

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G10L 15/26 (2006.01)

G06F 17/24 (2006.01)



[12] 发明专利说明书

专利号 ZL 02820209.0

[45] 授权公告日 2007 年 4 月 25 日

[11] 授权公告号 CN 1312657C

[22] 申请日 2002.10.7 [21] 申请号 02820209.0

[30] 优先权

[32] 2001.10.12 [33] EP [31] 01000535.3

[86] 国际申请 PCT/IB2002/004107 2002.10.7

[87] 国际公布 WO2003/034404 英 2003.4.24

[85] 进入国家阶段日期 2004.4.12

[73] 专利权人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 H·F·巴托斯克 K·拉吉

[56] 参考文献

CN1181825A 1998.5.13

US6195637B1 2001.2.27

US5031113 1991.7.9

审查员 刘亚斌

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 王岳 王忠忠

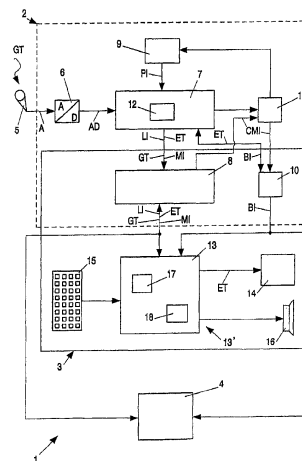
权利要求书 3 页 说明书 18 页 附图 5 页

[54] 发明名称

用于标注所识别文本的部分的转录设备和方法

[57] 摘要

在一种用于将口语文本(GT)转录成所识别文本(ET)以及用于编辑所识别文本(ET)中被非正确识别的部分的转录设备(1)中,标注装置(12,15,17)被提供,以便于被设置成用于对具有公共特征的口语文本(GT)部分和/或所识别文本(ET)部分进行部分自动标注和部分手动标注。结果是,具有公共特征的被标注的文本部分的后来的统一处理变为可能。



1. 一种转录设备(1)，其用于将口语文本(GT)转录成所识别文本(ET)并且用于编辑所识别文本(ET)中被非正确识别的部分，其中用于口语文本(GT)每部分的链接信息(LI)对相联系的所识别文本(ET)进行标记，

具有语音识别装置(2)，其用于将所接收的口语文本(GT)转录成所识别文本(ET)并且用于产生链接信息(LI)，以及

具有存储器装置(8)，其用于存储口语文本(GT)、所识别文本(ET)及链接信息(LI)，以及

具有标注装置(12, 15, 17)，其用于自动地确定标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项并且用于在存储器装置(8)中存储所确定的标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项，所述标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项标记具有为所有被标注的文本部分所共有的特定特征的部分所识别文本(ET)或部分口语文本(GT)，其中所述标注装置(12, 15, 17)具有用于手动录入标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项的输入装置(15, 17)，以允许也具有这个公共特征的所识别的或口语文本的另外部分被以这些标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项标注，以及

具有处理装置(13')，其用于统一处理被标注的文本部分。

2. 根据权利要求1的转录设备(1)，其中，为了允许语音识别装置(2)的参数(PI)得到适配，提供适配级(11)，该适配级被设置成用于评估被自动和手动加以标注的文本部分，所述被标注的文本部分具有公共的特征。

3. 根据权利要求1的转录设备(1)，其中再现装置(13)被连接到存储器装置(8)，并且被设置成当同步再现模式在转录设备(1)中被启动时用于口语文本(GT)的声学再现以及相联系的所识别文本(ET)的同步视觉标注，所述所识别文本(ET)由链接信息(LI)进行标记且从光学上被再现，以及其中被自动或手动加以标注的文本部分的声学或光学再现可以得到抑制。

4. 根据权利要求3的转录设备(1)，其中所述标注装置(12, 15, 17)被设置成用于自动确定且在存储器装置(8)中存储再现标注

信息(WMI),所述再现标注信息(WMI)标记当同步再现模式被启动时被再现装置(13)至少再现过一次的那些所识别文本(ET)部分和/或口语文本(GT)部分。

5. 根据权利要求3的转录设备(1),其中所述再现装置(13)被设置成用于光学再现得到抑制的被标注的所识别文本(ET)部分的替代文本。

6. 根据权利要求5的转录设备(1),其中所述再现装置(13)可以在替代文本的光学显示和被标注的文本部分的光学显示之间被切换。

7. 根据权利要求6的转录设备(1),其中对于形成文本模块的被标注的文本部分,可以进行到为下一个模块而定义的文本的光学显示的附加切换。

8. 根据权利要求1的转录设备(1),其中所述标注装置(12, 15, 17)被设置成:当一部分口语或所识别文本具有由标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项所标记的至少两个公共特征时,用于将这个被标注的文本部分标注有至少两个不同的标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项。

9. 一种将口语文本(GT)转录成所识别文本(ET)以及编辑所识别文本(ET)中被非正确识别的部分的方法,其中口语文本(GT)每部分的链接信息(LI)标记相联系的所识别文本(ET)并且其中下述步骤被加以执行:

将所接收的口语文本(GT)转录成所识别文本(ET)以及产生链接信息(LI);

存储口语文本(GT)、所识别文本(ET)和链接信息(LI);

自动确定标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项以及存储所确定的标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项,所述标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项标记具有为所有被标注的文本部分所共有的特定特征的所识别文本(ET)部分和/或口语文本部分;

手动录入标注信息(CMI, PMI, HMI, GMI, DMI, RMI, WMI)项以允许也具有这个公共特征的所识别的或口语文本的另外部分被

标注有这些标注信息 (CMI, PMI, HMI, GMI, DMI, RMI, WMI) 项;
以及

统一处理被标注的文本部分。

10. 根据权利要求 9 的转录方法, 其中下述进一步的步骤被加以执行:

当同步再现模式被启动时, 声学再现口语文本 (GT) 以及同步视觉标注由链接信息 (LI) 所标记且被光学再现的相联系的所识别文本 (ET), 同时自动或手动被标注的文本部分的声学或光学再现可以得到抑制。

11. 根据权利要求 9 的转录方法, 其中下述进一步的步骤被加以执行:

自动确定且存储再现标注信息 (WMI), 所述再现标注信息 (WMI) 标记在同步再现模式被启动之前所识别文本 (ET) 和/或口语文本 (GT) 中被再现过至少一次的那些部分。

12. 根据权利要求 9 的转录方法, 其中下述进一步的步骤被加以执行:

当一部分口语或所识别文本具有由标注信息 (CMI, PMI, HMI, GMI, DMI, RMI, WMI) 项所标记的至少两个公共特性时, 将这个被标注的文本部分以至少两个不同的标注信息 (CMI, PMI, HMI, GMI, DMI, RMI, WMI) 项来标注。

用于标注所识别文本的部分的转录设备和方法

技术领域

本发明涉及一种用于将口语文本转录成所识别文本以及用于对所识别文本的被非正确识别的部分进行编辑的转录设备，其中口语文本每个部分的链接信息标记（flag）相联系的所识别文本。

