

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200710086531.7

[43] 公开日 2007 年 11 月 28 日

[51] Int. Cl.

G11B 20/10 (2006.01)

G11B 27/10 (2006.01)

G11B 27/00 (2006.01)

[22] 申请日 2007.3.13

[21] 申请号 200710086531.7

[30] 优先权

[32] 2006.7.28 [33] US [31] 11/495,836

[71] 申请人 埃里克·路易斯·汉森

地址 加拿大新斯科舍省哈利法克斯市春季  
花园街哈利法克斯专业中心 645 号

[72] 发明人 埃里克·路易斯·汉森

[11] 公开号 CN 101079301A

[74] 专利代理机构 北京安博达知识产权代理有限公司

代理人 徐国文

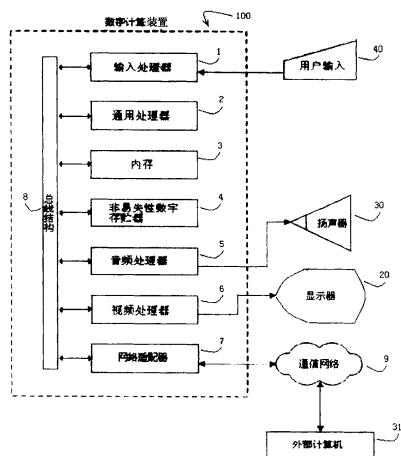
权利要求书 6 页 说明书 27 页 附图 3 页

[54] 发明名称

文本映射到音频并动画文本的设备及方法

[57] 摘要

本发明介绍了创建文本到音频时序映射的装置、方法和计算机可读取媒体。本发明还介绍了音频文本动画播放的装置、方法和计算机可读取媒体。映射器(10)将文本(12)以及与之对应的音频记录(11)作为输入，在用户帮助下将开始和结束时间(14)赋值到文本单元(15)中。播放器(50)将文本(15)、音频(17)和映射(16)作为输入，随着音频(17)的播放，同步制作文本动画(animate)并显示文本(15)。本发明可以用于在音频记录播放过程中给文本赋予生气；代替传统的重放控制器来控制音频播放；播放并显示语音记录的注释；在不用基础流协议的情况下实现流动音频的特征。



- 1、文本映射到音频的设备及方法，其中，包括至少一台计算机可读取多媒体，该多媒体包含着创建文本到音频记录时序映射的计算机程序指令，所述计算机程序指令执行：  
馈送步骤，作为文本到音频的映射设备及方法的输入，包括至少一台计算机，文本为计算机可读取格式的文本，相应音频记录为计算机可读取格式的音频记录；  
赋值步骤，将起始和终止时间赋值给对应此音频记录的文本单元，所述文本单元可以为文本粒度。
- 2、根据权利要求 1 所述的至少一台文本映射到音频的多媒体，其特征在于，所述文本粒度包括固定时间段音频、字母、音素、音节、单词、短语、句子和段落。
- 3、根据权利要求 1 中所述的至少一台文本映射到音频的多媒体，其特征在于，包括生成多个音频记录的步骤，通过对输入的音频记录在起始和终止时间边界上的差分，使得音频记录同对应的文本单元保持粒度上的一致。
- 4、根据权利要求 3 所述的至少一台文本映射到音频的多媒体，其特征在于，包括以下步骤：在不用基础流协议的情况下，利用所述多个音频记录实现音频流。
- 5、根据权利要求 1 所述的至少一台文本映射到音频的多媒体，其特征在于，上述文本的格式是以下格式之一：美国标准信息码 ASCII、统一码

Unicode、数字化乐器接口 MIDI 及在数字计算装置或者电子设备之间或之中发送数字编码音乐信息的格式。

- 6、根据权利要求 1 所述的至少一台文本映射到音频的多媒体，其特征在于，包括将注释赋值到上述单元的赋值步骤，其中：

注释的格式来自于一个格式组，该格式组包括文本、音频、图像、影像剪辑、统一资源定位器 URL 的格式和媒体格式；

注释可以是内容组中的内容，所述内容组由定义、翻译、脚注实例、参考、读音和用户对相应内容测验组成。

- 7、根据权利要求 1 所述的至少一台文本映射到音频的多媒体，其特征在于，包括将所述开始和结束时间以及所述单元以计算机可读取格式保存的步骤。

- 8、一种计算机实现的创建从文本到音频记录的时序映射方法，其特征在于，上述方法包括如下步骤：

馈送步骤，作为计算机实现映射模板的输入，文本为计算机可读取格式的文本，相应音频记录为计算机可读取格式的音频记录；

赋值步骤，将起始和终止时间赋值给对应此音频记录的文本单元，所述文本单元可以为文本粒度；

生成结构文本步骤，基于单元并进一步基于单元的开始和结束时间生成结构文本。

9、根据权利要求 8 所述的计算机实现的方法，其特征在于，其中的结构文本来自于超文本标记语言 HTML、可扩展标记语言 XML 和简单分界符文本；并且

结构文本指出的结构包括至少一个单元边界、不同粒度的单元层次、以及单元与单元的起始和终止时间之间的对应关系。

10、创建从文本同步映射到音频记录的设备，其特征在于，所述设备包括：

计算机上的映射模块，该模块将计算机可读取格式的文本以及对应于该文本的计算机可读取格式的音频记录作为输入；

赋值方法，将起始和终止时间赋值给对应此音频记录的文本单元，上述文本单元可以为文本粒度；及

交互选择方法，选择至少一个单元以及单元的颗粒值。

11、根据权利要求 10 所述的设备，其中，选择方法允许交互性改变、扩大和/或缩小颗粒值。

12、与音频记录同步动画 animating 文本并显示所述动画文本的设备，所述设备包括：

计算机实现的播放器模块，该模块将文本、对应于所述文本的音频记录以及文本与音频记录之间的时序映射作为输入，其中：

所述播放器模块动画文本，显示文本，并在音频记录播放时，同步显示文本；

所述动画使得被显示的文本随音频记录的播放而同步变化；并且，

所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。

13、根据权利要求 12 所述的设备，其中，所述文本是书面文本，并且所述音频记录是单词发声记录。

14、计算机实现的动画文本并与音频记录同步显示所述动画文本的方法，所述方法包括如下步骤：

馈送步骤，将文本、对应于所述文本的音频记录以及文本与音频记录之间的时序映射作为计算机实现的播放器模块的输入，其中：

所述播放器模块动画文本、显示文本，并在音频记录播放时，同步显示文本；

所述动画使得显示的文本随音频记录的播放而同步变化；

所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。

15、根据权利要求 14 所述的计算机实现的方法，其特征在于，还包括赋值给文本单元的注释的显示步骤，其中显示注释由用户依据文本单元基础交互触发或自动触发。

16、根据权利要求 15 所述的计算机实现的方法，其中：

注释由用户依据文本单元基础交互触发；并且

该基础是用户采用指示器或者输入设备来选择的字母、音素、音节、单词、短语、句子或者段落。

- 17、至少一台计算机可读取多媒体，包括动画文本并与音频记录同步显示所述动画文本的计算机程序指令，所述计算机程序指令完成下述步骤：  
    馈送，将文本、对应于所述文本的音频记录以及文本与音频记录之间的时序映射作为文本动画输出设备及方法的输入，其中：  
    所述文本动画输出设备及方法完成动画文本及显示文本，并在音频记录播放时，同步显示文本；  
    所述动画使得显示的文本随音频记录的播放而同步变化；并且  
    所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。
- 18、根据权利要求 17 所述的至少一台计算机可读取多媒体，其中，至少两个所述播放器模块、所述文本、所述音频记录和所述映射被集成在了一个可执行数字文件中。
- 19、根据权利要求 17 所述的至少一台计算机可读取多媒体，其特征在于，包括通过网络连接的传输步骤、至少一个所述播放器模块、所述文本、所述音频记录和所述映射。
- 20、根据权利要求 17 所述的至少一台计算机可读取多媒体，包括显示赋值给文本单元注释的步骤，其中，显示的注释由用户依据文本单元基础交互触发或自动触发。
- 21、根据权利要求 20 所述的至少一台计算机可读取多媒体，其中：  
    注释由用户依据文本单元基础交互触发；并且

