

《物联网与嵌入式技术》

大作业课题报告

台本获取和部分台本处理任务

学 生 姓 名：麦业_____

学科、 专业：材料科学与工程学院_____

学 号：32305070_____

完 成 日 期：2023.5.5_____

大连理工大学

Dalian University of Technology

目 录

1	课题背景	1
2	台本工具获取台本流程	2
2.1	提取字幕	2
2.2	将字幕转为 csv 文件	2
2.3	从手动标注的 csv 文件中提取特征	3
2.4	将特征文件自动分类台词所属人物	4
3	台本处理	4

1 课题背景

数字人形象的建立所用到的项目需要用到大量台本，因此需要从视频中提取人物的台本，并且将各个人物的台本分类，并与人物对应起来。本课题使用的台本工具为 ChatHaruhi 子项目的台本工具。其操作步骤可简单概括为以下四步：

- （1）从视频中提取字幕（使用 `whisper`）；
- （2）将字幕转为 `csv` 文件；
- （3）从手动标注的 `csv` 文件中提取特征；
- （4）通过第三步得到特征文件，再通过视频与字幕文件，自动分类台词所属人物。

2 台本工具获取台本流程

2.1 提取字幕

从视频中提取字幕，输出为 srt 文件。

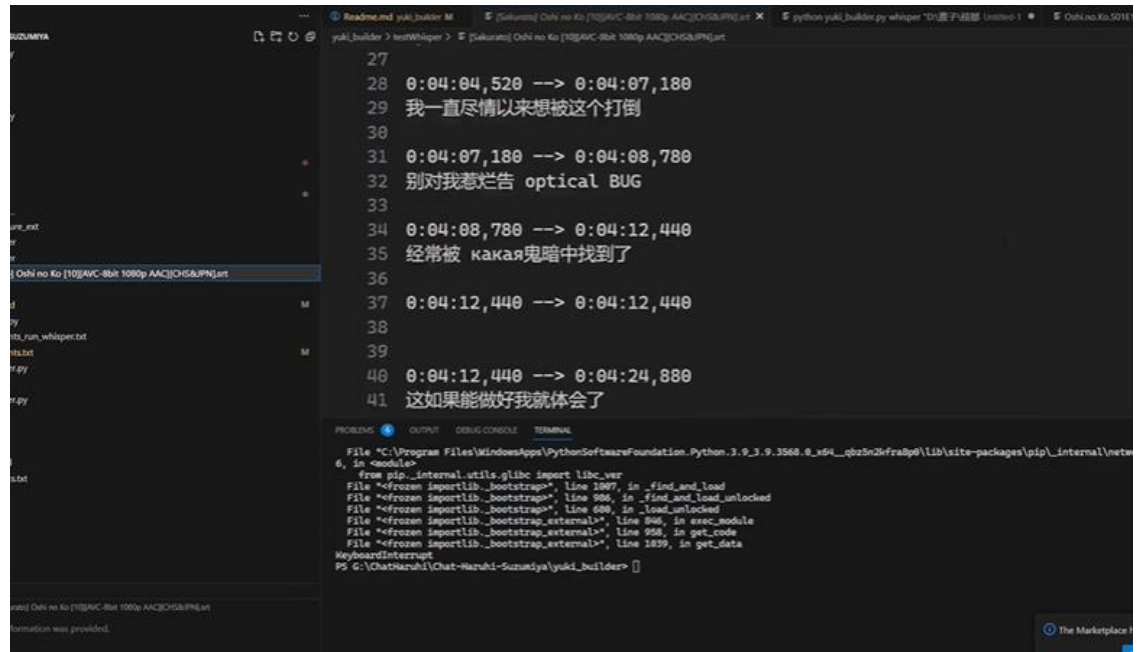


图 2.1 输出为 srt 文件

2.2 将字幕转为 csv 文件

将 srt 文件转为 csv 文件，用于后续使用，这个步骤会输出一个 csv 文件，这个文件中包含了每一句台词的开始时间，结束时间，台词内容。台词所属人物需要人工进行标注，作为后期人物的特征数据。