本发明还涉及一种用于将口语文本转录成所识别文本以及用于编辑所述所识别文本的被非正确识别的部分的转录方法，其中口语文本每个部分的链接信息标记相联系的所识别文本。

背景技术

这种类型的转录设备和转录这种方法从专利 US 5, 031, 113 中已知，其中公开了一种听写设备。已知的听写设备由计算机形成，所述计算机运行语音识别软件及文本处理软件。已知听写设备的使用者可以对着被连接到计算机的话筒说出口语文本。形成转录设备的语音识别软件执行转录方法并且通过这样做将所识别的字分配给口语文本的每个口语字，其结果是获得口语文本的所识别文本。同样，作为转录方法的一部分，链接信息被加以确定，即其标记针对口语文本的每个口语字而被识别的所识别文本的字。

已知的听写设备还包含用于自动确定和存储标注信息的标注装置。所识别文本的一个或更多个字可例如在这种情况下被识别为命令，所述命令例如在格式化所述所识别文本时将被执行。因此在所识别文本中的字序列“...next word bold”可以被翻译成用于格式化紧接在所识别文本后面的字的命令。这些已知的标注装置被加以设置用于命令标注信息的自动确定和存储，所述命令标注信息标注所识别文本中具有这样公共特征的那些字序列，即它们可以被用于格式化所识别文本的字的命令。

已知的听写设备还运行文本处理软件且因此形成校正设备，借此校正设备被非正确识别的字可以被编辑。出于这个目的，校正设备的使用者可以启动校正设备的同步再现模式，其中口语文本从声学上被

再现，并且与此同步由链接信息所标记的所识别文本的字从视觉上被加亮（即被标识）。同步再现模式在实践中已经被证实对于校正由语音识别设备所识别的文本尤其有利。

在实践中已经发现许多使用者难以记住一命令的正确字序列。因此例如他们说“...next word should be bold”（下个字应该为粗体），但是标注装置并没有将这个字序列识别为命令且因此对此并不标注。当校正这个所识别文本时使用者因此必须从所识别文本中删除整个字序列并且手动格式化紧接所述序列的字，其相当费时。还存在与同步再现模式相联系所出现的一个缺点，即在通过这种方法已经对所识别文本加以校正之后，已经被从所识别文本中删除的这个字序列的口语文本的链接信息或者指向所识别文本中非正确的字，或者根本并不指向其中的任何字，其引起任何后来的同步再现错误且让使用者迷惑。

发明内容

本发明的一个目的是提供在上面首段落中所定义类型的转录设备以及在上面第二段落中所定义类型的转录方法，其中上面所说明的缺点得以避免。

为了取得上面所指示的目的，在这种类型的转录设备中提供根据本发明的特点，以便于所述转录设备可以以下面所详细说明的方式被表征。

一种转录设备，其用于将口语文本转录成所识别文本以及用于对所识别文本的被非正确识别的部分进行编辑，其中口语文本每部分的链接信息标记相联系的所识别文本，

具有用于将所接收的口语文本转录成所识别文本且用于产生链接信息的语音识别装置，以及

具有用于存储口语文本、所识别文本和链接信息的存储器装置，以及

具有用于自动地确定标注信息且用于在所述存储器装置中存储所确定的所述标注信息的标注装置，所述标注信息标记具有为所有被标注的文本部分所共有的特定特征的所识别文本部分及/或口语文本部分，其中所述标注装置具有用于手动录入标注信息的输入装置，以允

许也具有这个公共特征的所识别或口语文本的另外部分被标注有这个标注信息；并且具有用于统一处理被标注的文本部分的处理装置。

为了取得上述所指示的目的，在上述类型的转录方法中提供根据本发明的特点，以便于所述转录方法可以以下面所详细说明的方式被表征。

一种转录方法，其用于将口语文本转录成所识别文本以及用于对所识别文本的被非正确识别的部分进行编辑，其中口语文本每部分的链接信息标记相联系的所识别文本，以及

其中下述步骤被加以执行：

将所接收的口语文本转录成所识别文本且产生链接信息；存储口语文本、所识别文本和链接信息；自动确定标注信息且存储所确定的所述标注信息，所述标注信息标记具有为所有被标注的文本部分所共有的特定特征的所识别文本部分及/或口语文本部分；手动录入标注信息，以允许也具有这个公共特征的所识别或口语文本的另外部分被标注有这个标注信息；以及统一处理被标注的文本部分。

根据本发明的特点取得：使用者能够手动地将部分所识别文本及/或口语文本标注有下述标注信息，所述标注信息与某些另外部分的所识别文本从前可能已经被自动标注的标注信息相同。在后来的处理操作中，那些被自动或手动加以标注的具有公共特征的文本部分可以通过处理装置以统一的方法被处理。被标注的文本部分的公共特征可例如在这种情况下是：在所识别文本中被标注的字序列被分配给一命令或被标记为头部；被标注的所识别文本部分被链接信息分配到包含犹豫声音、暂停、不清楚的口语文本部分或文本的冗余部分如被重复字的口语文本部分。

结果是取得下述优点，即已经被自动标注的文本部分可以利用手动标注被补充，这样具有特定特征的所有被标注的口语文本部分及/或所识别文本部分可以通过统一方法在后来的处理操作中被加以处理。因此，例如，由语音识别装置难以识别（即可能在识别中有大量错误）的所有文本部分以及由校正人员在同步再现模式中被手动识别为困难的所有文本部分可以通过统一方法被标注，以为了用于核查的目的使口语文本的作者参考这些被标注的文本部分。通过相同的方法，不同

的序列字可被标注有这样的标注信息，所述标注信息与标注装置已经用以自动标注被分配有一命令的序列字的标注信息相同，结果是所述命令也被分配给这些被手动标注的文本部分。

在上述的标注设备中，为了允许语音识别装置的参数得到适配，提供了适配级，其被设置成用于评估被自动和手动加以标注的文本部分，所述被标注的文本部分具有公共的特征。这种实施方式的优点在于，即已经被自动及手动标注的文本部分可以针对参数适配进行评估，借此参数适配所识别文本的质量可以得到改善。同样，通过手动标注命令标注信息，所识别文本中的字序列可以被分配给命令，然后在适配期间所述序列字被添加到标注装置中的命令集合中，并且下次将被自动标注。

当同步再现模式被启动时，声学再现口语文本以及同步视觉标注由链接信息所标记且被光学再现的相联系的所识别文本，同时自动或手动被标注的文本部分的声学或光学再现可以得到抑制。由此，提供了这样的优点，即例如，被自动以及还被手动标注为犹豫声音的所有文本部分在同步再现模式中从声学上并不被再现并且在所识别文本的光学再现期间是隐藏的。结果是所识别文本可以更高效地被显著校正，因为在声学再现时时间被节省且使用者可以专心于所识别文本的重要部分。

所述转录方法还自动确定且存储再现标注信息，所述再现标注信息标记在被启动的同步再现模式之前所识别文本和/或口语文本中被再现过至少一次的那些部分。由此，提供了这样的优点，即当同步再现模式被启动时，已经从声学上被再现且从视觉上被标注至少一次的那些所识别文本部分及/或口语文本部分通过标注装置被自动地标注有再现标注信息。为了使校正更容易，在同步再现已经被标注有再现标注信息的文本部分期间，转录设备因此有利地具有从声学上再现例如所有的犹豫声音和被重复字的能力。结果是根据本发明的校正设备的使用者可以更高效地显著校正所识别文本。

在本发明的转录设备中，再现装置被设置成用于光学再现得到抑制的被标注的所识别文本部分的替代文本。由此，提供了这样的优点，即替代文本被显示在所识别文本内其中已经得到抑制的被标注文本部