该基础是用户采用指示器或者输入设备来选择的字母、音素、音节、单词、短语、句子或者段落。

- 22、计算机实现的音频记录传输方法，其特征在于，所述方法包括下述步骤：客户端计算机请求服务器计算机向其发送较长音频记录中的音频段，所述音频段具有一定时长的时间间隔；作为对所述客户端计算机所述请求的响应，所述服务器计算机给所述客户端计算机发送所述音频段。
- 23、根据权利要求 22 所述的计算机实现的方法，其中：  
音频段的呈现形式为一组计算机文件；并且  
所述服务器计算机利用文件传送协议向所述客户端计算机发送所述音频段。
- 24、根据权利要求 22 所述的计算机实现的方法，其中：  
较长的音频记录包含讲话；并且  
音频段由被指定的音节、单个单词和/或一系列单词的开始和结束点设定。
- 25、根据权利要求 22 所述的计算机实现的方法，其特征在于，包括在不使用基础流协议的情况下，利用所述传输的音频段来实现音频流的特性的步骤。

## 文本映射到音频并动画文本的设备及方法

### 技术领域

本发明涉及音频分析领域，特别是诸如演讲等包含文本说明的音频。更具体地讲，涉及创建文本到音频映射的设备及相应处理过程，以及与播放音频信息同步的文字动画设备。与音频同步的文本动画方式来表示文本到音频的映射，所传达的信息量远远多于音频或者文本自身所表达出来的信息量，也多于音频和静态文字合在一起所表达出来的信息量。

以本发明首套实施方案为依据，我们提供一台设备（“声文映射器 10”）以及一套文本到音频映射的创建方法。

以本发明第二套实施方案为依据，我们提供了一台带有音频播放的文本动画设备（“声文播放器 50”）。

本发明映射器 10 和播放器 50，克服了过去技术上存在的缺陷。以前的技术阻碍了语音和文本表示的同步实现。通过克服这些不足之处，映射器 10 和播放器 50 为更好和更新颖地应用语音加文本表示形式，开辟了一条新的途径。

### 背景技术

在基于语言方面的第一次技术进步是简单发声法的开发，当时这些发声法只能在时间上孤立地进行意思传达。后来，人们按时间相位和相继次序对这些最初的发声法进行组合，形成了语音流（streams of speech）。再后来，人们发明了在洞壁或其他合适表面上绘制简单符号或图像，但它们只是在空间上孤立进行意思传达。后人及时将这些符号或图像与口语联系了起来。后来，人们把这些独立的与语言有关的图形在空间相位上按相继次序组合形成了书面语言或“文本”。具体地说，我们具有创新精神的祖先，开始对象形、表意或者有音素特征的字符进行顺序空间排序，这些字符对应并且部分表示实际言语中按时间顺序排列、用来传达意思的发声序列。这些二维空间字符既传达意思又与

---

发声有关，用这些字符表示的相继次序是一项非常关键的创新，它使得我们能够将瞬时动态语音流的部分表示“凝固”成静态可储存文本。

模拟语音处理的发明，进一步提高了人们通过说话和文本进行交流的能力。这项技术发明使得我们能够凝固并存储动态语音流的声音，而不是满足于把语音部分等效存储为文本。近年来，通过以下方式，人们通过语言进行交流的能力得到了进一步延伸：一是通过对语音和文字进行数字编码，存储，处理，然后对其进行解码处理；二是电脑化文本搜索技术的开发；三是通过交互式文本包括交互式文本注释和超文本的开发。最后，因特网分布语音录音和文本技术的开发，使得人们通过语言进行交流的能力大大提高。这项技术是通过因特网，将语音录音和文本分布到越来越盛行的可编程或专用数字计算装置上。

概括地说，两个相继次序的出现，使得口语和书面语言的交流成为可能：一是用来传达意思的说话发声的时间相继次序，二是表示说话发声的象形、表意或者有音素特征的字符的空间相继次序。虽然上述两个相继次序都分别是一种有力的语言沟通形式，但是，语音和文本的局部等同，使得我们利用其中一个来表示或替代另一个成为可能。已经有多种途径证明这个局部等同非常有用，这些途径包括人们对两种妨碍人际交流的残疾——耳聋和失明的征服。具体地说，听不见口语但可以看见并学过阅读的人，通过读出抄写的口语单词，至少可以理解讲话的部分意思。其次，那些看不见书面语言的人，可以通过倾听把书写下来的文字转换成的发声，或者通过聆听原始讲话录音，理解书面上写的是什么。

对于具有视力和听力者，在同时进行语音和文本表示时，语音和文本表示的配合可以创造出一种有力的混合式语言交流形式。具体地说，语音和文本的同时传达，使得听/读者在同一时间内同时通过两条语言交流途径，即听和看，来得到信息。语音加文本表示中的讲话部分，会支持并强化书面信息，而语音加文本表示中的文本部分，则支持并强化讲话信息。简而言之，语言加文本表示的效果，好于其各部分表示的总和。

例如，在熟悉的国歌响起时，看到同步显示“星条旗”歌词，可能会让人在一个全新角度上油然而生感激之情。同样，在聆听马丁路德金演讲录音的同时，阅读他著名的“我有一个梦想”的原文，会让人沉浸在语言加文本的全新体验中，这种体验，性质上决不同于简单的阅读原文或聆听讲演。

语音加文本表示，在教育领域也有广阔的应用。例如，学习阅读某个人的本国语言，涉及到书写字符与相应口语单词的联想。通过语音加文本的同步表示，可以使得这种联想学习法变得更加容易。

语音加文本表示在教育领域的另一项应用，是外语或“第二”语言的学习——也就是学习至少在开始以口头或书写形式不能理解的语言。例如，一位学习德语的学生，可以玩语音加文本版的卡夫卡的“变形”，在阅读文本的同时，可以聆听口语版的故事。在这种第二语言学习应用中，诸如书面译文等文本注释可以帮助学生以口语和书面语两种形式理解第二语言，并帮助学生获得口语和书面表达能力。口语翻译形式的文本注释可以增强外语资料的语音加文本表示效果，这些文本注释可以是清晰的发音，或是个别单词的读音，也可以是弹出式测验题。

这种语音加文本表示在工业教育领域的一项应用，是增加书面技术资料的音频版本。音像版企业培训手册或者航空机师指南，可以这样制作：在文本显示的同时播放音频，以便让人们更透彻地理解其中的术语。