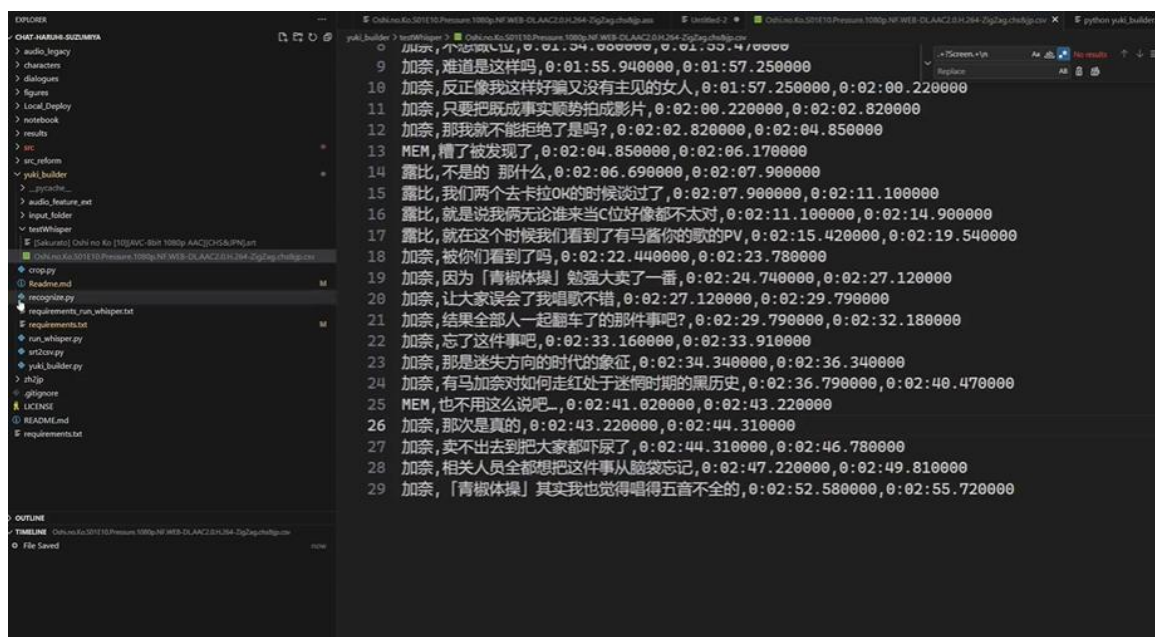


图 2.2 转为 csv 文件

2.3 从手动标注的 csv 文件中提取特征

根据已经手动标注好的 csv 文件来切割视频，提取特征，输出为 wav，pkl 等文件。

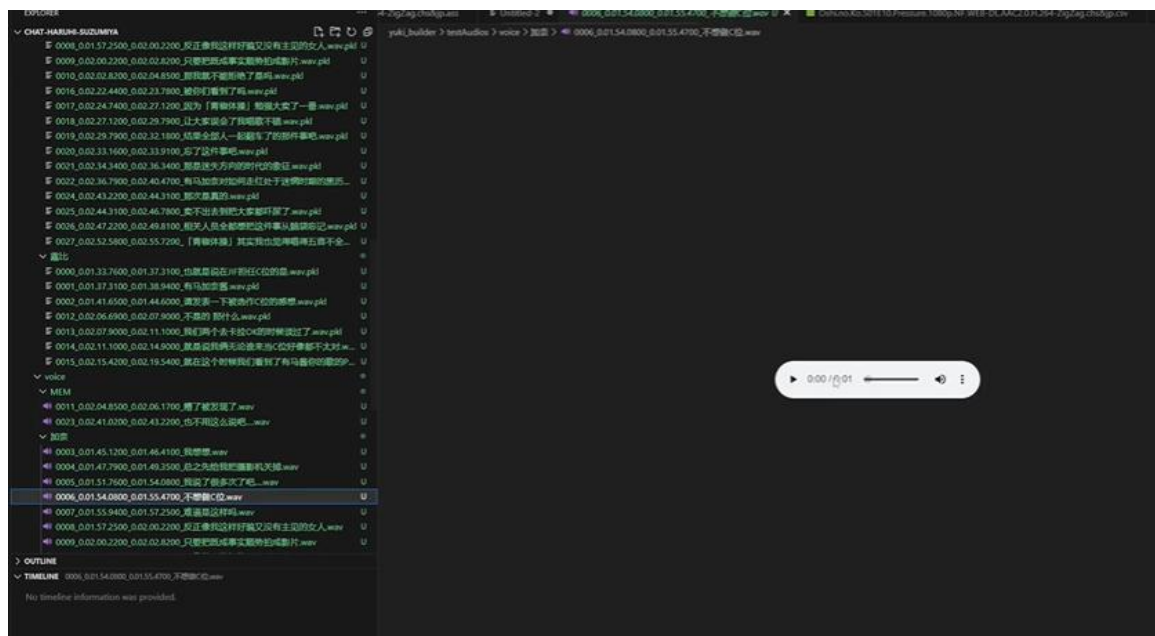


图 2.3 输出为 wav，pkl 等文件

2.4 将特征文件自动分类台词所属人物

通过第三步得到特征文件，再通过视频与字幕文件，自动分类台词所属人物，输出为 csv，txt 文件。

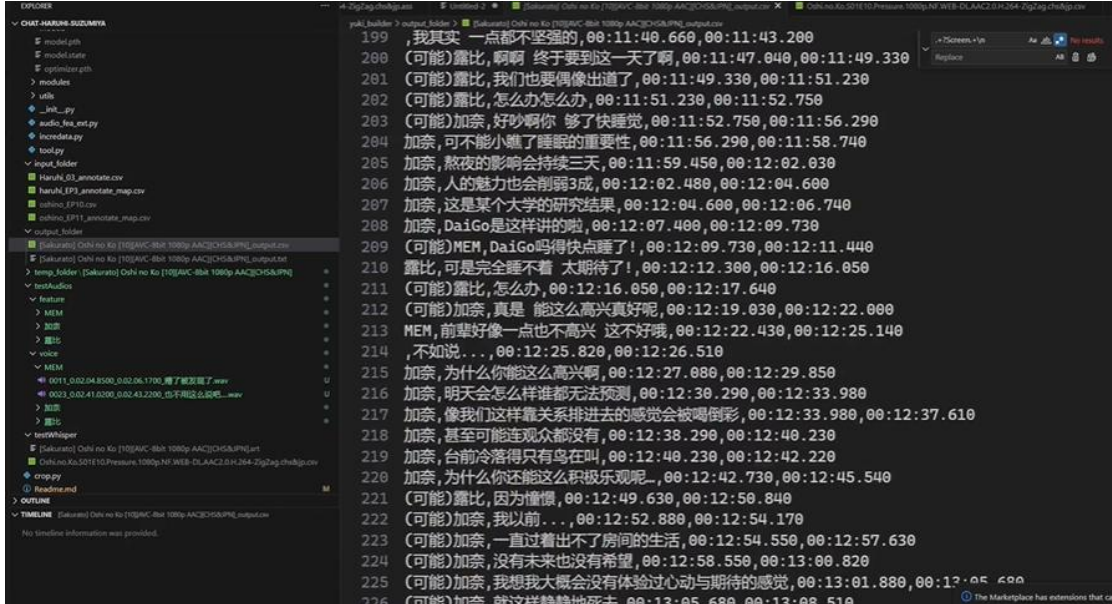


图 2.4 输出为 txt 文件

3 台本处理

获取的 txt 台本文件中会存在错误，需要手动进行订正。部分订正前后的台本如下表所示。

表 3.1 部分台本处理展示

订正前	订正后
哆啦 A 梦:「没有镜子的世界」 大雄:「每次看到镜子就觉得好讨厌」	哆啦 A 梦:「没有镜子的世界」 大雄:「每次看到镜子就觉得好讨厌」
哆啦 A 梦:「大熊,你怎么了」 哆啦 A 梦:「干嘛一直盯着镜子看」 大雄:「为什么我的脸像漫画人物一样」	哆啦 A 梦:「大熊,你怎么了」 哆啦 A 梦:「干嘛一直盯着镜子看」 大雄:「为什么我的脸像漫画人物一样」
哆啦 A 梦:「你不要在意这种事嘛」 哆啦 A 梦:「人的价值并不是脸」 哆啦 A 梦:「而是头脑跟实力」 哆啦 A 梦:「我是不是说错了什么」	哆啦 A 梦:「呵呵呵你不要在意这种事嘛」 哆啦 A 梦:「人的价值并不是脸」 哆啦 A 梦:「而是头脑跟实力」

