

一种实现人机协同的语音通讯系统及方法

本公开要求在 2020 年 4 月 28 日提交中国专利局、申请号为 202010346872.9、发明名称为“一种实现人机协同的语音通讯系统及方法”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

技术领域

本公开涉及智能通讯领域，更具体地说，它涉及一种实现人机协同的语音通讯系统及方法。

背景技术

随着 AI 和通信科技的快速发展，语音机器人已广泛应用各行各业。例如，当企业需要和潜在客户进行沟通时，语音机器人能够大幅度降低人工外呼成本，提高外呼效率，增加对潜在客户的覆盖；例如，客户有问题需要对接企业客服时，语音机器人能够快速回答已有模板的问题，效率高，减轻人工成本。

但是纯语音机器人存在外呼信息传达效率不高的问题，或是客户的问题问的复杂，超出模板外，语音机器人就会处理不了，达不成客户需求；并且鉴于对现有市场上语音机器人不够智能的认知，一般客户对于和语音机器人沟通这件事是排斥的，所以纯语音机器人很难独立工作。

相比于语音机器人，纯人工劳动力的优势在于常识积累和情绪识别，能够灵活应对更加复杂的状况，但工作量有限，情绪波动大，专业水平参差不齐。例如在人工坐席话务员进行电话销售时，真正的挑战往往不是专业度不足，而是不愿面对被客户拒绝和被挂电话的那份挫败感。

所以目前市场上更倾向于是人机协同的方式来运作，也就是先使用语音机器人对接客户，语音机器人不能解决了，再切换到人工。但是在人机切换后，话务员需要一定的时间去获取语音机器人和客户之间的通话信息，尤其是在客户面对语音机器人信息获取效率低时，还要给予话务员梳理通话内容的时间，就会造成客户体验差的不良后果，客户可能会产生反感甚至是厌恶的情绪，导致通话目的难以达成的问题的出现。

发明内容

本公开的目的是提供一种实现人机协同的语音通讯系统及方法，能够在通话时，人

工坐席和语音机器人通话无差别切换，人工介入后，话务员能快速了解通话意图，保证客户的通话体验良好。

本公开的上述技术目的是通过以下技术方案得以实现的。

第一方面，本公开提供一种实现人机协同的语音通讯系统，包括通讯模块，还包括：

5 接听与外呼模块，通话接通后，分配一个人工坐席；

音意理解模块，包含音意理解模型，用于理解通话过程中的意图；

声音克隆模块，包含声音克隆模型，用于将回复文本合成为对应话务员声音；

人机交互模块，由语音机器人调用音意理解模块得到客户意图并按照预先编好的话术，再调用声音克隆模块使用对应话务员声音与客户通话；

10 介入提示模块，包含量化分析子模块，对人机交互时的通话内容进行实时监测得到介入意向率，并根据预定义规则提示人工坐席；

人工介入模块，采用对应话务员语音回复客户和/或调用所述声音克隆模块将任一话务员的回复内容处理为对应话务员声音回复客户，还包含能够调用所述音意理解模块的检索提词子模块为任一话务员进行通话意图展示和通话提词。

15 第二方面，本公开提供一种实现人机协同的语音通讯方法，包括以下步骤：通话接通后，使用接听与外呼模块分配一个人工坐席；

由人机交互模块中的语音机器人调用音意理解模块得到客户意图并按照预先编好的话术，再调用声音克隆模块使用对应话务员声音与客户通话；

20 由介入提示模块的量化分析子模块对人机交互时的通话内容进行实时监测，得到介入意向率，并根据预定义规则提示人工坐席；

当人工坐席接收到介入提示时，由人工介入模块介入通话，采用检索提词子模块展示人机交互通话过程中的意图，采用对应话务员语音回复客户和/或调用所述声音克隆模块将任一话务员的回复内容处理为对应话务员声音回复客户，在通话过程中使用检索提词子模块为任一话务员进行通话提词。

25 可选的，所述量化分析子模块用于根据历史通话大数据和当前人工交互过程的对比分析，得到当前客户需要人工介入的介入意向率。

可选的，所述检索提词子模块在任一话务员选择人工介入时直接调用音意理解模块展示人机交互通话意图；所述检索提词子模块在人工介入通话过程中调用音意理解模块，得到最新通话意图，并从数据库中检索出参考信息生成回复文本并展示。