分所处存在的位置。如果这个替代信息标记已经得到抑制的被标注文本部分的特征，则其尤为有利。

在上面所述的转录设备中，再现装置可以在替代文本的光学显示和被标注的文本部分的光学显示之间被切换。由此，提供了这样的优点，即使用者可以以简单的方法选择当时更为有利的所识别文本的显示方式。

在上面所述的转录设备中，对于形成文本模块的被标注的文本部分，可以进行到为下一个模块而定义的文本的光学显示的附加切换。由此，提供了这样的优点，即有可以切换到显示文本模块的进一步有利的方式。

当一部分口语或所识别文本具有由标注信息项所标记的至少两个公共特性时，将这个被标注的文本部分以至少两个不同的标注信息项来标注。由此，提供了这样的优点，即具有多个特征的文本部分可以被标注有适当的不同项的标注信息。一部分文本可例如被标注为头部并且包括犹豫声音。在这种情况下犹豫声音将既被标注有头部标注信息又被标注有犹豫声音标注信息，并且因此将在后来的对口语和所识别文本中的头部进行编辑期间以及在对这些文本中的犹豫声音进行编辑期间被加以编辑。

现在参考附图将对本发明更详细地加以说明，所述附图示出单个实施例，但本发明并不被局限于此。

附图说明

图 1 示出一种用于将口语文本转录成所识别文本的转录设备，其中校正设备被提供以用于校正所述所识别文本。

图 2 以符号形式示出口语文本、由转录设备所识别的对应文本、以及标记针对口语文本每个音频段的所识别文本的相联系段的链接信息。

图 3 示出被存储在所述转录设备的命令存储器级 (stage) 中的命令表。

图 4 示出被存储在所述转录设备的存储器装置中的标注表。

图 5 以符号形式示出由所述转录设备的使用者所说的五个文本

实例。

图 6 以符号形式示出针对口语文本的五个实例由所述转录设备的语音识别装置所识别的文本，其中部分所识别文本已经被语音识别装置的第一标注级自动地标注且使用者利用第二标注级已经手动标注了文本的另外部分。

图 7 以符号形式示出针对五个实例由所述转录设备的校正设备所显示的文本，其中所识别文本的被自动和手动标注的文本作为替代文本被显示。

图 8 以符号形式示出当被标注的文本部分的所有替代文本被加以抑制时由校正设备所显示的文本。

图 9 以符号形式示出在口语及所识别文本的第一和后来第二同步再现期间哪些部分文本被加以再现。

具体实施方式

图 1 示出一种转录设备 1，其用于将口语文本 GT 转录成所识别文

本 ET 并且用于对未被正确地识别的所识别文本 ET 部分进行编辑。转录设备 1 由运行语音识别软件且形成语音识别设备 2 的第一计算机以及第二和第三计算机形成, 所述第二和第三计算机的每个运行文本处理软件且形成用来校正还一直未被正确识别的文本部分的第一校正设备 3 和第二校正设备 4。可提及的是: 根据本发明的转录设备还可以由形成语音识别设备及校正设备两者的仅一个计算机形成, 然后所述计算机将不得不运行语音识别软件和文本处理软件两者。

被连接到形成语音识别设备 2 的第一计算机上的是话筒 5, 从所述话筒 5 可以发射表示口语文本 GT 的音频信号 A。语音识别设备 2 包含 A/D 转换器 6、语音识别装置 7、存储器装置 8、参数存储器装置 9、命令存储器装置 10 和适配级 11。由话筒 5 所发射的音频信号 A 可以被馈送到 A/D 转换器 6, 所述 A/D 转换器将音频信号 A 转换成数字音频数据 AD。

表示口语文本 GT 的音频数据 AD 可以被馈送到语音识别装置 7 且由此可以被存储在存储器装置 8 中。当语音识别过程正在由语音识别装置 7 所执行时, 语音识别装置 7 被加以设置以确定所识别文本 ET, 在这样做时被存储在参数存储器装置 9 中的参数信息 PI 被加以考虑。参数信息 PI 在这种情况下包含词汇信息、语言模型信息及声学信息。

词汇信息包含可以由语音识别装置 7 以及相联系的音素序列所识别的所有字。语言模型信息包含在口语文本 GT 语言中常见的与字序列有关的统计信息。声学信息包含有关转录设备 1 使用者发音的特殊特点及有关话筒 5 和 A/D 转换器 6 的声学特征的信息。

US 专利 5, 031, 113 的公开被认为是被结合进本发明的公开内容以作为参考, 并且从这种类型的参数信息 PI 的观点来看由于语音识别过程的性能被公开在所述 US 专利中, 所以在此将不对其进行详细说明。语音识别过程的结果是包含所识别文本 ET 的文本数据可以被语音识别装置 7 存储在存储器装置 8 中。

在语音识别过程的执行期间, 链接信息 LI 也可以由语音识别装置 7 加以确定, 其中所述链接信息 LI 标记针对口语文本 GT 的每部分由语音识别装置 7 所识别的相联系文本 ET。链接信息 LI 的产生同样被公开在 US 专利 5, 031, 113 中, 为此在此未详细对其加以讨论。

在图 2 中,以符号形式示出沿着时间轴 t 由作者,即语音识别设备 2 的使用者对话筒 5 所说出的文本 GT。由语音识别装置 7 针对这个口语文本 GT 所识别的文本 ET 也被示出。当语音识别过程正在被执行时,口语文本 GT 被分成包含相联系的声学信息的音频段 AS。这种类型的相联系的声学信息可例如是字、两个字之间的稍长的语音暂停,即被称为犹豫声音如“aah”或“mm”,或噪音。

链接信息 LI 标记口语文本 GT 的每个音频段 AS 及所识别文本 ET 的相联系段 TS 的开始和结束。例如第一音频段 AS1 包含针对口语文本 GT 第一字“The”的持续 1.5 秒的声学信息,以及由链接信息 LI 所分配的第一文本段 TS1 包含由语音识别装置 7 所识别的字“The”的文本。

命令存储器级 10 存储被语音识别设备 2 识别为命令的字序列。被存储在命令存储器级 10 中的一部分命令表 BT 被示于图 3 中。通过参考所述命令表 BT,语音识别装置 7 将例如字序列“next word bold”识别为所识别文本 ET 中的下个字即将被显示成黑体的命令。命令号 BL12 被分配给命令表 BT 中的这个命令。字序列“insert text of module 1”(插入模块 1 的文本)同样被识别为命令,在这种情况下来自被存储在命令存储器级 10 内位置“2341”上的存储器位置处的文本模块的标准文本即将被插入进所识别文本 ET。

语音识别装置 7 进一步包括第一标注级 12,其被设置成用于自动确定标注信息 MI 的不同项且用于在存储器装置 8 中存储被如此标注的标注信息 MI 的项,所述标注信息 MI 的项标记具有为所有被如此标注的文本部分所共有的某些特征的所识别文本 ET 部分和口语文本 GT 部分。字“自动的”在这个环境中被理解为意味着“转录设备 1 的使用者没有任何动作”。第一标注级 12 被设置成用于自动标注带有暂停标注信息 PMI 的稍长语音暂停、带有犹豫声音标注信息 HMI 的犹豫声音、带有命令标注信息 CMI 的命令标注信息、带有重复标注信息 RMI 的被重复字、带有日期标注信息 DMI 的日期、以及带有噪音标注信息 GMI 的噪音。