<p>话」</p> <p>(可能)大雄:「头脑跟实力」</p> <p>(可能)大雄:「我两样都没有」</p> <p>(可能)大雄:「还是多拉雷梦妮好」</p> <p>大雄:「一看到你那个长相人家就死心」</p> <p>大雄:「为什么会有镜子呢」</p> <p>大雄:「如果谁都不知道自己的长相」</p> <p>大雄:「那就不会有人觉得自己有弱点」</p> <p>(可能)大雄:「也不会有什么自大的人了」</p> <p>哆啦 A 梦:「真的是这样吗」</p> <p>大雄:「如果世界上没有镜子的话」</p> <p>(可能)大雄:「那就好了」</p> <p>哆啦 A 梦:「好吧,那就试试看好了」</p> <p>哆啦 A 梦:「以前用过的如果电话亭」</p> <p>大雄:「用这个啊」</p> <p>大雄:「如果这个世界上日本人都看不到自己的长相」</p> <p>大雄:「好了,不知道会变得怎么样」</p> <p>(可能)大雄:「我们去看看」</p> <p>大雄:「妈妈正在化妆」</p> <p>大雄:「又照不出长相来」</p> <p>大雄妈妈:「这样化妆好难」</p> <p>大雄:「怎么样,妈妈,化好了没有」</p> <p>(可能)大雄妈妈:「你们来得正好」</p> <p>大雄妈妈:「你看,妈妈的口红有没有画歪啊」</p> <p>(可能)大雄:「要画右边一点」</p> <p>哆啦 A 梦:「还有,眉毛要画得浓一点」</p>	<p>哆啦 A 梦:「我是不是说错了什么话」</p> <p>大雄:「头脑跟实力」</p> <p>大雄:「我两样都没有」</p> <p>大雄:「还是哆啦 A 梦你好」</p> <p>大雄:「一看到你那个长相人家就死心」</p> <p>大雄:「为什么会有镜子呢」</p> <p>大雄:「如果谁都不知道自己的长相」</p> <p>大雄:「那就不会有人觉得自己有弱点」</p> <p>大雄:「也不会有什么自大的人了」</p> <p>哆啦 A 梦:「真的是这样吗」</p> <p>大雄:「如果世界上没有镜子的话」</p> <p>大雄:「那就好了」</p> <p>哆啦 A 梦:「好吧,那就试试看好了」</p> <p>哆啦 A 梦:「以前用过的如果电话亭」</p> <p>大雄:「用这个啊」</p> <p>大雄:「如果这个世界上任何人都看不到自己的长相」</p> <p>大雄:「好了,不知道会变得怎么样」</p> <p>大雄:「我们去看看」</p> <p>大雄:「妈妈正在化妆」</p> <p>大雄:「又照不出长相来」</p> <p>大雄妈妈:「这样化妆好难哦」</p> <p>大雄:「怎么样,妈妈,化好了没有」</p> <p>大雄妈妈:「你们来得正好」</p> <p>大雄妈妈:「你看,妈妈的口红有没有化歪啊」</p> <p>大雄:「要化右边一点」</p> <p>哆啦 A 梦:「还有,眉毛要画得浓一点」</p>
---	---