可选的，人工坐席介入的方式包括：

5 对应话务员直接语音回复客户；

任一话务员选择检索提词子模块生成的提词文本，由声音克隆模块生成语音回复客户；

任一话务员手动输入文本，由声音克隆模块生成语音回复客户；

10 任一话务员语音，使用声音克隆模块中的语音识别子模块将语音转换为文本，再由其中声音克隆模型生成语音回复客户。

可选的，所述音意理解模块中音意理解模型的生成和应用包括以下步骤：

使用大量带有意图标注的语音数据对基于深度学习的音意理解模型进行训练；

把训练后的音意理解模型部署到音意理解服务器；

15 对话过程中收到语音数据，并输入到音意理解服务器中调用音意理解服务，返回识别出的意图。

可选的，所述声音克隆模型的生成和应用的步骤为：

采集人工坐席对应话务员的声音语料，对初始声音克隆模型进行训练；

把训练后的声音克隆模型部署到声音克隆服务器。

将人工介入的回复文本合成为对应话务员声音播放给客户。

20 可选的，所述人工介入模块中还包含有质检模块，用于对任一话务员说话的语速、内容进行质检，并把质检结果实时反馈到人工坐席。

第三方面，本公开还提供一种实现人机协同的语音通讯方法，应用于前述第一方面所述的语音通讯系统，其中，包括：

将通话内容输入至音意理解模型，得到所述通话内容包含的通话意图；

获取与所述通话意图相匹配的预设回复内容；获取预先分配的人工坐席的声音特征；

将所述预设回复内容以及所述声音特征输入至声音克隆模型，得到克隆音频；

使用所述克隆音频答复通话；所述克隆音频为声音特征与所述人工坐席相匹配的音频；

- 5 将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率；所述当前介入意向率为产生意向客户的概率；

如果所述当前介入意向率大于预设阈值，调用所述人工坐席介入通话；

通过检索题词子模块获得历史通话意图以及预设通话提词文本；根据所述历史通话意图以及预设通话提词文本语音答复通话；

- 10 和/或，将参考回复内容输入至所述音频克隆模型，得到克隆音频；

使用所述克隆音频答复通话；所述参考回复内容为其他人工坐席回复的语音。

可选的，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤包括：

- 15 获取通话内容中包含的每一问询环节内容，所述问询环节为通话双方发起问询的过程；

将每一问询环节作为一个分支执行遍历操作；

在预设的介入意向率数据库中存在与遍历的分支相匹配的预定义关键分支的情形下，获取所述预定义关键分支的介入意向率，以作为当前通话的当前介入意向率。

- 20 可选的，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还包括：

获取通话内容中包含的每一问答环节内容，所述问答环节为通话双方的问询及回答的过程；

遍历所述介入意向率数据库；

- 25 在所述介入意向率数据库中存在与所述问答环节相匹配的预设关键词的情形下，获取所述预设关键词的介入意向率，以作为当前通话的当前介入意向率。

可选的，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还包括：

获取所述通话意图；

遍历所述介入意向率数据库；

- 5 在所述介入意向率数据库中存在与所述通话意图相匹配的预定义意图的情况下，获取所述预定义意图的介入意向率，以作为当前通话的当前介入意向率。

可选的，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还包括：

获取与所述通话意图相匹配的预设回复内容；

- 10 遍历所述介入意向率数据库；

在所述介入意向率数据库中存在与所述预设回复内容相匹配的预定义意向判断规则的情形下，获取所述预定义意向判断规则对应的介入意向率，以作为当前通话的当前介入意向率。

可选的，所述方法还包括：

- 15 设置预定义关键分支、预设关键词、预定义意图以及预定义意向判断规则；

将所述预定义关键分支、所述预设关键词、所述预定义意图以及预定义意向判断规则保存至预设的数据库中，以形成所述介入意向率数据库。

可选的，将所述预定义关键分支、所述预设关键词、所述预定义意图以及预定义意向判断规则保存至预设的数据库中的步骤后，还包括：