有多种原因导致难以理解一篇讲话，除了外语之外，还有例如，讲话录音中语音部分因本底噪声影响而模糊不清，讲话带有陌生口音，因与音乐伴奏混杂或节奏变化而导致的歌曲歌词难以听懂，或者在声乐中常见的因歌词或音节持续时间变化而导致的歌曲歌词难以听懂。所有这些问题，都可以通过将语音成分用书面和发声两种方式结合表示来解决。

对现存在的一些讲话以语音加文本形式进行记录，在保护濒危语言并存档保存这些语言方面，也可以起到建设性的作用。

一般地说，语音加文本的混合表示方式，借助于基于机器的文本搜索技术，

机器搜索这些表示中的语音成分的再现成为了可能。

我们首先提出从前的技术相对于本发明映射器 10 的不足之处，然后提出其相对于播放器 50 的不足之处。

现行的音频分析或者声音编辑程序，可以用来在用户选定位置录音时设置标记。然后可以输出这些标记，创建一个时间编码表。成对的时间编码可以按时间间隔译码。然而，用这样的方式创建的时间编码或时间编码间隔，并没有映射到文本信息中。该方法在录音和诸如讲话等可以以录音形式存在的文本表示之间形成一个映射。这就是以前的技术达不到本发明中映射器 10 的功能的原因。

现在，我们看看涉及到本发明播放器 50 以前的旧技术。在播放录音讲话的同时呈现该讲话的书面记录（或者在呈现文本的同时播放其发声转换版本）时，正在聆听的读者（或者是正在阅读的听众）会遇到几个问题：第一个问题是，对应于正在讲述的内容，怎样掌握文本中的文字进展到了何处？以前的技术有两种方法应对这个问题，在下文中我们会分析其不足之处。第二个问题是，在语音加文本表示中，组成文本的书写单词个体可以做成机器可搜索、可注释和交互式的，而音频部分的口语单词个体则不能。从前的技术尽管知道文本和音频之间的对应关系，但是未能使包含语音的音频做到机器可搜索、可注释和交互式。第三个问题是，音频成分的交互式传输需要制定一个流协议。从前的技术并没有涉及使用音频成分传送流协议的而引起的限制问题。

从前的技术曾试图以两种方法解决上述第一个问题，即“如何掌握文本中的文字进展”。

第一个方法是使语音加文本的分段保持简短。如果一段讲话比较简短，其相应的文本因此也会较短，因此播放的音频和显示的文本之间的关系可能会相对清晰一条件是正在聆听的读者明白语音加文本表示的语音和书面成分。同时显示的文本越长，受众理解讲话或书面文字（或二者）的难度越大，因此也越有可能不知道讲话在文本上进展到了何处。可是，正常人一般是以一个“行进

流”来讲话，并不局限于孤立的单词或者短语。此外，我们习惯阅读的文本是正常的连续文本，并不是那些为了方便显示而被分割成像单词或者短语那么长的一段段的文本。正常人的语音，包括声乐中的语音成分，如果其录音记录每次以单个单词或短语显示，然后快速变化以便跟上语音流，会显得很不自然。现有的伴读系统使用大块的文本或歌词，使得书面录音记录的表示更加自然，但同时也加大了受众在文本中搞不清讲话或歌词进展的可能性。

从前的技术曾试图用第二种方法，即与文本相关的动画法来解决跟上进展的问题。使用这种方法的实例有伴唱辅助系统，如一些较老的动画片中的“弹球”（bouncing ball），或者卡拉OK系统中的弹球或者其它的位置指示动画。画面上的球随着音乐从一个词移动到另一个词，随着音乐的继续提示歌词唱到了何处，或者是要唱的歌词。利用弹球或者等效物的移动，与文本有关的动画也增加了人们对其他静态文本的直观兴趣。

与语音同步的文本动画，显然有潜力以一种彻底、有效和令人喜爱的方法来将语音与其文本相联系。现有的技术实现了以视频记录或者电影的方式制作文本动画。但是以这种方式实现文本动画有很多弊病：

- 1、 这种影像的创作耗时较长，并且要求相关人员具有较高的技能。
- 2、 即使仅显示文本、播放音频，这种影像的创作也会形成大容量数据文件。这些大容量文件相应地占用大量带宽和数据存储空间，并因此对那些可以将语音加文本表示下载到可编程数字计算装置或者专用数字计算装置上的设备施加了很多限制。
- 3、 动画是固定式的。
- 4、 通常动画低于单词级粒度。
- 5、 除非作为视频的一部分，否则不能播放音频。

- 6、与音频之间的交互仅限于控制放像机。
- 7、音频不是机器可搜索或可注释的。
- 8、一旦制成视频，文本便无法更新或者改进。
- 9、文本不是机器可搜索或可注释的。
- 10、与文本自身不能交互。

## 发明内容

本发明在文本和音频之间建立连接，假定文本是录音语音的书面记录，或者语音是文本经发声转换而成的口语或歌唱形式。本发明：(a)定义了该种连接或者映射的创建过程；(b)提供了一套设备，以计算机程序的形式来帮助映射；(c)提供了另一个相关设备，也以计算机程序的形式，充分有效地证明了音频播放时文本和音频之间的连接。与音频播放同步的文本动画，说明了该连接的存在。

本发明采用的技术方案为：

文本映射到音频的设备及方法，其中，包括至少一台计算机可读取多媒体，该多媒体包含着创建文本到音频记录时序映射的计算机程序指令，所述计算机程序指令执行：

馈送步骤，作为文本到音频的映射设备及方法的输入，包括至少一台计算机，文本为计算机可读取格式的文本，相应音频记录为计算机可读取格式的音频记录；

赋值步骤，将起始和终止时间赋值给对应此音频记录的文本单元，所述文本单元可以为文本粒度。

在优选技术方案中，所述文本粒度包括固定时间段音频、字母、音素、音节、单词、短语、句子和段落。

在进一步的优选技术方案中，包括生成多个音频记录的步骤，通过对输入的音频记录在起始和终止时间边界上的差分，使得音频记录同对应的文本单元保持粒度上的一致。

在进一步的优选技术方案中，包括以下步骤：在不用基础流协议的情况下，利用所述多个音频记录实现音频流。

在进一步的优选技术方案中，上述文本的格式是以下格式之一：美国标准信息码 ASCII、统一码 Unicode、数字化乐器接口 MIDI 及在数字计算装置或者电子设备之间或之中发送数字编码音乐信息的格式。