图 4 示出被存储在存储器装置 8 中的标注表 MT,其中由第一标注级 12 所自动确定的标注信息 MI 的表项被第一标注级 12 录入。为了所识别文本 ET 中命令的自动标注,第一标注级 12 将被包含在命令表 BT

中的字序列与被包含在所识别文本 ET 中的字序列进行比较。当发现被包含在命令表 BT 中的字序列处在所识别文本 ET 中时，则第一标注级 12 将标识这个字序列的所识别文本 ET 中的那些文本段 TS、外加相联系 的命令号 BI 作为命令标注信息 CMI 录入在标注表 MT 中。这在下面参考应用实例被更详细地加以说明。

可提及的是：还有可能在每种情况下使口语文本 GT 的适当音频段 AS 被录入，而不是在标注表 MT 中标识所识别文本 ET 的某些文本段 TS。在每种情况下相联系的音频段 AS 和文本段 TS 可以借助于链接信息 LI 被加以确定。

语音识别装置 7 被设置成识别两个字之间的语音暂停（沉默）并且第一标注级 12 被设置成自动标注与带有标注表 MT 中暂停标注信息 PMI 的口语文本 GT 相对应的音频段 AS。

第一标注级 12 被设置成将被存储在 第一标注级 12 的犹豫声音（例如“aah”或“mhh”）与被包含在所识别文本 ET 中的字进行比较，且设置成将包含这种犹豫声音的所识别文本 ET 的文本段 TS 自动标注有犹豫声音标注信息 ZMI。

第一标注级 12 还被设置成将口语文本 GT 的音频段 AS 自动标注有噪音标注信息 GMI，如果这些音频段 AS 包含噪音和/或声音。出于这个目的，标注级 12 包含能够将噪音和/或声音区分于包含字的音频段的噪音探测器。

为了将所识别文本 ET 的被重复字标注有重复标注信息 RMI，第一标注级 12 被设置成比较所识别文本 ET 中彼此跟随的字或字序列。标注表 MT 还包含日期标注信息 DMI，其在下面参考应用实例被更详细地加以说明。

转录设备 1 的第一校正设备 3 包括再现装置 13，所述再现装置 13 连同存储器装置 8 和命令存储器级 10 由运行文本处理软件的第二计算机所形成。还被连接到第二计算机上的是监视器 14、键盘 15 和扬声器 16，其同样与第一校正设备 3 相联系。当同步再现模式在第一校正设备 3 中被启动时，再现装置 13 被设置成用于口语文本 GT 的声学再现且用于由链接信息 LI 所标记的相联系的所识别文本的同步视觉或光学标注。

在启动的同步再现模式中校正所识别文本 ET 再次被公开在 US 专利 5, 031, 113 中, 并且已经被证实在实践中非常有利。在这种情况下, 校正者, 即正在校正所识别文本 ET 的第一校正设备 3 的使用者, 可以同时聆听由作者对着话筒 5 所说的文本 GT, 并且检查或编辑由语音识别装置 7 为其所识别的文本 ET。被识别的文本 ET 从光学上被显示在监视器 14 上且由语音识别装置 7 针对刚刚从声学上被再现的口语字而被识别的字, 从光学上被再现装置 13 标注且相应地被显示在监视器 14 上。校正者可以启动、中断且停用同步再现模式并且利用键盘 15 编辑所识别的文本 ET。

第二校正设备 4 的结构基本上与图 1 中详细示出的第一校正设备 3 的结构相同, 为此第二校正设备 4 在图 1 中仅被示为一方框。然而, 第二校正设备 4 从物理上与语音识别设备 2 分开, 为此第二校正设备 4 此外具有存储器装置和命令存储器装置, 其中在利用第一校正设备 3 所校正的识别文本 ET 被编辑之前, 被存储在存储器装置 8 和命令存储器级 10 中的信息被存储。

第二校正设备 4 例如可被检查者, 即第二校正设备 4 的使用者使用, 在被存储在存储器装置 8 中的所识别文本 ET 已经由校正者加以校正之后, 所述检查者检查由校正者所完成工作的质量。出于这个目的检查者检查是否校正者已经错过所识别文本 ET 中的任何错误。这样的检查者主要由转录公司采用, 其通过随机检查被校正的所识别文本来确保被转录文本的质量。参考转录设备 1 的应用实例, 这被加以详细说明。

转录设备 1 的再现装置 13 进一步包括第二标注级 17, 所述第二标注级 17 连同键盘 15 和第一标注级 12 形成用于自动和手动标注部分口语文本 GT 或所识别文本 ET 的标注装置。对于第二标注级 17, 当校正者校正所识别文本 ET 时, 其有机会来手动地标注未被自动加以标注的文本的另外部分。

部分文本的这个手动标注可一方面被用来手动地标注具有与已经被自动标注的文本部分相同特征的部分所识别文本, 以允许参数信息 PI 或被存储在命令表 BT 中的信息借助于适配级 11 被加以适配。这个适配取得: 下次第一标注级 12 将能够自动标注已经被手动标注的文本

部分。此外，伴随着每个适配，语音识别设备 2 的识别率得到改善。结果是获得这样的优点，即转录设备可以减少随着每个进一步的使用校正者不得不做的越来越多的工作。

部分文本的这个手动标注另一方面被用来将部分文本标注成被删除，其随后将不再出现在被传送到作者的所识别文本中但尽管如此并没有被完全删除。将一部分文本标注为被删除具有这样的优点，即在较后阶段若需要，则这样的文本部分可以被重新包括在所识别文本中，并且检查者可以检查这些部分文本被删除是否是正确的。然而，特别有利地是，由口语文本 GT 和所识别文本 ET 之间的链接信息 LI 所提供的校正完全保持不变，因为实际上没有所识别文本部分被删除。因此，当同步再现模式被启动时，虽然被标注为被删除的部分口语文本 GT 从声学上被再现，但是替代文本而不是被删除的文本部分可以从光学上被标注，这在下面被详细加以说明。

利用手动标注具有相同特征的文本部分对自动标注部分文本的这个补充提供了进一步的优点，即经编辑的所识别文本 ET 可以进一步以特别高效的方法被编辑。因此，例如，被标注有日期的所有所识别文本部分可以特别高效地被均匀地修正，下面对此加以详细说明。

标注装置的第二标注级 17 也被设置成用于在存储器装置 8 中存储再现标注信息 WMI，所述再现标注信息 WMI 标记在启动的同步再现模式下被再现装置 13 已经至少再现一次的那些所识别文本 ET 部分和/或口语文本 GT 部分。

结果是，获得这样的优点，即通过评估被包含在标注表 MT 中的再现标注信息 WMI，再现装置 13 既能够从视觉上标注借助于同步再现模式已经被校正一次的所识别文本部分又能够从声学上标注相联系的口语文本 GT 部分。结果是根据本发明的校正设备的使用者能够显著更高效地校正所识别文本，其参考应用实例在下面被加以详细说明。

再现装置 13 进一步被设置成当同步再现模式被启动时用来抑制所不希望的口语文本 GT 部分的声学再现，所述所不希望的部分由被存储在存储器装置 8 中的抑制信息进行标记。这是这样的情况，即校正者可以使用键盘 15 来设定被包含在标注表 MT 中的标注信息 MI 的哪些项即将被用作抑制信息。使用者可，例如选择暂停标注信息 PMI 和犹豫

