

生成歌词、显示歌词的方法、装置、电子设备及存储介质

本申请要求于 2018 年 05 月 25 日提交的申请号为 201810513535.7、发明名称为“生成歌词、显示歌词的方法、装置、电子设备及存储介质”的中国专利申请优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及互联网技术领域，特别涉及一种生成歌词、显示歌词的方法、装置、电子设备及存储介质。

背景技术

随着互联网技术的发展，许多音乐播放器不仅支持海量歌曲的在线播放，还可以为用户提供 K 歌服务，K 歌服务是指由音乐播放器播放歌曲的音乐伴奏，并在当前播放界面上显示歌词，使得用户可以边看歌词边随着音乐伴奏唱歌。

目前，在当前播放界面上，终端通常以卡拉 OK 的形式，显示当前播放时刻对应的一句歌词，并通过字体颜色动态标注当前正播放到该句歌词中的哪个字符。在实际演唱歌曲时，为了更好的抒发歌曲表达的感情，原唱歌手通常将歌词内某些字符的原始读音更改为更为文艺的字符的读音进行演唱。而用户在不熟悉该歌曲的原唱时，往往会演唱该字符的原始读音，而与该歌曲应有的原唱不相符，导致用户演唱出错。因此，目前亟需一种显示歌词、生成歌词的方法，以保证用户可以准确演唱每个字符的原唱读音。

发明内容

本发明实施例提供了一种生成歌词、显示歌词的方法、装置、电子设备及存储介质，可以解决相关技术中用户唱错歌词中字符读音的问题。所述技术方案如下：

第一方面，提供了一种生成歌词的方法，所述方法包括：

获取目标歌曲的歌词和所述目标歌曲的音频文件；

根据所述歌词内多个字符中的待标注字符和所述音频文件中所述多个字符对应的音频片段，确定所述待标注字符的候选读音；

若所述待标注字符有一个候选读音时，则所述候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

若所述待标注字符有至少两个候选读音时，则与所述待标注字符的语义匹配的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件。

第二方面，提供了一种显示歌词的方法，所述方法应用在终端上，所述方法包括：

当接收到歌词显示指令时，获取目标歌曲的第一歌词文件，所述歌词显示指令用于显示所述目标歌曲的歌词；

从所述第一歌词文件中获取所述目标歌曲的歌词、所述歌词的多个字符中待标注字符在所述目标歌曲中对应的读音；

显示所述歌词的多个字符；

在显示所述待标注字符时，在所述待标注字符的目标位置上，标注所述待标注字符在所述目标歌曲中对应的读音，所述目标位置为所述待标注字符的上方。

第三方面，提供了一种生成歌词的装置，所述装置包括：

获取模块，用于获取目标歌曲的歌词和所述目标歌曲的音频文件；

确定模块，用于根据所述歌词内多个字符中的待标注字符和所述音频文件中所述多个字符对应的音频片段，确定所述待标注字符的候选读音；

所述确定模块，还用于若所述待标注字符有一个候选读音时，则所述候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

所述确定模块，还用于若所述待标注字符有至少两个候选读音时，则与所述待标注字符的语义匹配的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

生成模块，用于根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件。

第四方面，提供了一种显示歌词的装置，所述装置应用在终端上，所述装置包括：

获取模块，用于当接收到歌词显示指令时，获取目标歌曲的第一歌词文件，所述歌词显示指令用于显示所述目标歌曲的歌词；

所述获取模块，还用于从所述第一歌词文件中获取所述目标歌曲的歌词，所述歌词的多个字符中待标注字符在所述目标歌曲中对应的读音；

显示模块，用于显示所述歌词的多个字符；

标注模块，用于在显示所述待标注字符时，在所述待标注字符的目标位置上，标注所述待标注字符在所述目标歌曲中对应的读音，所述目标位置为所述待标注字符的上方。

第五方面，提供了一种电子设备，所述电子设备包括处理器和存储器，所述存储器中存储有至少一条指令，所述指令由所述处理器加载并执行以实现如第一方面所述的生成歌词的方法，或者如第二方面所述的显示歌词的方法所执行的操作。

第六方面，提供了一种计算机可读存储介质，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现如第一方面所述的生成歌词的方法，或者如第二方面所述的显示歌词的方法所执行的操作。

本发明实施例提供的技术方案带来的有益效果是：

本发明实施例中，服务器可以基于目标歌曲的歌词和目标歌曲的音频文件，确定歌词中每个待标注字符在该目标歌曲中对应的读音；并进一步根据多个字符和待标注字符的读音，生成目标歌曲的第一歌词文件，从而为每个待标注字符绑定了对应的读音，使得后续在显示歌词时，可以同步显示读音，保证用户可以基于正确演唱目标歌曲的每个字符的读音。并且，终端在显示歌词时，还可以在将读音标注在对应待标注字符的上方，使得该读音清晰可见，用户可以准确、快速的找到待标注字符对应读音，提高了显示歌词的准确性。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 是本发明实施例提供的一种实施环境的示意图；

图 2 是本发明实施例提供的一种生成歌词的方法的流程图；

图 3 是本发明实施例提供的一种显示歌词的方法的流程图；

图 4 是本发明实施例提供的一种歌词显示的界面示意图；

图 5 是本发明实施例提供的一种歌词显示的界面示意图；

图 6 是本发明实施例提供的一种生成歌词的装置的结构示意图；

图 7 是本发明实施例提供的一种显示歌词的装置的结构示意图；

图 8 是本发明实施例提供的一种终端的结构示意图；

图 9 是本发明实施例提供的一种服务器的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

图 1 是本发明实施例提供的一种实施环境的示意图，该实施环境包括：终端 101 和服务器 102。其中，该终端 101 和服务器 102 之间通过网络连接。该终端 101 上安装了应用程序，该终端 101 可以基于该应用程序从服务器 102 中获取目标歌曲，并进行目标歌曲的播放、歌词显示等。

其中，该服务器 102 可以事先基于目标歌曲的歌词内的多个待标注字符，以及该目标歌曲的音频文件，确定出该多个字符中待标注字符的读音，从而得到待标注字符在该目标歌曲中对应的读音，该服务器可以存储该歌词和该待标注字符在目标歌曲中对应的读音，并将该目标歌曲的歌词和待标注字符对应的读音发送至终端 101。该终端 101 在播放该目标歌曲时，同步显示该目标歌曲的歌词，并在待标注字符的上方标注待标注字符对应的读音。