- 20 获取历史通话数据，所述历史通话数据为通话双方的通话内容；

筛选出所述历史通话数据中有人工坐席介入的通话；

筛选出有人工介入的通话中存在所述预定义关键分支、所述预设关键词、所述预定义意图以及所述预定义意向判断规则的通话，以作为高意向率通话。

- 25 可选的，筛选出有人工介入的通话中存在所述预定义关键分支、所述预设关键词以及所述预定义意图的通话，以作为高意向率通话的步骤后，还包括：

统计有人工介入的通话的通话数量，以及所述高意向率通话数量；

计算所述高意向率通话数量与有人工介入的通话的通话数量之间的比值，得到所述预定义关键分支、所述预设关键词、所述预定义意图以及所述预定义意向判断规则的介入意向率。

- 5 综上所述，本公开可以做到，利用介入提示模块能够智能的给人工坐席提示，在最合适的时机提示话务员介入，避免因语音机器人不能很好回复客户，而人工没有及时介入，使得客户通话差的问题的发生；

本公开中的意图理解不是通过传统的 ASR 识别文本再用 NLP 识别意图，而是通过音意理解模型，直接从语音识别出用户意图，从效果、效率和成本上明显优于传统方法。

- 10 人工介入模块能够在通话时，有检索提词子模块来展示介入前的通话意图，无需话务员再去理解所有通话内容，直接了解客户意图，节省理解时间，使得客户的通话体验不受人机切换影响；在通话过程中，智能的识别客户意图并检索出相应的回复内容提示话务员，使得话务员在相关问题上响应更快，提高服务效率，保证客户通话体验。

- 15 利用声音克隆模块处理生成的人工坐席对应话务员声音，让客户感觉自己是一直在与人工坐席话务员通话，能实现人工和语音机器人的无差别切换，从而能够提供给客户更好的通话感受。

人工介入的方式不局限于特定的话务员语音回复，可以采用多人、换人的方式来服务客户，能够提供给客户更好的通话体验。

附图说明

- 20 为了更清楚地说明本公开的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，对于本领域普通技术人员而言，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 是本公开实施例的系统框图。

具体实施方式

- 25 以下将结合附图和具体实施例对本公开进行详细说明，显然，所描述的实施例仅仅只是本公开一部分实施例，而不是全部的实施例，基于本公开中的实施例，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，都属于本公开保护的范围。

如图 1 所示, 本公开提供一种实现人机协同的语音通讯系统, 包括通讯模块, 接听与外呼模块, 通话接通后, 分配一个人工坐席;

音意理解模块, 包含音意理解模型, 用于理解通话过程中的意图;

声音克隆模块, 包含声音克隆模型, 用于将回复文本合成为对应话务员声音;

- 5 人机交互模块, 由语音机器人调用音意理解模块得到客户意图并按照预先编好的话术, 再调用声音克隆模块使用对应话务员声音与客户通话;

介入提示模块, 包含量化分析子模块, 对人机交互时的通话内容进行实时监测得到介入意向率, 并根据预定义规则提示人工坐席;

- 10 人工介入模块, 采用对应话务员语音回复客户和/或调用声音克隆模块将任一话务员的回复内容处理为对应话务员声音回复客户, 还包含能够调用音意理解模块的检索提词子模块为任一话务员进行通话意图展示和通话提词。

对应的, 本公开还提供一种实现人机协同的语音通讯方法, 包括以下步骤: 通话接通后, 使用接听与外呼模块分配一个人工坐席;

- 15 由人机交互模块中的语音机器人调用音意理解模块得到客户意图并按照预先编好的话术, 再调用声音克隆模块使用对应话务员声音与客户通话;

由介入提示模块的量化分析子模块对人机交互时的通话内容进行实时监测, 得到介入意向率, 并根据预定义规则提示人工坐席;

- 20 当人工坐席接收到介入提示时, 由人工介入模块介入通话, 采用检索提词子模块展示人机交互通话过程中的意图, 采用对应话务员语音回复客户和/或调用声音克隆模块将任一话务员的回复内容处理为对应话务员声音回复客户, 在通话过程中使用检索提词子模块为任一话务员进行通话提词。