声音标注信息 HMI 作为抑制信息，这样当口语文本 GT 第一次被回放时如此被标注的文本部分得到抑制。参考应用实例这被详细加以说明。

现在将在下面参考图 3 至 9 所示的应用实例详细解释转录设备 1 的优点。图 5 示出由作者对着话筒 5 所说的五部分文本 GT。图 6 示出由语音识别装置 7 针对所述五部分口述文本 GT 所识别的文本 ET，其中部分所识别文本 ET 已经被第一标注级 12 自动加以标注。同样在图 6 中所示的还有借助于第二标注级 17 由校正者手动标注的文本部分。图 7 示出以如此形式的所识别文本 ET，其中被自动和手动标注的文本部分被示出由替代文本来代替。图 8 示出以其中所有被标注的文本部分已经得到抑制的形式的所识别文本 ET，所识别文本以所述形式被传送到作者。

在第一实例中，作者想要对着话筒说文本 “...company PHILIPS” 且同时将字 “PHILIPS” 标注为黑体。然而，在字 “company” 之后，作者简单地沉思到针对黑体格式化命令确切地用什么措词并且当他如此做时他发出犹豫声音 “aah”，正如许多作者一样。然后作者说出 “bold next”，但当他如此做时他想到这个命令的正确字序列是 “next word bold”，为此他说 “no”。然后他说出正确的命令 “next word bold”，并且继续带有字 “PHILIPS” 的文本。

正如从图 6 中可以看出，语音识别装置 7 识别音频段 AS3 的文本段 TS3= “aah” 并且在标注表 MT 的第四行中第一标注级 12 自动将这个犹豫声音标注有犹豫声音标注信息 HMI。替代文本 “<hes>” 而不是犹豫声音被示于图 6 中，其允许校正者看到犹豫声音在此刻已经得到标注。如果当编辑所识别文本 ET 时，校正者将再现装置 13 的文本光标放置在这个替代文本上面，然后由语音识别装置 7 所识别的犹豫声音被显示。结果是，取得这样的优点，即当校正时，校正者可以专心于所识别文本 ET 的实质部分，但是如果他想要了解在此作者发出什么犹豫声音以便于他可以校正相邻于所述犹豫声音的字时，他能够在任何时候看到这个声音。

同样，语音识别装置 7 现在已经为音频段 AS7-AS9 识别出由文本段 TS7-TS9 所形成的字序列 “next word bold”，在命令表 BT 中具有命令号 BI12 的命令已经被分配给所述文本段 TS7-TS9。第一标注级 12

有利地自动录入这些文本段以及在标注表 MT 第四行中作为命令标注信息 CMI 的这个命令号。在图 6 中，替代文本 “<com>” 而不是这些文本段被示出，其结果是获得上述提到的优点。同样，命令 BI12 得到执行，且形成针对音频段 AS10 被识别的文本段 TS10 的字 “PHILIPS” 被显示为黑体。

通过使用键盘 15 和第二标注级 17，校正者现在将文本段 TS4 和 TS5 标注为具有命令号 BI12 的命令，这样下次字序列 “bold next” 将被第一标注级 12 自动识别为命令。在图 7 中这部分文本同样由命令的替代文本 “<com>” 来表示。

通过使用命令号 BI12，第二标注级 17 将这个字序列 TS4+TS5 作为命令标注信息 CMI 录入在标注表 MT 的第五行中。当适应模式在语音识别设备 2 中被启动时，适配级 11 从存储器装置 8 中读取新的命令标注信息 CMI，并且将具有命令号 BI12 的命令进一步录入在命令表 BT 中。

这给出这样的优点，即校正者可以以特别的便利和效率将进一步的字序列定义为命令，在适配模式下所述字序列由语音识别设备 2 接收。当语音识别过程接下来被执行时，字序列 “bold next” 将因此被自动地识别为命令。

借助于键盘 15 和第二标注级 17，现在校正者继续将文本段 TS6 “no” 标注为被删除文本，因为他已经看出作者无心地说出这个字并且它将不被包括在传送给作者的最终文本中。这个被删除文本的替代文本 “<skip>” 被示于图 7 中。

可提及的是：校正者还可以将文本段 TS4-TS6 标注为被删除文本，因为具有命令号 BI12 的命令不应该出现两次。

将文本段 TS6 手动标注为被删除文本给出这样的优点，即甚至在例如利用第二校正设备 4 所执行的后来的同步再现模式期间，由链接信息 LI 所提供的相关性将被完全地保留，并且因此同步再现将没有任何错误地发生。

图 8 示出在第一实例中最终被传送到作者的所识别文本 “...company PHILIPS...”，尽管存在犹豫声音、一个无心的口语字及错误的命令，但是所述文本仍是作者所真正意味的。

在第二实例中，作者想要对着话筒说的文本是“...I fixed the left leg...”，但是在这种情况下在音频段 AS20 之后他沉思，并且保持安静三秒钟，这被第一标注级 12 自动识别为暂停且将其作为暂停标注信息 PMI 录入到标注表 MT 中。这个沉思暂停的替代文本“<sil>”被示于图 6 中。

紧接此暂停之后，作者重复字“T”，其被第一标注级 12 自动识别为被重复字且作为重复标注信息 RMI 被录入在标注表 MT 中。文本段 TS22 的替代文本“<rep>”被示于图 6 中。

紧接此被重复字之后，作者说出“fixed the left”，然后做出另一沉思暂停，并且最终说出“the left leg”。沉思暂停再次被第一标注级 12 自动标注，但是重复字“the left”不可能被自动加以识别和标注。校正者现在手动地将文本段 TS26-TS28 标注为被重复字，因此导致对应的重复标注信息 RMI 被录入在标注表 MT 中。

图 8 示出在第二实例中最终被传送到作者的所识别文本“...the left' leg”，并且尽管存在被重复字和沉思暂停，但是那由此仍是作者所真正意味的。除了自动标注以外所执行的手动标注的结果是：所有被重复的字被标注在所识别文本中，并且随着进一步的处理它们可能例如全部在显示器上被抑制或被用于适配模式。

在第三实例中，作者对着话筒 5 想要说的文本是“...and company PHILIPS will...”，但是在字“and”之后他不得不打喷嚏。因此音频段 AS51 包含当作者打喷嚏时所发卹的噪音。语音识别装置 7 无法识别这个音频段 AS51 的字，并且因为这第一标注级 12 自动地将这个音频段 AS51 标注为具有噪音标注信息 GMI 的噪音并将它如此录入到标注表 MT 中。替代文本“<non sp>”被示于图 6 中。

紧接他的打喷嚏之后，作者花费五秒才找到他的手帕且然后擤鼻子。音频段 AS52 反过来被自动地被标注为暂停且将被如此录入到标注表 MT 中。当作者擤鼻子时发出的噪音类似于噪音“tata”，并且为此语音识别装置 7 不正确地将字“that”识别为文本段 TS53。

有利地，当同步再现模式被启动时立即识别出这个错误的校正者能够手动地将文本段 TS53 标注为带有噪音标注信息 GMI 的噪音。由此获得这样的优点，即在第三实例中在所识别文本 ET 中的全部噪音事实