其中，上述就多个字符获取待标注字符的读音的过程，也可以由终端 101 执行，也即是，终端 101 从服务器 102 中获取目标歌曲的歌词后，自行对歌词

内待标注字符确定读音并标注。

需要说明的是，在一些歌曲中，在实际演唱歌曲时，为了更好的抒发歌曲表达的感情，通常将歌词内某些字符的原始读音更改为更为文艺的读音。例如，在一些日文歌曲中，“未来”常常会被演唱为“あす”，“あす”对应的汉字为“明日”，而不是未来的正确读音。为了歌曲艺术表达的需要，将“未来”演唱为“明日”的读音“あす”。

本申请中，将待标注字符在该目标歌曲在原唱时的读音，作为正确演唱的读音，写入该目标歌曲的歌词文件中进行存储，以便于后续基于该歌词文件显示该待标注字符在原唱中的读音。

其中，该应用程序可以为音乐播放器或者安装有音乐播放插件的应用程序，该终端可以为手机终端、PAD（Portable Android Device，平板电脑）终端或者电脑终端等。服务器 102 是该应用程序的后台服务器。服务器 102 可以为一台服务器，或者由若干台服务器组成的服务器集群，或者是一个云计算服务器中心。

图 2 是本发明实施例提供的一种生成歌词的方法的流程图。该发明实施例的执行主体为终端，参见图 2，该方法包括：

201、服务器获取目标歌曲的歌词和该目标歌曲的音频文件。

其中，该音频文件包括该歌词内多个字符对应的音频片段。该目标歌曲为任一需要生成歌词的歌曲，一般为歌词中包括汉字的歌曲，该汉字包括但不限于：简化汉字、繁体字等。该多个字符可以为目标语言中的文字符号，例如，日文中的汉字、振假名等。本发明实施例中，服务器可以基于该目标歌曲在原唱时的读音，确定歌词内各个待标注字符的读音，从而准确的得到与目标歌曲的原唱相符的读音。

本步骤中，服务器可以先获取该目标歌曲的第二歌词文件和原始音频文件，从该第二歌词文件中获取该目标歌曲的歌词，从该原始音频文件中提取该多个字符对应的音频片段，根据该多个字符对应的音频片段，生成该音频文件。

其中，该音频片段可以为入声频段的音频片段。本步骤中，服务器可以通过消除伴奏的方式，得到音频文件。服务器获取目标歌曲的音频文件的步骤可以为：服务器获取该目标歌曲的原始音频文件，该原始音频文件包括在采集的各个频段的音频片段，服务器从该原始音频文件中提取入声频段的音频片段，并根据该入声频段的音频片段，生成该目标歌曲的音频文件，从而将该目标歌

曲的伴奏消除。一般的，服务器从该原始音频文件中提取该目标歌曲在空间正前方的音频片段，从该音频片段中选择人声频段的音频片段。一般的，人声频段为中频、高频。

在一些歌曲的歌词文件中，往往还包括多个字符的显示时间，该显示时间用于指示每个字符在该歌曲播放过程中何时进行显示。并且，在歌词文件中，每个字符和该字符的显示时间是拼接存储的，字符与字符之间往往还存在显示时间，例如，歌词“雨の空港”在歌词文件中实际存储形式为“(39112)雨(39803)の(40356)空(40606)港(41176)”，其中括号内数字表示字符的显示时间，字符与字符之间均被显示时间切断；服务器可以获取该目标歌曲的第二歌词文件，从该第二歌词文件中提取该目标歌曲的歌词以及该歌词内多个字符的显示时间，并建立每个字符与每个字符的显示时间之间的索引关系，以保证后续显示时，还可以基于显示时间进行显示。其中，该第二歌词文件为该目标歌曲的原始歌词文件，该第二歌词文件包括该目标歌曲的歌词和该歌词的多个字符的显示时间。

其中，服务器可以以数组的形式分别存储字符与显示时间，以使字符与显示时间可以一一对应，在一种可能的设计中，服务器获取目标歌曲的歌词的步骤可以为：服务器获取该目标歌曲的第二歌词文件，建立第一数组和第三数组，该第一数组用于存储该歌词内的多个字符，该第三数组用于存储该多个字符的显示时间。服务器将该多个字符写入该第一数组，将该多个字符的显示时间写入该第三数组，将该第一数组内的多个字符确定为该目标歌曲的歌词。其中，在第三数组中，服务器可以按照第一数组中每个字符存储顺序，将每个字符的显示时间存储在第三数组中，基于存储顺序，使得第一数组中每个字符与第三数组中该字符的显示时间的一一对应，从而提高了歌词以及显示时间的准确性。

需要说明的是，服务器通过提取字符，以使后续可以基于字符进行读音的确定，同时，服务器通过建立每个字符和每个字符的显示时间的索引关系，在不破坏该第二歌词文件的原有内容的前提下，进行歌词的提取，保证了最终可以得到准确的歌词文件。

202、服务器将该多个字符中的汉字确定为该待标注字符。

本发明实施例中，服务器可以通过预设识别算法，识别出该多个字符中的汉字，将该汉字确定为该待标注字符。该汉字包括但不限于：简化汉字、繁体字等。其中，服务器一般存储该多个字符在编码后的编码数据，该预设识别算

法可以为基于该多个字符的编码方式的编码表，对该多个字符进行识别。当然，本发明实施例并不限定字符的编码方式，例如，该编码表可以为 Unicode（统一码）编码方式的编码表。

需要说明的是，服务器通过对该多个字符先进行识别，以识别出待标注字符，从而直接基于多个字符中待标注字符进行读音的确定，缩减了后续所要处理的字符的数量，进一步提高了确定读音的效率。并且，本发明实施例通过将汉字作为待标注字符，后续标注歌词中的汉字的读音，对于母语不是中文或者日文的用户，方便了用户通过唱歌的方式来学习一门外语，既满足了听觉乐趣，又达到学习外语的效果，极大的丰富了用户体验，从而大大提高了本方法的实用价值。

203、服务器根据该多个字符对应的音频片段，识别该多个字符的读音，根据该多个字符中的待标注字符，确定该待标注字符的候选读音。

在本发明实施例中，服务器可以基于该多个字符对应的音频片段，通过预设语音识别算法，识别出该多个读音，并根据多个字符中待标注字符和除待标注字符以外的其他字符之间的相对位置，确定出该待标注字符的候选读音。

然而，该音频文件内多个字符对应的音频片段中可能存在噪声，或者，在实际演唱时，某个字符的音节可能会被拉长，从而导致语音识别时，该多个字符的读音中，可能包括一些模糊的读音，服务器可以暂时用中间读音表示该未能唯一确定出的读音，因此，本步骤包括以下两种情况。

第一种情况、若基于音频片段识别的多个字符的读音中，不包括模糊的读音时，也即是，各个读音均是唯一确定时，服务器出根据该多个字符中的待标注字符，从该多个字符的读音中确定该待标注字符的候选读音。