通讯模块实现基础的通讯功能, 为外呼、接通、三方通话提供底层支持。

接听与外呼模块根据导入的号码呼叫客户或者接听客户打进的电话, 接通后给每个接听和呼叫任务分配一个人工坐席, 保证每一通电话都对应有一个人工坐席在服务。

- 25 音意理解模块, 包含有音意理解模型, 传统的音意理解模块大多是先通过语音识别技术获取文本再进行意图识别, 而基于文本的意图识别的准确率严重依赖于语音识别技术的准确率。而本公开中音意理解模型采用基于深度学习的音意理解技术, 直接用深度

神经网络从输入的语音数据中识别出相应的意图，达到了减少语音识别造成的信息缺失，提高语音意图识别准确率的效果；

音意理解模型的生成和应用包括以下步骤：

5 使用大量带有意图标注的语音数据对基于深度学习的音意理解模型进行训练；具体为：（1）首先接收声音序列数据输入，提取 MFCC 等语音特征；然后使用 transformer 神经网络，以这些语音特征数据作为输入，提取音素和字素特征；再把这些音素和字素特征输入到神经网络的输入层，对神经网络参数调优形成预训练模型；（2）在预训练模型的基础上，根据标注了意图的语音数据训练音意理解模型，首先收集包含用户意图的语音数据；再标注语音数据中所包含的用户意图类别，得到带有意图标签的用户语音；
10 然后将带有意图标签的用户语音，转换为语音意图特征的音素和字素；最后载入预训练模型，将前一步得到的音素和字素输入到神经网络的输入层，训练音意理解模型。

把训练后的音意理解模型部署到音意理解服务器；

对话过程中收到语音数据，并输入到音意理解服务器中调用音意理解服务，返回识别出的意图，具体如下：输入待识别语音；对输入语音进行降噪、特征提取处理，将待
15 识别语音转化为包含声音信息的特征向量；将该特征向量输入训练好的音意理解模型进行识别；输出识别后的语音意图。

声音克隆模块，包含有声音克隆模型和语音识别子模块，声音克隆模型的生成和应用的步骤为：采集人工坐席对应话务员的声音语料，对初始声音克隆模型进行训练；把
20 训练后的声音克隆模型部署到声音克隆服务器；将人工坐席介入的回复语音合成为对应话务员声音播放给客户。

其具体的步骤为：声音克隆模型由以下步骤得到：

（1）采集人工坐席对应话务员声音语料，语料采集模块为声音克隆模型的训练采集、
25 收集语音语料数据，数据主要是话务员本人的真实说话语音，因此训练出的声音克隆模型和对应话务员的声音具有较高的相似性，可以实现语音机器人和人工坐席语音的无缝切换。

语料的来源包括：话术市场中话术的录音，通话过程中话务员应答的语音音频语料，人工专门为克隆模型采集（例如在录音室环境下按照指定的文稿朗读录音）的语料。

采集到的语料需要先进行语料自检，剔除偏差较大、背景音明显、音色偏差大的语料。然后对语料进行自动降噪增强处理。

(2) 训练模型，基于上述对应话务员声音语料训练人工坐席专属的声音克隆模型。

对每个人工坐席，需要训练匹配对应话务员的声音克隆模型，以便在人工介入情况下尽可能地减小语音机器人声音和人工坐席话务员声音的差异。另外，由于底层模型的升级、优化，对已经上线的模型也需要支持反复的训练。

5 声音克隆模型的训练过程如下：

首先对采集的声音语料进行自检和降噪处理；采集到的语料一般都有对应的文字，但文字和音频经常有个别字词对应不准确的情况，所以需要音频和文字的对应进行人工校对，确保发音和文字精确对应，包括发音人朗读时额外加上的语气词等；然后对语料中的文本进行正则化处理，将数字、年月日、小数、单位符号等转换为汉字，利用自然语言处理(NLP)技术对文本进行分词，进行词性标注，识别出姓名、机构名称，对中文进行多音字消歧；再人工标注出句子的韵律边界、重音、停顿、变调等特征；然后把文字转换为拼音和音素，并进行统一编码；通过 NLP 模型（例如 BERT）进行字/词嵌入，把句子中的每个字或词转换成向量形式；对音频数据提取频谱特征；把文字嵌入向量和音频频谱特征一起输入声音克隆模型，运行反向传播算法进行迭代优化，直到模型收敛。