上被如此标注且在处理的进一步阶段期间可以用相同的方法被全部加以处理。第一标注级 12 的噪音探测器 12 可例如被适配于这些被标注的文本部分，以便于这样的噪音可以在将来被自动地加以识别。

在第四实例中，当作者实际上正在口述的同时，其希望向所识别文本 ET 中插入在命令存储器级 10 中以“module 1”的名义作为文本模块被存储的标准文本。出于这个目的作者口述“...is the best. Insert test module one. All...”。被识别的文本段 TS73-TS76 被识别为带有命令号 BI13 的命令且替代文本“<module 1>”被示于图 6 中。

这给出这样的优点，即文本模块已经以特别简单的方法被自动插入在所识别文本 ET 中。在这种情况下校正者或检查者可以有利地在三种显示类型之中加以选择。他可看到被实际识别的文本-文本段 TS73-TS76，或替代文本，或在图 8 中所看到的从命令存储器级 10 被添加的标准文本。

在第五实例中，作者对着话筒 5 所说的文本是“...tenth of October two thousand and one...”。本应该已经被第一标注级 12 自动识别且标注为日期的字序列本应该是“October tenth two thousand and one”。然而，所说的字序列没有被识别为日期，并且为此校正者将文本段 TS80-TS86 标注为带有日期标注信息量 DMI 的日期。

这给出这样的优点，即在所识别文本中已经被自动或手动标注为日期的所有文本部分可以在随后的处理操作中以特别简便和统一的方法被改变。为此，校正者可例如选择：被标注有日期标注信息 DMI 的全部日期即将以格式“MM.DD.YYYY”被显示。

现在将参考图 9 解释第一校正设备 3 的应用实例。在这个实例中假设校正者启动同步再现模式，基于此口语文本 GT 和所识别文本 ET 分别从音频段 AS1 和文本段 TS1 开始被第一次同步地再现。从符号上这由箭头 P1 来表示。校正者以如此方法配置再现装置 13，以便于在第一再现期间被标注有暂停标注信息 PMI、犹豫声音标注信息 HMI 和噪音标注信息 GMI 的那些文本部分的音频段从声学上没有被再现。这允许校正者特别快速地再现口语文本 GT 并且与此同时更多地校正所识别文本 ET。

在再现期间，标注表 MT 中的再现标注信息 WMI 被连续地更新。音

频段 AS53 (作者擤鼻子) 刚一从声学上被再现且文本段 TS53 刚一从视频上被标注时, 校正者就看出字 “that” 还一直没有被正确地加以识别。在此时文本光标的位置在图 9 中从符号上被指示为 P2。然而, 由于校正者并不肯定什么本应该真正地被识别, 所以他再次启动从音频段 AS50 开始的同步再现模式。在图 9 中这从符号上由箭头 P3 指示。

现在再现装置 13 从被录入在标注表 MT 第四行中的再现标注信息 WMI 中识别出音频段 AS1 至 AS53 已经在同步再现模式中被再次再现且因此从声学上再现所有的音频段 AS50 至 AS53。在图 9 中这从符号上被指示为箭头 P4。它仅从音频段 S54 的再现开始-见箭头 P5-即被标注的文本 (PMI, GMI, HMI) 将在声学再现期间再次被抑制。

这给出这样的优点, 即在其中校正者需要全部标注信息可用以使它能够正确地校正所识别文本 ET 的口述中的那些通路中, 口述的全部音频段 AS 被再现。在校正者可以通过聆听它们仅一次而进行校正的其它通路中, 不必要的音频段 AS 被加以抑制。

类似地, 当所识别文本 ET 被第一次再现时, 被标注的文本部分的替代文本可有利地被加以显示, 并且仅当再现被重复时使得自动切换到实际上被加以识别的文本 ET。

在同步再现模式中已经被再次再现的文本部分的自动标注产生进一步的显著优点。检查者的工作正是要通过随机取样来确定校正者对所识别文本 ET 加以校正做得如何。在这方面再现标注信息 WMI 对检查者非常有帮助。这是因为检查者可以立即识别到借助于同步再现模式哪些文本部分已经被检查并且哪些文本部分已经被跳过且因此有可能根本未被检查。检查者因此可以对没有被标注有再现标注信息 WMI 的文本部分进行特殊检查, 来看是否在所识别文本 ET 中存在任何错误。

如果校正者在工作时被中断在随后时刻想要继续进行, 再现标注信息 WMI 还是有利的。由再现标注信息 WMI 所标记的文本的声学标注 (例如作为口语文本背景的连续音调) 或视频标注 (例如以相反字符 (reverse characters) 所示的所识别文本) 的结果是: 校正者可以立即进行其工作。

第一校正设备 3 还具有定位装置 18, 其被加以提供以便于若同步再现模式被中断以校正被非正确识别的字时, 其允许标注文本输入位

置的文本光标被定位,所述定位装置 18 被设置成将文本光标定位在当同步再现模式被中断时刻所识别文本 ET 中被加以标注的字的上游的 N 个字处,如果所识别文本的这个部分已经被标注有再现标注信息 WMI,以及定位装置 18 被设置成将文本光标定位于当同步再现模式被中断时刻在所识别文本 ET 中被加以标注的字的上游的 M 个字处,如果所识别文本 ET 的这个部分还一直未被标注有再现标注信息 WMI。

所定义的数量可例如是 $M=3$ 及 $N=1$,其结果是当在所识别文本中第一次发现被非正确识别的字时则允许校正者有较长的沉思时间。这些安排是尤其有利的,因为一旦同步再现模式已经被中断则文本光标通常已经被定位在被非正确地识别的字上,并且手动定位文本光标所花的时间可以被节省。对本领域那些普通技术人员很显然地是:这将做出 M 和 N 有利值的许多其它组合。

如果定位装置 18 被设置成自教式则将尤为有利。在这种情况下在同步再现模式已经被中断之后定位装置 18 将从校正者所做出的定位录入中确定 N 和 M 和最佳值(即文本光标向前两个字或向后五个字)并且将连续地将它们适配于使用者的沉思时间。

适配级 11 还可被用来适配被存储在参数存储器装置 9 内的参数信息 PI。这给出这样的优点,即语音识别装置 7 的识别率被稳定地加以改善且较少错误被包含在所识别文本 ET 中。

正如从上述应用实例中可以看到,单独的音频段或文本段可以由标注信息 MI 的一个或甚至多个项进行标注。这允许所识别文本以有利的方式在所谓的等级下被加以处理。这是这样的情况,即例如被包含在命令中的犹豫声音可以与所识别文本 ET 中被标注为犹豫声音的所有其它部分相同的方法被编辑(即被抑制、被删除、替代文本被显示、所识别文本被显示)。这也并不干涉包含犹豫声音的命令的编辑。如果部分文本被标注有多项标注信息 MI,则因此有可能使这些等级被从光学上加以显示。

可提及的是:根据本发明的自动和手动的文本标注可利用标注信息 MI 的多个不同项被完成。这种类型的标注信息 MI 项可例如是:AutoPunctuation(自动标点), Silence(沉默), Nonspeech(没有语音), Noise(噪音), Music(音乐), Spelling(拼写), Hesitation

(暂停) , Insertion (插入) , NumberFormating (号格式化) , DateFormating (日期格式化) , HeaderFormating (头部格式化) , EnumerationFormating (枚举格式化) , QuantityFormating (数量格式化) , SelfCorrection (自校正) , PhraseRepetition (短语重复) , Stutter (结巴) , Discursiveness (离题的) , SpellingSequence (拼写序列) , RedundantPhrase (冗余短语) , NotUnderstood (不理解) , Remark (注释) , Deletion (删除) , Command (命令) 。