其中，服务器根据该多个字符中除待标注字符以外的其他字符在该多个读音中的位置，确定该待标注字符在该多个读音中对应的候选读音。当然，当该待标注字符的数目为多个时，服务器还可以根据与该待标注字符相邻的其他字符在该多个读音中的位置，确定该待标注字符在多个读音中对应的候选读音。

需要说明的是，本步骤中，服务器所确定出的待标注字符对应的候选读音的数目仅有一个。后续可以通过以下步骤 204，确定出该待标注字符在目标歌曲中对应的读音。

第二种情况、若基于音频片段识别的多个字符的读音中，包括模糊的读音时，服务器根据该多个字符中的待标注字符，从该多个字符的读音中确定该待

标注字符的中间读音，根据该中间读音，识别出该待标注字符的至少两个候选读音。

其中，该中间读音可以为罗马音，服务器可以基于该罗马音，从词典引擎中匹配出待标注字符对应的至少两个候选读音。该至少两个候选读音为用于表示日文中汉字读音的振假名。例如，以日文为例，对于歌词中“ペガサス幻想 そうさ夢だけは”对应的音频片段，语音识别出的多个读音可以为“ペガサス fantaji そうさゆめだけは”，其中，“fantaji”为罗马音，服务器可以基于该罗马音“fantaji”，确定待标注字符对应的3个候选读音分别为：“ふあんたじ”、“ファンタジ”、“ファンタジー”。

需要说明的是，本步骤中，服务器所确定出的待标注字符对应的候选读音的数目为至少两个。后续可以通过以下步骤205，确定出该待标注字符在目标歌曲中对应的读音。

需要说明的是，上述步骤202-203是“服务器根据该歌词内多个字符中的待标注字符和该音频文件中该多个字符对应的音频片段，确定该待标注字符的候选读音”具体实现方式，实际上，上述步骤是先确定歌词中的待标注字符，再基于音频文件进行语音识别的过程，然而，服务器还可以先基于音频文件进行语音识别，以确定多个字符的读音，再确定歌词中的待标注字符，从而确定候选读音，本发明实施例对此不做具体限定。

204、若该待标注字符有一个候选读音时，则服务器将候选读音确定为该待标注字符在该目标歌曲中对应的读音。

当服务器识别出待标注字符对应的候选读音是唯一确定时，对于每个待标注字符，服务器可以直接将该待标注字符对应的候选读音确定为该待标注字符在目标歌曲中对应的读音。

以日文为例，服务器可以用振假名表示该多个字符的读音。基于歌词中多个字符“誰でももし男女が人間で ありたいなら”，以及该多个字符对应的音频片段所识别出的多个读音“だれでももしひとがひとでありたいなら”，确定该多个字符中“誰”、“男女”、“人間”的读音的过程可以为：

a、将该多个读音与该多个字符进行匹配，分别确定出“誰でももし男女が人間で ありたいなら”中“でももし”、“が”、“で ありたいなら”在“だれでももしひとがひとでありたいなら”的位置。

b、根据该多个字符“誰”、“男女”、“人間”与已识别出的“でももし”、“が”、

“で ありたいなら”之间的相对位置，分别确定出该“誰”、“男女”、“人間”在该“だれでももしひとがひとでありたいなら”各自对应的读音：

“誰”对应读音为“だれ”；

“男女”对应读音为“ひと”；

“人間”对应读音为“ひと”；

需要说明的是，由于仅需确定日语歌曲中歌词的读音，因此，在基于音频片段进行语音识别时，仅需根据日文中各个振假名的读音进行识别，从而缩小了待识别的读音的范围，在容错率上，仅需基于日文范围内现有的各个振假名进行读音的识别，无需考虑其他语言，从而提高了识别出的多个读音的准确性。并且，服务器还可以基于各个待标注字符以及多个振假名，进一步限定出各个待标注字符在多个读音中对应的语音区间，从而提高了后续确定待标注字符的读音的准确性。

205、若该待标注字符有至少两个候选读音时，则服务器将与该待标注字符的语义匹配的候选读音确定为该待标注字符在该目标歌曲中对应的读音。

本步骤中，服务器可以事先建立并存储词典引擎，基于词典引擎来查询语音相匹配的候选读音。对于每个候选读音，服务器从第一词典引擎中查找该每个候选读音对应的中间字符，从第二词典引擎中查找该中间字符的语义，根据该待标注字符的语义，从该至少两个候选读音中选择中间字符的语义与该待标注字符的语义相匹配的候选读音，将选择的候选读音确定为该待标注字符在该目标歌曲中对应的读音。

其中，该第一词典引擎包括多个候选读音和多个候选读音对应的中间字符，该第二词典引擎包括多个中间字符和多个中间字符对应的多个语义。该待标注字符在该目标歌曲中对应的读音可以用日文中振假名表示。该中间字符可以为某一语言中的文字符号，例如，英文、中文、德文、日文、法文等语言中的文字符号，该第一词典引擎可以为日汉词典、日日词典、日英词典、日德词典、日法词典等，本发明实施例对此不作具体限定。

以日文为例，对于歌词中“ペガサス幻想そうさ夢だけは”对应的多个读音“ペガサス fantaji そうさゆめだけは”，以“幻想”为例进行说明，该确定读音的过程如下：

a、先将歌词内多个字符与该多个读音进行匹配，可知：“幻想”对应的读音为中间读音，罗马音“fantaji”；

b、确定罗马音“fantaji”对应的3个候选读音分别为：“ふあんたじ”、“ファンタジ”、“ファンタジー”；

c、对于每个候选读音，分别在第一词典引擎中查询每个候选读音对应的中间字符，并在第二词典引擎中查询该中间字符的语义，确定出候选读音“ファンタジー”，在日英词典中对应的英文字符为“Fantasy”；

d、从第二词典引擎中查找各个中间字符的语义，将每个中间字符的语义与“幻想”的语义进行语义匹配，确定出候选读音“ファンタジー”对应的英文字符“Fantasy”，在第二词典引擎中对应的语义可以为“幻想”，与日文中“幻想”的语义相同。

e、确定“幻想”对应的读音为“ファンタジー”。

通过上述步骤a-e，利用外文语言作为中间字符，通过查询事先建立的词典引擎，进一步从多个振假名中确定出与“幻想”语义相同的振假名“ファンタジー”。

需要说明的是，服务器利用外文语言作为中间字符，基于模糊不清的罗马音对应的多个可能的候选读音，先将待标注字符的读音过渡为中间字符，并查询中间字符对应的语义，从多个候选读音中匹配出与待标注字符语义最相近的候选读音，从而进一步分析出待标注字符在实际演唱时艺术演变后的读音，大大提高了确定该待标注字符对应的读音的准确性。