15 上述通过声音克隆模型合成语音的过程如下：

(1) 部署上线，把训练调优完成的声音克隆模型部署到声音克隆服务器。

(2) 合成语音，语音机器人服务或人工介入服务时，根据指定文本调用和人工坐席对应的声音克隆模型实时合成语音播放给客户。

语音合成的步骤如下：

20 首先对输入的待合成文本进行文本分析，通过文本正则化规则将数字、年月日、小数、单位符号等转换为汉字，利用自然语言处理(NLP)技术对文本进行分词，进行词性标注，识别出姓名、机构名称，对中文进行多音字消歧；再预测句子的韵律边界、重音、停顿、变调以及情感等影响发音的特征；然后把文字转换为拼音和音素，并进行统一编码；通过 NLP 模型（例如 BERT）进行字/词嵌入，把句子中的每个字或词转换成向量形式；嵌入向量输入训练好的声音克隆模型，运行前向传播算法，输出声学特征；再通过声码器（vocoder，如 wavenet, wavernn 等网络）把声学特征转换为语音波形，从而得到对应的语音。

人工交互模块实现了一个基本的 AI 对话语音机器人，也就是语音机器人，在预先编写好的话术基础上与客户进行语音交互。而语音机器人对客户意图的理解使用的是音意

理解模块中基于深度学习的音意理解模型，语音机器人和客户对话使用的音频是通过和人工坐席对应话务员声音相匹配的声音克隆模型合成得到的，且当前号码接通后，还会将语音机器人与客户的通话内容同步传输至当前人工坐席，以便人工介入时查看。

5 介入提示模块根据预定义规则和对话的动态信息，能够智能地给出人工坐席人工介入的提示，由人工坐席根据决定是否进行选择人工介入；

其中包含有量化分析子模块，用于根据历史通话大数据和当前人工交互过程的对比分析，得到当前客户需要人工介入的介入意向率，量化分析子模块会根据历史对话大数据分析各个流程节点，人工介入后产生意向客户的比率，称为“介入意向率”；

10 而预定义规则便是对介入意向率设定一个阈值，例如 50%，在语音机器人与客户对话的过程中，若发现匹配的介入意向率超过该阈值，就提示人工坐席在此时介入有较高产生意向客户的可能。

介入提示模块查询匹配介入意向率时，采用的规则包括下列规则中的一个或多个：

- (1) 对话进入预定义的关键分支。
- (2) 客户说的话触发特定的关键词。
- 15 (3) 识别出客户特定的意图。
- (4) 判断客户属于特定意向客户。

例如：贷款场景，在语音机器人与客户对话的过程中，客户说：“我最多能贷多少？贷的话有什么要求？”，进入了一个名为“回答额度”的流程节点，同时触发了关键词“什么要求”。查询统计数据，发现在“问额度”节点同时触发了关键词“什么要求”的介入意向率为 75%，超过了上述阈值（50%）。这时，就可以提示人工坐席“介入意向率 75%”，人工坐席点击该提示就可以进行人工介入。这里就同时用到了规则（1）和规则（2）。20

又如：在语音机器人与客户对话的过程中，根据音意理解模型识别出的意图为“问联系方式”，而意图为“问联系方式”时的介入意向率为 71%（超过 50%），则提示人工坐席“介入意向率 71%”。这里用到了规则（3）。

25 又如：在语音机器人与客户对话的过程中，根据话术中预先定义的意向判断规则，判定客户意向为 B，且对 B 类客户的介入意向率为 69%（超过 50%），则提示人工坐席“介入意向率 69%”。这里用到了规则（4）。

人工介入模块根据介入提示，选择介入人机交互，实现人工坐席、语音机器人和客户的三方通话，人工坐席介入可以为客户提供更复杂、灵活的服务。进入人工介入时，语音机器人停止自动语音交互，后续和客户的通话过程由话务员控制；