可提及的是：标注部分文本并不是如上面所解释必须借助于表格来完成，而是可作为选择地以树结构的形式发生。在这种情况下反过来包含进一步被标注的文本段的被标注文本段可以类似于树的形式被显示在分支上。

可提及的是：自动标注可甚至在语音识别过程正在被执行的同时发生而不是在其之后发生。

可提及的是：例如，仅由语音识别装置 7 识别起来困难（即，可能在识别中有大量错误）的所有文本部分可以被第一标注级 12 自动地标注，并且在同步再现模式中被校正者识别为有难度的所有文本部分可以被手动地标注，以为了用于检查目的警告检查者或口语文本的作者注意这些被标注的文本部分。这是传递与所识别文本有关信息的尤为有利的方法。

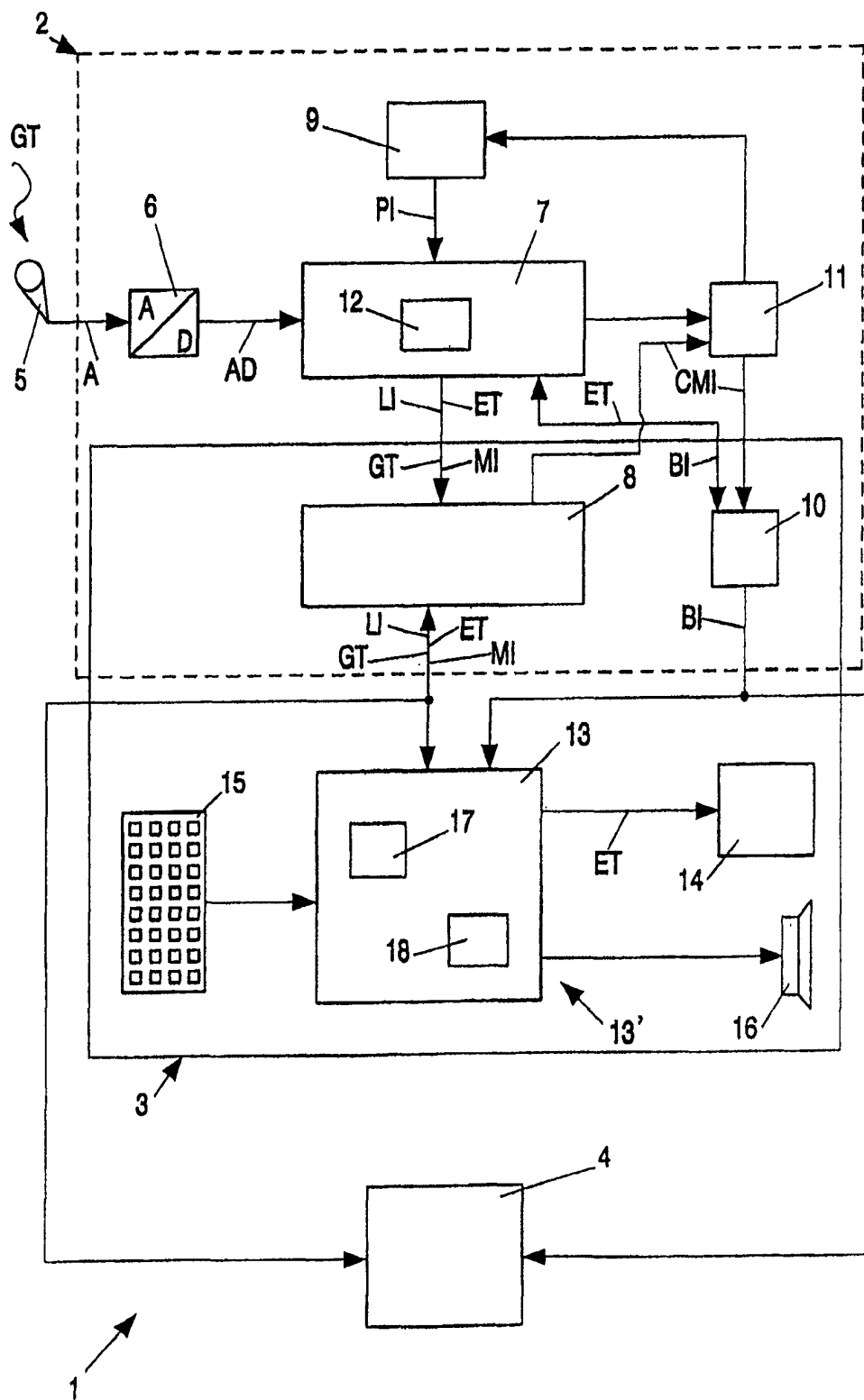


图 1

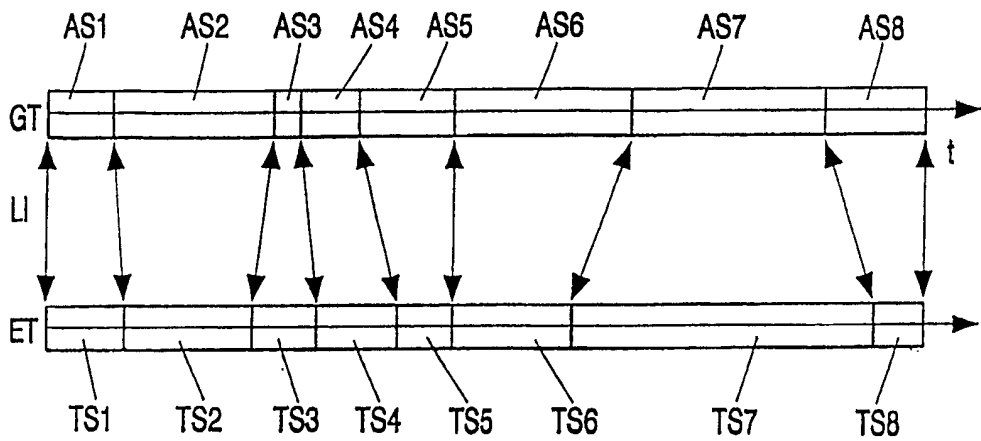


图 2

BI	ET	命令	文本模块
...	...	...	...
BI12	next word bold	bold	...
BI13	insert text module 1	insert memory position 2341	The patient is.... not now.
...	...	...	...

BT

图 3

CMI		PMI	HMI	GMI	DMI	RMI	WMI
ET	BI						
...	...	...	...				...
TS7-TS9	BI12	AS21	TS3	AS51	TS80-TS86	TS22	AS1-AS53
TS4-TS5	BI12	AS26		TS53		TS26-TS28	AS50-AS500
TS73-76	BI13	AS52					
...	...	...	...				...