206、服务器根据该多个字符和该待标注字符在该目标歌曲中对应的读音，生成该目标歌曲的第一歌词文件。

本步骤中，服务器创建第一歌词文件，将该多个字符和待标注字符在该目标歌曲中对应的读音存储在该第一歌词文件中，并建立待标注字符和待标注字符在该目标歌曲中对应的读音之间的索引关系。其中，服务器可以基于数组建立该字符和读音之间的索引关系，服务器建立第二数组，将该待标注字符在该目标歌曲中对应的读音写入该第二数组；将第一数组和该第二数组添加至第一歌词文件中。其中，对于每个待标注字符，服务器根据该待标注字符在该第一数组中的第一存储位置，将该待标注字符对应的读音存储在该第三数组的第二存储位置中，其中，该第一存储位置和该第三存储位置可以均为第一数组和第二数组中关联字节，例如，均为第一字节。

本发明实施例中，在步骤201中，服务器还可以获取该多个字符的显示时间，本步骤中，该服务器还可以将该显示时间添加到该第一歌词文件中。因此，

服务器根据多个字符和待标注字符在目标歌曲中对应的读音，生成目标歌曲的第一歌词文件的步骤还可以为：服务器创建该第一歌词文件，并建立第二数组，将待标注字符在目标歌曲中对应的读音写入第二数组；服务器将该第一数组、第二数组以及第三数组添加至该第一歌词文件中。

在一种可能的设计中，在一些语言中，可能存在一些字符数目与读音的数目并不相等的情况，例如，日文中的“^{かけ}欠片”包括的两个字符实际对应有三个读音，因此，当歌词内存在目标词时，服务器还可以根据目标词包括的至少两个相邻待标注字符的显示时间，确定目标词的显示时间，并根据该目标词的显示时间，更新该第三数组；最终，服务器将第一数组、第二数组和更新后的第三数组添加至第一歌词文件中；其中，目标词包括至少两个相邻的待标注字符，且包括的待标注字符的数目与目标词的读音的数目不相等。服务器确定该目标词的显示时间的方式为：将目标词包括的至少两个相邻待标注字符的显示时间进行合并，将合并后的显示时间确定为目标词的显示时间。

在日文中，许多词的读音与其包括的字符的数目不相等，例如，欠片、明後日、一昨日、黄昏、彷徨う、躊躇う、煙草、失人等。以欠片为例进行说明，更新第三数组之前，在第二歌词文件中存储的显示时间如下表 1 所示：

表 1

第三数组	39112	39803	40356	41176
第一数组	欠	片	の	色

显然,“欠”和“片”分别各自对应一个显示时间,然而,在实际显示时,由于欠片共同对应三个读音,因此,按照两个显示时间分开显示“欠”和“片”,显然会导致显示错误,通过对两个显示时间进行合并,得到更新后的第三数组中存储的显示时间如下表2所示:

表 2

第三数组	39112		40356	41176
第一数组	欠	片	の	色
第二数组	かけら			いろ

需要说明的是，服务器通过对目标词的内多个相邻字符的显示时间进行合并，使得多个相邻字符可以作为一个词，对应一个单独的显示时间，避免在实际播放歌曲时，基于每个字符对应一个显示时间进行显示，而导致各个字符的显示时间与实际所演唱的读音不匹配的问题，进而使得每个字符的显示时间与

该字符的实际读音相匹配，每个词的显示时间与该词的读音相匹配，使得第一歌词文件中的歌词的显示时间均能准确匹配到歌词的读音，进一步提升了最终所得到的第一歌词文件的准确性。

本发明实施例中，服务器可以基于目标歌曲的歌词和目标歌曲的音频文件，确定歌词中每个待标注字符在该目标歌曲中对应的读音；并进一步根据多个字符和待标注字符的读音，生成目标歌曲的第一歌词文件，从而为每个待标注字符绑定了对应的读音，使得后续在显示歌词时，可以同步显示读音，保证用户可以基于正确演唱目标歌曲的每个字符的读音。

图 3 是本发明实施例提供的一种显示歌词的方法的流程图。该发明实施例的执行主体为终端，参见图 3，该方法包括：

301、当接收到歌词显示指令时，终端获取目标歌曲的第一歌词文件。

本发明实施例中，该歌词显示指令用于显示该目标歌曲的歌词，当终端接收到歌词显示指令时，该终端可以根据目标歌曲的标识，从本地或服务器中获取该目标歌曲的第一歌词文件。其中，该歌词显示指令可以为用户触发终端播放目标歌曲时得到，或者，用户触发显示该歌词文件时得到。

需要说明的是，该第一歌词文件为事先基于歌词和多个字符的目标读音生成的歌词文件，具体生成过程如上述步骤 201-206 所示，该第一歌词文件至少包括该目标歌曲的歌词、该歌词的多个字符中待标注字符在该目标歌曲中对应的读音。

302、终端从该第一歌词文件中获取该目标歌曲的歌词、该歌词的多个字符中待标注字符在该目标歌曲中对应的读音。

本发明实施例中，在该第一歌词文件中，可以数组的形式分别存储歌词和待标注字符的读音。终端从该第一歌词文件中获取第一数组和第二数组，并从该第一数组中读取该歌词内的多个字符，从第二数组中读取该多个字符中待标注字符的读音。其中，该终端可以基于该第一数组中每个待标注字符的第一存储位置，确定该第二数组中关联的第二存储位置，从该第二存储位置中读取该待标注字符的读音。

在一种可能的设计中，该第一歌词文件中还可以包括该歌词中多个字符的显示时间，该终端还可以从第一歌词文件中获取该多个字符的显示时间。该过程可以为：终端从该第一歌词文件中获取第三数组，并从该第三数组中获取该

多个字符的显示时间。该多个字符的显示时间包括该歌词中目标词的显示时间以及除目标词以外的每个字符的显示时间。

其中，该目标词包括至少两个相邻的待标注字符，且包括的待标注字符的数目与该目标词的读音的数目不相等，该目标词的显示时间为该目标词包括的至少两个相邻待标注字符的显示时间合并后的显示时间。

303、终端显示该歌词的多个字符。

本发明实施例中，终端根据该目标词的显示时间，突出显示多个字符中当前播放的目标词；根据该除目标词以外的每个字符的显示时间，突出显示该除目标词以外的字符中当前播放的字符。其中，该终端可以通过字符的字体颜色来突出显示正在播放的字符或目标词。例如，将已经播放的字符或目标词渲染为第一颜色，将还未播放的字符或目标词渲染为第二颜色。