在进一步的优选技术方案中，包括将注释赋值到上述单元的赋值步骤，其中：

注释的格式来自于一个格式组，该格式组包括文本、音频、图像、影像剪辑、统一资源定位器 URL 的格式和媒体格式；

注释可以是内容组中的内容，所述内容组由定义、翻译、脚注实例、参考、读音和用户对相应内容测验组成。

在进一步的优选技术方案中，包括将所述开始和结束时间以及所述单元以计算机可读取格式保存的步骤。

在进一步的优选技术方案中，上述方法包括如下步骤：

馈送步骤，作为计算机实现映射模板的输入，文本为计算机可读取格式的文本，相应音频记录为计算机可读取格式的音频记录；

赋值步骤，将起始和终止时间赋值给对应此音频记录的文本单元，所述文本单元可以为文本粒度；

生成结构文本步骤，基于单元并进一步基于单元的开始和结束时间生成结构文本。

在进一步的优选技术方案中，其中的结构文本来自于超文本标记语言 HTML、可扩展标记语言 XML 和简单分界符文本；并且

结构文本指出的结构包括至少一个单元边界、不同粒度的单元层次、以及单元与单元的起始和终止时间之间的对应关系。

在进一步的优选技术方案中，所述设备包括：

计算机上的映射模块，该模块将计算机可读取格式的文本以及对应于该文本的计算机可读取格式的音频记录作为输入；

赋值方法，将起始和终止时间赋值给对应此音频记录的文本单元，上述文本单元可以为文本粒度；及

交互选择方法，选择至少一个单元以及单元的颗粒值。

在进一步的优选技术方案中，选择方法允许交互性改变、扩大和/或缩小颗粒值。

在进一步的优选技术方案中，所述设备包括：

计算机实现的播放器模块，该模块将文本、对应于所述文本的音频记录以及文本与音频记录之间的时序映射作为输入，其中：

所述播放器模块动画文本，显示文本，并在音频记录播放时，同步显示文本；

所述动画使得被显示的文本随音频记录的播放而同步变化；并且，所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。

在进一步的优选技术方案中，所述文本是书面文本，并且所述音频记录是单词发声记录。

在进一步的优选技术方案中，所述方法包括如下步骤：

馈送步骤，将文本、对应于所述文本的音频记录以及文本与音频记录之间的时序映射作为计算机实现的播放器模块的输入，其中：

所述播放器模块动画文本、显示文本，并在音频记录播放时，同步显示文本；

所述动画使得显示的文本随音频记录的播放而同步变化；

所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。

在进一步的优选技术方案中，还包括赋值给文本单元的注释的显示步骤，其中显示注释由用户依据文本单元基础交互触发或自动触发。

在进一步的优选技术方案中，注释由用户依据文本单元基础交互触发；并且

该基础是用户采用指示器或者输入设备来选择的字母、音素、音节、单词、短语、句子或者段落。

在进一步的优选技术方案中，包括动画文本并与音频记录同步显示所述动画文本的计算机程序指令，所述计算机程序指令完成下述步骤：

馈送，将文本、对应于上述文本的音频记录以及文本与音频记录之间的时序映射作为文本动画输出设备及方法的输入，其中：

所述文本动画输出设备及方法完成动画文本及显示文本，并在音频记录播放时，同步显示文本；

所述动画使得显示的文本随音频记录的播放而同步变化；并且所述动画和同步是在构成文本的字母、音素或者音节级别上的动画和同步，从而达到了与相应音频记录的同步播放。

在进一步的优选技术方案中，至少两个所述播放器模块、所述文本、所述音频记录和所述映射被集成在了一个可执行数字文件中。

在进一步的优选技术方案中，包括通过网络连接的传输步骤、至少一个所述播放器模块、所述文本、所述音频记录和所述映射。

在进一步的优选技术方案中，包括显示赋值给文本单元注释的步骤，其中，显示的注释由用户依据文本单元基础交互触发或自动触发。

在进一步的优选技术方案中，注释由用户依据文本单元基础交互触发；并且

该基础是用户采用指示器或者输入设备来选择的字母、音素、音节、单词、短语、句子或者段落。

在进一步的优选技术方案中，所述方法包括下述步骤：

客户端计算机请求服务器计算机向其发送较长音频记录中的音频段，所述音频段具有一定时长的时间间隔；

作为对所述客户端计算机所述请求的响应，所述服务器计算机给所述客户端计算机发送所述音频段。

在进一步的优选技术方案中，音频段的呈现形式为一组计算机文件；并且所述服务器计算机利用文件传送协议向所述客户端计算机发送所述音频段。

在进一步的优选技术方案中，较长的音频记录包含讲话；并且音频段由被指定的音节、单个单词和/或一系列单词的开始和结束点设定。

在进一步的优选技术方案中，包括在不使用基础流协议的情况下，利用所述传输的音频段来实现音频流的特性的步骤。

本发明具有如下特性：

- 1、 语音加文本表示的动画，能够充分有效地证明口语单词及其文本表示之间的暂时关系。

- 2、 创建语音加文本表示形式的效率较高，此技术无需具有专门技能，无需专门培训。
- 3、 表示语音加文本的数据存储文件较小，所需数据传输带宽较小，因此适于快速下载到便携式计算设备上。
- 4、 动画呈现形式易于修改。
- 5、 可全部或部分地播放音频而不依赖于动画或者文本的显示。
- 6、 与语音加文本表示形式的交互，不局限于传统上对音频与视频播放器的各种控制（即“播放”、“倒带”、“快速进带”和“重放”），包括适于该技术的各种控制（如“随机存取”、“重复上个短语”以及“翻译当前单词”）。
- 7、 本发明使得语音加文本表示能够做到可机器搜索、可注释和交互式实施。
- 8、 本发明允许播放音频注释，也允许显示文本注释。
- 9、 本发明允许在表示被创建后对文本中的成分进行纠正或改变。
- 10、 本发明允许交互式随机读取音频，而不使用基础流协议。
- 11、 本发明提供了一个灵活的文本动画和创作工具，该工具可用于创作栩栩如生的语音加文本表示来满足各种特殊应用需要，如文化培训、第二语言学习、语言翻译以及教育、培训、娱乐和营销等。

## 附图说明

在下文说明书中，还会更详尽地描述这些说明以及本发明其他更详细具体

的实物和特征。请参看附图，图中以各种角度放大显示了本发明，以方便理解。

图 1 是适于本发明的数字计算装置 100 的方框图；

图 2 是本发明的声文映射器（“Mapper”）10 的方框图以及相关装置和数据；

图 3 是本发明的声文播放器（“Player”）50 的方框图以及相关装置和数据。

### 具体实施方式

本发明可以各种不同的形式来具体化。因此，此处公布的细节与其说是限制，不如说是作为样本教导擅长该技术的人员在任何适当的系统或者结构上，或者以任何方式去使用本发明。

图 1 显示了本发明的数字计算装置 100。数字计算装置 100 的组成如下：

1. 输入处理器，2. 通用处理器，3. 内存，4. 非易失性数字存储器，5. 音频处理器，6. 视频处理器，7. 网络适配器，上述部件均通过总线结构 8 连接在一起。数字计算装置 100 可以装在标准个人电脑、手机、灵巧电话、掌上电脑、笔记本电脑、个人数字助理等装备有适当的输入、视频显示器以及音频硬件的设备内。也可用专用硬件和软件来实现。它们可以集成到消费电器和装置中。

在使用时，网络适配器 7 可以接到通信网络 9 上，如局域网、广域网、无线通信网、因特网等等。外部计算机 31 可通过网络 9 与数字计算装置 100 通信。

图 2 所示是声文映射器（“Mapper”）10，该装置用于创建文本与音频录音之间的时序映射。图 3 所示是声文播放器（“Player”）50，该装置用来使文本以栩栩如生的方式显示，并且使这些文本动画与音频播放同步。

在此描述的本发明的所有组件和模块可以用任何硬件、软件、和（或）固

件的组合来实现。用软件实现时，上述组件和模块可以嵌入在计算机可读取介质或者媒体上，如一个或多个硬盘、软盘、CD、DVD 等。

映射器 10 (在处理器 2 上执行) 从内存 3、非易失性数字存储器 4 和 (或) 通过网络适配器 7 从网络 9 接收输入数据。输入数据有两个组成部分，音频记录 11 和文本 12，它们通常作为分立文件实现。