<p>:「大概好了吧,有没有刮根去」</p> <p>(可能)大雄妈妈:「老公,你看看我化妆有没有化好啊」</p> <p>(可能)胖虎:「画的什么,跟女鬼一样啊」</p> <p>(可能)大雄:「什么,说我像女鬼一样」</p> <p>(可能)大雄:「那你自己又怎么样呢」</p> <p>:「我怎么样」</p> <p>大雄:「你看,连玻璃都照不出脸来」</p> <p>哆啦 A 梦:「对啊,而且连相机都不能照相」</p> <p>哆啦 A 梦:「所以谁都不知道自己长得怎么样」</p> <p>大雄:「不知道大家会变得怎么样」</p> <p>大雄:「我们出去看看」</p> <p>(可能)哆啦 A 梦:「小布」</p> <p>哆啦 A 梦:「求求你,帮我画素描」</p> <p>(可能)小夫:「有有我画」</p> <p>(可能)静香:「因为不画素描」</p> <p>(可能)静香:「根本不晓得自己长什么样子」</p> <p>哆啦 A 梦:「大家都说你画的素描跟本人很像」</p> <p>小夫:「好,画好了,这样可以吧」</p> <p>(可能)哆啦 A 梦:「好像」</p> <p>静香:「小布真的好会画画哦」</p> <p>(可能)大雄:「什么嘛,没有镜子可以看」</p> <p>大雄:「大家还不是很高兴」</p> <p>哆啦 A 梦:「你不是为了这个才不要镜子的吗」</p> <p>(可能)哆啦 A 梦:「可是这么一来」</p>	<p>大雄爸爸:「大概好了吧,有没有刮干净」</p> <p>大雄妈妈:「老公,你看看我化妆有没有化好啊」</p> <p>大雄爸爸:「啊啊啊,画的什么,跟女鬼一样啊」</p> <p>大雄妈妈:「什么,说我像女鬼一样」</p> <p>大雄妈妈:「那你自己又怎么样呢」</p> <p>大雄爸爸:「我怎么样」</p> <p>大雄:「你看,连玻璃都照不出脸来」</p> <p>哆啦 A 梦:「对啊,而且连相机都不能照相」</p> <p>哆啦 A 梦:「所以谁都不知道自己长得怎么样」</p> <p>大雄:「不知道大家会变得怎么样」</p> <p>大雄:「我们出去看看」</p> <p>路人女孩:「小夫」</p> <p>路人女孩:「求求你,帮我画素描」</p> <p>小夫:「又要我画啊」</p> <p>路人女孩:「因为不画素描」</p> <p>路人女孩:「根本不晓得自己长什么样子」</p> <p>路人女孩:「大家都说你画的素描跟本人很像」</p> <p>小夫:「好,画好了,这样可以吧」</p> <p>哆啦 A 梦:「好像」</p> <p>静香:「小布真的好会画画哦」</p> <p>大雄:「什么嘛,没有镜子可以看」</p> <p>大雄:「大家还不是很高兴」</p> <p>哆啦 A 梦:「你不是为了这个才不要镜子的吗」</p> <p>哆啦 A 梦:「可是这么一来」</p> <p>大雄:「反而让小夫受到大家的欢迎」</p>
--	--

<p>大雄:「反而让小布受到大家的欢迎」</p> <p>(可能)大雄:「真是气人」</p> <p>(可能)胖虎:「像这个样子,一摸就知道」</p> <p>(可能)胖虎:「我一定是一个长得非常帅气的男子汉」</p> <p>(可能)胖虎:「有高高的鼻子,紧闭的嘴唇」</p> <p>:「每天摸摸看都不会腻」</p> <p>(可能)胖虎:「对了,我要当明星」</p> <p>胖虎:「我要像亚兰德伦一样」</p> <p>(可能)胖虎:「将来演电影」</p> <p>大雄:「胖虎居然说他要像亚兰德伦一样」</p> <p>(可能)哆啦 A 梦:「因为没有镜子嘛」</p> <p>胖虎:「哎哟,说起来还真是可怜」</p> <p>胖虎:「世上竟然还有你们两个长得像漫画人物的人」</p> <p>(可能)胖虎:「一看到你们」</p> <p>(可能)胖虎:「我真的觉得很同情」</p> <p>(可能)大雄:「他说的什么话」</p> <p>(可能)大雄:「因为没有镜子」</p> <p>大雄:「所以大家看到的都应该一样才对」</p> <p>(可能)大雄:「既然如此」</p> <p>(可能)大雄:「那就让胖虎看一看镜子」</p> <p>哆啦 A 梦:「好不容易才变成没有镜子的世界」</p> <p>大雄:「可是从来没有看过镜子的人」</p> <p>(可能)大雄:「第一次看到镜子会怎</p>	<p>大雄:「真是气人」</p> <p>胖虎:「像这个样子,一摸就知道」</p> <p>胖虎:「我一定是一个长得非常帅气的男子汉」</p> <p>胖虎:「有高高的鼻子,紧闭的嘴唇」</p> <p>胖虎:「每天摸摸看都不会腻」</p> <p>胖虎:「对了,我要当明星」</p> <p>胖虎:「我要像亚兰德伦一样」</p> <p>胖虎:「将来演电影」</p> <p>大雄:「胖虎居然说他要像亚兰德伦一样」</p> <p>哆啦 A 梦:「因为没有镜子嘛」</p> <p>胖虎:「哎哟,说起来还真是可怜」</p> <p>胖虎:「世上竟然还有你们两个长得像漫画人物的人」</p> <p>胖虎:「一看到你们」</p> <p>胖虎:「我真的觉得很同情」</p> <p>大雄:「他说的什么话」</p> <p>大雄:「因为没有镜子」</p> <p>大雄:「所以大家看到的都应该一样才对啊」</p> <p>大雄:「既然如此」</p> <p>大雄:「那就让胖虎看一看镜子」</p> <p>哆啦 A 梦:「好不容易才变成没有镜子的世界」</p> <p>大雄:「可是从来没有看过镜子的人」</p> <p>大雄:「第一次看到镜子会怎么样呢」</p> <p>大雄:「我想这一定很有趣」</p> <p>哆啦 A 梦:「这样子啊」</p> <p>胖虎:「这是什么」</p> <p>哆啦 A 梦:「箱子里的镜子」</p> <p>哆啦 A 梦:「你走进去看一看会看</p>
--	---