304、在显示该待标注字符时，终端在该待标注字符的目标位置上，标注该待标注字符在该目标歌曲中对应的读音。

本发明实施例中，该目标位置为该待标注字符的上方，终端在显示该待标注字符时，根据该待标注字符的显示时间，突出显示当前播放的待标注字符和该当前播放的待标注字符在该目标歌曲中对应的读音。当然，该待标注字符中可能包括目标词，该终端可以根据目标词的显示时间，突出显示当前播放的目标词和该当前播放的目标词在该目标歌曲中对应的读音。

需要说明的是，在一些现有技术中，为了简化逻辑，对于歌词的独有读音，就使用括号的方式来标注读音，如图 4 所示，例如，誰でももし男女（ひと）が人間（ひと）で ありたいなら，而这样标注，会给用户造成一定的误解，比如会认为“女”单独发ひと的音，男的读音未知，特别是复合词，例如，時間旅行（タイムトリップ），用户浏览后无法分辨出括号内读音是旅行还是時間旅行。同时，在突出显示歌词时，无法保证字符和读音可以按照相同的显示节奏进行突出显示，例如，男女（ひと），男女只有两个音节，对应到一行文字就变成了 6 个（含括号），其中，在基于“男”的显示时间突出显示“男”时，需要将“男女（”渲染为第一颜色，显示“女”时对应的将剩下的部分“ひと）”渲染为第一颜色，从而导致显示错误，给用户带来较差的用户体验。

本发明实施例中，如图 5 所示，终端可以将该读音显示在待标注字符的上方，当用户浏览到该待标注字符时，可以清晰、准确的找到该待标注字符的读音，提高了用户获知待标注字符的读音的效率，并提高了歌词显示的准确性。

同时，用户基于该读音进行演唱，保证演唱就可以准确演唱每个字符读音。

本发明实施例中，当接收到歌词显示指令时，终端可以显示目标歌曲的歌词、该歌词的多个字符中待标注字符在该目标歌曲中对应的读音；从而保证用户演唱时可以准确的演唱每个字符的读音，并且，终端还可以在将读音标注在对应待标注字符的上方，使得该读音清晰可见，用户可以准确、快速的找到待标注字符对应读音，提高了显示歌词的准确性。

图 6 是本发明实施例提供的一种生成歌词的装置的结构示意图。参见图 6，该装置包括：获取模块 601、确定模块 602 和生成模块 603。

获取模块 601，用于获取目标歌曲的歌词和该目标歌曲的音频文件，该音频文件包括该歌词内多个字符对应的音频片段；

确定模块 602，用于根据该多个字符中的待标注字符和该多个字符对应的音频片段，确定该待标注字符的候选读音；

该确定模块 602，还用于若该待标注字符有一个候选读音时，则该候选读音确定为该待标注字符在该目标歌曲中对应的读音；

该确定模块 602，还用于若该待标注字符有至少两个候选读音时，则与该待标注字符的语义匹配的候选读音确定为该待标注字符在该目标歌曲中对应的读音；

生成模块 603，用于根据该多个字符和该待标注字符在该目标歌曲中对应的读音，生成该目标歌曲的第一歌词文件。

可选的，该确定模块 602，包括：

确定单元，用于将该多个字符汉字确定为该待标注字符；

识别单元，用于根据该多个字符对应的音频片段，识别该多个字符的读音；

该确定单元，还用于根据该多个字符中的待标注字符，从该多个字符的读音中确定该待标注字符的候选读音，或者，根据该多个字符中的待标注字符，从该多个字符的读音中确定该待标注字符的中间读音，根据该中间读音，识别出该待标注字符的至少两个候选读音。

可选的，该确定模块 602，包括：

查找单元，用于还用于对于每个候选读音，从第一词典引擎中查找该每个候选读音对应的中间字符，从第二词典引擎中查找该中间字符的语义，该第一词典引擎包括多个候选读音和多个候选读音对应的中间字符，该第二词典引擎

包括多个中间字符和多个中间字符对应的多个语义；

选择单元，用于根据该待标注字符的语义，从该至少两个候选读音中选择中间字符的语义与该待标注字符的语义相匹配的候选读音，将选择的候选读音确定为该待标注字符在该目标歌曲中对应的读音。

可选的，该获取模块 601，用于获取该目标歌曲的第二歌词文件，该第二歌词文件包括该目标歌曲的歌词和该歌词的多个字符的显示时间；建立第一数组和第三数组，将该多个字符写入该第一数组，将该多个字符的显示时间写入该第三数组，将该第一数组内的多个字符确定为该目标歌曲的歌词；获取该目标歌曲的原始音频文件，提取该原始音频文件中人声频段的音频片段，生成该音频文件。

可选的，该生成模块 603，用于建立第二数组，将该待标注字符在该目标歌曲中对应的读音写入该第二数组；当该歌词内存在目标词时，根据该目标词包括的至少两个相邻待标注字符的显示时间，确定该目标词的显示时间，该目标词包括至少两个相邻的待标注字符，且包括的待标注字符的数目与该目标词的读音的数目不相等；将该目标词的显示时间写入该第三数组；将第一数组、该第二数组和该第三数组添加至该第一歌词文件中；

其中，确定该目标词的显示时间的方式为：将该目标词包括的至少两个相邻待标注字符的显示时间进行合并，将合并后的显示时间确定为该目标词的显示时间。

可选的，该生成模块 603，用于建立第二数组，将该待标注字符在该目标歌曲中对应的读音写入该第二数组；将第一数组和该第二数组添加至该第一歌词文件中，该第一数组用于存储该歌词的多个字符。

本发明实施例中，服务器可以基于目标歌曲的歌词和目标歌曲的音频文件，确定歌词中每个待标注字符在该目标歌曲中对应的读音；并进一步根据多个字符和待标注字符的读音，生成目标歌曲的第一歌词文件，从而为每个待标注字符绑定了对应的读音，使得后续在显示歌词时，可以同步显示读音，保证用户可以基于正确演唱目标歌曲的每个字符的读音。

图 7 是本发明实施例提供的一种显示歌词的装置的结构示意图。该装置应用在终端上，参见图 7，该装置包括：获取模块 701、显示模块 702 和标注模块 703。

获取模块 701, 用于当接收到歌词显示指令时, 获取目标歌曲的第一歌词文件, 该歌词显示指令用于显示该目标歌曲的歌词;

该获取模块 701, 还用于从该第一歌词文件中获取该目标歌曲的歌词、该歌词的多个字符中待标注字符在该目标歌曲中对应的读音;

显示模块 702, 用于显示该歌词的多个字符;

标注模块 703, 用于在显示该待标注字符时, 在该待标注字符的目标位置上, 标注该待标注字符在该目标歌曲中对应的读音, 该目标位置为该待标注字符的上方。

可选的, 该显示模块 702, 包括:

第一显示单元, 用于根据该多个字符中目标词的显示时间, 突出显示多个字符中当前播放的目标词;

第二显示单元, 用于根据该多个字符中除目标词以外的每个字符的显示时间, 突出显示该除目标词以外的字符中当前播放的字符;

其中, 该目标词包括至少两个相邻的待标注字符, 且包括的待标注字符的数目与该目标词的读音的数目不相等, 该目标词的显示时间为该目标词包括的至少两个相邻待标注字符的显示时间合并后的显示时间。

可选的, 该第二显示单元, 用于在显示该待标注字符时, 根据该待标注字符的显示时间, 突出显示当前播放的待标注字符和该当前播放的待标注字符在该目标歌曲中对应的读音。

本发明实施例中, 当接收到歌词显示指令时, 终端可以显示目标歌曲的歌词、该歌词的多个字符中待标注字符在该目标歌曲中对应的读音; 从而保证用户演唱时可以准确的演唱每个字符的读音, 并且, 终端还可以在将读音标注在对应待标注字符的上方, 使得该读音清晰可见, 用户可以准确、快速的找到待标注字符对应读音, 提高了显示歌词的准确性。

上述所有可选技术方案, 可以采用任意结合形成本公开的可选实施例, 在此不再一一赘述。

需要说明的是: 上述实施例提供的生成歌词的装置在生成歌词时, 或者, 显示歌词的装置在显示歌词时, 仅以上述各功能模块的划分进行举例说明, 实际应用中, 可以根据需要而将上述功能分配由不同的功能模块完成, 即将设备的内部结构划分成不同的功能模块, 以完成以上描述的全部或者部分功能。另

外，上述实施例提供的生成歌词的装置与生成歌词的方法，以及显示歌词的装置与显示歌词的方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

图 8 是本发明实施例提供的一种终端的结构示意图。该终端 800 可以是：智能手机、平板电脑、MP3 播放器 (Moving Picture Experts Group Audio Layer III, 动态影像专家压缩标准音频层面 3)、MP4 (Moving Picture Experts Group Audio Layer IV, 动态影像专家压缩标准音频层面 4) 播放器、笔记本电脑或台式电脑。终端 800 还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

通常，终端 800 包括有：处理器 801 和存储器 802。

处理器 801 可以包括一个或多个处理核心，比如 4 核心处理器、8 核心处理器等。处理器 801 可以采用 DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field-Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。处理器 801 也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理器，也称 CPU (Central Processing Unit, 中央处理器)；协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中，处理器 801 可以在集成有 GPU (Graphics Processing Unit, 图像处理器)，GPU 用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中，处理器 801 还可以包括 AI (Artificial Intelligence, 人工智能) 处理器，该 AI 处理器用于处理有关机器学习的计算操作。

存储器 802 可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器 802 还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器 802 中的非暂态的计算机可读存储介质用于存储至少一个指令，该至少一个指令用于被处理器 801 所执行以实现本申请中方法实施例提供的生成歌词的方法、显示歌词的方法。

在一些实施例中，终端 800 还可选包括有：外围设备接口 803 和至少一个外围设备。处理器 801、存储器 802 和外围设备接口 803 之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口 803

相连。具体地，外围设备包括：射频电路 804、触摸显示屏 805、摄像头 806、音频电路 807、定位组件 808 和电源 809 中的至少一种。

外围设备接口 803 可被用于将 I/O (Input /Output, 输入/输出) 相关的至少一个外围设备连接到处理器 801 和存储器 802。在一些实施例中，处理器 801、存储器 802 和外围设备接口 803 被集成在同一芯片或电路板上；在一些其他实施例中，处理器 801、存储器 802 和外围设备接口 803 中的任意一个或两个可以在单独的芯片或电路板上实现，本实施例对此不加以限定。

射频电路 804 用于接收和发射 RF (Radio Frequency, 射频) 信号，也称电磁信号。射频电路 804 通过电磁信号与通信网络以及其他通信设备进行通信。射频电路 804 将电信号转换为电磁信号进行发送，或者，将接收到的电磁信号转换为电信号。可选地，射频电路 804 包括：天线系统、RF 收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路 804 可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于：城域网、各代移动通信网络 (2G、3G、4G 及 5G)、无线局域网和/或 WiFi(Wireless Fidelity, 无线保真) 网络。在一些实施例中，射频电路 804 还可以包括 NFC (Near Field Communication, 近距离无线通信) 有关的电路，本申请对此不加以限定。

显示屏 805 用于显示 UI (User Interface, 用户界面)。该 UI 可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏 805 是触摸显示屏时，显示屏 805 还具有采集在显示屏 805 的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器 801 进行处理。此时，显示屏 805 还可以用于提供虚拟按钮和/或虚拟键盘，也称软按钮和/或软键盘。在一些实施例中，显示屏 805 可以为一个，设置终端 800 的前面板；在另一些实施例中，显示屏 805 可以为至少两个，分别设置在终端 800 的不同表面或呈折叠设计；在再一些实施例中，显示屏 805 可以是柔性显示屏，设置在终端 800 的弯曲表面上或折叠面上。甚至，显示屏 805 还可以设置成非矩形的不规则图形，也即异形屏。显示屏 805 可以采用 LCD(Liquid Crystal Display, 液晶显示屏)、OLED(Organic Light-Emitting Diode, 有机发光二极管)等材质制备。

摄像头组件 806 用于采集图像或视频。可选地，摄像头组件 806 包括前置摄像头和后置摄像头。通常，前置摄像头设置在终端的前面板，后置摄像头设置在终端的背面。在一些实施例中，后置摄像头为至少两个，分别为主摄像头、

景深摄像头、广角摄像头、长焦摄像头中的任意一种，以实现主摄像头和景深摄像头融合实现背景虚化功能、主摄像头和广角摄像头融合实现全景拍摄以及VR（Virtual Reality，虚拟现实）拍摄功能或者其它融合拍摄功能。在一些实施例中，摄像头组件 806 还可以包括闪光灯。闪光灯可以是单色温闪光灯，也可以是双色温闪光灯。双色温闪光灯是指暖光闪光灯和冷光闪光灯的组合，可以用于不同色温下的光线补偿。

音频电路 807 可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波，并将声波转换为电信号输入至处理器 801 进行处理，或者输入至射频电路 804 以实现语音通信。出于立体声采集或降噪的目的，麦克风可以为多个，分别设置在终端 800 的不同部位。麦克风还可以是阵列麦克风或全向采集型麦克风。扬声器则用于将来自处理器 801 或射频电路 804 的电信号转换为声波。扬声器可以是传统的薄膜扬声器，也可以是压电陶瓷扬声器。当扬声器是压电陶瓷扬声器时，不仅可以将电信号转换为人类可听见的声波，也可以将电信号转换为人类听不见的声波以进行测距等用途。在一些实施例中，音频电路 807 还可以包括耳机插孔。