图 4

MT

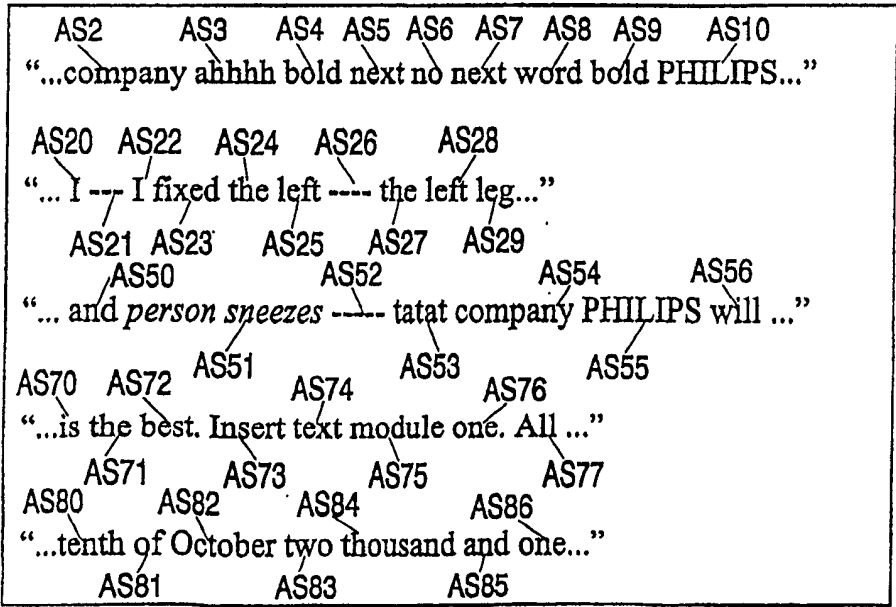


图 5

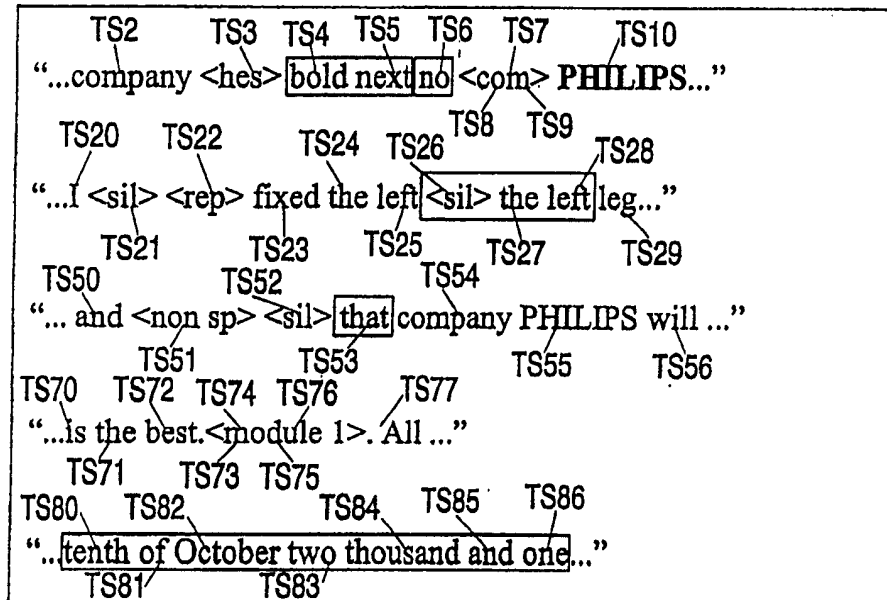


图 6

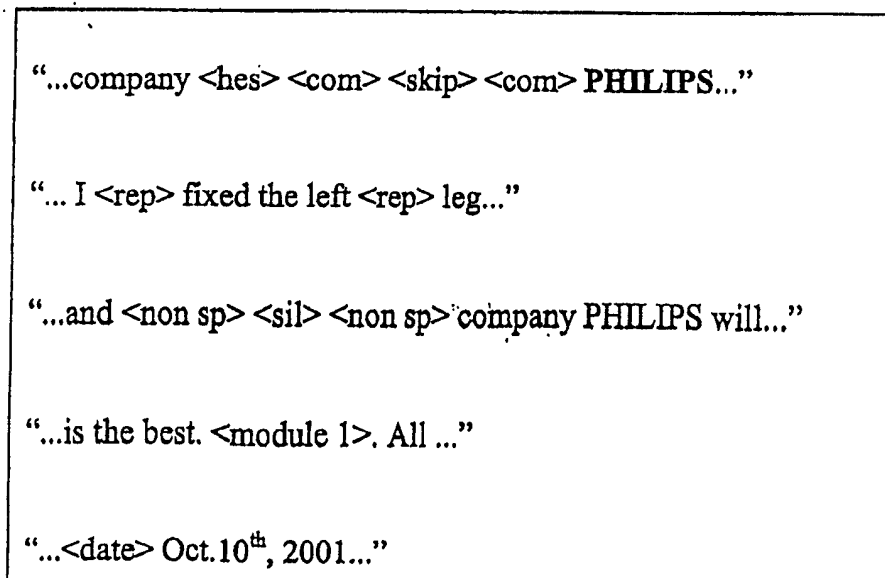


图 7

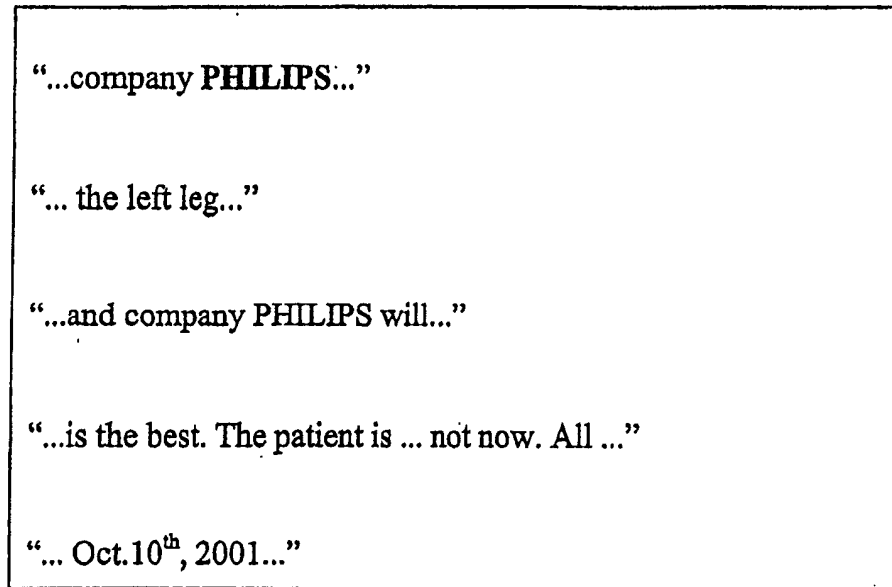


图 8

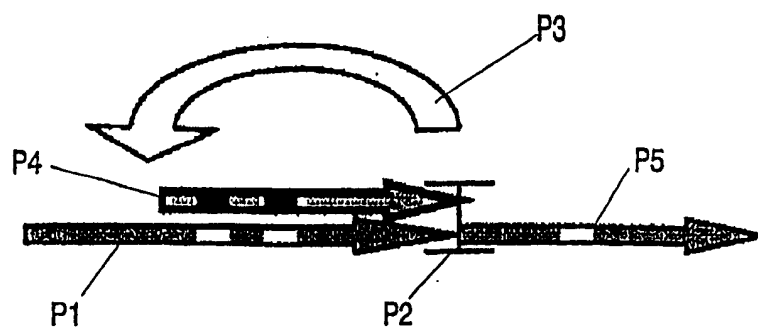


图 9