人工坐席介入的方式包括：

5 对应话务员直接语音回复客户；

任一话务员选择检索提词子模块生成的提词文本，由声音克隆模块生成语音回复客户；

任一话务员手动输入文本，由声音克隆模块生成语音回复客户；

10 任一话务员语音，使用声音克隆模块中的语音识别子模块将语音转换为文本，再由其中声音克隆模型生成语音回复客户。

具体的，检索提词子模块在任一话务员选择人工介入时直接调用音意理解模块展示人机交互通话意图；检索提词子模块在人工介入通话过程中调用音意理解模块，得到最新通话意图，并从数据库中检索出参考信息生成回复文本并展示。

15 检索提词子模块针对对话过程的动态信息，实时检索相关数据，对坐席给出提示或回复建议，根据对话的过程触发的预设关键词，结合音意理解模型识别出的用户意图，从数据库中检索出相关信息，例如常见问题、用户关心的数据；或者针对客户的话查找数据库中合适的回复建议。以上检索到的数据实时显示在人工坐席屏幕，供话务员介入时参考或直接选择使用，可以大大提高人工坐席的响应速度。

20 例如：贷款场景，在语音机器人或人工坐席与客户对话的过程中，客户说：“我最多能贷多少？贷的话有什么要求？”，音意理解模型识别出的意图为“问额度”，同时触发了关键词“贷多少”和“什么要求”，根据这些条件，依据话术中预先定义的规则，进一步地通过查询 API 接口从贷款业务数据库查询出贷款利率、额度等信息，动态地构造出提示语句“借贷利率 5%，个人借款金额至少 1 万元，借款要提前 1-3 天告知机构”，显示到人工坐席，以便需要人工介入时参考。配合提示信息，还可以从常见问题库 (FAQ)

25 中查询出历史总结出的经验和答复语句，告知人工坐席介入后话务员应如何回复客户，如何引导客户意向，并对客户给出指导性建议，确保最终的成单。

具体的，人工介入模块中还包含有质检模块，用于对任一话务员说话的语速、内容进行质检，并把质检结果实时反馈到人工坐席。

质检模块，能够在通话中，实时提醒话务员，实现优化客户通话体验的效果。

内容质检的方法是：根据分词结果和敏感词库进行匹配，匹配方法使用正则表达式和预定义规则库，对匹配结果进行统计分析。

5 语速质检方法是：先用语音识别技术把语音识别成文字，在根据语音时长和字数计算语速；把计算出的当前语速和历史平均语速相比，如果显著高于历史平均语速，则认为语速过快；

例如，如果话务员说话情绪不对，或有超出服务范围的语言时，给予提示警告，确保客户体验。

10 例如：人工坐席介入时，话务员由于情绪激动，辱骂客户，质检模块根据内置的敏感词库匹配了脏话的敏感词，说明坐席人员态度恶劣。该模块就把这一记录推送到主管，主管就可以对他进行批评教育，提出警告。又如：人工坐席介入时，质检模块用通话结束前坐席说的最后一句话去匹配内置的礼貌用语词库，可以检查话务员是否使用了“再见”、“谢谢”等礼貌性词语结束对话，就可以帮助改善服务态度。

15 本公开还提供一种实现人机协同的语音通讯方法，应用于上述实施例所述的语音通讯系统。具体的，包括：

S1：将通话内容输入至音意理解模型，得到所述通话内容包含的通话意图；

在本实施例中，所述音意理解模型能够直接根据通话内容识别出通话内容中包含的通话意图。例如，当通话人说出：电话号码是多少？音意理解模型即识别出通话人的通话意图是“问联系方式”。以音意理解模型为基础识别通话内容的方法，准确率高。

20 S2：获取与所述通话意图相匹配的预设回复内容；获取预先分配的人工坐席的声音特征；

本公开提供的技术方案中，在通话接通后，首先为通话分配一个人工坐席，人工坐席对应一个人工话务员，一个人工话务员可以对应多个人工坐席。人工话务员在通话接通后，可以用于监听通话人及语音机器人之间的通话。

25 S3：将所述预设回复内容以及所述声音特征输入至声音克隆模型，得到克隆音频；

在本公开的技术方案中，得到预设回复内容后，需要将预设回复内容的文本转换为语音答复通话人，并且，答复的语音必须与分配的人工坐席的声音特征相匹配。为如果后续由人工坐席介入通话，生成克隆音频这一步骤，可以减少介入前与介入后通话的差