定位组件 808 用于定位终端 800 的当前地理位置，以实现导航或 LBS（Location Based Service，基于位置的服务）。定位组件 808 可以是基于美国的 GPS（Global Positioning System，全球定位系统）、中国的北斗系统、俄罗斯的格雷纳斯系统或欧盟的伽利略系统的定位组件。

电源 809 用于为终端 800 中的各个组件进行供电。电源 809 可以是交流电、直流电、一次性电池或可充电电池。当电源 809 包括可充电电池时，该可充电电池可以支持有线充电或无线充电。该可充电电池还可以用于支持快充技术。

在一些实施例中，终端 800 还包括有一个或多个传感器 810。该一个或多个传感器 810 包括但不限于：加速度传感器 811、陀螺仪传感器 812、压力传感器 813、指纹传感器 814、光学传感器 815 以及接近传感器 816。

加速度传感器 811 可以检测以终端 800 建立的坐标系的三个坐标轴上的加速度大小。比如，加速度传感器 811 可以用于检测重力加速度在三个坐标轴上的分量。处理器 801 可以根据加速度传感器 811 采集的重力加速度信号，控制触摸显示屏 805 以横向视图或纵向视图进行用户界面的显示。加速度传感器 811 还可以用于游戏或者用户的运动数据的采集。

陀螺仪传感器 812 可以检测终端 800 的机体方向及转动角度，陀螺仪传感

器 812 可以与加速度传感器 811 协同采集用户对终端 800 的 3D 动作。处理器 801 根据陀螺仪传感器 812 采集的数据,可以实现如下功能:动作感应(比如根据用户的倾斜操作来改变 UI)、拍摄时的图像稳定、游戏控制以及惯性导航。

压力传感器 813 可以设置在终端 800 的侧边框和/或触摸显示屏 805 的下层。当压力传感器 813 设置在终端 800 的侧边框时,可以检测用户对终端 800 的握持信号,由处理器 801 根据压力传感器 813 采集的握持信号进行左右手识别或快捷操作。当压力传感器 813 设置在触摸显示屏 805 的下层时,由处理器 801 根据用户对触摸显示屏 805 的压力操作,实现对 UI 界面上的可操作性控件进行控制。可操作性控件包括按钮控件、滚动条控件、图标控件、菜单控件中的至少一种。

指纹传感器 814 用于采集用户的指纹,由处理器 801 根据指纹传感器 814 采集到的指纹识别用户的身份,或者,由指纹传感器 814 根据采集到的指纹识别用户的身份。在识别出用户的身份为可信身份时,由处理器 801 授权该用户执行相关的敏感操作,该敏感操作包括解锁屏幕、查看加密信息、下载软件、支付及更改设置等。指纹传感器 814 可以被设置终端 800 的正面、背面或侧面。当终端 800 上设置有物理按键或厂商 Logo 时,指纹传感器 814 可以与物理按键或厂商 Logo 集成在一起。

光学传感器 815 用于采集环境光强度。在一个实施例中,处理器 801 可以根据光学传感器 815 采集的环境光强度,控制触摸显示屏 805 的显示亮度。具体地,当环境光强度较高时,调高触摸显示屏 805 的显示亮度;当环境光强度较低时,调低触摸显示屏 805 的显示亮度。在另一个实施例中,处理器 801 还可以根据光学传感器 815 采集的环境光强度,动态调整摄像头组件 806 的拍摄参数。

接近传感器 816,也称距离传感器,通常设置在终端 800 的前面板。接近传感器 816 用于采集用户与终端 800 的正面之间的距离。在一个实施例中,当接近传感器 816 检测到用户与终端 800 的正面之间的距离逐渐变小时,由处理器 801 控制触摸显示屏 805 从亮屏状态切换为息屏状态;当接近传感器 816 检测到用户与终端 800 的正面之间的距离逐渐变大时,由处理器 801 控制触摸显示屏 805 从息屏状态切换为亮屏状态。

本领域技术人员可以理解,图 8 中示出的结构并不构成对终端 800 的限定,可以包括比图示更多或更少的组件,或者组合某些组件,或者采用不同的组件

布置。

图 9 是本发明实施例提供的一种服务器的结构示意图，该服务器 900 可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上处理器（central processing units, CPU）901 和一个或一个以上的存储器 902，其中，所述存储器 902 中存储有至少一条指令，所述至少一条指令由所述处理器 901 加载并执行以实现上述各个方法实施例提供的生成歌词的方法。当然，该服务器还可以具有有线或无线网络接口、键盘以及输入输出接口等部件，以便进行输入输出，该服务器还可以包括其他用于实现设备功能的部件，在此不做赘述。

在示例性实施例中，还提供了一种计算机可读存储介质，例如包括指令的存储器，上述指令可由终端中的处理器执行以完成上述实施例中的生成歌词的方法或者显示歌词的方法。例如，所述计算机可读存储介质可以是 ROM、随机存取存储器（RAM）、CD-ROM、磁带、软盘和光数据存储设备等。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种生成歌词的方法，其特征在于，所述方法包括：

获取目标歌曲的歌词和所述目标歌曲的音频文件；

根据所述歌词内多个字符中的待标注字符和所述音频文件中所述多个字符对应的音频片段，确定所述待标注字符的候选读音；

若所述待标注字符有一个候选读音时，则所述候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

若所述待标注字符有至少两个候选读音时，则与所述待标注字符的语义匹配的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件。

2、根据权利要求1所述的方法，其特征在于，所述根据所述歌词内多个字符中的待标注字符和所述音频文件中所述多个字符对应的音频片段，确定所述待标注字符的候选读音包括：

将所述多个字符中的汉字确定为所述待标注字符；

根据所述多个字符对应的音频片段，识别所述多个字符的读音；

根据所述多个字符中的待标注字符，从所述多个字符的读音中确定所述待标注字符的候选读音，或者，根据所述多个字符中的待标注字符，从所述多个字符的读音中确定所述待标注字符的中间读音，根据所述中间读音，识别出所述待标注字符的至少两个候选读音。

3、根据权利要求1所述的方法，其特征在于，所述当所述待标注字符有至少两个候选读音时，将与所述待标注字符的语义匹配的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音包括：

对于每个候选读音，从第一词典引擎中查找所述每个候选读音对应的中间字符，从第二词典引擎中查找所述中间字符的语义，所述第一词典引擎包括多个候选读音和所述多个候选读音对应的中间字符，所述第二词典引擎包括多个中间字符和多个中间字符对应的多个语义；