<p>么样呢」</p> <p>(可能)大雄:「我想这一天很有趣」</p> <p>(可能)大雄:「这样子啊」</p> <p>胖虎:「这是什么」</p> <p>哆啦 A 梦:「箱子裡的镜子」</p> <p>哆啦 A 梦:「你走进去看一看会看到很有趣的东西」</p> <p>胖虎:「是真的吗」</p> <p>(可能)胖虎:「我来看一下」</p> <p>(可能)胖虎:「有人在里面」</p> <p>(可能)大雄:「是吗,你看他长得怎么样」</p> <p>胖虎:「这个家伙长得像大猩猩一样」</p> <p>(可能)胖虎:「你看到我敢笑」</p> <p>胖虎:「好小子,你给我出来」</p> <p>(可能)胖虎:「好动啊,可恶,居然躲在玻璃后面」</p> <p>(可能)小夫:「那个箱子里面」</p> <p>小夫:「有一个长得像大猩猩的家伙」</p> <p>(可能)小夫:「一点都没错」</p> <p>(可能)小夫:「一看就知道是个大坏蛋」</p> <p>小夫:「对,不过不像大猩猩」</p> <p>(可能)小夫:「反而像只狐狸」</p> <p>胖虎:「你胡说,哪有那种狐狸」</p> <p>大雄:「我来看一下」</p> <p>大雄:「你们两个真是的」</p> <p>大雄妈妈:「是个可爱的女孩子吗」</p> <p>(可能)哆啦 A 梦:「你们是个可爱的女孩子」</p> <p>(可能)哆啦 A 梦:「你们是个可爱的女孩子」</p>	<p>到很有趣的东西」</p> <p>胖虎:「是真的吗」</p> <p>胖虎:「我来看一下」</p> <p>胖虎:「有人在里面」</p> <p>大雄:「是吗,你看他长得怎么样」</p> <p>胖虎:「这个家伙长得像大猩猩一样」</p> <p>胖虎:「你看到我敢笑」</p> <p>胖虎:「好小子,你给我出来」</p> <p>胖虎:「可恶」</p> <p>胖虎:「好痛哦,可恶,居然躲在玻璃后面」</p> <p>小夫:「那个箱子里面」</p> <p>小夫:「有一个长得像大猩猩的家伙」</p> <p>小夫:「一点都没错」</p> <p>小夫:「一看就知道是个大坏蛋」</p> <p>小夫:「对,不过不像大猩猩」</p> <p>小夫:「反而像只狐狸」</p> <p>胖虎:「你胡说,哪有那种狐狸」</p> <p>静香:「我来看一下好了」</p> <p>静香:「你们两个真是的」</p> <p>静香:「是个可爱的女孩子嘛」</p> <p>胖虎:「一看就是大猩猩嘛」</p> <p>静香:「是个可爱的女孩子」</p> <p>路人:「啊,这是」</p> <p>路人:「这不是十年前我那个离家出走的哥哥吗」</p> <p>路人:「这到底是怎么回事啊」</p> <p>路人:「为什么会在这种地方呢」</p> <p>路人:「我明白了」</p> <p>路人:「你一定是被关在玻璃做的笼子里面」</p> <p>路人:「是谁那么狠心做出这种事情</p>
--	---