音频记录 11 是任意长度声音的数字表示，该数字表示用 MP3、OGG 或者 WAV 格式编码。音频记录 11 通常包括口语讲话。

文本 12 是书面文本或者字形的数字表示，该数字表示用 ASCII 或者 Unicode 格式进行编码。文本 12 还可以是乐器数字接口 MIDI(Music Instrument Digital Interface) 表示或者任何其他在数字计算装置或者电子装置之间发送音乐数字编码信息的格式。文本 12 通常由自然语言的书面文字组成。

音频记录 11 和文本 12 有一个内在的对应。示例之一是一段讲话的音频记录 11 与该段讲话文字稿的文本 12。另一个示例是一首歌曲的音频记录 11 与该首歌曲歌词的文本 12。还有一个示例是多个鸟语的音频记录 11 以及这些鸟的种类的文本名称 12。一个按时序对应的映射 (项(jana)明细表 16) 反映了这种内在对应。

标记(Marko)表 14 定义为开始和结束时间对 (标记开 (mark-on) 、标记关 (mark-off))，其单位是秒或其他时间单位。举例来说，数字对 2.000:4.500 定义了音频记录 11 中的音频数据，该音频记录开始于 2.000 秒，结束于 4.500 秒。

对标记 (Marko) 14 的限制包括数字对中的第二个数始终大于第一个数，并且标记 (Marko) 14 不重叠。

令牌表 15 是对应标记 (Marko) 14 的文本或者符号表示表。

成对出现的标记 (Marko) 14 和对应此标记 (Marko) 14 的文本或符号表示 15 称作项 (jana) 16 ( jana 发音为 yaw-na )。举例来说，单词 “hello”的音频在音频记录 11 中开始于 2.000 秒，结束于 4.500 秒，该单词的音频由 Marko2.000:4.500 来规定。Marko2.000:4.500 和令牌 “hello” 确定一个具体的项 (jana) 16。注意项 (jana) 16 只是数字和令牌 15 的对 14，项 (jana) 16 并未包含实际音频数据 11。

项 (jana) 表 16 是标记 (Marko) 表 14 和令牌表 15 的组合。项 (jana) 表 16 定义音频记录 11 和文本 12 之间的时序映射。

Mishcode (mishmash code 混杂编码) 定义为令牌 15 是符号而不是文本的项 (jana) 16。可能表示为 mishcode 的音频段实例是寂静、欢呼、咳嗽、仅有乐器声的音乐或者任何选定为不由文本表示的其他声音。例如，在音频记录 11 中的欢呼声开始于 5.200 秒并终止于 6.950 秒，该记录由 marko5.200:6.950，与之成对的是令牌 “<mishcode>” ，在此 “<mishcode>” 指的是一个具体的 mishcode。注意一个 mishcode 是一类项 (jana) 16。

带有文本表示的 mishcode 不再是 mishcode。例如，欢呼声可能通过文本 “掌声”、“欢呼”或者“观众突然欢呼”来表示。用该文本置换 “<mishcode>” 令牌后，它不再是一个混杂编码，但仍是项 (jana) 16。同理，用令牌 “<mishcode>” 来取代文本表示后，带有文本表示的项 (jana) 16 转换成一个 mishcode。

各项(jana)表示的音频，可以另存为独立的音频记录 17，通常计算机文件称为分离文件(split files)。表 14-16 和文件 17 可以存储在非易失性数字存储器 4 上。

显示器 20 连接在视频处理器 6 上，该显示器为用户提供数字计算装置 100 的视觉反馈。与音频处理器 5 连接的扬声器 30，为用户提供音频反馈。利用连接在输入处理器 1 和映射器 10 的用户输入 40，比如鼠标和(或)键盘，用户可以控制映射器 10。

在一套实施方案中，映射器 10 在显示器 20 上显示以下四个窗口：标记(Marko)窗格 21、令牌窗格 22、控制器窗格 23 和音量图窗格 24。在其他实施方案中，映射器的功能可以分散在不同数量的窗格中，该数量或多于、或少于 4 个。

标记(Marko)窗格 21 显示标记(Marko)14，每行显示一个。窗格 21 是可滚动的，该项可选。窗格 21 还可有交互控制功能。

令牌窗格 22 显示令牌 15，每行显示一个。窗格 22 也是可滚动的，该项可选。窗格 22 也可以有交互控制功能。

控制器窗格 23 显示编辑、播放、保存、加载和程序控制的控制器。

音量图窗格 24 显示音频记录 11 的某一段的音量图。窗格 24 也可以有交互控制功能。

图 2 所示系统的操作描述如下：

音频记录 11 通过映射器 10 接收，映射器生成一个初始标记 (Marko) 表 14，并在标记 (Marko) 窗格 21 中显示标记 (Marko) 表 14。初始标记 (Marko) 表 14 的创建方法有两种，一是利用音频记录 11 的声学分析由映射器 10 来创建；二是由映射器 10 把音频记录 11 划分为任意预选持续时间段的固定间隔。

可以根据预先选定的时间长度的音频 11 的音量是高于还是低于预先选定的音量阈值来进行声学分析。

在声学分析扫描中考虑三种情况：（1）等于或长于持续时间 D1 内小于音量阈值 V1 的音频记录的音频段为“平静” (lull) 类；（2）等于或长于持续时间 D2 内开始和结束音量大于阈值 V2 且不包含平静 (lull) 的音频段 11，为“声音” (sound) 类；（3）未包括在上述两类中的音频 11 为“模棱两可” (ambiguous) 类。

参数 V1 和 V2 规定音量，或者更准确地讲，规定声功率水平，比如以瓦或者分贝为单位测得的声功率水平值。参数 D1 和 D2 规定时间间隔，其测定单位为秒或者其它时间单位。四个参数 (V1、V2、D1 和 D2) 都是用户可选参数。

利用映射器 10，把模棱两可类的音频分解成接近的声音类或者平静类。这一点可由映射器 10 在声学分析完成后利用逻辑法则自动进行，也可由用户在控制器窗格 23 中人工干预进行。该步结束后，会形成一个标记 (Marko) 14 表，该表定义音频记录 11 中的各个声音；该表在标记 (Marko) 窗格 21 中显示。

利用任意持续时间的固定间隔创建初始标记 (Marko) 表 14，要求用户在控制器窗格 23 中选择一个时间间隔。标记 (Marko) 14 是选定的重复时间间隔，以

---

覆盖音频记录 11 的整个持续时间。表中的最后一个标记 (Marko) 14，可以比选定的时间间隔短。

文本 12 由映射器 10 接收，初始令牌表 15 则由映射器 10 生成并在令牌窗格 22 内显示。在标点、文字或者诸如 HTML 标签等中间数据的基础上，通过将文本 12 分离成单元（令牌），可以创建出初始令牌表 15。

下一步是一个交互式过程，通过此过程，用户可以创建标记 (Marko) 14 和令牌 15 个体之间的对应。

用户可以从标记 (Marko) 窗格 21 中选择一个标记 (Marko) 14 个体，并利用控制器窗格 23 从音频记录 11 播放对应的音频。从扬声器 30 中可以听到声音，并且声音的音量图显示在音量图窗格 24 内。标记 (Marko) 窗格 21 和令牌窗格 22 则显示标记 (Marko) 14 和令牌 15 之间的大致对应。通过下述操作，用户可以交互式地“提纯”这种对应。

