中。

[0054] 在庭审过程中,公众可以对可公开审理的案件进行旁听,因此法院的审判庭中一般都提供旁听席,便于公众对案件审理过程进行监督,但是公众在庭审过程中没有发言的权利,也不得违反法庭纪律。在可公开审理的案件中,在将诉讼参与人一方的视音频发送给其他方的同时,发送给用于网络直播的服务器端,服务器端对其处理后,进行网络直播。也可以在庭审结束后,将存储的所有视音频文件整合成一段完整的视音频文件,在互联网上公布,用于公众在线观看和下载。这样法院不但不用提供旁听席,而且也避免了旁听者对案件当事人的影响,以及违反法庭纪律的旁听者对案件审理过程的影响。

[0055] 在庭审过程中,还需要对所有的视音频文件进行备份,以便法院方保留存根,便于以后查阅该案件的审理过程,或法院之间交流、学习使用,还可以将其加工后发布到互联网上,供公众观看和下载。

[0056] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其同等技术的范围之内,则本发明也意图包含这些改动和变型在内。

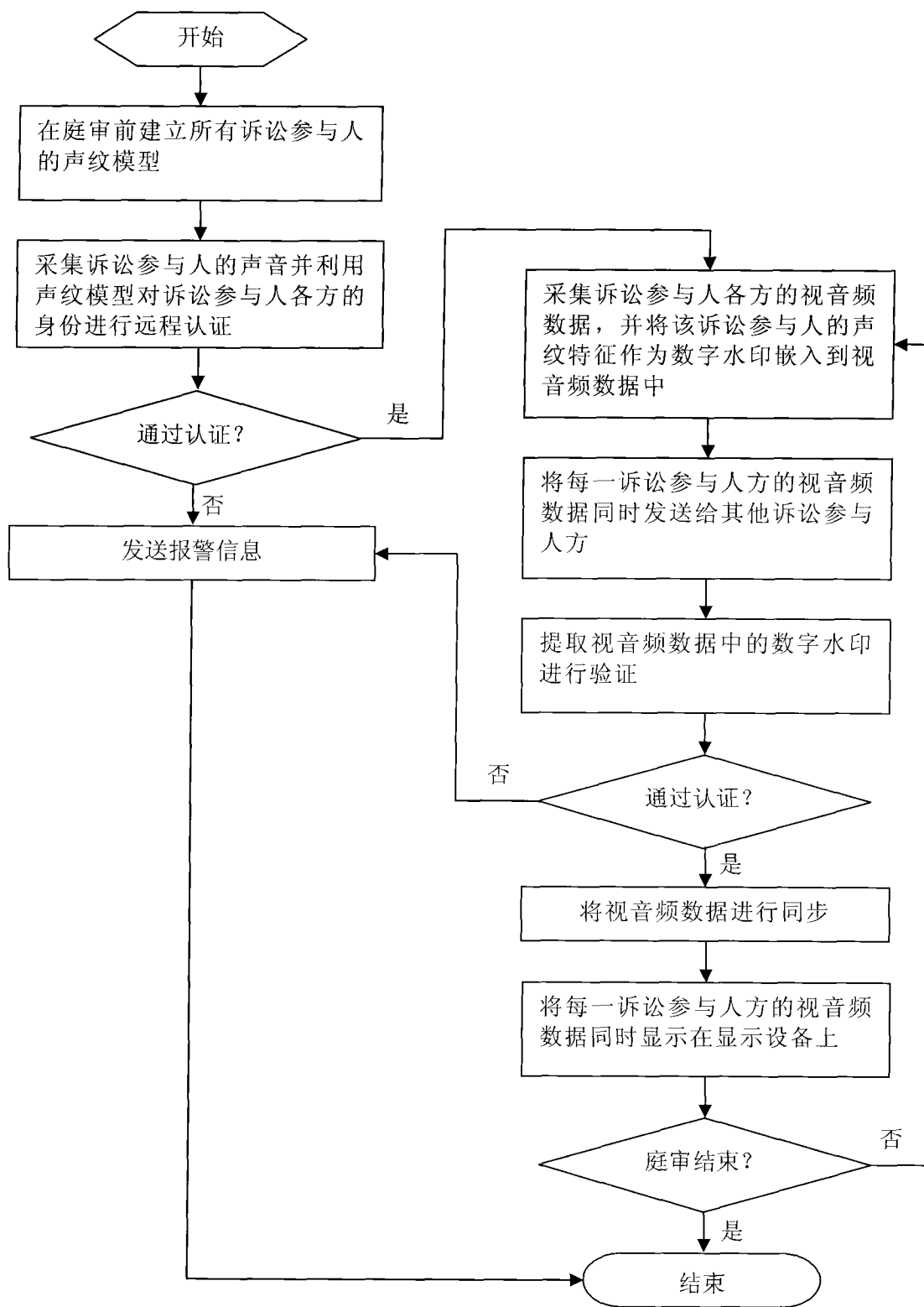


图 1

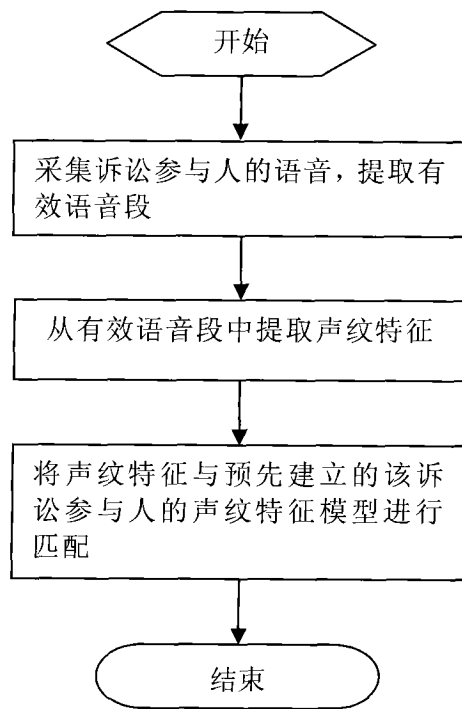


图 2

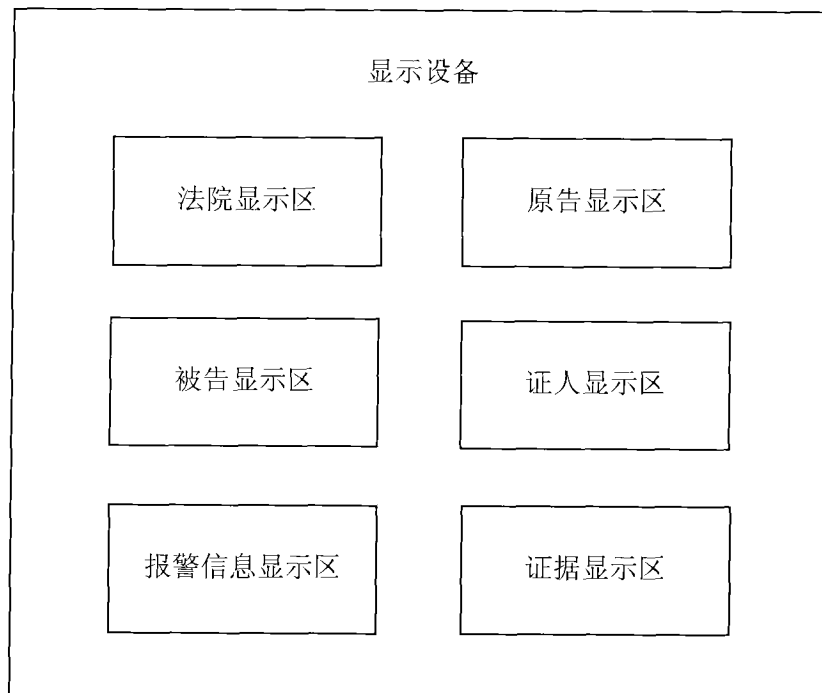


图 3