异，提高通话质量。

具体的，生成克隆音频的步骤可以包括：

S3001：将预设回复内容作为待合成文本输入，通过文本正则化规则以及自然语言处理技术对所述待合成文本进行预处理；

5 S3002：预测所述待合成文本的发音特征，根据预测到的发音特征，将所述待合成文本转换为拼音及音素，并进行统一编码；

S3003：通过 NLP 模型对统一编码后的所述待合成文本进行向量转换；

S3004：将向量转换后的所述待合成文本输入所述声音克隆模型，运行前向传播算法，得到所述待合成文本的声学特征；

10 S3005：利用声码器将所述声学特征转换为语音波形，以得到克隆音频。

本公开提供的技术方案，还可以包括训练所述声音克隆模型，具体包括以下步骤：

S3006：采集所述人工坐席的声音语料，提取所述声音语料中的频谱特征；

S3007：将所述声音语料与对应的文字文本进行校对；

15 S3008：通过文本正则化规则以及自然语言处理技术对校对后的所述文字文本进行预处理；

S3009：对预处理后的所述文字文本进行发音特征的标注，根据标注将所述声音语料转换为拼音及音素，并进行统一编码；

S3010：通过 NLP 模型对统一编码后的所述文字文本进行向量转换；

20 S3011：将所述频谱特征以及向量转换后的所述文字文本输入所述声音克隆模型，运行反向传播算法进行迭代优化至所述声音克隆模型收敛，以得到训练后的所述声音克隆模型。

在本实施例中，还可以包括对声音语料进行降噪处理，以提高准确率。

S4：使用所述克隆音频答复通话；所述克隆音频为声音特征与所述人工坐席相匹配的音频；

25 S5：将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意

向率；所述当前介入意向率为产生意向客户的概率；

在实际应用中，通话内容及其包含的通话意图，能够体现出通话人是否能够成为意向客户，此时可以由人工坐席介入通话，进一步与通话人进行沟通。所述介入意向率即为当前通话能够产生意向客户的概率。

- 5 具体的，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤包括：

S5001：获取通话内容中包含的每一问询环节内容，所述问询环节为通话双方发起问询的过程；

S5002：将每一问询环节作为一个分支执行遍历操作；

- 10 S5003：在预设的介入意向率数据库中存在与遍历的分支相匹配的预定义关键分支的情形下，获取所述预定义关键分支的介入意向率，以作为当前通话的当前介入意向率。

在本实施例中，问询环节的示例如下，语音机器人发起问询：是否有在 A 地购房的意愿。如果预定义关键分支中包括“问询购房意愿”，即可表明通话进入关键分支。此时，以所述预设关键分支的介入意向率作为当前介入意向率。在实际应用中，所述预设关键分支可以根据实际情况预先设置，本公开不作具体限定。

15

将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还包括：

S5004：获取通话内容中包含的每一问答环节内容，所述问答环节为通话双方的问询及回答的过程；

- 20 S5005：遍历所述介入意向率数据库；

S5006：在所述介入意向率数据库中存在与所述问答环节相匹配的预设关键词的情形下，获取所述预设关键词的介入意向率，以作为当前通话的当前介入意向率。

- 在本实施例中，所述预设关键词可以为：想购房、购房资格等。可以根据实际情况进行设计。问答环节示例如下：语音机器人发起问：是否有在 A 地购房意愿。此时通话人的回答可以包括以下几种情况：（1）肯定回答：我想在 A 地购房；（2）反问回答：我是否有购房资格；（3）否定回答：我不想买房。上述答复方式中，进入了预设关键分支，并且肯定回答及反问回答触发预设关键词。相应的，即使否定回答未触发预设关键词，通话也进入了预设关键分支。
- 25