标记 (Marko) 操作包括“拆分”、“结合”、“删除”、“裁剪”和“播放”。令牌操作包括“拆分”、“结合”、“删除”和“剪辑”。符号令牌的唯一规定操作是“删除”。根据具体的实施方案不同，可以通过标记 (Marko)、控制器和音量图窗格（分别为窗格 21、23 和 24）的组合来执行标记 (Marko) 操作，或者通过其他的用户输入 40 来执行标记 (Marko) 操作。根据具体的实施方案不同，可以通过令牌窗格 22 和控制器窗格 23 的组合来执行令牌操作，或者通过其他的用户输入 40 来执行令牌操作。

标记 (Marko) 拆分是将标记 (Marko) 窗格 21 中的标记 (Marko) 转换成两个有序的 marko X 和 marko Y，其中拆分点可以是原标记 (Marko) 14 开始和结束之

---

间的任何一个点。Marko X 始于原标记 (Marko) 的开始点, Marko Y 终于原标记 (Marko) 的终点, 且 Marko X 的终点与 marko Y 的始点相同。该点就是拆分点。用户可以参考音量图窗格 24 来帮助确定出合适的拆分点。音量图窗格 24 显示与当前的项 (jana) 16 对应的音频记录 11 部分的音量图。

标记 (Marko) 结合是将标记 (Marko) 窗格 21 中的两个有顺序的标记 (Marko), 即 marko X 和 marko Y 转换成一个单个标记 (Marko) 14, 其开始点是 marko X 的始点, 终点是 marko Y 的终点。

标记 (Marko) 删除是从显示在标记 (Marko) 窗格 21 中的标记 (Marko) 表 14 中去掉一个标记 (Marko)。

标记 (Marko) 裁剪是从标记 (Marko) 14 的始点或终点去除额外的信息。这相当于将标记 (Marko) 14 拆分成两个标记 (Marko) 14, 并废弃表示额外信息的那个标记 (Marko) 14。

标记 (Marko) 播放是播放与标记 (Marko) 14 对应的音频记录 11 部分。播放时, 该部分音频记录 11 在扬声器 30 上产生, 音量图在音量图窗格 24 上显示, 而对应于播放标记 (Marko) 14 的令牌 15 则在令牌窗格 22 上突出显示。在这种情况下“突出显示”指的是任何一种直观强调方式。

标记 (Marko) 操作还定义为标记 (Marko) 组操作: 一个标记 (Marko) 14 可以拆分为多个标记 (Marko), 多个标记 (Marko) 14 可以按相同的量来剪裁, 而多个标记 (Marko) 14 可以被结合、删除或播放。

“令牌拆分”是将令牌窗格 22 内显示的令牌 15 转换成两个有顺序的令牌,

---

即令牌 X 和令牌 Y，在此拆分点是字母、字符或象形字对之间的一个点。

“令牌结合”是将令牌窗格 22 内显示的两个顺序排列的令牌，即令牌 X 和令牌 Y，转换成一个单一令牌 15。转换的方式是将令牌 Y 以文本附加的方式附加到令牌 X 上。

“令牌编辑”指的是修改令牌 15 的文本，例如纠正拼写错误。

“令牌删除”是从令牌窗格 22 中显示的令牌表 15 中去除一个令牌。

交互过程完成后，各个标记 (Marko) 14 都有一个与之对应的令牌 15；这个标记 (Marko)、令牌对被称为项 (jana) 16，项 (jana) 16 的类集则被称为项 (jana) 表 16。

用户可以利用控制器自动生成在音频记录 11 中所有时间间隔的 mishcode，这些时间间隔未包括在音频记录 11 的项 (jana) 表 16 的任何一个标记 (Marko) 14 中。

项 (jana) 表 16 可以由映射器 10 保存，保存格式是计算机可读取格式，通常是一个或多个计算机文件。在一套实施方案中，项 (jana) 表 16 另存为两个分离文件，即标记 (Marko) 表 14 和令牌表 15。在另一个实施方案中，二者都保存在一个单一的项 (jana) 表 16 内。

标记 (Marko) 表 14 和令牌表 15 合并成一个单一项 (jana) 文件 16 包括如下方法：（1）表 14、15 单元的成对拼接 (concatenation)；（2）一个表 15 在另一个表 14 终点的拼接；（3）为标记 (Marko) 14 和令牌 15 单元定义 XML

---

或者其他中间数据标签。

映射器 10 的一个可选功能是创建各个项(jana)16 的分离音频记录 17。这些记录通常存储为计算机文件集，这些文件集通称为拆分文件 17。在不使用基础流协议的情况下，拆分文件实现了流的仿真。

为了解释其工作原理，下面对流作一个简要论述。在音频内容较多的常见流中，服务器和客户之间必须有一个通用的流协议。客户从服务器请求具体的内容片断。服务器开始利用达成一致的协议传输内容。服务器传送完一定量的内容后，通常是传送的量足以填满客户端的缓存器时，客户端开始播放该内容。用户快进该内容，由客户端向服务器发送一个请求来启动，该请求包括一个时间编码。然后，服务器中断流的传输，重新开始某一位置传输，这一位置由从客户端接收的时间编码规定。在这个点是，客户端缓存器开始重新填充。

流的本质是：（1）客户端向服务器发送一个请求；（2）服务器开始向客户端传输；（3）客户端缓存器充填；（4）客户端开始播放。

下面论述本发明如何对流进行仿真。客户（在此是外部计算机 31）请求从服务器（在此为处理器 2）传输一个内容片断的项(jana)表 16。服务器 2 利用任何一种文件传送协议按文本文件传输项(jana)表 16。客户端 31 向服务器 2 发送连续请求，请求传输有序的拆分文件 17 个体。服务器 2 利用任何一种文件传送协议将请求的文件 17 传输到客户端 31。请求的发送和对应拆分文件 17 的接收可以同时且非同步进行。通常，首个拆分文件 17 完成下载后，客户端 31 可以开始播放内容。

本发明可以实现音频流的正常要求。该流仿真方法的本质是：（1）客户端

31 向服务器 2 发送一个请求；（2）服务器 2 开始向客户端 31 传输；（3）客户端 31 至少接收一个单一的拆分文件 17；（4）客户端 31 开始播放拆分文件 17。

该音频传输方法有利于流的传输，具体优点有以下四点：

(1) 本发明使内容供应商不必再购买或者使用专门的流服务器软件，其原因是所有的内容传输均由文件传送协议而不是流协议来处理。网络服务器通常包括文件传递手段。因此，本发明适用于多数甚至全部网络服务器，不再需要任何流协议。

(2) 本发明允许以项(jana)16 或者其中多个项(jana)16 的颗粒度来播放不同范围的音频。注意项(jana)16 通常较小，时间跨度仅有几秒。流协议不能分离播放一块或者一个范围的音频，它们从某个给定点开始向前播放，所以，一旦客户端已经接收了用户期望的内容范围，客户端必须单独请求服务器停止传送。