根据所述待标注字符的语义，从所述至少两个候选读音中选择中间字符的

语义与所述待标注字符的语义相匹配的候选读音，将选择的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音。

4、根据权利要求1所述的方法，其特征在于，所述获取目标歌曲的歌词和所述目标歌曲的音频文件包括：

获取所述目标歌曲的第二歌词文件，所述第二歌词文件包括所述目标歌曲的歌词和所述歌词的多个字符的显示时间；

建立第一数组和第三数组，将所述多个字符写入所述第一数组，将所述多个字符的显示时间写入所述第三数组，将所述第一数组内的多个字符确定为所述目标歌曲的歌词；

获取所述目标歌曲的原始音频文件，提取所述原始音频文件中人声频段的音频片段，生成所述音频文件。

5、根据权利要求4所述的方法，其特征在于，所述根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件包括：

建立第二数组，将所述待标注字符在所述目标歌曲中对应的读音写入所述第二数组；

当所述歌词内存在目标词时，根据所述目标词包括的至少两个相邻待标注字符的显示时间，确定所述目标词的显示时间，所述目标词包括至少两个相邻的待标注字符，且包括的待标注字符的数目与所述目标词的读音的数目不相等；

将所述目标词的显示时间写入所述第三数组；

将第一数组、所述第二数组和所述第三数组添加至所述第一歌词文件中；

其中，确定所述目标词的显示时间的方式为：将所述目标词包括的至少两个相邻待标注字符的显示时间进行合并，将合并后的显示时间确定为所述目标词的显示时间。

6、根据权利要求1所述的方法，其特征在于，所述根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件包括：

建立第二数组，将所述待标注字符在所述目标歌曲中对应的读音写入所述第二数组；

将第一数组和所述第二数组添加至所述第一歌词文件中，所述第一数组用于存储所述歌词的多个字符。

7、一种显示歌词的方法，其特征在于，所述方法应用在终端上，所述方法包括：

当接收到歌词显示指令时，获取目标歌曲的第一歌词文件，所述歌词显示指令用于显示所述目标歌曲的歌词；

从所述第一歌词文件中获取所述目标歌曲的歌词、所述歌词的多个字符中待标注字符在所述目标歌曲中对应的读音；

显示所述歌词的多个字符；

在显示所述待标注字符时，在所述待标注字符的目标位置上，标注所述待标注字符在所述目标歌曲中对应的读音，所述目标位置为所述待标注字符的上方。

8、根据权利要求7所述的方法，其特征在于，所述第一歌词文件还包括所述多个字符的显示时间，相应的，所述显示所述歌词的多个字符包括：

根据所述多个字符中目标词的显示时间，突出显示多个字符中当前播放的目标词；

根据所述多个字符中除目标词以外的每个字符的显示时间，突出显示所述除目标词以外的字符中当前播放的字符；

其中，所述目标词包括至少两个相邻的待标注字符，且包括的待标注字符的数目与所述目标词的读音的数目不相等，所述目标词的显示时间为所述目标词包括的至少两个相邻待标注字符的显示时间合并后的显示时间。

9、根据权利要求8所述的方法，其特征在于，所述在显示所述待标注字符时，在所述待标注字符的目标位置上，标注所述待标注字符在所述目标歌曲中对应的读音包括：

在显示所述待标注字符时，根据所述待标注字符的显示时间，突出显示当

前播放的待标注字符和所述当前播放的待标注字符在所述目标歌曲中对应的读音。

10、一种生成歌词的装置，其特征在于，所述装置包括：

获取模块，用于获取目标歌曲的歌词和所述目标歌曲的音频文件；

确定模块，用于根据所述歌词内多个字符中的待标注字符和所述音频文件中所述多个字符对应的音频片段，确定所述待标注字符的候选读音；

所述确定模块，还用于若所述待标注字符有一个候选读音时，则所述候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

所述确定模块，还用于若所述待标注字符有至少两个候选读音时，则与所述待标注字符的语义匹配的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音；

生成模块，用于根据所述多个字符和所述待标注字符在所述目标歌曲中对应的读音，生成所述目标歌曲的第一歌词文件。

11、根据权利要求 10 所述的装置，其特征在于，所述确定模块，包括：

确定单元，用于将所述多个字符中的汉字确定为所述待标注字符；

识别单元，用于根据所述多个字符对应的音频片段，识别所述多个字符的读音；

所述确定单元，还用于根据所述多个字符中的待标注字符，从所述多个字符的读音中确定所述待标注字符的候选读音，或者，根据所述多个字符中的待标注字符，从所述多个字符的读音中确定所述待标注字符的中间读音，根据所述中间读音，识别出所述待标注字符的至少两个候选读音。

12、根据权利要求 10 所述的装置，其特征在于，所述确定模块，包括：

查找单元，用于对于每个候选读音，从第一词典引擎中查找所述每个候选读音对应的中间字符，从第二词典引擎中查找所述中间字符的语义，所述第一词典引擎包括多个候选读音和所述多个候选读音对应的中间字符，所述第二词典引擎包括多个中间字符和多个中间字符对应的多个语义；

选择单元，用于根据所述待标注字符的语义，从所述至少两个候选读音中

选择中间字符的语义与所述待标注字符的语义相匹配的候选读音，将选择的候选读音确定为所述待标注字符在所述目标歌曲中对应的读音。

13、一种显示歌词的装置，其特征在于，所述装置应用在终端上，所述装置包括：

获取模块，用于当接收到歌词显示指令时，获取目标歌曲的第一歌词文件，所述歌词显示指令用于显示所述目标歌曲的歌词；

所述获取模块，还用于从所述第一歌词文件中获取所述目标歌曲的歌词、所述歌词的多个字符中待标注字符在所述目标歌曲中对应的读音；

显示模块，用于显示所述歌词的多个字符；

标注模块，用于在显示所述待标注字符时，在所述待标注字符的目标位置上，标注所述待标注字符在所述目标歌曲中对应的读音，所述目标位置为所述待标注字符的上方。

14、一种电子设备，其特征在于，所述电子设备包括处理器和存储器，所述存储器中存储有至少一条指令，所述指令由所述处理器加载并执行以实现如权利要求 1 至权利要求 6 任一项所述的生成歌词的方法，或者权利要求 7 至权利要求 9 任一项所述的显示歌词的方法所执行的操作。

15、一种计算机可读存储介质，其特征在于，所述存储介质中存储有至少一条指令，所述指令由处理器加载并执行以实现如权利要求 1 至权利要求 6 任一项所述的生成歌词的方法，或者权利要求 7 至权利要求 9 任一项所述的显示歌词的方法所执行的操作。