将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还包括：

S5007：获取所述通话意图；

S5008：遍历所述介入意向率数据库；

- 5 S5009：在所述介入意向率数据库中存在与所述通话意图相匹配的预定义意图的情况下，获取所述预定义意图的介入意向率，以作为当前通话的当前介入意向率。

在实际应用中，通过遍历得到与所述通话意图相匹配的预定义意图这一步骤，能够识别出通话人特定的意图。

- 10 在本公开部分实施例中，将所述通话内容以及所述通话意图输入至量化分析子模块，以计算当前介入意向率的步骤还可以包括：

S5010：获取与所述通话意图相匹配的预设回复内容；

S5011：遍历所述介入意向率数据库；

- 15 S5012：在所述介入意向率数据库中存在与所述预设回复内容相匹配的预定义意向判断规则的情形下，获取所述预定义意向判断规则对应的介入意向率，以作为当前通话的当前介入意向率。

在本实施例中，预设回复内容能够体现通话人为哪一种类的意向客户，因此，通过设置预定义意向判断规则，可以明确通话人的意向种类。

- 20 在实际应用中，当前通话的当前意向率可以由上述步骤共同决定，例如，以上述步骤计算出的当前介入意向率的总和为最终介入意向率。具体计算方式可以根据实际情况进行设计。

在本公开部分实施例中，所述方法还包括：

S5013：设置预定义关键分支、预设关键词、预定义意图以及预定义意向判断规则；

S5014：将所述预定义关键分支、所述预设关键词、所述预定义意图以及预定义意向判断规则保存至预设的数据库中，以形成所述介入意向率数据库。

- 25 将所述预定义关键分支、所述预设关键词、所述预定义意图以及预定义意向判断规则保存至预设的数据库中的步骤后，还包括：

S5015: 获取历史通话数据, 所述历史通话数据为通话双方的通话内容;

S5016: 筛选出所述历史通话数据中有人工坐席介入的通话;

S5017: 筛选出有人工介入的通话中存在所述预定义关键分支、所述预设关键词、所述预定义意图以及所述预定义意向判断规则的通话, 以作为高意向率通话。

- 5 在实际应用中, 高意向率通话可以为包括高意向率事件的通话, 高意向率事件可以为触发预定义关键分支、预设关键词、预定义意图以及预定义意向判断规则的事件, 可以根据实际情况进行设计。

筛选出有人工介入的通话中存在所述预定义关键分支、所述预设关键词以及所述预定义意图的通话, 以作为高意向率通话的步骤后, 还包括:

- 10 S5018: 统计有人工介入的通话的通话数量, 以及所述高意向率通话数量;

S5019: 计算所述高意向率通话数量与有人工介入的通话的通话数量之间的比值, 得到所述预定义关键分支、所述预设关键词、所述预定义意图以及所述预定义意向判断规则的介入意向率。

- 15 在本公开的技术方案中, 计算介入意向率示例如下: 统计的通话数量为 100 个, 其中包括 10 次有人工坐席介入的通话, 人工坐席介入的通话中, 包括 5 个触发预设关键词的高意向率事件, 则预设关键词对应的客户意向率为 $5/10=50\%$ 。

S6: 如果所述当前介入意向率大于预设阈值, 调用所述人工坐席介入通话。

在实际应用中, 所述预设阈值可以为 50%, 可以根据实际情况进行设计。

- 20 S7: 通过检索题词子模块获得历史通话意图以及预设通话提词文本; 根据所述历史通话意图以及预设通话提词文本语音答复通话;

和/或, 将参考回复内容输入至所述音频克隆模型, 得到克隆音频;

使用所述克隆音频答复通话; 所述参考回复内容为其他人工坐席回复的语音。

在本实施例中, 人工坐席介入通话后, 人工坐席的人工话务员可以根据历史通话意图及预设通话提词文本进行语音答复通话人。

- 25 由以上技术方案可知, 本公开提供的实现人机协同的语音通讯系统及方法, 能够根据通话人的通话内容准确获知通话意图, 根据通话意图匹配预设回复内容, 并通过声音

克隆模块生成克隆音频答复通话。还可以计算介入意向率，保证人工坐席顺利介入通话。同时，答复通话的灵活性强，通话流畅度高。

5 以上实施例仅用以说明本公开的技术方案而非限制，尽管参照较佳实施例对本公开进行了详细说明，本领域的普通技术人员应当理解，可以对本公开的技术方案进行修改或者等同替换，而不脱离本公开技术方案的宗旨和范围，其均应涵盖在本公开的权利要求范围当中。本公开未详细描述的技术、形状、构造部分均为公知技术。