<p>:「这是」</p> <p>:「这不是十年前我那个离家除走了哥哥吗」</p> <p>(可能)小夫:「这到底是怎么回事啊」</p> <p>:「为什么会在这种地方呢」</p> <p>(可能)胖虎:「我明白了」</p> <p>:「你一定是被关在玻璃座的笼子里面」</p> <p>:「是谁那么狠心做出这种事情来」</p> <p>(可能)胖虎:「哥哥,我马上会把你救出来的」</p> <p>(可能)哆啦 A 梦:「你要等我」</p> <p>哆啦 A 梦:「箱子不见了」</p> <p>(可能)大雄:「对啊,刚才明明还摆在这里的」</p> <p>哆啦 A 梦:「奇怪了,会不会被谁搬走了」</p> <p>(可能)大雄妈妈:「奇怪了,会不会被谁搬走了」</p> <p>(可能)小夫:「警察先生」</p> <p>(可能)胖虎:「什么,你搁在箱子里面」</p> <p>:「他是被人关进去的,你快救他」</p> <p>(可能)胖虎:「好,我知道了」</p> <p>:「快点把那个箱子打开」</p> <p>(可能)胖虎:「有个警察」</p> <p>(可能)胖虎:「从来没有看过,不知道是哪个分局的」</p> <p>:「他一定是个假警察」</p> <p>:「他很可能是绑架犯啊」</p> <p>(可能)胖虎:「难怪长得一脸凶相」</p> <p>大雄:「一定是有什么人觉得他很稀奇就拿走了」</p>	<p>来」</p> <p>路人:「哥哥,我马上会把你救出来的」</p> <p>路人:「你要等我」</p> <p>哆啦 A 梦:「箱子不见了」</p> <p>大雄:「对啊,刚才明明还摆在这里的」</p> <p>哆啦 A 梦:「奇怪了,会不会被谁搬走了」</p> <p>路人:「警察先生」</p> <p>警察:「什么,你哥在箱子里面」</p> <p>路人:「他是被人关进去的,你快救他」</p> <p>警察:「好,我知道了」</p> <p>警察:「快点把那个箱子打开」</p> <p>警察:「嗯？」</p> <p>路人:「嗯？」</p> <p>警察:「有个警察」</p> <p>警察:「从来没有看过,不知道是哪个分局的」</p> <p>路人:「他一定是个假警察」</p> <p>路人:「他很可能是绑架犯啊」</p> <p>警察:「难怪长得一脸凶相」</p> <p>大雄:「一定是有什么人觉得他很稀奇就拿走了」</p> <p>哆啦 A 梦:「但愿不要发生什么事才好」</p> <p>大雄:「看到了」</p> <p>大雄:「在警局那边」</p> <p>警察:「你为什么不肯放人家的哥哥呢」</p> <p>警察:「好,我要逮捕你」</p> <p>警察:「可恶,你居然也拔枪啊」</p> <p>警察:「你想杀警察是不是啊」</p>
--	---

<p>哆啦 A 梦:「但愿不要发生什么事才好」</p> <p>大雄:「看到了」</p> <p>(可能)大雄:「在警局那边」</p> <p>(可能)胖虎:「你为什么不肯放人家的哥哥呢」</p> <p>(可能)胖虎:「好,我要逮捕你」</p> <p>(可能)胖虎:「可恶,你居然也拔枪啊」</p> <p>(可能)胖虎:「你想杀警察是不是」</p> <p>大雄妈妈:「糟糕了」</p> <p>哆啦 A 梦:「快点变回原来的世界」</p> <p>胖虎:「我为什么会对着镜子开枪啊」</p> <p>(可能)胖虎:「真是奇怪了」</p> <p>:「我为什么会背一块镜子来这呢」</p> <p>(可能)哆啦 A 梦:「如果没有镜子」</p> <p>(可能)大雄:「未必是一件好事」</p> <p>哆啦 A 梦:「就是啊,还是像平常一样」</p>	<p>哆啦 A 梦:「糟糕了」</p> <p>哆啦 A 梦:「快点变回原来的世界」</p> <p>警察:「啊?」</p> <p>路人:「啊?」</p> <p>警察:「我为什么会对着镜子开枪啊」</p> <p>警察:「真是奇怪了」</p> <p>路人:「我为什么会背一块镜子来这呢」</p> <p>大雄:「如果没有镜子」</p> <p>大雄:「未必是一件好事」</p> <p>哆啦 A 梦:「就是啊,还是像平常一样最好」</p>
---	---