(3) 在本发明中，快进和随机访问是设计的内在要素。服务器 2 不需要内容的内部结构知识来实现这些功能性单元，这一点与常见的流协议不同，那些协议要求服务器具有内部结构的详细知识。在本发明中，客户 31 通过发送有顺序的拆分文件 17 请求来实现快进和随机访问，其开始点是与音频中重放的开始点相对应的拆分文件 17。通过参考项(jana)表 16 来确定该点，具体地讲，就是(先前传递到客户端 31 的)项(jana)表 16 中的标记(Marko)14。执行文件传输的所有服务器 2，都可以实现本发明。

(4) 客户端 31 和服务器 2 之间的数据传输速度不足以跟上客户端 31 的音

频重放时，讲话录音重放中会有跳动现象，本发明改进了讲话录音重放中的这种跳动现象。在流协议中，音频重放会在音频流中某个不可预知的点上暂停，以充填客户端缓存器。从统计学意义上讲，在流型讲话（streaming speech）中，这类点可能发生在单词内。而在本发明中，这类点仅在项(jana)16边界上出现。至于讲话，项(jana)16符合自然讲话界限，这些边界通常定义音节、单词或者短的单词序列的开始和结束点。

播放器50（在处理器2上执行）从内存3、非易失性数字存储器4和（或）通过网络适配器7从网络9接收输入数据。输入数据至少有两个成分，通常作为文件，即项(jana)表16和一组拆分文件17实现。输入数据可以选择性地包括一组注释文件和索引56。

项(jana)表16是如上所述的时序映射。拆分文件17则是如上所述的音频记录。表16和文件17也许已经由图2所示装置产生，也许尚未由其产生。

注释文件集和索引56是由注释加一个索引组成的中间数据。注释格式可以是任意一个媒体格式，包括文本、音频、图像、影像剪辑（video clip）和（或）URL；可以有任意内容，包括定义、翻译、脚注、示例、参考、清晰注明的发音、交替的发音和测验（其中用户接受内容测验）。令牌15、令牌组、文本单元或者各注释个体所属的时间编码14等在索引中规定。在一套实施方案中，注释自身也可以有注释。

显示器20连接在视频处理器6上，该显示器为用户提供视觉反馈。与音频处理器5连接的扬声器30，为用户提供声频反馈。用户输入40，比如鼠标和（或）小键盘，连接在输入处理器1上，提供用户控制器。

---

播放器 50 在显示器 20 上显示窗口窗格。在一套实施方案中，窗口窗格有三个组成部分，即文本区 61、控制器 62 和可选滚动条 63。在其他实施方案中，播放器的功能可以扩展到多个数量不等的视觉组成部件中。

文本区 61 显示令牌 15，令牌 15 的格式符合用户选定的标准，包括文本单元的颗粒度，颗粒度可以为单词，短语，句子，或者段落。格式类型实例包括每行一个令牌 15、每行一个单词、按歌曲或诗歌的诗句或者按书本的段落。组成部分 61 也可以有交互式控制器。

控制器组件 62 显示各种控制器，比如音频播放、停止、回倒、快进、加载、动画类型、显示格式和注释弹出。

如果认为有必要或者希望滚动文本区 61，可以利用可选式滚动条 63。

图 3 所示系统的操作描述如下：

播放器 50 请求项 (jana) 表 16 的具体内容片断、相关注释文件以及索引 56 (如果有)。项 (jana) 表 16 由播放器 50 接收，并且显示文本区 61 和控制器 62。对应的令牌表 15 显示在文本区 61 内。

播放器 50 可以配置成在启动时自动开始重放，也可以配置成等待用户启动重放。两种情况下，播放器 50 都播放一个或一组项 (jana) 16。短语“项 (jana) 组”包括如下几种情况：整个项 (jana) 表 16 (从始到终)；从某个特定的项 (jana) 16 到最后一个项 (jana) 16 (当前位置到终点)；任意两个项 (jana) 16 之间。

可用以下方式启动重放，播放对应的项 (jana) 16 或者一组项 (jana) 16：

---

(1) 由用户启动播放整个项(jana)表16的起始控制器；(2) 启动播放当前项(jana)16到终点的起始控制器；(3) 利用鼠标、小键盘或者其他输入设备40，在文本区61内选择任意一个令牌15或者一组令牌播放。

项(jana)16的播放，通过播放相对应的拆分文件17实现。播放器50从以下三个地方得到所需的拆分文件17：正在运行的播放器50的处理器2；另一台计算机；内存3，条件是先前已经得到了拆分文件17并缓存在内存中。

如果需要多个拆分文件17，并且那些文件17未在超高速缓存器3，则播放器50开始连续请求必需的拆分文件17。

重放的启动，会同时启动一个（连接在播放器上的）实时时钟，该实时时钟已经根据被播放的项(jana)16内的标记(Marko)14的开始时间进行了初始化。

实时时钟被设定到了与音频重放同步。例如，如果音频重放停止，实时时钟也同步停止；如果音频重放速度放慢、加快或跳跃播放，实时时钟也随之调整。

动画文本按实时时钟进行。具体地说，当时实时时钟在项(jana)的标记(Marko)时间间隔内，播放此项(jana)16的令牌15的动画。另外，如果文本区61内正在播放项(jana)16的文本是不可视的，文本区61会自动滚动，以便使文本可见。

文本动画包括各种情况，其中文本变化的直观表示与音频重放同步。动画和同步可以在单词、短语、句子或段落水平上实现，也可以在构成文本的字母、

---

语音或音节水平上实现，以便与相应音频记录的重放达到平滑、精密的同步。

文本动画包括文本或者背景的运动幻影和(或)颜色、字体、透明度和(或)可见度的变化。运动幻影可以是逐字式(即一个字跟一个字)，比如卡拉OK中的“弹球”，弹出式，或升离基准线式。运动幻影也可连续发生，比如显示条随文本移动，或“彩带”效应。可单独或组合采用不同的动画方法。

如果当前项(jana)表16已有注释文件和索引56，那么便可以显示、播放或者弹出有关注释。包括文本、音频、图像、影像剪辑(video clip)、URL等的注释文件和索引56，是根据需要进行请求的。

注释的显示、播放或者弹出，要么由用户启动，要么自动操作。

用户触发的注释显示通过用户与令牌15上的文本区61或文本单元基础的交互来实现。举例来说，用户触发的注释，其调用方法包括采用鼠标、小键盘或者其它输入设备40来选择单词、短语或者句子。

如果激活了自动注释，则可以利用间隔定时器，由实时时钟从外部激励源或者随机激活。自动注释实例包括幻灯、文本区背景或者音频、视频或文本注解。

三个具体的注释实例是：(1)在文本区61中单词“埃佛勒斯峰”上单击鼠标右键，弹出一个埃佛勒斯峰图像；(2)在文本区61中，当单词“你好”突出显示时，按下翻译按钮则显示法语翻译“bonjour”；(3)在播放歌曲“老麦克唐纳”时，会适时自动出现农家院家畜的插图。

在一套实施方案中，播放器50、项(jana)表16、拆分文件17和(或)注

---

释文件和索引 56 等都综合在一个单一可执行数字文件内。上述文件可以通过网络适配器 7 传输到装置 100 之外。

最后应当说明的是：以上实施例仅用以说明本发明的技术方案而非对其限制，尽管参照上述实施例对本发明进行了详细的说明，所属领域的普通技术人员应当理解：依然可以对本发明的具体实施方式进行修改或者等同替换，而未脱离本发明精神和范围的任何修改或者等同替换，其均应涵盖在本发明的权利要求范围当中。

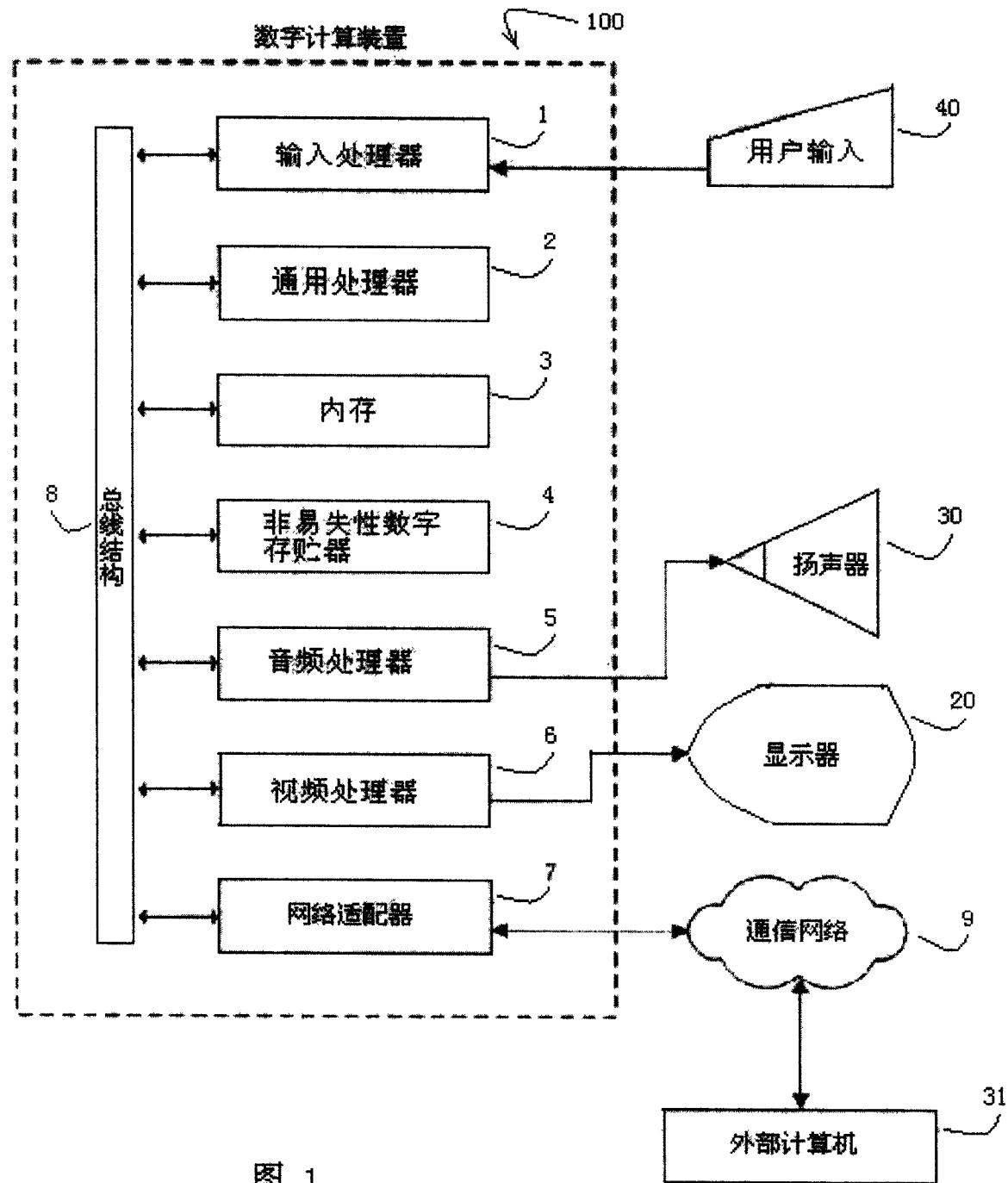


图 1

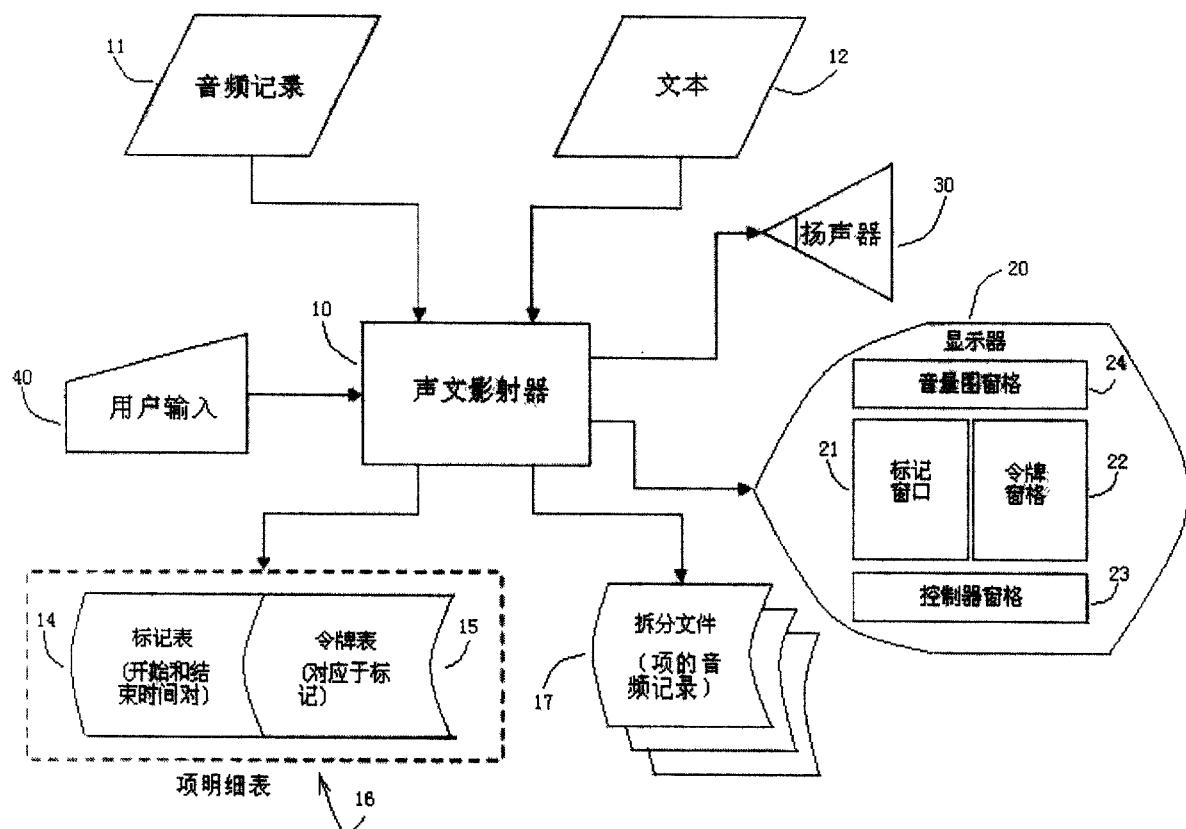


图 2

